The 31st Architectures and Mechanisms for Language Processing 2025

Book of Abstracts

Editor: Jan Chromý
Faculty of Arts, Charles University
Prague
https://amlap2025.ff.cuni.cz/

Contents

Preface 4

Sponsors . . . 4

Acknowledgements . . . 4

Organization 5

Organization team . . . 5

Student Assistants . . . 5

Reviewer committee 6

Important Information 7

Venue . . . 7

Registration . . . 7

Poster presentations . . . 7

Coffee breaks . . . 7

Internet . . . 7

Social program 8

Welcome reception . . . 8

Conference dinner . . . 8

Scientific Program 9

Selection Criteria . . . 9

Keynotes . . . 9

Satellite Workshops . . . 9

Schedule 10

List of Accepted Posters 15

Satellite Workshops 31

Big Data in Psycholinguistics . . . 31

Multilingual Experience: From Individuals to Society . . . 35

Keynote Talks 39

Oral Presentations 44

Session 1 45
Session 2 56
Session 3 67
Session 4 76
Session 5 85
Session 6 92
Session 7 103
Session 8 114

Poster sessions 123

Thursday Morning Posters 124
Thursday Afternoon Posters 251
Friday Morning Posters 377
Friday Afternoon Posters 503
Saturday Morning Posters 633

Preface

We are pleased to welcome you to Prague for the 31st edition of the conference *Architectures and Mechanisms for Language Processing*. We hope you will enjoy the conference itself as well as the satellite workshops, which take place the day before the main event.

This volume contains essential information about the conference, along with the abstracts of the keynote talks, contributed talks, and posters. The abstracts are reproduced here in the format submitted by the authors.

Sponsors

We would like to sincerely thank the Faculty of Arts, Charles University, for their financial, technical, and administrative support of the conference. The satellite workshops are supported by the European Regional Development Fund project *Beyond Security: Role of Conflict in Resilience-Building* [CZ.02.01.01/00/22_008/0004595].

Acknowledgements

We would like to thank Hannah Rohde, the organizer of AMLAP 2024, for sharing her experiences and providing many useful tips that greatly helped us in the organization process. We are also very grateful to Pia Knoeferle for sharing the AMLAP 2018 LaTeX template, which we adapted in preparing this brochure.

Enjoy the conference!

Jan Chromý (Conference Chair)

Organization

Organization team

Jan Chromý (Conference Chair), Markéta Ceháková, Kateřina Chládková, Anna Chromá, Maroš Filip, Radim Lacina, Nikola Paillereau, Mikuláš Preininger, and Filip Smolík

Student Assistants

Lucie Doležal Nováková, Ondřej Drobil, Martina Dvořáková, Lucie Guštarová, Michaela Hájková, Najma Jaufeerally, Lukáš Jiřička, Ivan Kraus, Nina Mallia, Barbora Musilová, Eva Pospíšilová, Filipa Rychterová, and Táňa Šimková.

Reviewer committee

We greatly appreciate the work of the AMLaP 2025 reviewing committee; their careful and timely reviews played a crucial role in creating a strong and engaging program.

Anne Abeillé, F.-Xavier Alario, Manabu Arai, Jane Aristia, Sudha Arunachalam, Emily Atkinson, Esperanza Badaya, Nadine Bade, Markus Bader, Chiara Barattieri di San Pietro, Ioli Baroncini, Paloma Batista Cardoso, Esti blanco-elorrieta, Ivan Paul Bondoc, Mara Breen, Jonathan Brennan, Andrea Bruera, Marc Brysbaert, Wednesday Bushong, Spencer Caplan, Katy Carlson, Markéta Ceháková, Charles Clifton, Derya Cokal, Aymeric Collart, Saveria Colonna, Eva Commissaire, Carla Contemori, Ruth Elizabeth Corps, Ian Cunnings, Julien Diard, Mariapaola D'Imperio, Jakub Dotlacil, Katerina Drakoulaki, John Duff, Shaohua Fang, Maroš Filip, Dušica Filipović Đurđević, Steven Foley, Francesca Foppolo, Alice Foucart, Stefan L. Frank, Raquel Freitag, Martin Fuchs, Zuzanna Fuchs, Alison Gabriele, Alan Garnham, Edward Gibson, Margaret Grant, Jeffrey J. Green, Jiuzhou Hao, Charlotte Hauser, Fredrik Heinat, Julia Heine, Barbara Hemforth, Yuko Hijikata, Masako Hirotani, Jet Hoek, Holger Hopp, Edward Matthew Husband, Kateřina Chládková, Youngon Choi, Kiel Christianson, Jan Chromý, Kin Chung Jacky Chan, Aine Ito, Cassandra L. Jacobs, Edith Kaan, Elsi Kaiser, Anna Kamenetski, Yuki Kamide, Serpil Karabuklu, Kalliopi Katsika, Gerrit Kentner, Maayan Keshev, Eva Klingvall, Alina Konradt, Maria Korochkina, Franziska Kretzschmar, Helene Kreysa, Li-Chuan Ku, Anuenue Kukona, Dave Kush, Nayoung Kwon, Radim Lacina, Sol Lago, Hyun Kyung Rachel Lee, Robin Lemke, Lisa Levinson, Keng-Yu Lin, Shane Lindsay, Maria Lobo, Madeleine Long, Paula Luegi, Oana Lungu, Jun Lyu, Marcus Maia, Kyla McConnell, Kevin McManus, Natalia Meir, Aya Meltzer-Asscher, Michael Meng, Barbara Mertins, Katherine Messenger, Stephan Meylan, Evelyn Milburn, Jelena Mirkovic, Natalia Mitrofanova, Holger Mitterer, Savithry Namboodiripad, Alexandra Navarrete-González, Shinri Ohta, Rachel Ostrand, Dario Paape, Elena Pagliarini, Nikola Paillereau, Dan Parker, Elisa Passoni, Umesh Patil, Michelle Perdomo, Jasmin Pfeifer, Grusha Prasad, Mikuláš Preininger, Anna Pryslopska, Yanina Prystauka, Péter Rácz, Ingo Reich, Francois Rigalleau, Leah Roberts, Hannah Rohde, Margaret Ryan, Mikel Santesteban, Kelsey Sasaki, Gyu-Ho Shin, Merel CJ Scholman, Florian Schwarz, Mieke Slim, Shayne Sloggett, Giuditta Smith, Filip Smolík, Jina Song, Christoforos Souganidis, Kate Stone, Anastasia Stoops, Mikihiro Tanaka, Gabriel Thiberge, Guillaume Thierry, Ellen Thompson, Norbert Vanek, Shravan Vasishth, Margreet Vogelzang, Titus von der Malsburg, Laurence White, Hans A. Wilke, Eva Wittenberg, Masaya Yoshida, Kaitlyn Leigh Zavaleta, Likan Zhan, Linmin Zhang

Important Information

Venue

AMLaP 2025 takes place at the Faculty of Arts, Charles University (nám. Jana Palacha 2, Prague 1, 110 00, Czech Republic). The talks will take place in the room n. 131, poster sessions will take place in the corridors on the first and second floor.

Registration

On Wednesday and Thursday mornings, the registration desk will be located at the main entrance of the Faculty of Arts (nám. Jana Palacha 2), where the conference takes place. At other times, registration will be available by the coffee-break tables on the second floor of the Faculty of Arts. You can also register during the welcome reception on Wednesday evening at Hybernská 4. Please register promptly and pick up your badge, as well as your dinner ticket (if applicable).

Poster presentations

Each poster stand carries a number that you can also find in the program. Please put up your poster at the corresponding poster stand (check your number in the program) and do not forget to remove your poster once the poster session ends.

Coffee breaks

Coffee breaks will take place on the 2nd floor of the building. Please note that another event, the Prague Summer School of Chinese Poetry 2025, is being held on the 1st floor. They also have refreshments available, so we kindly ask you to make sure not to take food or drinks intended for their participants.

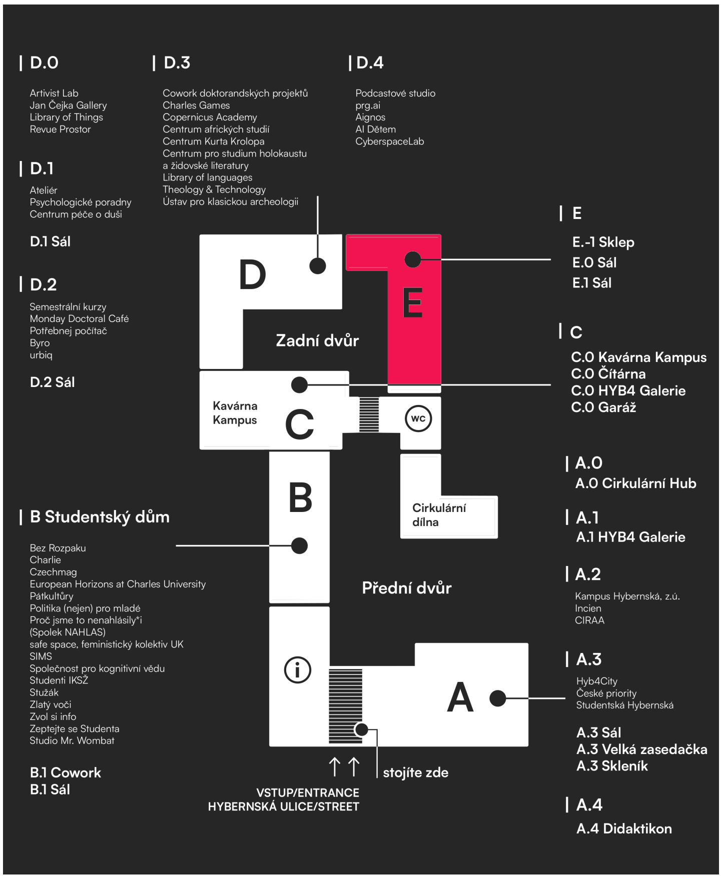
Internet

Free Internet connection is available at the venue through your Eduroam account. Unfortunately, we are not able to offer an alternative connection.

Social program

Welcome reception

The welcome reception will follow the satellite workshop on Wednesday, September 3, and will be held at Kampus Hybernská (Hybernská 4, Praha 1). The reception is open to all conference participants, and on-site registration will also be available. Guests can enjoy food and beverages, accompanied by live music from the band Der Šenster Gob. The map of the Kampus is shown on the next page. The reception will take place in section E (marked with red color).



Conference dinner

The conference dinner will take place on Friday, September 5, 2025, at Vinohradský pivovar (Vinohrady Brewery; Korunní 2506/106, 101 00 Praha 10–Vinohrady), starting at 7 PM. The venue is within walking distance of Flora metro station (direct line from Staroměstská which is the metro station next to the conference venue). The closest public transport stop is Orionka tram station. The dinner is reserved for registered participants who have purchased a ticket. Your dinner ticket will be attached to your badge and handed to you at registration.

Scientific Program

Selection Criteria

AMLaP 2025 received a total of 361 submissions, of which 13 had to be rejected. Based on reviewers' scores and their recommendations for oral presentation, 31 papers were accepted as talks, while the remaining 317 submissions were accepted as posters.

Keynotes

We are delighted to welcome our four invited speakers: Linda Drijvers, Vera Demberg, Michael Ramscar, and Guillaume Thierry.

Satellite Workshops

On September 3, we organize two satellite workshops: Big Data in Psycholinguistics and Multilingual Experience: From Individuals to Society.

Schedule

Wednesday 3rd September

08:30–09:00 Registration

9:00–14:30 Workshop 1: Big Data in Psycholinguistics

09:00–09:15 Opening

09:15–10:25 Talk 1: Penny Pexman: Leveraging Big Data to Map Lexical-Semantic Space

10:25–10:45 Coffee break

10:45–11:55 Talk 2: Erin Buchanan: Designing Big Science: Lessons from a 25,000-Participant, 30-Language Semantic Priming Study

11:55–12:15 Coffee break

12:15–13:25 Talk 3: Emmanuel Keuleers: Understanding Measures of Word Occurrence in Psycholinguistics through a Network of Static and Dynamic Actors

13:25–14:30 Lunch (on your own)

14:30–18:00 Workshop 2: Multilingual Experience: From Individuals to Society

14:30–15:00 Opening talk: Kateřina Chládková: Researching and communicating multilingualism

15:00–15:20 Coffee break

15:20–16:30 Talk 2: Sharon Unsworth: Using Public Engagement to ‘Normalize’ Multilingualism

16:30–16:50 Coffee break

16:50–18:00 Talk 3: Natalie Boll-Avetisyan: Born into a Multilingual Society: Early Language Acquisition in Ghana

19:00–22:00 Welcome reception, Kampus Hybernská

Thursday 4th September

08:00–08:50 Registration

08:50–09:00 Opening

09:00–10:00 Keynote 1: Vera Demberg: Towards personalized models of processing difficulty: Modelling individual differences in working memory capacity and background knowledge

10:00–11:30 Coffee & Thursday Morning Posters

11:30–12:50 Session 1 (Chair: Filip Smolík)

Eunice G. Fernandes & Yanina Prystauka & Foyzul Rahman & Helene Slaattelid Øya & Allison Wetterlin & Katrien Segaert & Linda Wheeldon: Age-related effects of language proficiency and use on language switching

Jade Sandstedt & Hjalmar T. Eiksund & Mizuki Tanigawa: Processing multilectal grammatical microvariation: Mapping individual differences with self-paced reading

Hannah G. Treadway & Souad Kheder & Jorge Valdes-Kroff & Edith Kaan: Cognitive control adaptation in code-switching: An ERP study

Dušica Filipović Đurđević & Lara Perić & Viktorija Jovanović: Phonological ambiguity effect: new kid on the block switching paradigm

12:50–14:20 Lunch (on your own)

14:20–15:40 Session 2 (Chair: Radim Lacina)

Leonardo Concetti: Encoding and Reactivating Syntactic Nodes: insights from Coordination

Li Kloostra & Rick Nouwen & Jakub Dotlacil: Memory Retrieval in Discourse with ‘again’: Eye-tracking and acceptability studies

Chengjie Jiang & Walter van Heuven & Ruth Filik: Is world knowledge activation exhaustive or selective during language comprehension? Evidence from bidirectional self-paced reading

Pia Schoknecht & Dario Paape & Shravan Vasishth: The time course of local coherence effects in reading times and event-related potentials

15:40–16:00 Coffee break

16:00–17:20 Session 3 (Chair: Nikola Paillereau)

Serge Minor & Natalia Mitrofanova & Gillian Ramchand: “Quantum leaps” in grammar acquisition: Evidence from child Russian

Julia Ercse & Peter Hendrix: The temporal dynamics of word learning: a time-to-event analysis of age-of-acquisition

Elena Marx & Hanna Shine & Eva Wittenberg & Jesse Snedeker: Preschoolers use event dynamics to infer temporal order in language

Barbora Skarabela & Mitsuhiko Ota & Filip Smolík: Children’s sensitivity to animacy constraints in possessive noun phrases: Priming induces non-adult-like structures in Czech preschoolers

17:20–18:50 Snacks & Thursday Afternoon Posters

Friday 5th September

09:00–10:00 Keynote 2: Guillaume Thierry: The Bilingual Multiverse: Real-Time Effects of Language of Operation on Categorical Perception and Abstract Conceptualisation

10:00–11:30 Coffee & Friday Morning Posters

11:30–12:50 Session 4 (Chair: Filip Smolík)

Gustavo Lopez Estivalet & Katie L. McMahon & Joanne Arciuli & Greig I. de Zubicaray: Phonological typicality in the distribution of nouns and verbs in French: Statistical relations between form and grammar

Jéssica Gomes & Sol Lago & João Veríssimo: Ageing yields improvements in morphosyntactic prediction

Danny Dixon & Lisa Levinson: Sensitivity to Verb Bias as a Continuous Variable in L1 and L2 Processing

Anna Fiona Weiss & Markus Bader & Michael Meng: The task dependence of misinterpretation effects: A comparison of L1 and L2 speakers of German

12:50–14:20 Lunch (on your own)

14:20–15:20 Session 5 (Chair: Radim Lacina)

Katja Haeuser: Predictive processing adapts to prediction error in a non-linear fashion – that's why adaptation effects are so difficult to detect using LMERs

Amalia Spyromilio & Holly Jenkins & Elizabeth Wonnacott & Michael Ramscar: How gender information influences spontaneous speech in context

Susanne Eisenhauer & David Hernández-Gutiérrez & Simona Mancini: Linguistic complexity measures (surprisal, entropy, semantic similarity, syntactic node counts) differentially impact initial vs. re-reading

15:20–16:50 Coffee & Friday Afternoon Posters

16:50–17:50 Keynote 3: Michael Ramscar: What is language processing anyway?

19:00–23:00 Conference dinner (Vinohradský pivovar), only with dinner registration

Saturday 6th September

09:00–10:00 Keynote 4: Linda Drijvers: Towards a multimodal view on the neurobiology of language

10:00–11:30 Coffee & Saturday Morning Posters

11:30–12:50 Session 6 (Chair: Filip Smolík)

Fabio Marson & Giulia Loca & Marco Ciapparelli & Marco Marelli: Context-based encoding of novel meanings after minimal exposure to natural text: an EEG study on integration of linguistic chimeras

Harshada Vinaya & Sean Trott & Seana Coulson: Vision Language Model Representations Predict EEG Response to Visual and Auditory Attributes in Property Verification

Svetlana Mnogogreshnova & Sol Lago & Esther Rinke & Petra Schulz & Clara Vilà Dolado: Comprehension of Pragmatically Licensed Sentential Negation and Its Influence on Memory Retention

Shiyu He & Petar Milin & Dagmar Divjak: Where L2 Still Looks L2: Spatial Undershoot and Logographic Saccadic Programming in Advanced L1 Chinese/L2 English readers

12:50–14:20 Lunch on your own

14:20–15:40 Session 7 (Chair: Nikola Paillereau)

Tamara Butigan & Norbert Vanek & Robert Greenberg: Language planning and lexical competition: evidence from lexical processing for sociopragmatic differentiation of absolute synonyms

Marc Brysbaert & Javier Conde & Pedro Reviriego & Gonzalo Martínez: Augmenting the psycholinguistic toolbox with AI-generated word characteristics

Sophie Repp & Heiko Seeliger & Sven David Weber: Pronoun accentuation produces interference effects in memory for alternatives

Charles Lin & Zeping Liu & Xiao Dong: Building Structures Left to Right and Bottom Up: The Production and Perception of Syntactic Branching by L1 and L2 Users of a Tone Language

15:40–16:00 Coffee break

16:00–17:20 Session 8 (Chair: Radim Lacina)

Elliot Schwartz & Griffin Pion & Jake Quilty-Dunn & Eric Mandelbaum & Spencer Caplan: Thinking about nothing: the processing and mental representation of lexical ambiguity

Isabella Fritz & Joshua Booth & Aditi Lahiri: Redefining psycholinguistic cognates: Linguistic and historical considerations

Ana Bautista & Clara Martin: Not that cloze: Semantic sentence constraint is influenced by language background and dominance

John Duff & Laura Pissani: Jointly modeling maze RT and accuracy using diffusion models: A first case study

17:20–17:40 Closing remarks

List of Accepted Posters

Thursday Morning

(4 September, 10:00–11:30)

2 Jun Lyu: Acquisition of the blocking effect in L2 Chinese by L1 Japanese speakers

11 Tess Fitzpatrick: Finding, sharing, and losing words: word associations and the mental lexicon

13 Robiatu Al Addawiyah & Cristiano Chesi & Adriana Belletti: A Computational Perspective on the Stage of Acquisition of Grammatical Competence: Testing the Growing Tree Approach

24 Yung Han Khoe & Gerrit Jan Kootstra & Stefan L. Frank & Rob Schoonen & Edith Kaan: Shared syntax in bilingual humans and cognitive models: Code-switching increases cross-language structural priming

34 Yichi Serena Zhang & Xufeng Duan & Zhenguang Cai: The Causal Role of Supplementary Motor Area (SMA) in Orthographic Retrieval During Chinese Character Handwriting

37 Leona Polyanskaya & Mikhail Ordin: Emergence of suffixing bias: Affixation patterns in L1 and sequence processing by statistical learning mechanisms

40 Leigh B. Fernandez & Muzna Shehzad & Lauren V. Hadley: Does prediction require executive resources?

44 Anne Abeille & Emma Kious: Extraction out of wh-clauses depends on the construction: evidence from French

49 Anne Neveu & Emma Libersky & Margarita Kaushanskaya: Novel word learning over different time scales: A comparison of paired-associate and cross-situational paradigms

58 Dominic Schmitz: The processing costs of generic and specific singular they: A self-paced reading study

62 Michelle Suijkerbuijk & Naomi Tachikawa Shapiro & Peter de Swart & Stefan L. Frank: The success of Neural Language Models on syntactic island effects is not universal: strong wh-island sensitivity in English but not in Dutch

64 Cecilia Husta & James Trujillo & Judith Holler & Linda Drijvers & Antje Meyer: Utterances with Decreasing Entropy Facilitate Speech Comprehension and Concurrent Planning

74 Hannah Bou-Lai Lam & Johanne Paradis: It’s not all Chinese to them: Differential heritage bilingual processing and rating of classifiers in Cantonese and Mandarin

77 Jolana Treichelová & Anna Chromá & Filip Smolík: Beyond Familiar Verbs: Czech-learning Children’s Comprehension of Noncanonical OVS Word Order

81 Haoyu Zhou & Fabienne Chetail & Louisa Bogaerts: Reliable measures of orthographic statistical learning predict spelling but not reading skill

92 Bernard A J Jap & Yu-Yin Hsu: When cues collide: The role of contextual and classifier-based prediction in Mandarin comprehension

93 Antje Lorenz & Anna-Lisa Döring & Lara Mundt & Pienie Zwitserlood & Rasha Abdel Rahman: On the lexical representation(s) of compounds: Evidence from continuous naming in young and older healthy speakers

94 Koyel Mukherjee & Bidisha Som & Abhishek Shrivastava: Spatial Order and Cognition Difficulties: An Eye Tracking Study of Comic Panel Layouts

99 Nitzan Trainin & Einat Shetreet & Aya Meltzer-Asscher: Online Generalization of Speaker-Specific Lexical Preferences

108 Dinah Baer-Henney & Alexander Clemen: Tracing the development of German number cues: A case study from the LEO corpus

110 Natalia Mitrofanova & Serge Minor & Nadine Kolb & Christina Athanasiadi & Marit Westergaard: The role of heritage and societal languages in L3 aspect processing: Evidence from eye-tracking

115 Felipe von Hausen & Lucía Castillo & Mauricio Aspé & Ernesto Guerra: Sight translation of non-canonical structures: Eye movement patterns and individual differences

118 Juliana Gerard & Adina Camelia Bleotu: (Dis)agreement across languages: Cues to control in English and Romanian

127 Yunju Nam & Sun-Young Lee & Hyeonjeong Jeong & Juno Baik: Word-order or Truth-value? Dominant cues during Korean incremental processing with the picture-sentence verification task

130 Ngoc-Anh Tran & Kazimierz Garstecki & Giovanni Cassani: A cute horgous meets a scary timfil: how do we interpret novel words in context?

131 Raya Mezeklieva & Peter Hendrix: Can two words mean exactly the same? Insights from a distributional semantics approach.

134 Holly Jenkins & Elizabeth Wonnacott & Michael Ramscar: The role of contextual alignment in artificial grammar learning

140 Andrea Hofmann & João Veríssimo & Isabell Wartenburger: The relationship between perceptual abilities and speech-to-speech synchronization: A Bayesian mixture modeling approach

144 Giulia Bovolenta & John N. Williams: Declarative memory effects in L2 morphology learning reflect explicit rule acquisition

152 Stefan Blohm & Mathias Barthel: When 'yes' sounds like 'maybe': Inferences about respondents in offers and requests

156 Pepita Alex & Marta Brzeska & Julia Schwarz & Benjamin W. Tatler & Agnieszka E. Konopka & Anastasia Klimovich-Gray: Conditioned Delusions: Belief Updating During Naturalistic Reading is Modulated by Individual Cognitive Profile

157 Marie Christin Walch: Context Effects on the Interpretation of Bare Numerals: Evidence from Event Uncertainty and Roundness

172 Liliana Nentcheva & Andrea Santi: Testing the Specificity of Human Parser Predictions during ‘Hyper-Active’ Gap Filling

188 Kirill Chuprinko & Artem Novozhilov & Arthur Stepanov: Modeling Acceptability in Free Word Order Languages: The Role of Dependency Distance and Projectivity

201 Benedek Bartha & Eva Wittenberg & Christophe Heintz & Jennifer Culbertson: Conceptual similarity, but not informativeness, shapes evidential systems during learning

227 Maroš Filip & Kateřina Chládková: Neural speech tracking in a bilingual cocktail party: Does language identity matter?

228 Sebastian Walter & Lennart Fritzsche: Headnods don’t always mean ‘yes’: Ambiguity in gestural responses to negative questions

239 Foteini Karkaletsou & Gunnar Jacob & Shanley E. M. Allen: Cross-linguistic structural priming of reciprocal innovations in French-English bilinguals

243 Clara Seyfried & Yuki Kamide: Between conflict and causality: the connective “but” in discourse processing and recall

248 Christopher Allison & Falk Huettig & Thomas Lachmann: Visuospatial cognitive load disrupts predictive gaze behavior but not prediction

249 Maryam Meghdadi & John Duff & Vera Demberg: Integrating language model embeddings into ACT-R

252 Suzy Park: Contrastive Prosody and Pragmatic Meaning: Evidence from Korean L2 Speakers of English

257 Rosa Zaaijer & Caitlin Meyer & Marieke Schouwstra & Monique Flecken: Language and Line Dancing: the Role of Linguistic Labels in Action Learning

270 Marina Sokolova: Parsing effect of structural prediction in sentences with code-switching

272 Kanika Sachdeva & Himanshu Yadav: No Evidence for Syntactic or Semantic Interference in Hindi Subject-Verb Processing

273 Franziska Kretzschmar & Sandra Hansen & Christian Lang: On the suitability of LLM output as an experimental data source in German: Evidence from GPT-4o, LLaMa 3.1 70B and LLaMa 3.1 8B

283 Constantijn van der Burght & Antje Meyer: Working memory capacity predicts sensitivity to prosodic structure

284 Roberto Petrosino & Jon Sprouse & Diogo Almeida: No pseudo-morphological decomposition during lexical access, but actual morphological analysis in the lexicon: Meta-analytical evidence from seven new replicated masked stem priming experiments

290 Francesca Penoncelli & Nino Grillo & Giuliano Bocci: Implicit Prosody and Pseudo Relative availability independently modulate RC-attachment

293 Demi Zhang & Emiliana Pulido & Maria Josefina Estrada & Souad Kheder & Edith Kaan: Producing code-switches: Adaptation of cognitive control in code-switching

295 Natalia Slioussar: Teasing apart productivity and defaultness in time-frequency responses: an EEG on Russian

299 Mercedes Martinez Bruera & Matilde Calmejane & Carolina Gattei & Carlos J. Alvarez & Horacio A. Barber & Daniel Weingärtner & Andrea Listanti & João Veríssimo & Sol Lago: Similar meaning does not always mean similar processing

310 Michael Vrazitulis & Pia Schoknecht & Shravan Vasishth: A Progress Report on Ongoing Benchmark Data Collection for German Sentence Processing: Eye-Tracking and Self-Paced Reading

326 Anna Runova & Zuzanna Fuchs: Perception and production of gender-marking vowels in heritage Russian

328 Ren Li & Walter van Heuven & Chen Zhao: English sentence planning differences between English L1 and Chinese-English L2 Speakers: Evidence from eye-tracking

332 Mikael André Albrecht & Katrien Segaert & Eunice G. Fernandes & Allison Wetterlin & Linda Wheeldon: Fluency and complexity in speech production: effects of healthy ageing

333 Astha Singh & Evgeny Chukharev & Mark Torrance: Brief lookback cues content generation in spontaneous multi-sentence text production

352 Patricia Fuente-García & Julián Villegas & Irene de la Cruz-Pavía: Shifting, Inhibition and Updating in younger and older Basque-Spanish bilinguals

Thursday Afternoon

(4 September, 17:20–18:50)

8 Verónica García-Castro & Norbert Vanek: Syntactic engagement of newly learned words: a garden-path method applied to track emerging sensitivity to structural ambiguity

9 Adam Ussishkin & Jessica Nieder: Language specific differences in morphological processing: The role of semantics in Maltese vs. Hebrew lexical access

10 Titus von der Malsburg & Sebastian Padó: Transformers fail to predict consistent effects for agreement attraction configurations

15 Hao Zeng & Aine Ito: Investigating the resolution of conflicting predictions from global and local contexts: An eye-tracking sentence reading study

23 Anna Teresa Porrini & Veronica D'Alesio & Matteo Greco: The processing and interpretation of Expletive Negation in children and adolescents

27 Saveria Colonna & Paul Lejeune & Elif Mutlu: Lexical Recognition of Gender-Fair Contracted Forms in Typical and Dyslexic Readers

46 Francesca Foppolo & Valeria Galimberti & Dongpeng Pan & Francesca Panzeri & Stephanie Durrleman: Some and all in the visual world of preschoolers

54 Anuenue Kukona: Distinguishing the mechanisms that support predictive sentence processing: Evidence from associations and speech rate

67 Areti Kotsolakou & Frank Wijnen & Sergey Avrutin: Input entropy affects frame-based category-learning

68 Dandan Li & Pia Knoeferle & Agnes Villwock & Katja Maquate: How age modulates the ability to benefit from sensory-situated semantic congruence: an ERP study

73 Alice Eddyshaw: Speaker social characteristics and the resolution of linguistic ambiguities: A self-paced reading task study

82 Bahareh Yousefzadeh & Cassandra L Jacobs: The role of Ezafe in the typed production of Persian compound words

90 Arrate Isasi-Isasmendi & Roberto Zariquiey & Balthasar Bickel & Caroline Andrews: Neural Signatures of Dependency Processing: Distinguishing Syntax and Semantics

95 Junhua Ding & Siyu Chen & Chen Feng & Su Li: Occipitotemporal and frontal regions are crucial for Chinese children's reading development

97 Zehua R. Jiang & Mingyuan Yang: Frequency Modulates Phonetic but Not Semantic Radical Effects in Chinese Character Recognition

103 Andreas Opitz & Denisa Bordag & Hans-Georg Berulava: Asymmetry in the Retention of Content and Surface Linguistic Information During Reading in L1 and L2: An Eye-Tracking Study

104 Philine Link & Leendert van Maanen & Jakub Dotlacil: Similarity Comes at a Cost: Novel Evidence for Associative Memory Retrieval

107 Christina Papoutsi & Elli Tourtouri & Vitória Piai & Antje S. Meyer: What drives word choices and naming latencies? Examining the roles of semantic and lexical variables in modal and alternate word production

111 Eva Pospíšilová & Ondřej Drobil & Anna Marklová & Jiří Milička: Humans are bad at recognizing AI - but they can learn it from feedback

117 Vera Heyer & Holger Hopp & Regina Hert & E Jamieson & Barbara Köpke & Monika S. Schmid: How Bilinguals Use Grammatical Cues to Make and Revise Predictions: Effects of Age of Onset and Cross-Linguistic Influence

119 Yoana I. Dancheva & Margreet Vogelzang & Ianthi M. Tsimpli: Is code-switching effortless? A look at processing and production costs

124 Yixin Cui & Lavinia Salicchi & Yu-Yin Hsu: How Large Language Models Evaluate Embedded Wh-Questions: A Cross-Linguistic Comparison of Chinese and English

132 Xueyi Yao & Natalia Jardon & Jonathan Kominsky & Eva Wittenberg: Remembering times ahead: The effect of linguistic framing on representational momentum in state-change events

137 Yunju Nam & Geon Kim & Gaeun Lim: An Eye-tracking Study on the Presupposition Processing of Korean L2 Learners of German: Focusing on "wieder (again)"

139 Lena Wieland & Ingo Reich: Figurative Meaning Is Recoverable: Idiom Comprehension, Preference, and Processing Constraints in Adult Low-Literacy Readers

161 Doina-Irina Giurgea & Veronica Diveica & Penny M. Pexman & Richard J. Binney: The role of social experience and motivated cognition in the representation of concepts: a behavioral and functional neuroimaging study

163 Lucie Guštarová & Jan Chromý: Immediate Recall, Later Word Recognition, and Information Congruency in Reading and Listening Comprehension

170 Xinyue Jia & Christoph Aurnhammer & Torsten Kai Jachmann & Francesca Delogu & Heiner Drenhaus & Matthew W. Crocker: The Influence of Linearization on Expectation: Evidence from SPR and ERP Studies on Lossy Context Surprisal

175 Vera Kempe & Marta Brzoska & Hajar Benharraf & Neil W. Kirk: The Emergence of Sociolinguistic Competence in Scottish Children: Social Registers Are Acquired Before Regional Dialects

179 Ernesto Guerra & Andrea Helo & Carlos Rojas & Bernardo Riffo: Bridging inference costs in late adulthood: Eye-tracking evidence from third- and fourth-age readers

182 Joshua Hartshorne & Tobias Gerstenberg & Noah Goodman: Good explanations fit prior knowledge

184 Cristian Rivera & Morten H. Christiansen: Comparing natural language statistical learning and human intuition for chunking language

186 Maria Grabovskaya & Anastasia Vyrenkova & Natalia Slioussar: L2 acquisition of verb and noun paradigms: a study on Russian

200 Helene Slaattelid Øya & Jens Roeser & Gary Jones & Mark Torrance: How do you spell "hånd" in English: Does knowing another language affect word retrieval

220 Dorotea Bevivino & Barbara Hemforth & Giuseppina Turco: Priming cooperating prosodic phrasing increases reading times: An eye-tracking study

221 Zhimin Hu & Eduardo Navarrete & Yao Yao: Language and Script Effects on Information Credibility in a Triliteral Context

229 Markéta Ceháková & Jan Chromý: Cloze, Frequency, Surprisal, or Plausibility? A Comparative Analysis of Predictors for Local Ambiguity Resolution

233 Jennifer Keller & Ingo Plag: Discriminative learning of number interpretation of German pseudo-nouns

235 Ricarda Scherer & Robin Lemke & Ingo Reich & Heiner Drenhaus & Lisa Schäfer: Having one or three uncles: equally acceptable. A study about number mismatches in nominal Right-Node-Raising in German

237 Qingyuan Gardner & Vasiliki Chondrogianni & Peng Li & Holly P Branigan: Morphophonological Effects on Morphosyntactic Processing During L2 English Real-time Comprehension

240 Mikuláš Preininger & Filip Smolík & Nikola Paillereau: Early sensitivity to gender morphology in Czech infants

246 Xu Ji & Dawei Jin: Resumption in Anaphoric Dependencies: A Case Study of Mandarin Topicalization

250 Elise Oltrogge & Eun-Kyoung Rosa Lee & Sol Lago: Can planned words trigger interference during real-time sentence production?

255 Fabrizio Luciani & Federico Frau & Paolo Canal & Riccardo Venturini & Luca Bischetti & Valentina Bambini: A key to interpreting late effects in the brain response to metaphors: priming figurative (but not literal) meaning

258 Alaa M. Salem & Daniel Gallagher & Emi Yamada & Shinri Ohta: Two Sites, Two Languages: tDCS and EEG Evidence for Argument-Structure and L1 Feature Transfer

276 Charles Redmon & Aditi Lahiri: The acquisition of noun-verb stress alternation by Bengali learners of English

277 Runchen Liu & Suhas Arehalli: L2 English speakers exhibit native-like garden path difficulty across constructions

285 Giulia Li Calzi & Antje Meyer & Constantijn van der Burght: Lexical stress precedes syllable structure during speech planning – evidence from EEG multivariate pattern analysis

287 Zuzanna Fuchs & Anna Runova: A grammatical animacy agreement feature: evidence from processing in Polish

288 Yufen Wei & Guillaume Thierry: Languages of Power: Metaphorical Grounding of Perceived Power in the Bilingual Mind

296 Yourdanis Sedarous & Savithry Namboodiripad: Resumptive pronouns are grammatical in English

302 Crystal Jemy & Roberto Petrosino & Diogo Almeida: Dissociating sublexical and lexical masked priming effects: Morphological decomposition interacts with prime lexicality.

304 Panagiota Rassia & Natalja S. Peiseler & Torgrim Solstad & Oliver Bott: Complement coercion revisited: Reassessing the psycholinguistic and the information-theoretic approach

309 Işın Tekin & Duygu Özge Sarısoy: Incremental processing of context during metaphor interpretation in preschool children: Evidence from a visual world eye-tracking study

313 Harrison Albert Paff & Alissa Melinger & Sheila Cunningham & Josephine Ross: To thine native self be true: Exploring the link between self, emotion and language

334 Jinbiao Yang: Rethinking Reasoning: When Next-Token Prediction Mimics Thought

354 Samuele Bruzzese & Buhan Guo & Shayne Sloggett: Memory and Focus: How Bound Focus Affects Illusions of Plausibility

371 Erin Buchanan: ManyLanguages: A global network for Big Team Language Science

Friday Morning

(5 September, 10:00–11:30)

3 Jun Lyu: The processing of Chinese reflexives as plain anaphors and intensifiers

14 Ioannis Iliopoulos: Bilinguals’ Neurocognitive Profiles in L1 and L2: N400 vs. P600 Dominance Reflects Divergent Processing of Filler-Gap Dependencies

20 Radim Lacina & Mojmír Dočekal: Sentential negation causes both NPI and NCI illusions in Czech

22 Alexander Kilpatrick & Rikke Bundgaard-Nielsen: Say what you mean: Linguistic vividness and information theory

25 Katerina Stoumpou & Ghada Khattab & Faye Smith: The role of Morphological Skills as a Compensatory Mechanism in adult Developmental Dyslexia

26 Sasha Kenjeeva & Giovanni Cassani & Noortje Venhuizen & Afra Alishahi: Does multimodal pre-activation influence linguistic expectations in LLMs and humans?

28 Ana Bautista & Francesca Branzi & Clara Martin: Does overt production facilitate language prediction in challenging situations only?

30 Anna Gupta & Carsten Eulitz: Processing morphologically complex words: Insights from Russian

35 Daiwen Gong & Aine Ito: The markedness effect on form-based predictions of sound and number: Evidence from a visual-world study

39 Aini Li & Lacey Wade: Tacit knowledge of stylistic variation: Evidence from (ING) perception in native and non-native listeners

41 Camilla Masullo & Beatrice Giustolisi: Do code-switching and sociolinguistic environment modulate the processing of ambiguous pronouns? Insights from Italian-English bilinguals

51 Mizuki Yoshio & Toshimune Kambara: Linguistic conditioning to change the emotional and gustatory meanings of new words

52 Kaiying Kevin Lin: Do Mandarin speakers retain categories for unaccusativity?

53 Sara Møller Østergaard & Bruno Nicenboim: A Corpus of Joint EEG and Self-Paced Reading of Natural Dutch Texts

57 Weijia Hu & Huanhuan Yin & Martin J. Pickering: How do Mandarin Chinese speakers prepare the form and content of their answers in turn-taking conversation?

59 Dominic Schmitz: Polysemy and acoustic duration: Different senses come with different durations

76 Hsin-Ju Wu & Chia-Hsuan Liao: The influence of context on the processing of (a)typical thematic relations in Mandarin Chinese

83 Yiwei Si & Aditi Lahiri & Isabella Fritz: Phonology in morphological priming: Evidence from German complex verbs

87 Katrin Odermann & Renate Delucchi Danhier & Barbara Mertins: Processing of Homonyms in Bilingual Children: A Visual World Eye-Tracking Study

88 Anastazja Rosanoff & Peter Hendrix: A matter of time and meaning: a time-to-event analysis of response times in a semantic categorization task

106 Sophie Slaats & Alexis Hervais-Adelman: Patterns fast and slow: The structure and statistics of language shape high- and low frequency neural signals

109 Giulio Massari & Fanny Meunier & Raphaël Fargier: Features all the way down: visual masking interacts with age of acquisition and iconicity in picture naming

113 Jessie S. Nixon & Erdin Mujezinovic & Ruben van de Vijver: What drives incremental sequence learning?

129 Katja Maquate & Angela Patarroyo & Angelina Ioannidou-Tsiomou & Pia Knoeferle: Age differences in spoken language comprehension: verb-argument and formality-register congruence influence real-time sentence processing

135 Laia Colina Fortuny & Li Kloostra & Johan Bos & Jakub Dotlacil: Semantics in reading-time corpora

138 Tiziana Srdoc & Elena Marx & Anna Viola Sáfrány & Eva Wittenberg: Event construal through social verbs in English, German, and Hungarian: The LISADA corpus

141 Yu-Yin Hsu & Anqi Xu: When Focus Overrides Form: Prosodic Rephrasing in Mandarin complex nominals

153 John Cristian Borges Gambôa & Shaiban Alshaibani & Christopher Allison & Leigh B. Fernandez & Shanley E. M. Allen: Divergence Point Analysis: does it really establish the precise timepoint of divergence?

159 Binger Lu & Julie Boland & Robert J. Hartsuiker: Does language similarity affect second language prediction in discourse comprehension? Evidence from visual-world Eye-tracking

164 Gerakini Douka & Despina Papadopoulou: Relative clause processing and comprehension in Greek: Effects of academic background

167 Jens Roeser & Pablo Aros Munoz & Mark Torrance: "Write here, write now"': Spelling difficulty disrupts parallel planning in sentence production

168 Siddharth Gupta & Alessandro Lopopolo & Milena Rabovsky: Semantic Update as a Predictor of Reading Time: Moving Beyond Word-Level Surprisal

176 Yimin Zhu & Caterina Donati: Transferring islands across languages

177 Xuetong Yuan & Minjae Joh & Ming Xiang: Predicting scalar diversity with crowdsourcing QUD in naturalistic discourse

183 Ting-Wu Lee & Shiao-hui Chan: Action imagination, and verb semantics in Mandarin Chinese influences neural responses beyond somatotopic mapping: An fMRI study

189 Edmundo Kronmuller & Ernesto Guerra: Robust mutual exclusivity in multiparty conversations: Contextual adaptation without speaker-specific effects

206 Alessandro Lopopolo & Milena Rabovsky: Surprisal is Influenced by Syntax and Semantics, but not Equally across Language Models

208 Jon Lapresa Serrano & Marianne Hundt & Fernando Zúñiga: Nominalised adjectives in Basque: experimental evidence from a self-paced reading experiment

210 Jéssica Gomes & João Veríssimo & Dan Parker & Sol Lago: Eyes on delay: Revisiting the timecourse of spoken word recognition in L1 and L2 speakers

216 Kaidi Lõo & Anton Malmi & Benjamin V Tucker: Introducing the Estonian Auditory Lexical Decision database

222 Alice Rees: Aligning to what I don’t say: structural alignment and pragmatic inferencing

223 Yi-ching Su & Antonella Sorace & Ming-Lei Chen: Binding Principle C in Online Processing of Mandarin Cataphoric Pronoun Resolution

230 Raffaella Folli & Juliana Gerard & Heidi Harley & Balthazar Lauzon & Morgan Macleod: Animacy and null objects in English

232 Monika Kučerová & Kateřina Chládková: The impact of multi-accent and L1-accented input on preschoolers' perceptual adaptation to L2 vowels

254 Hoekeon Choi & Haeun Ko & Ha-a-yan Jang & Jonghyun Lee & Sung-Eun Lee: Exploring ERP Components in Emotional Word Processing: Participant Subjectivity and Embodiment

261 Alaa M. Salem & Shinri Ohta: Investigating Cerebellar-Language Network Alterations in Parkinson's Disease Using Open Data

267 Yuko Hijikata & Masumi Ono & Haruka Shimizu & Yuko Hoshino & Yuji Ushiro: Understanding intertextual relations and numerical processing in L2 multiple-text reading: An eye-tracking study

286 Nilanjana Chowdhury & Bidisha Som & Sukumar Nandi: Cognitive Load and Language Dominance: Bilingual Performance in a Dual-Task Paradigm

300 Michaela Svoboda & Natálie Kikoťová & Kateřina Chládková: Cross-Modal Activation in Hearing-Impaired Preschoolers: An fNIRS Study of Speech and Sign Processing

301 Da Thao Anh Ngo & Nino Grillo: Universal parsing biases: Small Clauses drive RC attachment in Vietnamese

316 Pia Schoknecht: Task adaptation in web-based self-paced reading

320 Bálint József Ugrin & Péter Rácz & Ágnes Lukács: Vocabulary Size as Prediction Error: A New Method for Lexical Assessment in Adults

345 Hailin Hao & Zuzanna Fuchs: Effects of Surprisal and Contextual Entropy on L2 and Heritage Language Processing

350 Klára Matiasovitsová & Filip Smolík: Sentence imitation and its relation to working memory and language skills

351 Iza Škrjanec & Irene Elisabeth Winther & Merit Huisman & Vera Demberg & Sybrine Bultena & Stefan L. Frank: Slower reading on interlingual homographs can be a surprisal effect

355 David Pagmar & Asad B. Sayeed: Local context in quantifier scope ambiguity resolution in Swedish

358 Lily Arrom & Samantha Wray: Untangling musical and linguistic processing using low-resolution EEG

370 Jan Chromý & Markéta Ceháková & Michael Ramscar: Reading Tiramisu in Czech and English: Robust Processing Speed Differences in Translation Equivalent Stimuli

Friday Afternoon

(5 September, 15:20–16:50)

18 Lu Li & Jiayi Lu & Jueyao Lin & Changsheng Li & Zhengqin Liu & Cehao Yu & Caicai Zhang: Sleep Patterns and Language Acquisition in Cantonese-Speaking Preschoolers: Preliminary Evidence for the Role of Sleep Regularity

19 Oleksandra Osypenko & Aina Casaponsa & Silke Brandt: The (Non-)Effect of Grammatical Gender on Early Perception: ERP Study in Simultaneous Bilinguals

50 Angèle Brunellière & Laurent Ott & Solène Kalénine & Martin Pickering: Interacting with someone shapes prediction in spoken-language comprehension

55 Dongpeng PAN & Kilian Seeber: The effect of visual cuing during simultaneous interpreting

56 Naomi Nota & Muzna Shehzad & Ruth Corps & Martin Pickering & Graham Naylor & Lauren Hadley: The effect of speech rate on two prediction stages in older adults with and without hearing loss

66 Sophie Repp & Heiko Seeliger & Judith Schlenter & Petra B. Schumacher: How information structure, prosodic prominence, and speech act affect reference resolution: Evidence from eye-tracking

84 Kohei Haneda & Anja Schüppert & Roel Jonkers: Visual Cues Not Only Facilitate Online Sentence Comprehension But Also License Ellipsis Resolution: A Self-Paced Reading Study of English Verb Phrase Ellipsis

85 Michael Vrazitulis: The Role of Task Framing and Context Source in Scalar Implicature Detection

91 Nan Kang & Satoru Saito: Consistency and Frequency Effects in Japanese Kanji Nonword Reading by L1-Chinese Speakers

101 Nitzan Trainin & Einat Shetreet: The effects of perceived cooperativeness of lexical alignment, memory, and social judgments

112 Sarah Michel & Céline Pozniak & Saveria Colonna: The Mid-dot in Gender-Inclusive French: A Reading Study

114 Sara Božić & Dušica Filipović Đurđević: The neglected role of sensorimotor information in the processing and representation of polysemous words

120 Ana Zarwanitzer & Santiago Estremero & Gala Esperanza Coronas & Carlos Gelormini-Lezama: Inclusive language, then and now: a self-paced reading experiment in Argentina

147 Emma Kious & Gabriel Thiberge & Anne Abeillé & Céline Pozniak & Heather Burnett: An Emerging Non-Binary Stereotype? An Experimental Assessment of the NB-ness of French Nouns

166 S Shalu & R.Muralikrishnan & Kamal Kumar Choudhary: Does the verb type modulate the ERPs for Thematic Reversal Anomalies? The case of Subject and Object experiencer verbs in Malayalam.

169 Emma Libersky & Kimberly Crespo & Margarita Kaushanskaya: Speech disfluencies and implicit word learning: Fluency shapes preference, not performance

173 Julia Muschalik: Velocity is key: Morphological structure affects planning and execution stages in typing

174 Vera Yunxiao Xia & Johanne Paradis & Juhani Järvikivi: The role of expectedness in L1 and L2 ditransitive prediction in Mandarin-English late bilinguals and heritage bilinguals

193 Oliwia Iwan & Eva Wittenberg: Compositional Parsing in Adjective-Noun Phrases: The Role of Adjective Semantics

194 Panpan Bi & Cheng Wang & Chen Feng: The Impact of Semantic Distance on Multiple Phonological Activation in Chinese Speech Production: Evidence from a Picture-Word Interference Study

204 Harriet Yates & Corien Bary & Bob van Tiel & Peter de Swart: Evidentiality and Speaker Commitment: An fEMG Study

205 Janika Stille & Anne Wienholz & Annika Herrmann & Ivo Weber & Barbara Hänel-Faulhaber: Sign language processing in deaf early signing children – an ERP study

207 Wonil Chung & Keonwoo Koo & Myung-Kwan Park: Focus Shifts in Contextual and Lexical Cue Interactions in GPT Models

211 Ryoko Uetomi & Leah Roberts & Heather Marsden: Online cataphoric pronoun resolution in L1-, L2- and L3-Mandarin: the Maze task

212 Opangienla Kechu & Bidisha Som: Not Native, Yet Dominant: The Role of Language Context and Social Value in Multilingual Language Processing

217 Thomas Lieber & Giovanni Cassani & Emmanuel Keuleers & Peter Hendrix: Exploring semantic priming effects using piece-wise additive mixed models

218 Juan Haro & Daniel Huete-Pérez & Miguel Ángel Pérez-Sánchez & José Antonio Hinojosa & Pilar Ferré: Characterising the affective content of sentences and its role in reading and memory

224 Yvonne Portele & Sebastian Walter: Pronoun interpretation in German speech reports

231 Lion Oks & Francesca Foppolo & Carlo Cecchetto & caterina donati: Parsing strategies in Hebrew and Italian Relative Clauses: Shall I Avoid Gaps?

241 Daria Antropova & Natalia Slioussar & Elizaveta Galperina & Olga Kruchinina: Grammatical gender, number and case in processing: experimental studies on Russian

242 Sarah Cameron & Natalia Kartushina & Björn Lundquist & Sendy Caffarra: ERPs reveal differential processing of three types of gender violations in Norwegian

245 Agnieszka E. Konopka & Evita Ahmed Hashmi & Martina Italia & Keir Lawley & Joost Rommers & Brian Mathias: Prediction Updating During Novel Word Learning: Evidence from Cerebellar TMS

247 Bohyun Tak & Jihun Im & Ha-a-yan Jang & Jungmin Moon & Sung-Eun Lee: Neural Decoding of Pragmatic Inferential Processing in First and Second Language

260 Joshua Hartshorne: World knowledge without world knowledge: Winograd meets the Jabberwocky

264 Hening Wang & Jia Ren & Michael Franke: Interpreting Plural Predication in Visual Contexts: Cover-Based Resolution of NP Structures

268 Inbal Kuperwasser & Einat Shetreet: Processing of novel metaphors in an intergroup context

274 Fabian Istrate & Laia Mayol & Gabriela Bîlbîie & Barbara Hemforth: The role of information structure for subject position: evidence from Romance languages

280 Teresa Quesada & Jacopo Torregrossa & Cristóbal Lozano: The role of distance on pronoun resolution: Evidence for a two-stage model

282 Chiara Battaglini & Federico Frau & Veronica Mangiaterra & Luca Bischetti & Paolo Canal & Valentina Bambini: When girls are pearls, are they pleasant or are they beautiful? Capturing variation in metaphor interpretation via Intersubject Representational Dissimilarity

289 Ingmar Brilmayer & Petra B. Schumacher: Referential Resolution in Naturalistic Contexts: Audio-Visual Integration in the N400/P300 Window

291 Noelia Ayelén Stetie & Gabriela Mariel Zunino: Blame it on the verb: Implicit causality verbs and its incidence in relative clause attachment

292 Ondřej Drobil & Jan Chromý: The Effect of Adjective Position on Information Recall in Czech

303 Gabriel P. Moya & Ernesto Guerra: Irony Processing in Reading: Eye-Tracking Evidence on the Predictive Role of Mentalizing and Vocabulary

307 Martina Dvořáková & Natálie Kikoťová & Josef Urbanec & Antonia Goetz & Kateřina Chládková: Detecting foreign rhythm in native-language speech at birth

318 Vera Heyer: Seeing the Little Things: Context Effects on the Processing of Inflectional Affixes on Novel Words in Late Second Language Learners

319 David Pagmar & Yuval Marton & Asad B. Sayeed: Animacy cues and word order in language acquisition and dialogue corpora

321 Buhan Guo & Andrea Santi & Shayne Sloggett & Giuseppina Turco & Sven Mattys & Nino Grillo: Reanalysis as Last Resort: Coercion in Tense Harmony Violations

323 Liliia Terekhina: The three way relationship among sleep quality, bilingualism and cognition

325 Nikonova Yana & Alexeeva Svetlana: Letter computation in Russian: further exploration

327 Hannah S. Rognan & Shohini Bhattasali: Modelling Temporal Connective Processing with LLMs: Insights from English & Norwegian

329 Robin Lemke: Investigating crosslinguistic processing constraints on preposition omission under ellipsis

337 Cui Ding & Shan Gao & Ethan Wilcox & Lena Ann Jäger: When Half a Word Is Enough: How Lexical Expectations Modulate Visual Uncertainty in English and Chinese

343 Spencer Caplan & Douglas Richard Guilbeault & Charles Yang: A unified threshold for individual learning and convergence across populations

347 Shiyu Li & Jordan Gallant & Gary Libben & Gonia Jarema: Chinese Compound Word Production in Typing to Dictation

348 Miyuki Rachel Oshima & Yasunori Morishima: Dilemmas and Language: Observing the decision-making process between first and second language using fNIRS

349 Miriam Schulz & Masato Nakamura & Matthew W. Crocker: Earlier and stronger effects of prediction through production

359 Fang Yang & Holly P. Branigan & Martin J. Pickering: Relative Activation of Competing Event Roles in Mandarin Discourse Development

366 Joana Miguel & Catarina Barbosa & Susana Cardoso & João Veríssimo: Learning morphological rules across typical and atypical development

Saturday Morning

(6 September, 10:00–11:30)

12 Nikolaos Ntagkas & Despina Papadopoulou: Morphological processing in Modern Greek: A form-then-meaning, dual-route account

33 Alexander Kilpatrick & Rikke Bundgaard-Nielsen: Language Processing Insights from Average Phonemic Bigram Surprisal

43 Emma Corbeau & Céline Pozniak & Heather Burnett: Experimental approach to advice-giving in French L1 and Japanese L1

47 Julia Chauvet & Andrea E. Martin & Ardi Roelofs & Frank H. Guenther & Antje S. Meyer: Bridging models of linguistic planning and speech production: The case of lexical stress in English

48 Nevena Klobucar & Esther Rinke & Raffaella Folli & Chrisina Sevdali & Juliana Gerard: Online and offline pronoun comprehension by German-speaking children and adults

63 Diane Mézière & Titus von der Malsburg: Predicting Reading Comprehension from Eye-Tracking Measures with Random Forests

69 Shiyu He & Dagmar Divjak & Petar Milin: The Cost of L2 Fluency: Eye-Movement Evidence for Reduced L1 Reading Automaticity

71 Kamila Kuishibekova & Valentina Apresyan: What Word-Guessing Reveals About Your Brain: Patterns of Lexical Storage and Processing

72 Yoko Nakano & Chunxia Hu & Atsushi Yuhaku: Interference Triggered by Syntactic and Semantic Similarities in L1 and L2 Japanese

79 Elena Marx & Zofia Kordas & Hannah Grobauer & Eva Wittenberg: The causal chain in English conditionals depends on event structure

100 Katharina Spalek & Merel CJ Scholman & Vera Demberg: The effect of the focus particle `only' on discourse expectations and discourse marking

121 Jiaxuan Li & Kayla Keyue Chen & Anne Wang & Yuhan Shen & Yijia Luo & Richard Futrell & Wing-Yee Chow: The good-enough listener: A visual world paradigm reveals the interaction between prediction and bottom-up input

128 Danning Sun & Aine Ito: Cue weighting in prediction: context and classifier effects in English-Chinese bilinguals

136 Lara Kelly-Iturriaga & Mitsuhiko Ota & Martin Pickering: The effect of language distance on bilingual lexical processing

143 Annett B. Jorschick & Yuan Zhou: Not All Vowels Are Learned Alike: The Limits of L2 Experience in Cross-Linguistic Vowel Perception

148 Valeria Galimberti & Beatrice Giustolisi & caterina donati & Francesca Foppolo: Formal and semantic cues in gender assignment to novel words in Italian

149 Rupali Limachya & Steven Frisson & Federica Degno & Simon P. Liversedge & Kevin B. Paterson & Ascensión Pagán: Investigating prediction error cost during natural reading in young and older adults: Evidence from eye movements and fixation-related potentials

150 Salma Gilani & Katrien Segaert & Linda Wheeldon & Evelien Heyselaar: Structure dependent differences in the persistence of syntactic priming

151 Inês Cardoso Ferreira & Leona Polyanskaya & Mikhail Ordin: Tracking Affixation: ERP and Behavioural Insights into the Suffixing Bias

160 Sara Košutar & Judith Schlenter & Natalia Mitrofanova & Serge Minor: Cross-linguistic influence in bilingual minds: A Visual World eye-tracking study on grammatical aspect processing in Croatian-German and Croatian-Italian children

171 Cassandra L Jacobs & Loïc Grobol & Alvin Tsang & Ryan J. Hubbard: Semi-automatic selection of semantic substitutes in sentence comprehension stimuli

180 Chi Hou Pau & Grant Goodall: The absence of D-linking effects in Cantonese Wh-islands

185 Junghwan Maeng & Hyun Kyung Rachel Lee & Samuel Sui Lung Sze & Joey Zhiyin Zhu & Yoonsang Song: EEG time-frequency analysis of syntactic unification in Cantonese and English

187 Mikihiro Tanaka: The Production of Coercion in Japanese: Evidence from Priming

195 Bram De Keersmaecker & Rob Hartsuiker & Aurélie Pistono: The role of attentional resources on errors and disfluency in speech production across different degrees of speech rate restriction.

197 Kyla McConnell & Berit Reise & Antje S. Meyer: The pervasive role of linguistic knowledge in verbal fluency tests: How individual differences in language skills shape the mental lexicon

199 Anna Viola Sáfrány & Anna Kamenetski & Tiziana Srdoc & Attila Balla & Eva Wittenberg: Marking Aspect in Social Events: The Hungarian Verbal Prefix Meg- Increases Perceived Mutuality

202 Yun Feng & Shinyi Li & Ming Xiang & Yao Yao: Gender stereotype in auditory sentence processing: effects of talker and listener gender

213 Kate Stone & Milena Rabovsky & Henning Holle: Immediate sensitivity to thematic role constraints in a lexical decision task

219 Chia-Hsuan Liao: When “Mayor apologized citizens” becomes acceptable: ERP investigations on the transitive use of intransitive verbs in Mandarin

225 Yao-Ying Lai & Maria Pinango & Hiromu Sakai & Michiru Makuuchi: Task-dependent neural modulation during sentential meaning computation

236 Fengyun Hou & Alexander Anderson & Nina Kazanina: Automatic processing of relational structure in language: A frequency-tagging EEG study of Chinese compounds

244 Fenja Plate-Güneş & Jana-Elina Jordan & Markus Bader & Sascha Bargmann & Gert Webelhuth: Influence of the syntactic function on the production of negated sentences in German

256 Kristof Strijkers: Language in the Dyad: Linking Linguistic and Neural Alignment.

259 Owen Kapelle & Monique Flecken & Conrado Bosman & Paul Boersma: Linguistic and Non-Linguistic Cues during Colour Discrimination May Function as Differential Cues for Expectations: An Approach Using ERP and Oscillatory Analyses

262 Anna Cameron & Alexandra Cleland & Agnieszka E. Konopka: Sound-Symbolism Effects in Novel Word Generation

266 Angelina Ioannidou-Tsiomou & Katja Maquate & Sarah Creel & Pia Knoeferle: Mozart’s Concert in da Hood: A VWP Eye-tracking study on the effects of music as formality-context on online comprehension of register nuances.

278 Regina Hert & Barbara Köpke & E Jamieson & Vera Heyer & Monika S. Schmid & Holger Hopp: T(w)o Gender(s) or not t(w)o Gender(s) – Gender Assignment in German in English-German and French-German Bilinguals

281 Linh Pham & Zuzanna Fuchs & Elsi Kaiser: Dynamic language transfer in bilingualism: How L1 Vietnamese L2 English speakers process filler-gap dependencies in English

294 Jens Schmidtke: Exposure Frequency and Native Language Interference in Early Second Language Auditory Word Learning

308 Gabriela Mariel Zunino & Carmela Tomé Cornejo & Gloria Gagliardi & Raquel Freitag & Noelia Ayelén Stetie & Sofía Tzinavos Muñoz & Emanuele Miola: Female butchers meet male babysitters: a multilingual maze on the effects of gender stereotypes and grammatical gender during sentence processing

312 Nino Grillo & Buhan Guo & Keir Moulton & Shayne Sloggett: Composition-sensitive predictions: Incremental Processing of Experiential Perfects

314 Andrés Contreras & Anita Tobar-Henríquez & Ernesto Guerra & Edmundo Kronmüller: Dissociating Speaker-Specific Effects on Referential Precedent Interpretation

317 Yixia Wang & Peter Hendrix & Emmanuel Keuleers: Network Properties of Chinese Characters and their Effect on Processing

322 Eva Pospíšilová & Jan Chromý: Sentence Processing and Memory: Immediate Recall of Information from Adjectives with Different Syntactic Status in Czech Adult Speakers

336 Robin Lemke: Predicting ellipsis usage with a game-theoretic model informed by production data

338 Valerio Pepe & Joshua Hartshorne: A large-scale investigation of pronoun interpretation biases in LLMs

340 Ebony Pearson & Van Rynald T. Liceralde & Duane Watson: Listeners without the pin-pen merger find 'pin' and 'pen' ambiguous: Evidence for a parallel activation account of dialect processing

341 Sumin Jo & Yunju Nam & Jaewon Choi: The role of emotional valence of head-NP in the Korean relative clause attachment

344 Hailin Hao & Elsi Kaiser: Revisiting Uniform Information Density and *that*-reduction in English Complement Clauses

353 Hannah Krueger & Samantha Wray: Effects of written but unpronounced morphemes on auditory word recognition

361 Xiao Ke & Silke Brandt & Katherine Messenger: The Influence of Conceptual and Syntactic Interaction on Syntactic Structure Selection in Chinese Speakers' Language Production

363 Paula Luegi & Isabel Falé & Jéssica Gomes & Amália Mendes: Processing (dis)continuous explicit and implicit discourse relations in European Portuguese

364 Filip Smolík & Maroš Filip: Learning articles in an artificial minilanguage: error-based learning or propose-but-verify

365 Fang Yang & Holly P. Branigan & Martin J. Pickering: Communicative Goals Influence Conceptual but Not Structural Alignment

368 Onur Keleş & Nazik Dinctopal Deniz: Task effects in good-enough parsing in Turkish: Human and LM comprehension of thematic roles

369 Onur Keleş & Zeynep Irem Bayrakli Keles & Nazik Dinctopal Deniz: How good-enough is L2 sentence comprehension? Processing of thematic role (reversal)s in L2 Turkish

Satellite Workshop I: Big Data in Psycholinguistics

The workshop is financed from the project Beyond Security: Role of Conflict in Resilience-Building (CoRe; Reg. Nr. CZ.02.01.01/00/22_008/0004595). The participation in both workshops is free of charge.

Keynote 1: Leveraging big data to map lexical-semantic space

Penny Pexman

Western University

London, Canada.

The development of large-scale word norms and behavioural megastudies has enabled consideration of multiple lexical and semantic dimensions and exploration of the possibility that these dimensions have simultaneous and interacting effects on behaviour. These tools have also supported evaluation of multiple representation theories, which posit that word meanings are represented via a combination of properties derived from sensorimotor, affective, social, cognitive, linguistic, and other experiences. I will describe a series of studies in which my colleagues and I have tested predictions of these theories for acquisition and processing of word meanings, exploring how these multiple dimensions might be mapped in semantic space.

Keynote 2:
Designing Big Science: Lessons from a 25,000-participant, 30-language semantic priming study

Erin Buchanan
Harrisburg University
Pennsylvania, USA.

This talk shares insights from the Semantic Priming Across Many Languages (SPAML) project, one of the largest cross-linguistic studies in cognitive science to date. I will walk through the process of designing and managing this global big team science effort, highlighting challenges and solutions in cultural adaptation, translation, IRB coordination across 126 labs, data harmonization, and adaptive sampling. The talk will offer practical lessons for others building multilingual, high-volume, open science collaborations.

Keynote 3: Understanding measures of word occurrence in psycholinguistics through a network of static and dynamic actors

Emmanuel Keuleers

Tilburg University

Tilburg, Netherlands.

I propose a framework for understanding language as a network structure that is theoretically rooted in a usage-based approach to language and introduces formalizations from network science. While a traditional way of modeling language as a network, is to represent language users as nodes while edges represent communicative interactions, the framework introduces a crucial distinction between two types of nodes:

1. Dynamic actors: Entities capable of processing, storing, and producing language autonomously (e.g., humans, advanced AI systems).

2. Static actors: Language artifacts that store information but cannot process it independently (e.g., books, databases).

Interpreting different measures of word occurrence in psycholinguistics (word frequency, contextual diversity, word prevalence) in such a framework shows misalignment with what researchers think they are measuring. I explore the consequences of this misalignment and propose some alternatives. Taking a wider view, I will discuss how changes in the network composition and connectivity can fundamentally change our operationalizations of participants' language environment.

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Satellite Workshop II: Multilingual Experience: From Individuals to Society

The workshop is financed from the project Beyond Security: Role of Conflict in Resilience-Building (CoRe; Reg. Nr. CZ.02.01.01/00/22_008/0004595). The participation in both workshops is free of charge.

CoRe | Beyond Security
Role of Conflict in Resilience Building

Opening Talk:
Researching and communicating multilingualism

Kateřina Chládková

Faculty of Arts, Charles University

Prague, Czechia.

The way people speak and the language they use are crucial parts of their identity. Language helps individuals identify with a community, but it also signals when they stand outside of another. In today's world, where it is easier than ever to connect across countries and cultures, one's accent or language may sometimes hinder social integration. Rather than attempting to eliminate accents (and thus part of a person's identity), the way forward is to teach societies and communities to value and promote linguistic variation. I will briefly present some of the studies from our current project, in which we investigate language development in multilingual and multi-accent settings, the origins of accent- and language-based social biases as well as the factors that modulate them. The ultimate goal of the project is to put this knowledge into practice so that schools, social services, healthcare providers, and the general public become more aware of the benefits that multilingualism brings and learn to embrace it. The purpose of this AMLaP satellite workshop is to learn about successful research and science-transfer approaches to multilingualism across cultures and societal settings, which the two invited talks will showcase from different perspectives.

Keynote 1:
Using public engagement to 'normalize' multilingualism

Sharon Unsworth

Radboud University

Nijmegen, Netherlands.

Most of the world's children are growing up with more than one language and yet, many educational systems are designed with monolinguals in mind. Teachers and parents regularly lack understanding of what 'normal' multilingual acquisition looks like, and their views are often informed by popular misconceptions and personal beliefs. In this talk, I'll consider how we can use public engagement to 'normalize' multilingualism. In the first part, I'll present a (very!) brief overview of some of our recent work on cross-linguistic influence, lexical processing and language distance and consider what this means for expectations concerning multilingual language development. In the second part, I'll showcase our public engagement initiative Kletskoppen ('Chatterboxes'). Kletskoppen organises activities for bilingual and monolingual children promoting language, language science and multilingualism. I'll present evidence from impact evaluations demonstrating how these activities can change children's attitudes to multilingualism, as well as providing some more general tips on engaging the public with language science.

Keynote 2: Born into a multilingual society: Early language acquisition in Ghana

Natalie Boll-Avetisyan
University of Potsdam
Potsdam, Germany.

Most infants worldwide grow up in multilingual societies in the Global South. Yet, for over half a century, research on infant language acquisition has focused almost exclusively on babies raised in monolingual Western contexts, acquiring one or at most two languages. This lack of diversity in study populations poses significant challenges for developing comprehensive theories of language acquisition and multilingualism. In this keynote, I will present research addressing this gap by focusing on infants growing up in Ghana, a highly multilingual society. I will first present survey data documenting the degree of multilingualism in Ghanaian infants' language input. Next, I will present experimental work on multilingual infants' speech processing, which includes a study of word recognition in code-mixed sentences and a study on the use of speech segmentation cues. I will demonstrate how we adapted classical psycholinguistic methods, using mobile equipment, to the West African context, and I will highlight where our field needs to become more flexible in embracing modifications to standard methodologies. I will conclude by discussing how our research findings relate to current theories of language acquisition.

Main Event: Keynote Talks

Keynote 1: Towards personalized models of processing difficulty: Modelling individual differences in working memory capacity and background knowledge

Vera Demberg

Saarland University

Saarbrücken, Germany.

Today's LLMs are able to model language with unprecedented accuracy. But are they also useful as tools for cognitive modelling? In my talk, I will argue that LLMs enable us to test more precise hypotheses about human language processing. Specifically, I will show that the lossy context surprisal model, which incorporates memory constraints, can be used to model the effects of differences in human working memory capacity on eye-movements during reading, and I will demonstrate that surprisal models that reflect the specific experience and background knowledge of a human can better predict their reading behaviour. Modelling the variability in human comprehenders cannot only help us gain a deeper understanding of cognitive mechanisms by identifying the effect of specific cognitive capacities, but also enable educational and clinical applications.

Keynote 2:
The Bilingual Multiverse: Real-time effects of language of operation on categorical perception and abstract conceptualisation

Guillaume Thierry
Bangor University
Bangor, UK.

This talk examines neuroscientific evidence showing that language is deeply entangled with perception, emotion, and conceptual thought, challenging the modular view of language as an isolated cognitive system. Across diverse linguistic and cultural contexts, the brain constructs meaning in ways that reflect not a single universal model, but a semantic multiverse: a plurality of coexisting conceptual frameworks grounded in bodily experience and shaped by linguistic structure.

Pre-attentive brain responses reveal that language-specific terminology influences early perceptual processes—for example, modulating sensitivity to colour distinctions (Thierry et al., 2009). Speakers of different languages also experience language-dependent categorical perception for object shapes, offering strong evidence for a double Whorfian dissociation (Casaponsa et al., 2024). Abstract domains such as time are flexibly encoded: Chinese-English bilinguals shift between horizontal and vertical spatial metaphors depending on the active language, indicating that even the flow of time is linguistically and bodily grounded (Li et al., 2019; Li et al., 2023).

Emotion and morality, too, are shaped by language. Reading in a second language reduces emotional resonance (Wu & Thierry, 2012, Jończyk et al., 216, 2024) and inhibits native-language lexical access under negative affect (Zhang et al., 2023). Strikingly, individuals are more likely to accept foreign cultural norms when they are expressed in a second language, pointing to the powerful role of language context in moral and cultural judgment (Hu et al., 2025).

Together, these findings suggest that language is not merely a tool for expressing thought but a mechanism for shaping it—neurologically, emotionally, and socially. The semantic multiverse is thus not a metaphorical stunt, but an empirical reality: the human brain, far from being modular, operates as a massively interconnected network, forming meaning through situated, embodied interactions with the world.

Keynote 3:
What is language processing anyway?

Michael Ramscar

University of Tübingen

Tübingen, Germany.

Language science has traditionally assumed that language processing (somehow) revolves around an inventory of discrete form elements associated with discrete meanings, and the mechanisms that build them into larger signals, such that research involves identifying and classifying these elements, and describing the inductive processes of composition and decomposition they support. By contrast, formal theories of communication have adopted discriminative (deductive) models based on systems: information theory does not treat “information” as being a property of individual transmitted signals, but rather as a function of all of the symbols that could have potentially been sent in the system.

In this talk, I will describe how languages are shaped by information theoretic principles (and the constraints imposed by the fact they must be learned), introduce some of the socially evolved structures that support human communicative processes (and language learning), and explain why these processes are best thought of as serving to reduce communicative uncertainty within a shared, probabilistic system, rather than supporting the transfer of meaning.

In doing so, I will focus on some ubiquitous aspects of human communication that either appear redundant (grammatical gender) or else problematic and random (personal names) from the traditional perspective. I will show how, across languages, an information theoretic perspective helps reveal a remarkable amount of functional structure in these domains, and describe how the same communicative structures can be seen across linguistic domains at different levels of description. While these abstract statistical structures can seem baffling from a traditional perspective, they are both consistent with, and predicted by, a discriminative, information theoretic approach to language processing.

Keynote 4:
Towards a multimodal view on the neurobiology of language

Linda Drijvers

Radboud University

Nijmegen, Netherlands.

Face-to-face communication involves auditory signals, such as speech, and visual signals, such as visual speech and hand gestures. Despite the abundance of visual expressions in language, most models and theories on the neurobiology of language are based on characteristics of (clear) speech and text, and they rarely consider multimodal signals. In this talk, I will argue that we need a multimodal view on the neurobiology of language, and that these visual signals are often taken on board immediately by listeners in creating and shaping an interpretation of the linguistic input. This talk will center around the question of how we, as language users, integrate auditory and visual signals into a coherent message, how this is orchestrated within and between brains, and how we do this in both clear and adverse listening conditions.

Oral Presentations

Session 1

Age-related effects of language proficiency and use on language switching

Eunice G. Fernandes[1], Yanina Prystauka[2], Foyzul Rahman[3], Helene Slaattelid Øya[4], Allison Wetterlin[5], Katrien Segaert[3], Linda Wheeldon[5]

[1] School of Psychology, University of Minho; [2]Department of Linguistic, Literary and Aesthetic Studies, University of Bergen; [3]School of Psychology, University of Birmingham, [4]Department of Psychology, Nottingham Trent University, [5]Department of Foreign Languages and Translation, University of Agder

eunice.fernandes@psi.uminho.pt

Background: L2 proficiency and use have been claimed to modulate bilingual performance in language switching tasks: High proficient L2 speakers show symmetrical switching into L1 and L2 [1] and frequent switchers show smaller switching costs [2]. However, these effects were reported for young adults, and studies with older bilinguals have not investigated such modulations, though they have shown larger switching and mixing costs [3], and reduced reversed dominance [4], in older compared to younger bilinguals. We tested young and older bilinguals on a language switching picture naming task and measured both switching and mixing costs, which we analysed as a function of individual differences on participants' L2 proficiency and frequency of language switching.

Method: *Participants*: We tested 138 older (OAs; 80F; Age M=68.47, SD=5.74) and 80 younger (YAs; 57F; Age M=23.06, SD=3.20) adult Norwegian(L1)-English(L2) bilinguals. *Materials/ Design*: There were two single language blocks at the beginning and end of the experiment, and four intervening mixed language blocks. 24 experimental pictures with non-cognate names appeared in every block. Half of these were to be named in L1 and half in L2. In mixed blocks, half of the trials were 'stay' (same language) and half were 'switch' trials (other language), relative to the preceding trial. Participants had to name the picture, as quickly as possible, in the language cued by a coloured frame. We measured reaction times (RTs) from the onset of the picture to speech onset. L2 proficiency was measured by a vocabulary task (% of correct answers) and Switching frequency was self-rated on a scale from 1 to 4, where 1=rarely, 2=monthly, 3=weekly and 4=daily (collapsed in Low (1,2) and High (3,4) levels of switching frequency).

Analyses, Results and Discussion: We fitted linear mixed models (using *blmer* function in *R*) to RTs on correct trials (91.87% of data), after z-scoring and removing RTs<250ms and beyond 2SD from each participant mean (5.66% and 6.32% of data for switching and mixing costs, respectively). Switching costs compared 'stay' to 'switch' trials in mixed blocks, and mixing costs compared 'stay' trials on mixed to 'stay' trials on single language blocks. L2 proficiency and Switching Frequency were added as covariates.

The switching data (Figure1; Table1) replicated findings for reversed dominance (slower RTs in L1 than L2) and symmetric costs (in L1 and L2) [1]. In addition, we found that switching was modulated by L2proficiency in older adults only: Increasing proficiency led to faster RT on switching to L2 and slower RT on switching to L1, which is compatible with stronger reliance on control (and L1 inhibition) for high proficient bilinguals from single/ dual language contexts [5]. The mixing data (Figure2; Table2) showed larger costs for OAs [3]. In addition, mixing was modulated by Switching Frequency: Frequent switchers showed reduced mixing costs in L1 and more reversed dominance in single language blocks, compared to rare switchers, which reflects modulations of L1 inhibition with varying language control demands associated with the experiment and the predominant conversational contexts of bilinguals [5]. These findings highlight the importance of testing across the age span to better understand the mechanisms of language control.

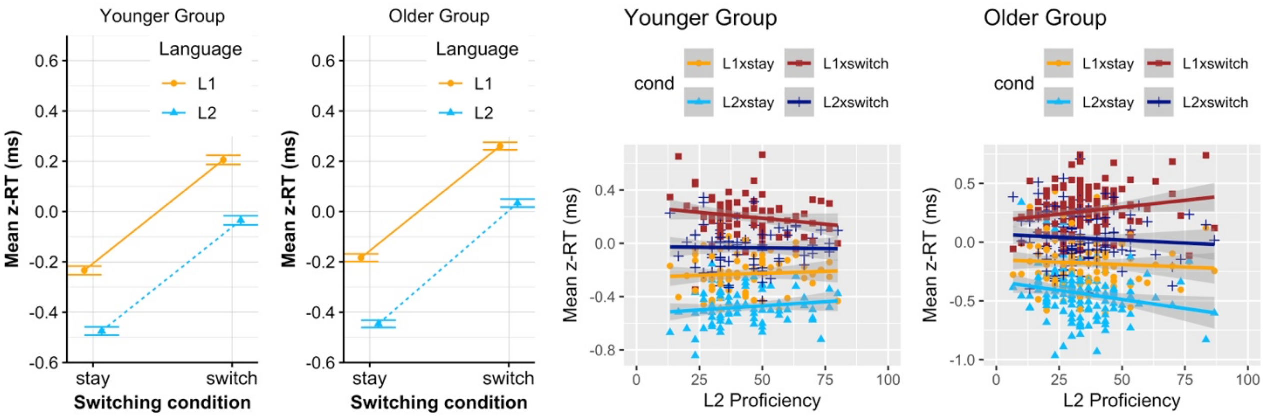


Figure 1. (z-scored) RTs on mixed language blocks

Table 1. Model output (only sig. effects) for the switching costs. (Zscored-) RT to target				
Predictors	Est.	SE	t	p
(Intercept)	-0.10	0.03	-2.83	<.01
trialType [stay, -0.48; switch, 0.52]	0.47	0.01	36.47	<.01
Language [L1, -0.5; L2, 0.5]	-0.24	0.02	-14.53	<.01
trialType:ageGroup:L2Proficiency	0.00	0.00	-2.26	0.02

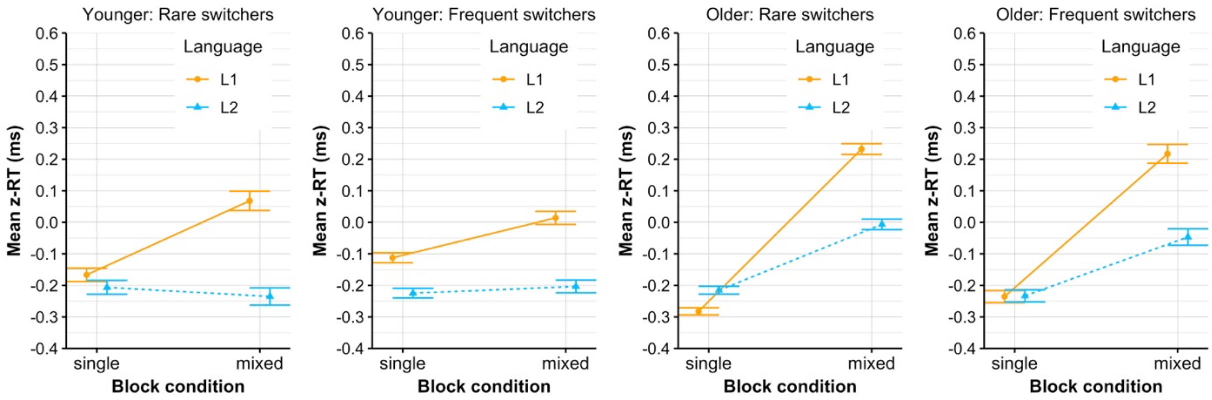


Figure 2. (z-scored) RTs on stay trials in single and mixed language blocks

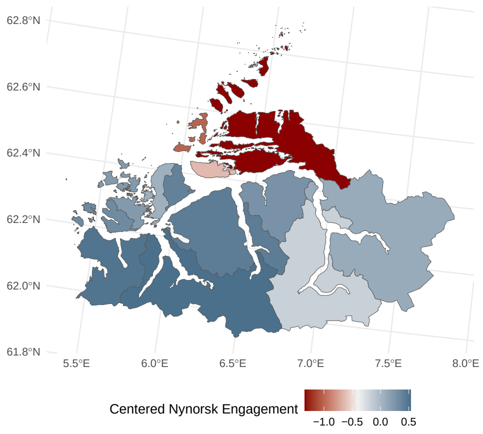
Table 2. Model output (only sig. effects) for the mixing costs. (Zscored-) RT to target				
Predictors	Est.	SE	t	p
(Intercept)	-0.11	0.04	-3.11	0.00
mixing [mixed, -0.66; single, 0.33]	-0.24	0.02	-14.06	<1e-04
Language [L1, -0.49; L2, 0.51]	-0.08	0.02	-4.78	<1e-04
mixing:Language	0.25	0.02	13.17	<1e-04
mixing:ageGroup	0.25	0.04	6.88	<1e-04
Language:ageGroup	-0.09	0.04	-2.48	0.01
mixing:Language:SwitchFreq [low, -0.58; high, 0.42]	0.09	0.04	2.31	0.02
mixing:Language:ageGroup	-0.10	0.04	-2.43	0.02

[1] Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *JML*, 50(4), 491–511; [2] Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *J. International Neuropsychological Society,* 17(4), 682–691; [3] de Bruin, A., Kressel, H., Hemmings, D. A. (2023). A comparison of language control while switching within versus between languages in younger and older adults. *Scien Rep*, 13(1):16740; [4] Stasenko, A.,Kleinman, D., & Gollan, T. H. (2021). Older bilinguals reverse language dominance less than younger bilinguals: Evidence for the inhibitory deficit hypothesis. *Psychology and Aging, 36*(7), 806–821; [5] Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cog Psych*, 25(5), 515–530

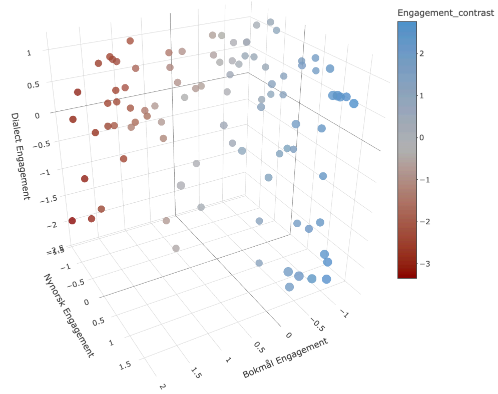
Processing multilectal grammatical microvariation: Mapping individual differences with self-paced reading

Jade Sandstedt[1], Hjalmar T. Eiksund[1], Mizuki Tanigawa[2]
[1] Volda University College; [2] The University of Tokyo
jade.jorgen.sandstedt@hivolda.no

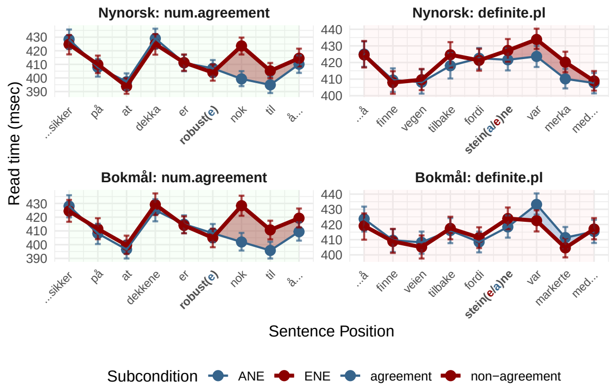
Background: A growing body of evidence suggests that speakers are highly sensitive to microvariation in their language/s, using cues such as gender, ethnicity, dialect, and social context to shape real-time comprehension [1]. This sensitivity extends to multilectalism, where closely related varieties – spoken and written – are processed differently given sufficient engagement and exposure [2, 3]. Norways dual Norwegian written standards, Bokmål and Nynorsk, offer a uniquely controlled setting for investigating multilectal language processing. While mutually intelligible, the two written standards exhibit systematic lexical and grammatical differences, enabling precise investigation of grammatical microvariation during sentence processing. Although formally equal in status, Bokmål functions as the de facto majority variety, with widespread use by most people and institutions nationwide. Nynorsk, by contrast, is a minority standard with strong regional support in parts of western Norway. This sociolinguistic landscape provides an ideal context for examining how inter- and intra-individual variation in language engagement and exposure shapes the processing of fine-grained grammatical contrasts.
Method: Our study was conducted in western Norway, in the Sunnmøre dialect region – an area characterized by significant variation in Nynorsk/Bokmål engagement (Figure 1ab). The study employed an online self-paced reading task across two separate Nynorsk (n = 116) and Bokmål (n = 105) counterbalanced sessions (minimum one-week interval), followed by a language social background questionnaire measuring degrees of Nynorsk, Bokmål, and Sunnmøre dialect engagement, exposure, and proficiency. Stimuli included non-contrastive (plural predicate number agreement) and contrastive conditions (e.g., Nynorsk- and Bokmål-specific definite allomorphy) to isolate how identical inflectional forms are interpreted differently depending on the language mode.
Results: At the group level, Nynorsk/Bokmål contrastive (definite) sentences elicited weaker and more variable reading-time effects than the non-contrastive control condition (Figure 1c), suggesting that exposure to Nynorsk/Bokmål grammatical variation may attenuate processing sensitivity. Exploratory factor analysis identified latent factors representing degrees of engagement and exposure with each standard. Mixed-effects modeling revealed that individual differences in engagement and exposure predicted sensitivity to grammatical violations within the minority Nynorsk standard mode (Figure 1d): higher Nynorsk engagement correlated with larger reading-time costs for inflectional errors in the Nynorsk mode, whereas in the Bokmål (majority standard) mode, responses to violations remained consistent regardless of individual Nynorsk/Bokmål engagement.
Discussion: These results suggest that multilectal speakers form distinct grammatical representations of closely related varieties, shaped by individual experience. Sensitivity to Nynorsk-specific violations increased with higher engagement, suggesting that active use is needed to sustain robust representations of the minority variety. Echoing findings from bilingualism [4], this highlights that multilectal processing is experience-dependent and influenced by sociolinguistic differences. These findings support the view that multilectalism is a true sub-case of multilingualism and underscore the need to consider inter- and intra-individual experiential factors when modeling grammatical processing, even in contexts and with groups traditionally considered “unilingual”.



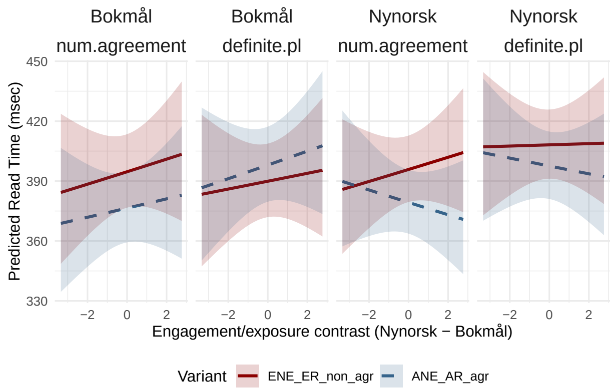
(a) Average self-reported Nynorsk engagement by municipality across the Sunnmøre region ($n = 422$).



(b) Factor scores for Nynorsk, Bokmål, and dialect (writing) engagement between Nynorsk and Bokmål users ($n = 95$).



(c) Group-level reading times across language modes and grammatical conditions (Nynorsk = 116, Bokmål = 105).



(d) Predicted reading times in the critical word region and spillover, as a function of engagement contrast (Nynorsk Bokmål).

Figure 1: Geospatial, individual, and group-level evidence for mode-specific processing of grammatical microvariation. Reading-time sensitivity to contrastive Nynorsk/Bokmål forms was modulated by individual engagement in the Nynorsk (minority) mode, but not in the Bokmål (majority) mode – suggesting an asymmetric influence of language experience on the processing of minority versus majority standards.

References

[1] Casasanto, L. S. (2008). Does social information influence sentence processing? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*, 799–804. https://escholarship.org/uc/item/8dc2t2gf

[2] Kubota, M., Alonso, J. G., Anderssen, M., Jensen, I. N., Luque, A., Soares, S. M. P., Prystauka, Y., Vangsnes, Ø. A., Sandstedt, J. J. M., & Rothman, J. (2023). Bilectal exposure modulates neural signatures to conflicting grammatical properties: Norway as a natural laboratory. *Language Learning: A Journal of Research in Language Studies*. https://doi.org/https://doi.org/10.1111/lang.12608

[3] Sandstedt, J. J., Kubota, M., Anderssen, M., Darby, J. A., Helset, S., Tavakoli, E., Vangsnes, Ø. A., & Rothman, J. (2025). Bidialectal language representation and processing: Evidence from Norwegian ERPs. *Journal of Memory and Language*, *140*, 104557. https://doi.org/10.1016/j.jml.2024.104557

[4] Bayram, F., Kubota, M., & Pereira Soares, S. M. (2024). The next phase in heritage language studies: Methodological considerations and advancements. *Frontiers in Psychology*, *15*, 1392474.

Supplementary Language Notes

Norways linguistic landscape is marked by widespread variation and two official written standards, Bokmål and Nynorsk, which have shared equal judicial status since 1885. Bokmål developed after Norways independence from Denmark in 1814, diverging gradually from Danish, while Nynorsk was originally created by the linguist Ivar Aasen in the 19th century, based on a systematization of rural Norwegian dialects into a single written standard.

Despite their formal equality, Bokmål overwhelmingly dominates written use today: in 2023, 89.6% of elementary students used Bokmål as their main written language (Statistisk sentralbyrå). Nynorsk remains strongest in Western Norway, including Sunnmøre, where this study is situated. Nonetheless, Bokmål is ubiquitous across Norway through national media, education, and urban influence.

Control and target conditions

In the control condition (1), we tested plural predicate number agreement, where all three varieties – Bokmål, Nynorsk, and Sunnmøre dialect – require an adjective marked for plural (e.g., *robust-e*) to match a plural subject. The uninflected bare stem (singular form) *robust-* is ungrammatical.

The target condition (2) examines variety-specific definite plural morphology. Nynorsk and Sunnmøre dialects maintain a three-gender system, with distinct plural endings for each gender: masculine -*ane* (e.g., *stein-ane* "stone-MASC.DEF.PL"), feminine -*ene* (e.g., *jent-ene* "girl-FEM.DEF.PL"), and neuter -*a* (e.g., *dekk-a* "(car) tire-NEUT.DEF.PL"). In contrast, Bokmål uses a generalized -*ene* for all three genders (e.g., *stein-ene*, *jent-ene*, *dekk-ene*).

For the definite condition, we specifically focused on masculine nouns, where -*ane* is grammatical in Nynorsk but ungrammatical in Bokmål, and vice versa for -*ene* (grammatical in Bokmål but ungrammatical in Nynorsk). This design allows us to test how multilectal speakers process the same inflectional forms differently, depending on the variety mode. Such differences in processing reflect distinct Nynorsk- and Bokmål-specific grammatical representations, supporting the view that even subtle morphological differences between closely related varieties can be differentially acquired, represented, and processed, contingent on the speaker's level of engagement and exposure.

(1) **Common predicate plural number agreement**

Sjåføren er sikker på at dekka er robust(e) nok til å tåle vintervegane. **Nynorsk**
Sjåføren er sikker på at dekkene er robust(e) nok til å tåle vinterveiene. **Bokmål**

'The driver is confident that the tires are robust(-PL/*-∅) enough to handle the winter roads.'

(2) **Contrastive plural definite allomorphy**

Det var lett å finne vegen tilbake, fordi stein(**a**/***e**)ne var merka med maling. **Nynorsk**
Det var lett å finne veien tilbake, fordi stein(**e**/***a**)ne var merket med maling. **Bokmål**

'It was easy to find the way back, because the stones(-M/C) were marked with paint.'

Cognitive control adaptation in code-switching: An ERP study

Hannah G. Treadway[1], Souad Kheder[2], Jorge R. Valdés Kroff[1], Edith Kaan[1]
[1]University of Florida, [2]University of Algiers
treadwayhannah@ufl.edu

Background: Temporary ambiguity resolution in sentence processing is tentatively linked to the upregulation of domain general cognitive control.[1] Real-time comprehension of intrasentential code-switching (CS) presents a compelling testbed for cognitive control in language processing as it may require cross-linguistic conflict resolution. Previous work finds that CS detection induces an upregulation of cognitive control that extends to succeeding nonlinguistic (e.g., flanker) tasks; the conflict effect, or the increased time to resolve incongruent flanker trials, was reduced after reading sentences with CS.[2] However, this adaptation has not been consistently replicated. Some studies observe either no change in or a larger conflict effect after CS relative to after unilingual sentences.[3,4] Such results favor the predictions of the Control Processes Model (CPM), which posits that some CS types require open control, increasing vulnerability to interference.[5] This study examines the contributions of upregulation and interference accounts and assesses the impact of individual CS experience by characterizing the neural signatures underlying CS cross-task paradigms.
Method: Early Spanish-English bilinguals (*n*=39) participated in a cross-task experiment during EEG recording. In critical trials, participants heard a sentence (unilingual Spanish or with multiple switches, Table 1) and then responded to a flanker trial (congruent or incongruent; *n*=32 per sentence type) (Figure 1), with effects of sentence type and congruency on ERPs predicted 200–800ms post-flanker onset. Participants also completed the Bilingual Code-Switching Profile (BCSP) to measure CS engagement and experience.[6]
Results: A permutation-based cluster mass analysis [7] (200–800ms post-flanker onset) revealed a significant late (496ms–800ms) centro-posterior positive cluster for the main effect of congruency [$p<0.001$]; incongruent flanker trials elicited greater positivity. No significant clusters emerged for the main effect of sentence type or the congruency × sentence type interaction. A post-hoc linear mixed-effects model on the cluster revealed a main effect of BCSP score [$b=0.09$, $SE=0.04$, $p=0.01$]; increased CS experience predicted larger late positive amplitudes. A significant congruency × BCSP interaction [$b=-0.02$, $SE=0.01$, $p=0.032$] indicated greater CS experience was associated with a smaller conflict effect (Figure 2). A separate exploratory analysis evaluated switch effects during sentence processing. A linear mixed effects model for central electrodes (400–600ms post-switch) showed a main effect of switch [$b=-0.52$, $SE=0.09$, $p<0.001$], with switch trials eliciting a more negative N400, and a switch × BCSP interaction [$b=0.01$, $SE=0.01$, $p=0.027$], with increased CS experience predicting reduced N400 amplitude (Figure 3).
Discussion: Switch effects captured during sentence listening did not spill over to subsequent flanker trials; neither upregulation nor interference adaptation accounts are supported. Instead, a late (496–800ms) centro-posterior positivity indexing a robust conflict effect sensitive to individual differences in CS experience was observed; frequent code-switchers experienced smaller conflict effects regardless of preceding sentence type, suggesting generalized cognitive control efficiency rather than adaptation benefits (i.e., upregulation at the trait, not state, level). Frequent code-switchers also exhibited facilitated CS integration, reflecting efficiency in linguistic conflict resolution. Results highlight the role of bilingual language experience in shaping control processes. Future analyses will discern whether theta-band activity, linked to cognitive control, is differentially elicited for (in)congruent flanker resolution as a function of sentence type.[8]

References

[1] Ness, T., Langlois, V., Kim, A., & Novick, J. (2023). The state of cognitive control in language processing. *Perspect. on Psychol. Sci.*, 1–22.

[2] Adler, R. M., Valdés Kroff, J.R., & Novick, J. (2020). Does integrating a code-switch during comprehension engage cognitive control? *J. Exp. Psychol. Learn. Mem. and Cogn.*, *46*(4), 741–759.

[3] Kheder, S., Medina, R., Valdés Kroff, J., & Kaan, E. (2024). Investigating the Effects of Code-Switch Types on Cognitive Control [Preprint]. https://doi.org/10.31219/osf.io/q9kzn.

[4] Bosma, E., & Pablos, L. (2020). Switching direction and cognitive control in bilingual reading: An ERP study. *J. Neurolinguist.*, 55.

[5] Green, D. W. (2018). Language control and code-switching. *Languages*, *3*(2), 8.

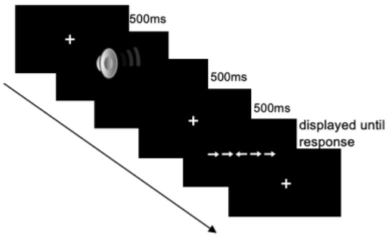
[6] Olson, D. J. (2024). The bilingual code-switching profile (BCSP) Assessing the reliability and validity of the BCSP questionnaire. *Ling. Appr. to Biling.*, *14*(3), 400-433.

[7] Fields, E. C. (2017). *Factorial mass univariate ERP toolbox [Computer software]*.

[8] Ness, T., Langlois, V. J., Novick, J. M., & Kim, A. E. (2024). Theta-band neural oscillations reflect cognitive control during language processing. *J. Exp. Psychol. Gen., 153*(9), 2279.

Table 1 (left). *Critical sentences by sentence type*; **Figure 1 (right).** *Critical trial structure for main task.*

Sentence Type	Example	Translation
Unilingual Spanish	*En la fiesta, la joven pidió una copa de vino de la anfitriona.*	“At the party, the young woman asked for a glass of wine from the hostess.”
CS	*Un* recent study *reveló que* the popular diet *realmente perjudica la salud*.	“A recent study revealed that the popular diet actually harms health.”



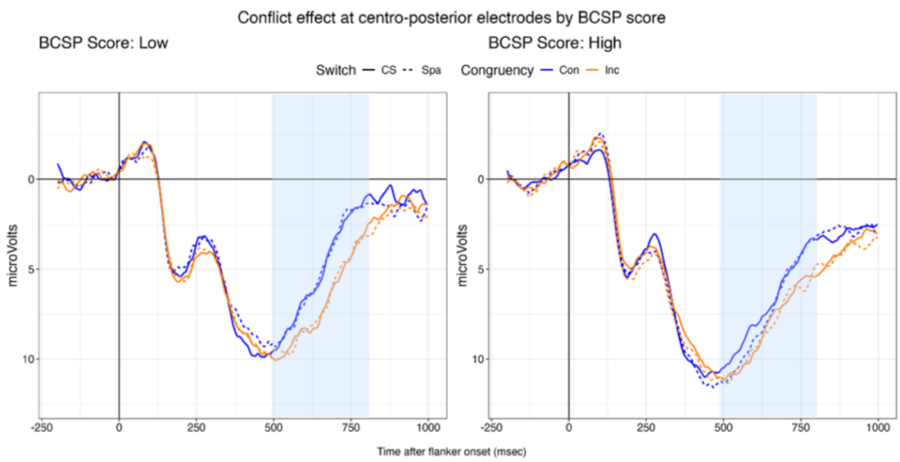


Figure 2. *Grand average ERP signatures time-locked to post-sentence flanker trials for centro-posterior electrodes by BCSP score. The 496ms–800ms time window is shaded. Trials following CS sentences are plotted by solid lines and trials following unilingual sentences are plotted by dotted lines. Congruent flanker trials are in blue and incongruent in orange. Low BCSP score is on the left and high BCSP score is on the right. Negative is plotted up.*

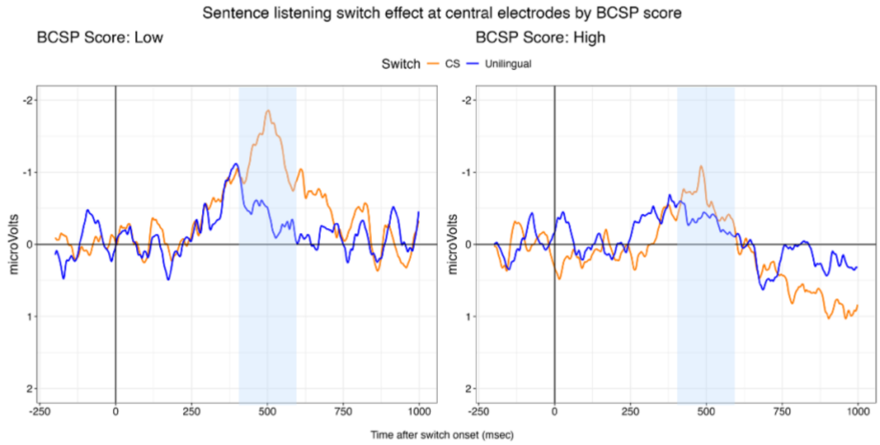


Figure 3. *Grand average ERP signatures time-locked to switch onset during sentence listening for central electrodes by BCSP score. The 400–600ms time window is shaded. CS sentences are plotted in orange and unilingual sentences are plotted in blue. Low BCSP score is on the left and high BCSP score is on the right. Negative is plotted up.*

Phonological ambiguity effect: new kid on the block switching paradigm

Dušica Filipović Đurđević, Lara Perić, Viktorija Jovanović
Faculty of Philosophy, University of Belgrade
dusica.djurdjevic@f.bg.ac.rs

Background: Language control is well documented in bilingual language use [1]. Typical experimental design involves two single language blocks (e.g. L1 in the first block and L2 in the second block) followed by a mixed-language block. The switching cost represents the difference in latencies between items on switch-trials (L1-L2 or L2-L1) and on repeat-trials (L1-L1 or L2-L2) in a mixed-language block, and is interpreted as an indicator of transient control (reactive control in response to the encountered conflict between languages). The mixing cost represents the difference between repeat-trials in the mixed block and trials in the single language blocks, and is interpreted as an indicator of sustained control (prolonged need to resolve the conflict). The longer processing in mixed block and on language switch trials both fit the hypothesis that language interference is resolved by inhibiting the unintended language that persists into the following trials [2]. The two forms of language control were recently documented in comprehension during within-language alphabet switching [3]. The testbed language was Serbian, which uses Cyrillic (C) and Roman (R) scripts (for details, see Figure 2 and Table 1). The indicator of interference between the alphabets was the size of the phonological ambiguity effect (PAE), i.e. the difference in visual lexical decision (VLD) latencies between the phonologically ambiguous and the phonologically unambiguous form of the same lexeme, with larger PAE indicating greater interference [4, 5]. However, the novelty of the design prevented a more direct comparison between bilingual language switching and within-language alphabet switching. Therefore, to enable generalization of the findings to a more general framework of code-switching, in this study, we presented two alphabets in a traditional block-switching paradigm. In experiment 1 (E1) the only conflict was that of alphabets, whereas in experiment 2 (E2), an additional conflict of grapheme-to-phoneme mapping was introduced.

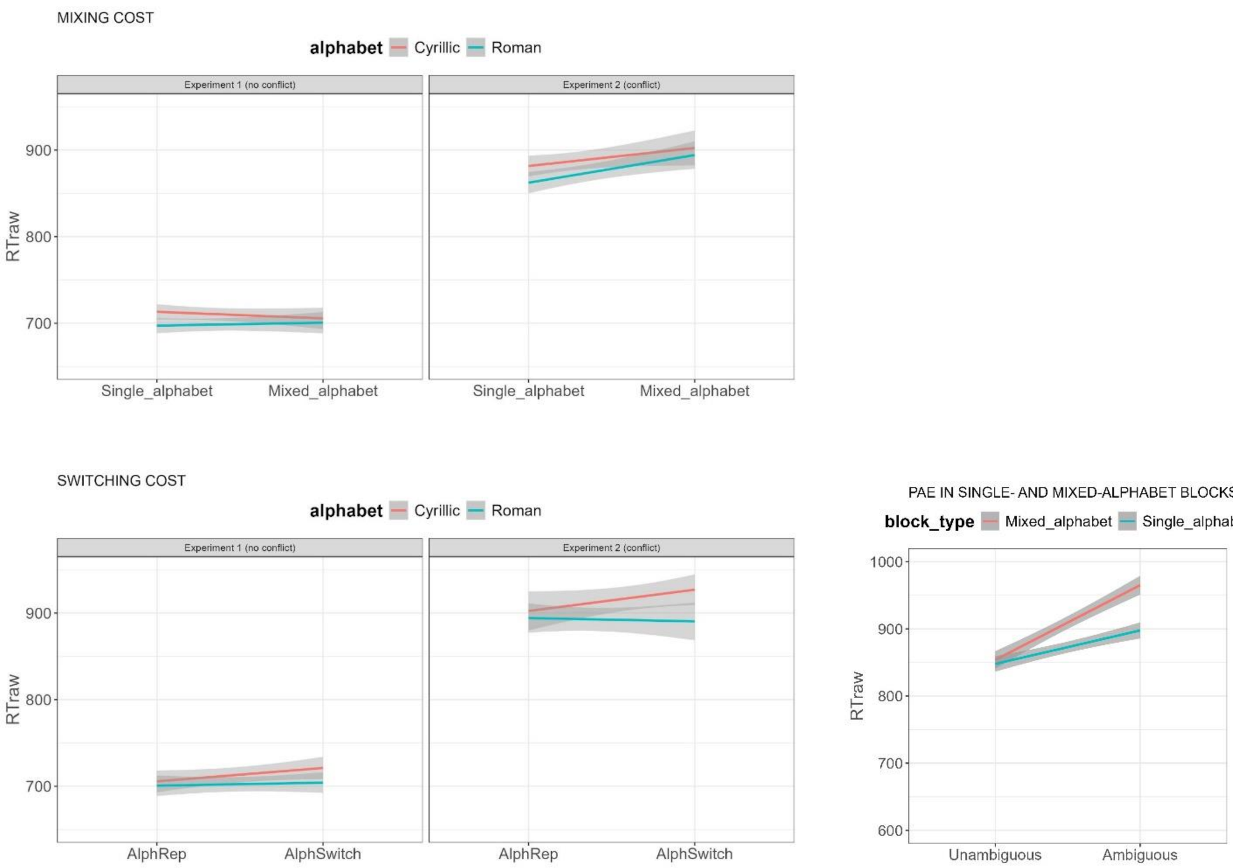
Method: We presented native speakers with the VLD task in two experiments in Serbian employing the block-switching paradigm. The first two blocks included single-alphabet items (C then R, or R then C), followed by a mixed-alphabet block (R and C). In E1 (N=87) all items (120 words and 120 pseudowords) had unique phonological mapping (e.g. MAČKA or МАЧКА). In E2 (N=111; novel set of 120 words and 120 pseudowords), half of the items were presented in their phonologically unambiguous form (e.g. SAJAM, ПЕТАК), and half were presented in phonologically ambiguous form (CAJAM, PETAK). Alphabet was counterbalanced across items and blocks, and the words presented in E1 and E2 were matched for several relevant lexical variables.

Results: In both experiments, linear mixed-effects regression revealed significant switching cost, i.e. the effect of switched vs. repeat-trials in the mixed block (E1: β=.03, CI[.01-.04], p<.001; E2: β=.03, CI[.01-.05], p<.001), and marginal mixing cost (i.e. the effect of repeat-trials in mixed vs. single-alphabet blocks). Finally, in E2, PAE was significant in single-alphabet blocks (β=.07, CI[.05-.09], p<.001) and significantly larger in mixed-alphabet block (β=.13, CI[.12-.15], p<.001).

Discussion: The observed results corroborated the presence of a robust effect of transient control in within-language alphabet switching, and a marginal effect of sustained control, as typically observed with language switching, thus indicating that alphabet switching and language switching can be seen as cases of more general switching of codes. PAE was present in single-alphabet context and larger in mixed-alphabet context, suggesting that phonological conflict introduces an additional dimension and a potential for novel insights.

Figure 1

Observed reaction latencies from Experiment 1 (left hand column, row 1 and 2) and Experiment 2 (middle and right hand column, row 1 and 2) illustrating mixing cost (row 1), switching cost (row 2), and phonological ambiguity effect (row 2, right hand column).



References

[1] Declerck, M., & Philipp, A.M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin and Review, 22*, 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

[2] Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67-81. https://doi.org/10.1017/S1366728998000133

[3] Filipović Đurđević, D. & Feldman, L. B. (2024). Do readers exert language control when switching alphabets within a language?. *Journal of memory and language, 139*. https://doi.org/10.1016/j.jml.2024.104546.

[4] Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 288-298.

[5] Filipović Đurđević, D., Milin, P., & Feldman, L. B. (2013). Bi-alphabetism: A window on phonological processing. Psihologija, 46(4), 419-436. https://doi.org/10.2298/PSI1304421F

Serbian orthography

Highly transparent orthography of Serbian can be transcribed in two alphabets – Cyrillic and Roman, and native speakers are frequent and highly fluent alphabet switchers. The two alphabets are not typically mixed within the same word or sentence, but the speakers are capable of reading effortlessly even in such conditions. Typically, the alphabets are mixed across discourse (e.g. one could read a textbook in one alphabet, then switch to another alphabet when reading e-mails, etc.). The letters of the two alphabets partially overlap (Figure 2), thus enabling the creation of various visual verbal stimuli (Table 1).

Figure 2

The structure of the Serbian writing system (from Fiipović Đurđević & Feldman, 2024).

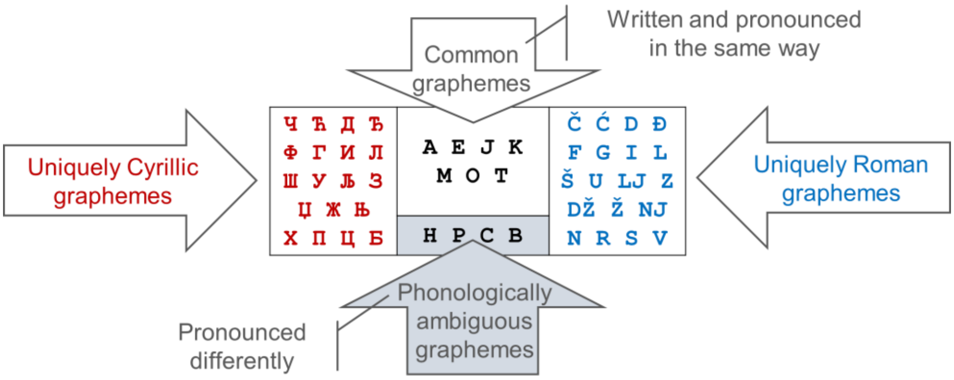


Table 1

Examples of the orthographic sequences in Serbian that can be constructed by combinations of unique, common and ambiguous graphemes (from Fiipović Đurđević & Feldman, 2024).

Composition	Word	Alphabet	Phonemic interpretation	Meaning
Unique and Common graphemes	SAJAM	Roman	/sajam/	*fair*
		Cyrillic	/	/
	ПETAK	Roman	/	/
		Cyrillic	/petak/	*Friday*
Common graphemes	ATOM	Roman	/atom/	*atom*
		Cyrillic	/atom/	*atom*
	ATOK	Roman	/atok/	/
		Cyrillic	/atok/	/
Ambiguous and Common graphemes	KACA	Roman	/katsa/	*barrel*
		Cyrillic	/kasa/	*cash register*
	PETAK	Roman	/petak/	*Friday*
		Cyrillic	/retak/	/
	CAJAM	Roman	/tsajam/	/
		Cyrillic	/sajam/	*fair*
	ATOP	Roman	/atop/	/
		Cyrillic	/ator/	/

Session 2

Encoding and Reactivating Syntactic Nodes: insights from Coordination

Leonardo Concetti
University of Siena
leonardo.concetti@student.unisi.it

Background: A growing body of research highlights the importance of syntactic information for memory mechanisms subserving sentence comprehension [e.g., 1, 2, 3]. The first goal of the present study is to offer a formal discussion on syntactic encoding, built on cue-based retrieval theory [4]. We argue that encoding-retrieval operations target fully specified syntactic nodes, rather than lexical items. The second goal is to empirically investigate the reality of an understudied component of the model: activation boost due to successive retrievals. Both goals are pursued by leveraging the hierarchical structure of coordination [5], which is displayed in Fig.1 along with the relevant memory representations.

Method: Two reading time experiments conducted on Italian investigate the role of reactivation mechanisms in modulating retrieval times. Both experiments are based on the following 2-condition design (example of Italian stimuli in Appendix A, along with some language-specific considerations):

(a) RC-plur: *The assistant and the president* [that __ *have*.PL done the interview with the reporters from the newspaper] quickly **return**.PL to the office.

(b) RC-sing: The assistant and *the president* [that __ *has*.SG done the interview with the reporters from the newspaper] quickly **return**.PL to the office.

In both conditions, retrieval of the full coordination ("the assistant and the president") is necessary at the matrix verb ("return"). Manipulating the number marking on the Relative Clause (RC) verb (singular vs plural) forces different RC attachments in the two conditions: the whole coordination is retrieved and interpreted as the subject of the RC in (a), while only the second conjunct (*the president*) is retrieved in (b). If silent reactivation boosts activation, retrieval at the matrix verb should be facilitated more in (a) than in (b), leading to faster reading times. Therefore, we predict higher reading times at the matrix verb in condition (b) with respect to (a).
Experiment 1: 240 native speakers of Italian, recruited through Prolific, completed a moving-window self-paced reading task on PCIbex. *Experiment 2:* 107 native speakers of Italian completed an eye-tracking while reading task, which served as a lab-based replication of Experiment 1, with a different participant pool and a more naturalistic task. Both experiments employed the same stimuli (32 experimental items; 64 fillers).

Results: Reading times on the matrix verb were analyzed with Bayesian linear mixed models, with the RC verb number as fixed effect, and a full random effects structure. Alongside 95% Credible Intervals for the parameter of interest, formal hypothesis testing was carried out using Bayes Factor. Due to their sensitivity to the prior specification, Bayes Factors are reported for a range of priors for the parameter of interest [6].
Experiment 1 showed a slowdown on the matrix verb in condition (b): mean difference 24ms, 95% CrI [7, 40] and BF_{10} ranging from 3 to 12. Experiment 2 replicated this slowdown in *total time*: mean difference 37ms, 95% CrI [0, 74] and BF_{10} ranging from 2 to 6.

Discussion: This study offers direct empirical evidence for activation boosts due to silent reactivation and promotes a syntactically explicit framework for understanding memory operations during real-time sentence comprehension, contributing to a more integrated view of syntax and memory in language processing.

References

[1] Franck, J., & Wagers, M. (2020). Hierarchical structure and memory mechanisms in agreement attraction. *PLoS One*, *15*(5), Article e0232163. http://dx.doi.org/10.1371/journal.pone.0232163.

[2] Arnett, N., & Wagers, M. (2017). Subject encodings and retrieval interference. *Journal of Memory and Language*, *93*, 22–54. http://dx.doi.org/10.1016/j.jml.2016.07.005.

[3] Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103. http://dx.doi.org/10.1016/j.jml.2013.04.003.

[4] Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375-419. https://doi.org/10.1207/s15516709cog0000_25.

[5] Zhang, N. N. (2010). *Coordination in syntax*. Cambridge University Press.

[6] Schad, D. J., Nicenboim, B., Bürkner, P. C., Betancourt, M., & Vasishth, S. (2022). Workflow techniques for the robust use of Bayes factors. *Psychological Methods*, *28*(6), 1404–1426. http://dx.doi.org/10.1037/met0000472.

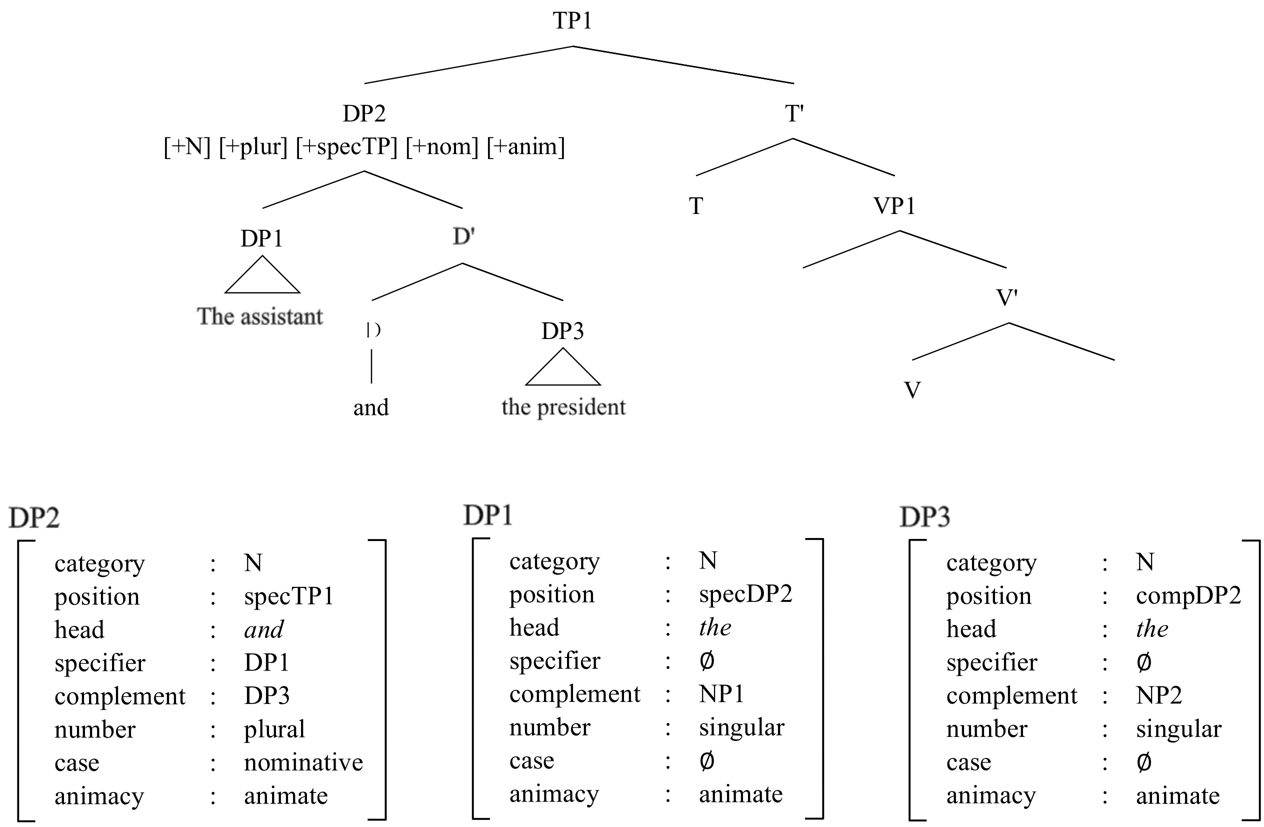


Fig.1 An example of how the hierarchical structure of coordination is assumed to be encoded in working memory. Structural features like position are used to identify syntactically licensed candidates for retrieval. Each node is encoded by a separate memory representation, leading to two different representations being reactivated in (a) and (b) of the experiments presented here.

Appendix A – example item in Italian and some linguistic considerations

(a) RC-plur: L'assistente e il presidente che hanno sostenuto l'intervista con i giornalisti del quotidiano tornano velocemente in ufficio.
(The assistant and the president that gave.$_{PL}$ the interview with the reporters from the newspaper quickly return to the office.)

(b) RC-sing: L'assistente e il presidente che ha sostenuto l'intervista con i giornalisti del quotidiano tornano velocemente in ufficio.
(The assistant and the president that gave.$_{SG}$ the interview with the reporters from the newspaper quickly return to the office)

1. Due to the agreement patterns of Italian, the different interpretation of the relative clause in (a) vs (b) remains the same with multiple verb forms. The choice of the present perfect form ("ha/hanno sostenuto") is due to it being more natural in the context described by the sentence, with respect to other verb forms such as present simple or past simple. The English version of the two sentences, on the other hand, may feel more natural with the use of past simple verbs inside the relative clause ("…that *gave* the interview"). This version would not allow the same experimental manipulation, due to the lack of number agreement which makes the RC attachment ambiguous.
2. The only possible interpretation of (a) is that both the assistant and the president gave the interview, while the only possible interpretation of (b) is that only the president gave the interview.
3. The mandatory number agreement with the matrix verb "torn-ano" ("return" / "go back") makes it clear that the verb refers to a plural entity, i.e., both the assistant and the president, excluding every other potential interpretation.

Memory Retrieval in Discourse with ‘again’: Eye-tracking and acceptability studies

Li Kloostra[1], Rick Nouwen[1], Jakub Dotlačil[1]

[1] Utrecht University, Utrecht, The Netherlands

l.kloostra@uu.nl

Background: Memory retrieval for dependency processing is commonly postulated to rely on a cue-based retrieval (CBR) mechanism[1]. Under the CBR framework, elements that share cues with a target can cause interference during dependency resolution, even when their syntactic position should make them unavailable for resolution. Despite successes of CBR, the theory faces several challenges. First, it is tested only on small subset of cases, mainly agreement and reflexives[1], but we know little about its role beyond the sentence domain. Second, the small magnitude of the established interference effects in combination with the noisy nature of data makes investigating the theory’s properties prohibitive and prone to spurious findings[1,2]. We address both concerns by considering additive presuppositions, following recent research on cue-based retrieval in discourse[3]. Consider “We met last year. I hope we meet again soon”. Here, the presupposition is triggered by “again”, requiring a discourse dependency between the information in the second clause and the proposition in the first clause (i.e., we met previously). We argue that such presuppositions can elicit interference effects similar to those predicted by CBR and in line with prior findings[3]. Moreover, in our design, we observe effects compatible with CBR, which are much stronger than those reported to date.

Method: We designed two experiments in Dutch to study the effect of (mis)matching information and interference on presupposition resolution in short discourses. Following an introduction sentence, the second sentence introduced a target direct object (DO) and a distractor DO. Either could (mis)match the presupposition in the third sentence, leading to a 2x2 design (T-MATCH vs. T-MIS, D-MATCH vs. D-MIS). Crucially, the distractor appeared under the scope of negation (Exp1) or was linked to a different discourse referent (Exp 2) which made it inaccessible for presupposition resolution. See (**1**) and (**2**) below for example items. We ran a separate acceptability study for either experiment to pretest the items, and an eye-tracking-while-reading study on Exp 1. Eye-tracking-while-reading for Exp 2 is planned but not done yet.

Results + discussion: For Exp 1 (n=56, items=36) the eye-tracking results were partially consistent with CBR (**Fig.1**): we found slowdown effects for T-MIS and D-MIS in early measures (95% CI regression path duration: T-MIS [-0.07;-0.01], D-MIS [-0.05;-0.01]; right bounded similar). Faster reading times for T-MIS D-MATCH compared to T-MIS D-MIS are also consistent with the interference effect predicted by CBR, however, the comparable difference in the T-MATCH condition requires additional explanation, perhaps encoding interference[4]. For Exp 2, the off-line acceptability study showed an interaction between Target and Distractor and the nested model (**Fig. 2**) already revealed a strong interference effect in line with CBR: D-MATCH increased acceptability in T-MISMATCH conditions while the interference is trending towards decreased acceptability in T-MATCH. This demonstrates that (i) people are sensitive to information relevant for resolution and (ii) their reading times show interference effects of inaccessible DOs, especially when the accessible DO mismatches. We believe that the experiments provide evidence that memory models can and should be applied beyond syntactic domain. Furthermore, discourse dependencies provide unique and valuable opportunity to collect novel data that is more easily obtained. Given our data, we argue that size of the effect reflects not just a contrast between discourse and syntax, but between the ways information is encapsulated in memory and tied to other information in a discourse.

Supplementary material

(1) Exp 1, example item in all four conditions. Regions of interest are indicated with square brackets. Regions 6 and 7 are considered the critical and post-critical regions.

T-MATCH D-MATCH	[Megan had honger maar gelukkig stond er een volle fruitschaal in de lobby van het hotel.]$_1$ *Megan was hungry but luckily there was a full fruit bowl in the lobby of the hotel.*	[Megan heeft een gele **appel** gegeten]$_2$ [maar ze heeft geen groene **appel** gegeten.]$_3$ *Megan ate a yellow apple but she did not eat a green apple.*	[De ochtend erna heeft Megan]$_4$ opnieuw]$_5$ [een **appel** gegeten]$_6$ [want ze wilde]$_7$ [meer fruit eten.]$_8$ *The morning after, Megan again ate an apple because she wanted to eat more fruit.*
T-MATCH D-MIS		[Megan heeft een gele **appel** gegeten]$_2$ [maar ze heeft geen groene **peer** gegeten.]$_3$ *Megan ate a yellow apple but she did not eat a green pear.*	
T-MIS D-MATCH		[Megan heeft een gele **peer** gegeten]$_2$ [maar ze heeft geen groene **appel** gegeten.]$_3$ *Megan ate a yellow pear but she did not eat a green apple.*	
T-MIS D-MIS		[Megan heeft een gele **peer** gegeten]$_2$ [maar ze heeft geen groene **peer** gegeten.]$_3$ *Megan ate a yellow pear but she did not eat a green pear.*	

Exp 2, example item in four conditions. An additional four versions were created where the order of referents was reversed, i.e., 'Jan' and the distractor DO precede 'Megan' and the target DO. The final sentence was the same in all versions. Regions 6 and 7 are again the critical and post-critical regions.

T-MATCH D-MATCH	[Megan had honger maar gelukkig stond er een volle fruitschaal in de lobby van het hotel.]$_1$ *Megan was hungry but luckily there was a full fruit bowl in the lobby of the hotel.*	[Megan heeft een gele **appel** gegeten]$_2$ [en Jan heeft een groene **appel** gegeten.]$_3$ *Megan ate a yellow apple but Jan ate a green apple.*	[Megan heeft de ochtend erna]$_4$ [opnieuw]$_5$ [een **appel** gegeten]$_6$ [want ze wilde]$_7$ [meer fruit eten.]$_8$ *The morning after, Megan again ate an apple because she wanted to eat more fruit.*
T-MATCH D-MIS		[Megan heeft een gele **appel** gegeten]$_2$ [en Jan heeft een groene **peer** gegeten.]$_3$ *Megan ate a yellow apple but Jan ate a green pear.*	
T-MIS D-MATCH		[Megan heeft een gele **peer** gegeten]$_2$ [en Jan heeft een groene **appel** gegeten.]$_3$ *Megan ate a yellow pear but Jan ate a green apple.*	
T-MIS D-MIS		[Megan heeft een gele **peer** gegeten]$_2$ [en Jan heeft een groene **peer** gegeten.]$_3$ *Megan ate a yellow pear but Jan ate a green pear.*	

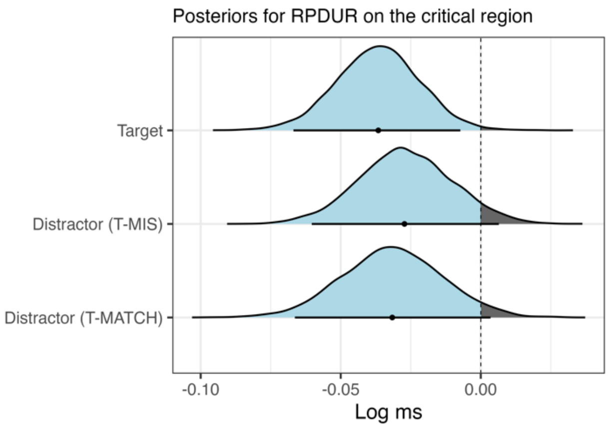


Fig. 1. Exp 1 Regression path duration with effects of the Distractor nested inside the two Target conditions. The horizontal lines denote the 95% credible intervals. Matching Distractors increased reading times in both T-MIS and T-MATCH conditions.

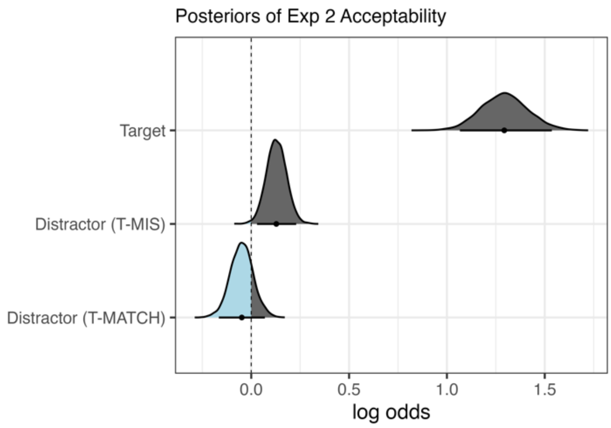


Figure 2. Acceptability study. Effects of the Distractor nested inside the two Target conditions. The horizontal lines the 95% credible intervals. Matching Distractors increased acceptability only in the T-MIS condition.

References

[1] Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316-339.

[2] Schad, D. J., Nicenboim, B., Bürkner, P. C., Betancourt, M., & Vasishth, S. (2023). Workflow techniques for the robust use of bayes factors. *Psychological methods*, *28*(6), 1404.

[3] Schmitz, T., Nouwen, R., & Dotlačil, J. (2025). Memory retrieval in discourse: Illusions of coherence during presupposition resolution. *Journal of Memory and Language*, *143*, 104637.

[4] Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language*, *129*, 104400.

Is world knowledge activation exhaustive or selective during language comprehension? Evidence from bidirectional self-paced reading

Chengjie Jiang, Walter van Heuven, Ruth Filik
School of Psychology, University of Nottingham
chengjie.jiang@nottingham.ac.uk (C.J.)

Background: It has been extensively demonstrated that world knowledge (WK) can be readily activated and influence real-time language comprehension, but few studies have recognised that an individual may hold conflicting knowledge from multiple cultural backgrounds simultaneously. For example, though *"the trains are white"* disrupts processing when presented to Dutch participants in Dutch (as Dutch trains are yellow, not white [1]), it may not be the case if presented to Chinese participants in Mandarin, as white trains are common in China. This raises an important question: when bilinguals have knowledge of multiple cultures (e.g., knowledge about trains in both the Netherlands and China), is WK activation culturally exhaustive or selective? On the one hand, because lexical semantics are closely linked with their corresponding real-world concepts, WK may be stored and activated in a way similar to lexical semantics, which is exhaustive [2]. On the other hand, WK activation may be selective based on situational relevance [3]. For example, people may only activate knowledge about the country where the experiment is conducted. Alternatively, given the strong link between language and culture, bilinguals may automatically access culture-specific knowledge based on the language in which they are reading.

Method: We conducted two bidirectional self-paced reading experiments in the UK. All participants were late and proficient Mandarin-English bilinguals, who were born and lived in Mainland China for at least 18 years, and had lived in the UK for at least six months at the time of the experiment. Both Exp1 ($n = 76$) and Exp2 ($n = 60$) manipulated the written language (Mandarin vs. English) and WK (consistent vs. inconsistent with the culturally specific WK for each language, e.g., English-consistent conditions were consistent with UK WK but not China WK and vice versa) in a 2 × 2 design (Table 1, 24 targets, 48 fillers). In Exp 1, the stimuli did not specify which country was being discussed. If WK activation is selective to the language of the stimuli, the WK consistency effect (longer RTs in inconsistent, than consistent, conditions) should be found in both Mandarin and English stimuli; if WK activation is selective to the country of the experiment, the consistency effect should be found in English stimuli but reversed in Mandarin stimuli; if WK activation is exhaustive, there should be no consistency effect in either language. The stimuli in Exp 2 included specification of the relevant country (e.g., "In the UK…") as a control for Exp 1, to confirm that WK consistency effects can be observed in a specified scenario independent of written language using the current methodology.

Results: A combined analysis was conducted using Bayesian mixed models for three reading measures (first, go-past, and total reading times) across seven regions of interest (from "bad fortune" to "of all kinds", Table 1). Analyses provided evidence for a WK × Experiment interaction for all three measures in the critical (BF_{10}s > 6.38), spill1 (BF_{10}s > 78.71), spill2 (BF_{10}s > 1000), and spill5 regions (BF_{10}s > 49.94). Specifically, there is evidence for the WK consistency effect in Exp2 BF_{10}s > 5.81, but not in Exp 1, BF_{10}s < 0.25.

Discussion: The current results challenge the notion that only the information most relevant to the discourse is activated during comprehension [3]. Instead, they suggest that when the cultural context is not explicitly specified, WK is activated in an exhaustive manner similar to lexical semantics [2] — all knowledge is retrieved from comprehenders' long-term memory, irrespective of situational relevance.

Table 1. Exp 1 & Exp 2 experimental conditions and exemplar stimuli

Conditions	Examples
Mandarin-consistent	[在中国，]Exp 2 only*四*常被视作*一个*<u>**不吉利**</u>*的数字。*不同人*可能会有*各种不同的*迷信。 [In China,]Exp 2 only four is often perceived as an <u>**unlucky**</u> number. Different individuals may have all kinds of different superstitions.
Mandarin-inconsistent	[在中国，]Exp 2 only*十三*常被视作*一个*<u>**不吉利**</u>*的数字。*不同人*可能会有*各种不同的*迷信。 [In China,]Exp 2 only thirteen is often perceived as an <u>**unlucky**</u> number. Different individuals may have all kinds of different superstitions.
English-consistent	[In the UK,]Exp 2 only the number 13 * is often * associated with * <u>**bad fortune**</u> * by many people. * Different individuals * may have * other superstitions * of all kinds.
English-inconsistent	[In the UK,]Exp 2 only the number 4 * is often * associated with * <u>**bad fortune**</u> * by many people. * Different individuals * may have * other superstitions * of all kinds.

Note. Stimuli of Exp 2 were modified from those used in Exp 1 by adding a specification of the discussed country at the beginning (“In the UK” was added to all English stimuli, and “In China” to all Mandarin stimuli).

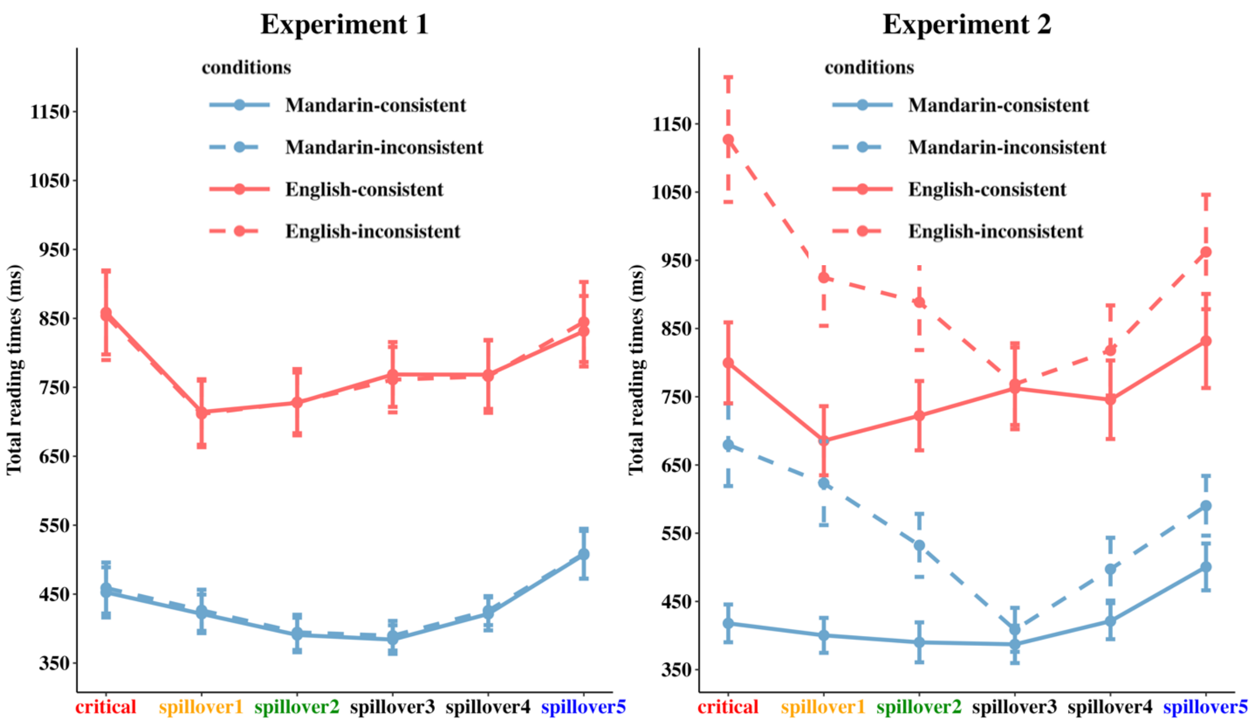


Figure 1. Exp 1 & Exp 2 mean total reading times (ms) by condition and by region (the error bars present the 95% confidence intervals)

References

[1] Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.

[2] Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197.

[3] Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, *26*(2–3), 159–190.

The time course of local coherence effects in reading times and event-related potentials

Pia Schoknecht[1], Dario Paape[1], Shravan Vasishth[1]
[1] University of Potsdam
pia.schoknecht@uni-potsdam.de

Background: In sentences like “The coach smiled at the player tossed a frisbee,” the string “the player tossed a frisbee” cannot be an active subject-verb-object (SVO) clause given the preceding context; yet, comprehenders seem to entertain this incorrect parse, at least momentarily. Behaviorally, this momentary mis-parse is expressed as greater difficulty during and after the SVO phrase is read [1].

This phenomenon, called local coherence effect, has important implications for sentence processing theories that treat grammar as a strict filter during incremental sentence processing [e.g. 2, 3]: Under such a strict filter, local coherence effects should never occur. Several studies report the existence of local coherence effects in languages like English, German, and Hindi [e.g. 1, 4, 5], but one question remains unanswered: at what moment is the local coherence effect triggered, and how quickly does grammar override the mis-parse?

Method: We investigate the time course of local coherence effects through two experiments in German (self-paced reading, SPR, and electroencephalography, EEG). Items with locally coherent SVO substrings (*einer der Köche marinierte Forellen*, ’one of the chiefs marinated trouts’) were created using lexical ambiguity and word order differences in German main and subordinate clauses (see page 3 of this abstract).

The SPR results are based on the data of 135 native speakers of German who were recruited via Prolific and read 26 critical items in a standard moving-window format. The EEG results are based on the data of 97 native speakers of German who read 80 critical items in RSVP format while their EEG was recorded.

We estimated local coherence effects using Bayesian hierarchical mixed effects models and used Bayes factors for hypothesis testing. To statistically take into account differences due to different words between conditions, we included surprisal in all our analyses [3].

Results: The only effect observed in the SPR experiment was that the object of the locally coherent SVO chunk (*Forellen*, ’trouts’) was read more slowly than the same word in the control condition (see Figure 1). In the EEG experiment, processing difficulty was observed even earlier than in SPR, suggesting that the reading times effect at the object may have been a spillover effect from the preceding ambiguous adjective, i.e., would-be verb: Compared to the control, the would-be verb of the locally coherent parse (*marinierte*, ’marinated’) elicited a more positive P600. The two following words, the object and the clause-final verb (*grillte*, ’barbecued’) , elicited reduced event-related potentials, indicating facilitated processing (see Figure 2).

Discussion: Our data indicate that the conflict between locally coherent and global parse immediately affects processing adversely at the word where the locally coherent parse arises. A novel finding in our data is that this increase in processing difficulty is short-lived: Our experiments showed no effect of the local coherence manipulation downstream from the locally coherent chunk. On the contrary, event-related potentials showed facilitated processing after the locally coherent chunk. These findings suggest that the resolution of the conflict between parses strengthened the structural prediction of the global parse.

A broader implication of our findings is that although grammar is not a strict a-priori filter, it rapidly steps in to correct incremental sentence structure building.

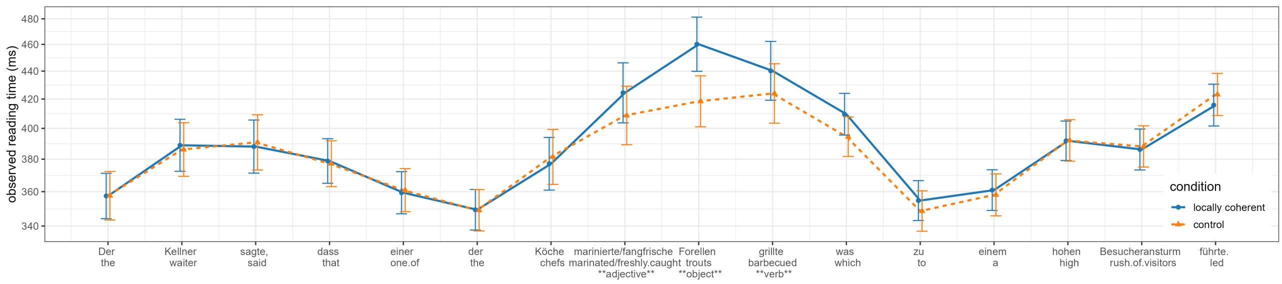


Figure 1: Observed self-paced reading times with 95% confidence intervals.

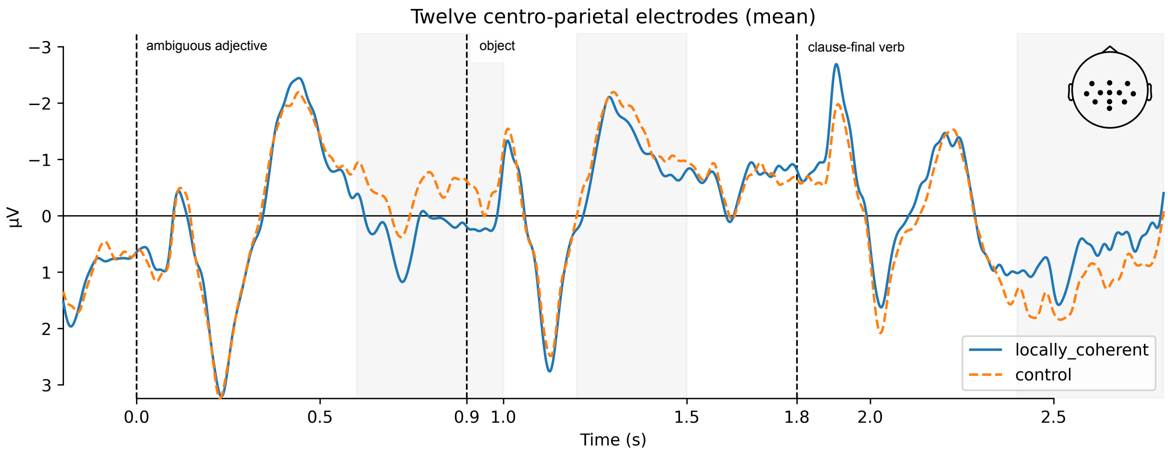


Figure 2: ERPs (average of the centro-parietal electrodes which are shown on the upper right). Onset of the ambiguous adjective (*marinierte/fangfrische*, “marinated/freshly caught”) was at 0 s, onset of the object (*Forellen*, ‘trouts’) was at 0.9 s, onset of the clause-final verb (*grillte*, ‘barbecued’) was at 1.8 s. Critical time windows are shaded in lightgrey.

References

[1] Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370.

[2] Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 559–586). Lawrence Erlbaum Associates.

[3] Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second meeting of the North American chapter of the Association for Computational Linguistics*.

[4] Paape, D., & Vasishth, S. (2016). Local coherence and preemptive digging-in effects in German. *Language and Speech*, *59*(3), 387–403.

[5] Bhatia, S., & Husain, S. (2023). Preverbal syntactic complexity leads to local coherence effects. *Language, Cognition and Neuroscience*, *38*(3), 359–389.

[6] Paape, D., Vasishth, S., & Engbert, R. (2021). Does local coherence lead to targeted regressions and illusions of grammaticality? *Open Mind*, *5*, 42–58.

[7] Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567.

[8] Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627–645.

Additional information:

In line with work by Paape and colleagues [4, 6], we used lexical ambiguity and word order differences in German main and subordinate clauses to create sentences with locally coherent substrings. In the examples shown below, the word *marinierte*, “marinated,” is a homograph. In (1a), it is used as an adjective and in (1b), it is used as a finite verb.

(1) a. Marinierte Forellen sind lecker.
Marinated trouts are delicious.

b. Einer der Köche marinierte Forellen.
One.of the chefs marinated trouts.

Although the word order in German main clauses is SVO, German subordinate sentences have SOV word order when headed by *dass*, “that,” as shown in (2). The difference in word orders between main and subordinate German clauses does not create measurable processing difficulty for native speakers of German. This is exemplified by robust monitoring of open verb phrases in sentences with multiple center-embeddings in German [7], and even facilitatory anti-locality effects in verb-final clauses with additional intervening material [8].

Now consider example (2c). This sentence is created by inserting the ambiguous adjective *marinierte* into the SOV structure shown in (2b). Note that (1b) and (2c) contain the identical words *einer der Köche marinierte Forellen*, forming a merely locally coherent chunk in (2c).

(2) a. Der Kellner sagte, dass einer der Köche Forellen marinierte.
The waiter said that one.of the chefs trouts marinated
‘The waiter said that one of the chefs marinated trouts.’

b. Der Kellner sagte, dass einer der Köche Forellen grillte.
The waiter said that one.of the chefs trouts barbecued
‘The waiter said that one of the chefs barbecued trouts.’

c. Der Kellner sagte, dass einer der Köche marinierte Forellen grillte.
The waiter said that one.of the chefs marinated trouts barbecued
‘The waiter said that one of the chefs barbecued marinated trouts.’

Session 3

“Quantum leaps” in grammar acquisition: Evidence from child Russian

Serge Minor[1], Natalia Mitrofanova[1], Gillian Ramchand[2,1]

[1] *UiT – The Arctic University of Norway*, [2] *University of Oxford*

Introduction. How do children acquire grammatical categories of the target grammar? Constructivist theories of language acquisition emphasize the importance of gradual item-by-item learning, whereas generativist theories posit abstract rules and categories from early on. We focused on the acquisition of grammatical aspect in Russian — a category expected to be particularly challenging for children given the complexity of form-meaning mappings in this domain ([1],[2],[3]). We report on a cross-sectional study testing the comprehension of Perfective and Imperfective verbs used to describe episodic accomplishment events at a specific time point in the past. In this context, Russian aspect typically marks the contrast between ongoing (Imperfective) and completed (Perfective) events. The study combined offline and online measures (picture selection and eye-tracking). **Participants.** 174 Russian-speaking children were sampled from five age groups: 3 y.o. (*n*=33), 4 y.o. (*n*=34), 5 y.o. (*n*=36), 6 y.o. (*n*=33), and 7-8 y.o. (*n*=38). **Materials and procedure**. Each trial included an audio preamble which located the narrative in the past (e.g. *It was a sunny day*), followed by a sentence-picture matching task. Participants were presented with a pair of pictures on a screen: one representing an Ongoing Event (OE), i.e. an action in progress (Fig. 1), and one representing a Completed Event (CE), i.e. the result after the action was completed (Fig. 2). While looking at the pictures, the participants heard an audio recording of a sentence in the past tense with either a perfective or an imperfective verb (ex. 1), and were asked to choose the picture best matching the sentence. The experiment included 24 test trials, 2 practice trials and 11 fillers. **Online results**. To investigate the effect of aspect on the gaze patterns, we conducted a cluster-based permutation analysis ([4]) comparing looks to the OE picture in the two aspectual conditions (Fig. 3). Significantly more looks to the OE picture were observed in the Imperfective condition compared to the Perfective across all age groups. **Offline results**. In offline responses, OE pictures were coded as targets in the Imperfective condition, and CE pictures as targets in the Perfective condition. Mean accuracy scores for the different age groups are given in Table 1. A mixed-effects logistic model revealed significant differences in mean accuracy between the 5- and 4-year-olds (p = 0.024) and between the 6- and 5-year-olds (p = 0.048). Mean accuracy in all age groups was significantly above chance. We further analyzed the distribution of participants' accuracy scores within the age groups (Fig. 4). To identify the presence of distinct sub-populations (clusters) within each age group, we fit a series of intercept-only binomial finite mixture models ([5]). No distinct sub-clusters were found in the 3 y.o. group. However, in each of the other age groups the analysis identified two separate clusters of participants: one performing at 60-70% accuracy, and one performing at ceiling (>90%) accuracy on average (Table 2). **Discussion**. Both eye tracking data and offline results indicate sensitivity to the grammatical aspect distinction already in the youngest age group (3 y.o.), with average effect size increasing with age ([6], [7], [8]). However, already from age 4, children fall into two distinct sub-groups: one performing slightly above chance, and the other performing at ceiling. This pattern is sustained in all the older age groups (4, 5, 6, and 7-8 y.o.), with the proportion of participants in the higher-performing sub-group steadily increasing with age (Table 2). This suggests a two-phase process of aspect acquisition: children begin with gradual verb-by-verb acquisition which allows them to achieve up to 70% accuracy on the comprehension task, which is then followed by a “quantum leap” to above 90% accuracy. We hypothesize that this leap is triggered by the emergence of aspect as a discrete and obligatory verbal category in the children's grammar. Our results indicate broad variability in the age at which this occurs ([9], [10]).

Ex. 1. Babuška **saža-la / po-sadi-la** bʲelyj cvʲetok
grandma **plant.IMP-PAST / PVF-plant-PAST** white flower
‘Grandma was planting a white flower.’

Age group	Mean Accuracy
3 y.o	64.3%
4 y.o.	68.1%
5 y.o.	77.5%
6 y.o.	85.9%
7-8.y.o.	86.9%

Table 1. Accuracy on the picture selection task

	Cluster 1		Cluster 2	
	N of participants	*Estimated mean accuracy*	*N of participants*	*Estimated mean accuracy*
3 y.o.	38 (100%)	64.1	0 (0%)	—
4 y.o.	22 (65%)	56.6	12 (35%)	90.9
5 y.o.	18 (50%)	61.2	18 (50%)	94.2
6 y.o.	12 (36%)	70.6	21 (64%)	94.2
7-8 y.o.	14 (37%)	70.1	24 (63%)	96.8

Table 2. Sub-populations of participants based on offline accuracy scores within the age groups.

Figures

Fig 1. The ongoing event of ‘grandma planting a flower’.

Fig 2. The completed event of ‘grandma planting a flower’.

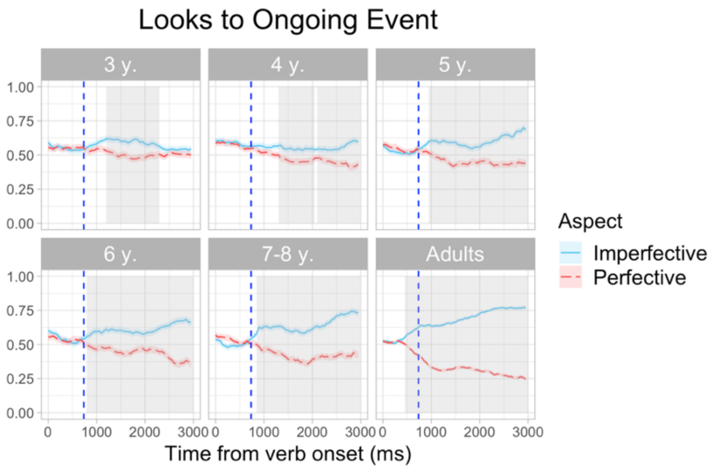


Fig. 3. Looks to the OE picture by Aspect for the five child age groups. Adult data (n=124, mean age = 22) from a previous study using the same design included for comparison. Plots start at verb onset. Dashed vertical lines represent average verb offset. Shading marks the localization of significant effects.

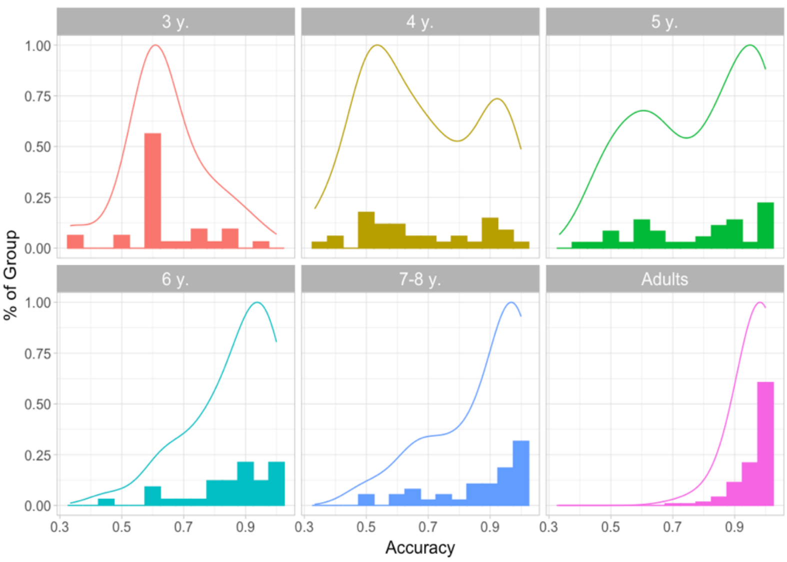


Fig. 4. Distributions of accuracy scores (histograms and densities) for the different age groups. Once again, adult results included for comparison.

Selected references: [1] **Forsyth (1970).** *A grammar of aspect: Usage and meaning in the Russian verb*. CUP. [2] **Zaliznjak & Šmelev (2000).** *Vvedenie v russkuju aspektologiju.* [Introduction to Russian aspectology]. Moscow: Jazyki russkoj kul'tury. [3] **Janda & Lyashevskaya (2011).** Aspectual Pairs in the Russian National Corpus. *Scando-Slavica*. [4] **Maris & Oostenveld (2007).** Nonparametric statistical testing of EEG-and MEG-data. *J of Neurosc Methods*. [5] **McLachlan et al. (2019).** Finite mixture models. *Annu. Rev. Stat. Appl.* [6] **Vinnitskaya & Wexler (2001).** The role of pragmatics in the development of Russian aspect. *First Language*. [7] **Stoll (1998).** The Role of Aktionsart in the Acquisition of Russian Aspect. *First Language*. [8] **Kazanina & Philips (2007).** A developmental perspective on the Imperfective Paradox. *Cognition*. [9] **Maratsos (2000).** More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *J of Child Lang.* [10] **Yang (2002).** *Knowledge and Learning in Natural Language.* MIT working papers in linguistics. OUP.

The temporal dynamics of word learning: a time-to-event analysis of age-of-acquisition

Julia Ercse[1], Peter Hendrix[1]
[1]Tilburg University
j.ercse@tilburguniversity.edu, p.h.g.hendrix@tilburguniversity.edu

Background: Understanding why some words are acquired earlier than others is a central question in the language acquisition literature. Prior research has identified a range of lexical and distributional properties that predict age-of-acquisition ratings, including word frequency, phonological neighborhood density, and concreteness [see, e.g. 1, 2, 3, 4]. These studies have been instrumental in revealing which properties are associated with earlier or later word learning. However, they do not provide insight into *when* these properties exert their influence during the acquisition process. Here, we address this gap by applying a time-to-event analysis to the age-of-acquisition data provided by Kuperman et al. (2012) [1], using a piecewise additive mixed model (PAMM; [4]) to investigate the temporal dynamics of lexical acquisition.

Method: We fit a PAMM to the age-of-acquisition data provided by Kuperman et al. (2012) [1], modeling the influence of lexical-distributional variables on the instantaneous probability of a word being acquired over time. To ensure model interpretability, we restricted the tensor product smooths used to model interactions between time and predictors to fourth-order nonlinearities. The PAMM framework accommodates non-linear predictor effects that may themselves evolve non-linearly over time. As such, it allows for detailed insight into the temporal development of predictor effects during the language acquisition process.

Results: The results of the PAMM analysis are shown in Figure 1. Consistent with previous research [1] we observed an inhibitory effect of word length, with a lower instantaneous probability of acquisition for longer words. In addition, the PAMM analysis revealed that this effect is confined to the early stages of language acquisition: by the age of 8, word length no longer significantly influences the acquisition process. Similarly, we found an effect of word frequency [cf. 1, 2, 4], with a higher instantaneous probability of acquisition for more frequent words. While this effect was most prominent during the early part of the analysis window, it remained significant throughout, persisting up to 14 years of age.

We furthermore observed effects of two neighborhood density measures: phonological neighborhood density (PND) and semantic neighborhood density (SND). Words from dense phonological neighborhoods (i.e., lower PND values) and dense semantic neighborhoods (i.e., higher SND values) were associated with a higher instantaneous probability of acquisition. The effect of SND was early and short-lived (ages 3 to 7), whereas the effect of PND emerged later and persisted longer (significant from ages 5 to 13). Finally, we found effects of two additional semantic predictors: concreteness and valence. More concrete words and words with higher valence showed a consistently higher probability of acquisition across the analysis window, although the effect of concreteness was most pronounced during the early stages of development.

Discussion: The results reported here demonstrate that the influence of lexical and semantic predictors on the word acquisition process is time-dependent. Some effects, like word length and semantic neighborhood density, are limited to early development, while others - such as word frequency and phonological neighborhood density – emerge later or persist longer. The current work thus highlights the value of time-sensitive models for uncovering not just *whether*, but also *when* predictors shape acquisition. This, in turn, reveals how the relative importance of different cues shifts as children's linguistic and cognitive capacities grow, offering a window into the evolving mechanisms that support vocabulary learning.

Figure

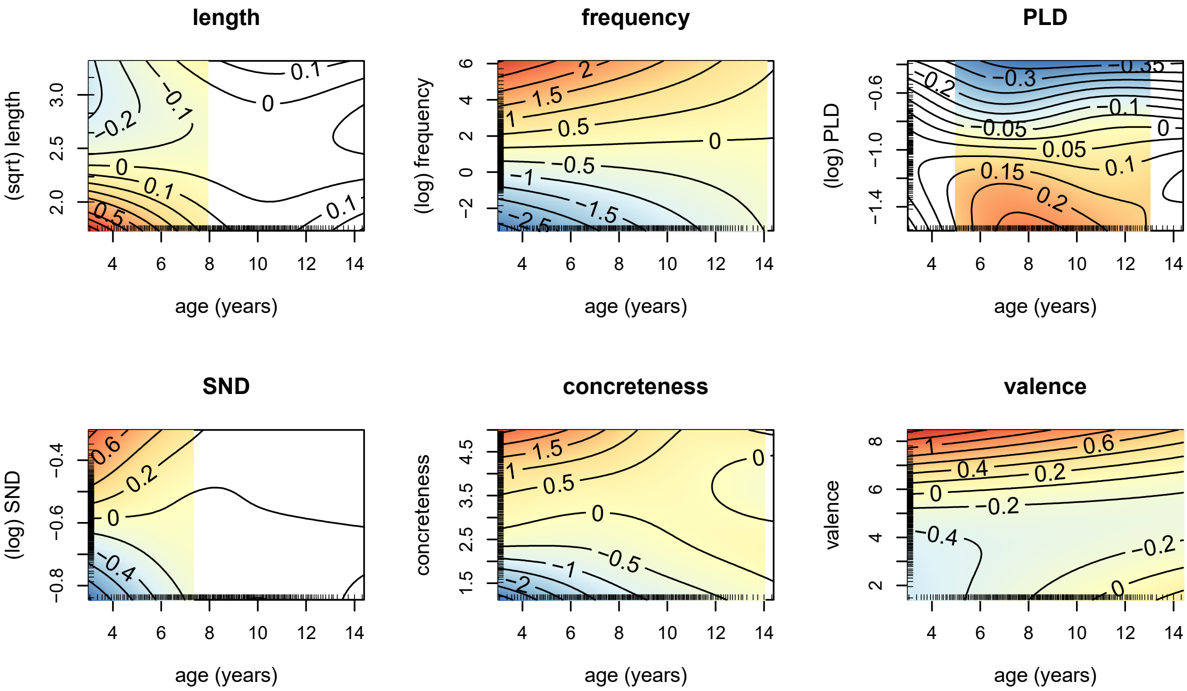


Figure 1. Results of the PAMM analysis on the age of acquisition norms [1]. Shown are predictor-related adjustments to the main effect of time. Warmer colors indicate higher hazard rates. PLD refers to phonetic Levenshtein distance; SND to semantic neighborhood density.

References

[1] Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990.

[2] Brysbaert, M., & Ellis, A. W. (2016). Word frequency and age of acquisition in language processing. In A. W. Ellis & D. A. Balota (Eds.), *The Oxford handbook of reading* (pp. 286–299). Oxford University Press.https://doi.org/10.1093/oxfordhb/9780199600491.013.017.

[3] Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency-related, partly frequency-independent. *Visual Cognition*, *13*(7–8), 992–1011.

[4] Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 528–559.

[5] Bender, A., Groll, A., & Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, *18*(3-4), 299–321.

Preschoolers use event dynamics to infer temporal order in language

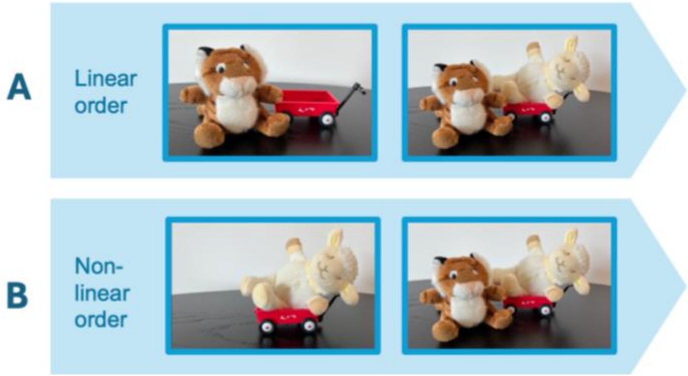
Elena Marx*[1], Hanna Shine[2], Eva Wittenberg[1], Jesse Snedeker[2]
[1] Central European University, Vienna, AUT [2] Harvard University, Cambridge, MA, USA
*marx_elena@phd.ceu.edu

Language helps children to learn about situations that they themselves did not experience, including how those situations are ordered in time [1]. While temporal order is sometimes explicitly encoded (e.g., *before*/*after*-clauses*)*, often the order of situations is left underspecified: In a sentence like *Tiger danced next to the wagon that Sheep sat on*, did the dancing happen before, during, or after the sitting? Recent studies have shown that adults use the dynamic properties of situations to infer temporal order in such cases [2-4]: static states like *sitting* are inferred to be the backdrop (i.e. the Ground) against which dynamic events like *dancing* (i.e., the Figure) unfold. Whether children follow the same strategy is an open question.

In the present pre-registered study, we address this question, using an act-out task to test preschoolers' comprehension of relative clauses like the one above. We considered two hypotheses: Children might either (a) interpret the linear order of mention as iconically reflecting temporal order in the world, a strategy known to guide their interpretation of *before*/*after*-clauses [5-6], or (b) rely on the same strategy as adults [7-9], treating stative situations as stable backgrounds of more dynamic events, and thus inferring states should happen first in time.

Methods: We asked English speaking 4-year-olds (N=28, planned 36) to act out sentences with toys (8 critical + 8 fillers + 4 *before*/*after* sentences). Critical sentences were relative clause constructions, crossing the dynamic properties between main (MC) and relative (RC) clauses (see 1-2). Enactments beginning with the MC situation were coded as linear order responses (Fig.1A) and enactments beginning with the RC situation as non-linear order responses (Fig.1B). **Results** (mixed-effects logistic regression model): As predicted by a Figure-Ground strategy in (b), children were more likely to follow the linear order of a sentence when the MC encoded a state (M=0.77, SD=0.42), but tended to reverse the order when the RC encoded a state (M=0.36, SD=0.48, main effect: z=-4.72, p<0.001, Fig.2). Overall, they systematically interpreted states as preceeding events and they performed critical trials accurately, using appropriate characters in the appropriate actions ($M_{ACCURACY}$=0.83, SD=0.38, Fig.3A). However, children's performance on post-test trials with explicit temporal order information (*before/after* sentences) was poor ($M_{ACCURACY}$=0.52, SD=0.50): Here, trials where linear order matched temporal order were performed significantly more accurately than those with non-linear order (Fig. 3B, M_{LINEAR}=0.68, SD=0.47, z=2.86, p=0.004), in line with previous findings on a linearity bias in (a).

In sum, our findings provide first developmental evidence that children use event dynamicity as a cue to infer temporal order, extending prior findings with adults [2-4]: Children's inferences are not at random or guided by linear, iconic order, but reflect broader principles of conceptual organization that shape their understanding of how events unfold in time. These results also highlight the importance of investigating children's temporal understanding in contexts that lack explicit temporal order information, to fully capture the scope of their temporal reasoning abilities in language.



(1) eventive MC – stative RC
Tiger ***danced*** *next to the wagon that Sheep* ***sat*** *on*

(2) stative MC – eventive RC
Tiger ***sat*** *next to the wagon that Sheep* ***danced*** *on*

Figure 1 Coding scheme for example sentences in (1-2) as linear order enactment with the main clause enacted first (A) and non-linear order enactment with the relative clause enacted first (B).

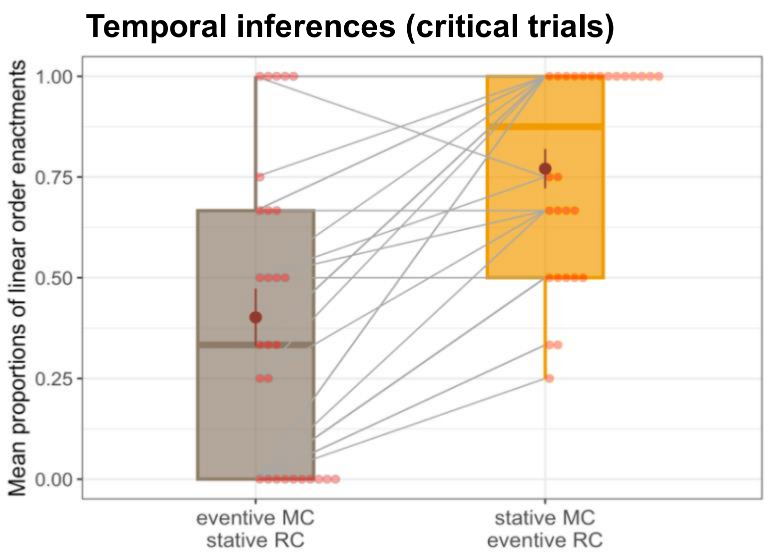


Figure 2 Mean first enactments of main clauses per condition. Boxplots summarize the means per participants, individually plotted as lightred dots and grey lines. Darkred dots and error bars represent overall means and the standard error of the mean.

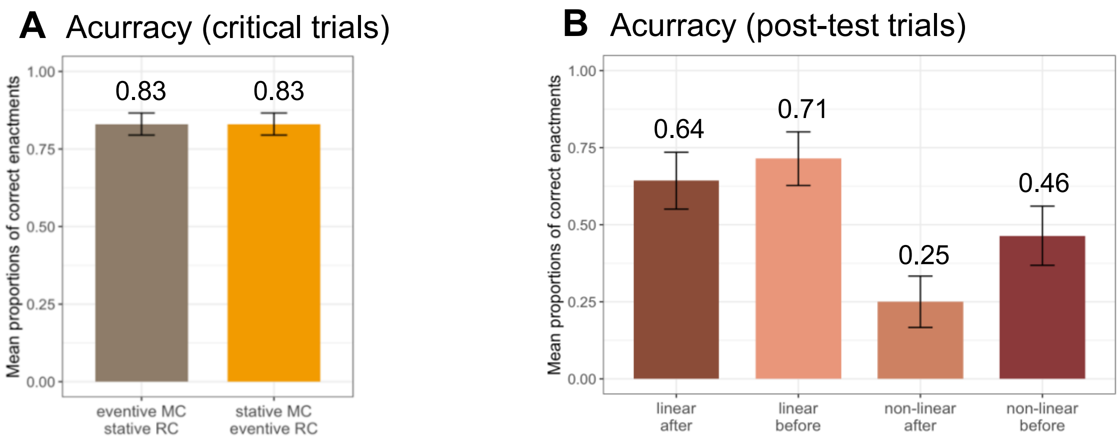


Figure 3 Mean accuracy in the critical conditions (A) and post-test conditions (B). Errorbars represent the standard error of the mean. Incorrect enactments in (A) included incomplete enactments, incorrect placement of characters, inappropriate characters or actions. Incorrect enactments in (B) added incorrect temporal order as a possible error type.

References

[1] Zhang & Hudson, *Frontiers in Psychology*, 2018; [2] Marx & Wittenberg, *Cognition,* 2025; [3] Marx, Iwan & Wittenberg, *Cognitive Science,* 2025; [4] Marx & Wittenberg, *Glossa Psycholinguistics,* 2024; [5] de Ruiter, Theakston, Brandt & Lieven, *Cognition*, 2018; [6] Blything & Cain, *Developmental Psychology,* 2016; [7] Gleitman, Gleitman, Miller & Ostrin. *Cognition,* 1996; [8] Talmy, *Annual Meeting of the Berkeley Linguistics Society*, 1975; [9] Hopper, *Discourse and Syntax,* 1979.

Children’s sensitivity to animacy constraints in possessive noun phrases: Priming induces non-adult-like structures in Czech preschoolers

Barbora Skarabela[1,2], Mitsuhiko Ota[2], Filip Smolík[1]
[1] Institute of Psychology, Czech Academy of Sciences; [2] School of Philosophy, Psychology and Language Sciences, University of Edinburgh
b.skarabela@ed.ac.uk

Prior exposure to syntactic structures in adults facilitates subsequent processing and increases the likelihood of their repetition [1]. In children, the tendency to repeat recently encountered structures offers insight into their underlying syntactic knowledge [2]. In this study we exploit this phenomenon to examine children’s knowledge of syntactic structures that are constrained by semantic factors. Specifically, we focus on Czech children’s expression of possessive relationships, using either **possessive adjectives**—restricted to human possessors (e.g., *pirátův klobouk* / "the pirate’s hat," but not **kabátova kapsa* / "the coat’s pocket")—or **the genitive case**, which accommodates both human and non-human possessors (e.g., *klobouk piráta* / "the hat of the pirate" and *kapsa (od) kabátu* / "the pocket of the coat"). Are Czech four-year-olds sensitivite to animacy constraints in possessive adjectives?

In syntactic priming experiments with a prime-target animacy mismatch, participants described images featuring two inanimate entities (e.g., "coat" and "pocket") after being exposed to descriptions with a human possessor and an inanimate noun (e.g., "pirate" and "house"), presented in either the possessive adjective or genitive case form in a between-subject design. The experimenter took turns with the participant in describing pictures across three phases: (1) the pretest, when the participants described pictures before exposure to the prime; (2) the test phase, when the participants in the PossAdj group described pictures immediately upon hearing the prime with the possessive adjective, or the participants in the Genitive group gave the description immediately after hearing the prime with the genitive; and (3) the post-test, when the participants described pictures without hearing the prime. Each phase included eight opportunities for the target noun-noun description.

Experiment 1 established the baseline for animacy-based morphosyntactic constraints in adult speakers (N=40). The results confirmed that Czech-speaking adults consistently apply the genitive case when referring to inanimate entities, regardless of the prime, in strict adherence to established grammatical norms. Experiment 2 examined the sensitivity of preschool-aged children (N=40, age range: 3;11–4;7) to these constraints. A binomial mixed-effects model analysis of the data confirmed that, compared to the pre-test, the children primed with the possessive adjectives produced significantly more possessive adjectives in the test phase and the post-test. Thus, when primed, preschoolers used the possessive adjectives for inanimate possessors—an innovation not observed in adult usage (e.g., *kabátova kapsa*/ “the coat’s pocket”).

Although the children predominantly used the genitive case in their descriptions, demonstrating sensitivity to animacy constraints in possessive adjectives, they were, unlike adult speakers, susceptible to priming with the structure, occasionally producing unattested examples. These findings provide novel insights into morphosyntactic acquisition and underscore the complexity of mastering structures conditioned by animacy, highlighting the gradual refinement of grammatical knowledge over time. Implications for models of structural priming and theoretical accounts of morphosyntactic development are discussed.

Table 1. Mean proportions of possessive adjective (vs. genitive) responses (Experiment 1: Adults).

Condition	Phase	Mean poss. adj. proportion (SD)
Genitive (N=20)	pretest	0.00 (0.00)
	prime	0.00 (0.00)
	post	0.00 (0.00)
Possessive Adjective (N=20)	pretest	0.00 (0.00)
	prime	0.02 (0.06)
	post	0.00 (0.00)

Table 2. Mean proportions of possessive adjective (vs. genitive) responses (Experiment 2: Children).

Condition	Phase	Mean poss. adj. proportion (SD)
Genitive (N=20)	pretest	0.03 (0.09)
	prime	0.02 (0.05)
	post	0.04 (0.10)
Possessive Adjective (N=20)	pretest	0.02 (0.09)
	prime	0.16 (0.25)
	post	0.13 (0.26)

Model specification: possAdj ~ condition * phase + (1|id)

NOTE:

- The response variable (possAdj) was binary between 1 (possessive adjective) and 0 (otherwise, i.e., genitive). Condition was dummy coded as 1 (possessive adjctive prime) and 0 (genitive prime). Phase was separated into two dummy coded variables with the pretest as the reference level, one with the prime phase as 1 (and else 0) and another with the post-test phase as 1 (and else 0).
- This final model was obtained through reduction of the full model by first removing the random slopes for by-participant and by-item random effects, and then the random intercept by items until convergence was obtained. Models were run using the glmer function on R with the family 'binomial'.

Statistical results

```
                  Estimate Std. Error z value Pr(>|z|)
Intercept         -6.4020     1.4639  -4.373 1.22e-05 ***
condition         -0.9140     1.5761  -0.580   0.5620
phase(prime)      -0.2955     1.0898  -0.271   0.7863
phase(post         0.4042     0.9606   0.421   0.6739
condition:post     2.4742     1.2594   1.965   0.0495 *
condition:prime    3.3310     1.3603   2.449   0.0143 *
```

References

[1] Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

[2] Messenger, K. (Ed.) (2022). Syntactic priming in language acquisition: Representations, mechanisms and applications. *Trends in Language Acquisition Research,* vol. 31. John Benjamins Publishing Company.

Session 4

Phonological typicality in the distribution of nouns and verbs in French: Statistical relations between form and grammar

Gustavo L. Estivalet[1,2], Katie L. McMahon[3], Joanne Arciuli[4], Greig I. de Zubicaray[1]

[1] School of Psychology and Counselling, Queensland University of Technology, Australia; [2] National Council for Scientific and Technological Development, Brazil; [3] School of Clinical Sciences, Queensland University of Technology, Australia; [2] College of Nursing and Health Sciences, Flinders University, Australia.
gustavo.lopezestivalet@qut.edu.au

Background: While arbitrariness has been considered an inherent characteristic of the relation between word sound and meaning, recent research has shown that form systematicity plays an important role in word representation and processing [1]. Early research on phonological typicality of lexical category was restricted to Germanic languages and monosyllabic words [2] but few studies have addressed Romance languages and polysyllabic words. Aiming to fulfill these gaps, this research has three objectives: investigate the phonological typicality of the distribution of nouns and verbs in French, extend the methods to mono- and polysyllabic words, and analyze the prediction of the distributional typicality on the lexical processing in behavioral visual and auditory lexical decision experiments.

Method: We selected ambiguous words which can be nouns or verbs from the *Lexique* database and matched two samples in frequency, length, and neighborhood (*MatchIt* R package), one with higher prevalence of frequency for nouns and another with higher prevalence of frequency for verbs. We extracted 73 phonological characteristics from each word, the length in phonemes and syllables, the overall number of phonological features, and the information from the initial and final phonemes. Then, we calculated the ratio between noun and verb frequencies [3], excluded collinear variables, and determined the best subset of predictors (*caret* and *leaps* R packages) to calculate the phonological typicality of lexical category using a linear regression with robust standard errors [4]. We also calculated the phonological discrepancy as the absolute difference between the noun/verb ratio and the distributional typicality (Figure 1). Finally, we analyzed the prediction of these and other control lexical variables in the French behavioral megastudy MEGALEX with visual and auditory lexical decision experiments [5] through relative importance analysis [6].

Results: The matched sample of ambiguous noun/verb French words included 2,768 words. The best model contained 19 predictors and explained 12% of the overall variance of noun/verb ratio, with number of rounded and final front vowels as the strongest noun- and verb-like predictors, respectively. We found a significant difference between the distributional typicality of nouns and verbs: Welch's $F(1, 2766) = 288.71$, $p < .001$. The distributional typicality was a significant predictor ($p < .001$), explaining 1.4% and 8% of the latency variance in the visual and auditory lexical decision experiments, respectively (Figure 2). The form discrepancy was also a significant predictor ($p < 0.01$), explaining ~1% in both experiments.

Discussion: Our results in French are aligned with results in English with greater discrepancy and noun-like words yielding significant slower latencies [3], suggesting that ambiguous word forms are modulated by the distributional typicality. These results favor accounts in which lexical processing is based on distributional information provided by word systematicity founded on phonological features in keeping with a statistical approach to language processing. We extended English findings in the visual modality [7], replicating the results in the auditory modality in which the distributional typicality explained even more variance.

Figures

Figure 1. Distributions of the phonological typicality and discrepancy of nouns and verbs.

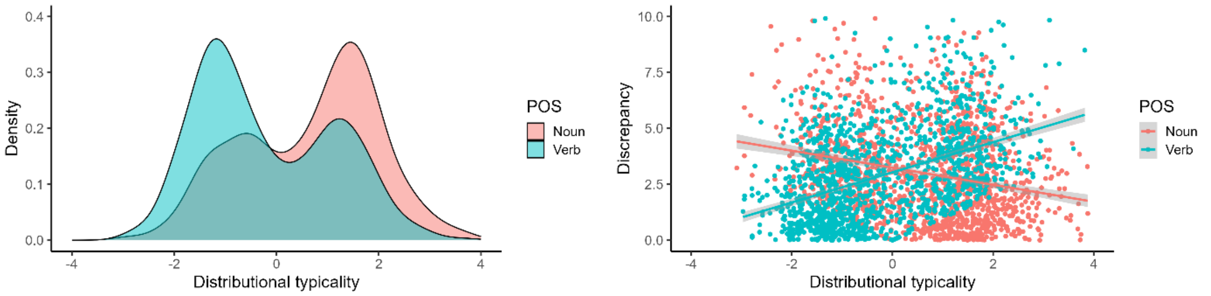
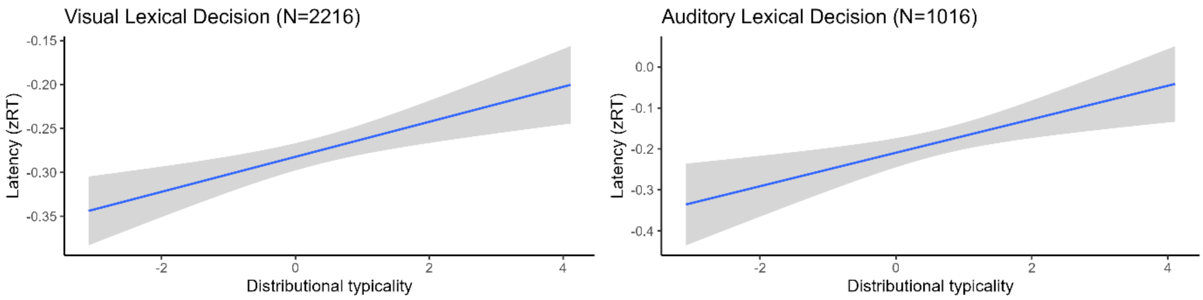


Figure 2. Latencies in functional of the distributional phonological typicality of nouns and verbs.



References

[1] Haslett, D. A., & Cai, Z. G. (2024). Systematic mappings of sound to meaning: A theoretical review. *Psychonomic Bulletin & Review, 31*(2), 627–648. https://doi.org/10.3758/s13423-023-02395-y

[2] Monaghan, P., Christiansen, M., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

[3] Sharpe, V., & Marantz, A. (2017). Revisiting form typicality of nouns and verbs. *The Mental Lexicon, 12*(2), 159–180. https://doi.org/10.1075/ml.17004.sha

[4] de Zubicaray, G. I., & Hinojosa, J. A. (2024). Statistical Relationships Between Phonological Form, Emotional Valence and Arousal of Spanish Words. *Journal of Cognition, 7*(1). https://doi.org/10.5334/joc.366

[5] Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., Dufau, S., Mathôt, S., & Grainger, J. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods, 50*(3), 1285–1307. https://doi.org/10.3758/s13428-017-0943-1

[6] Wilcox, R. R. (2018). Robust regression: an inferential method for determining which independent variables are most important. *Journal of Applied Statistics, 45*(1), 100–111. https://doi.org/10.1080/02664763.2016.1268105

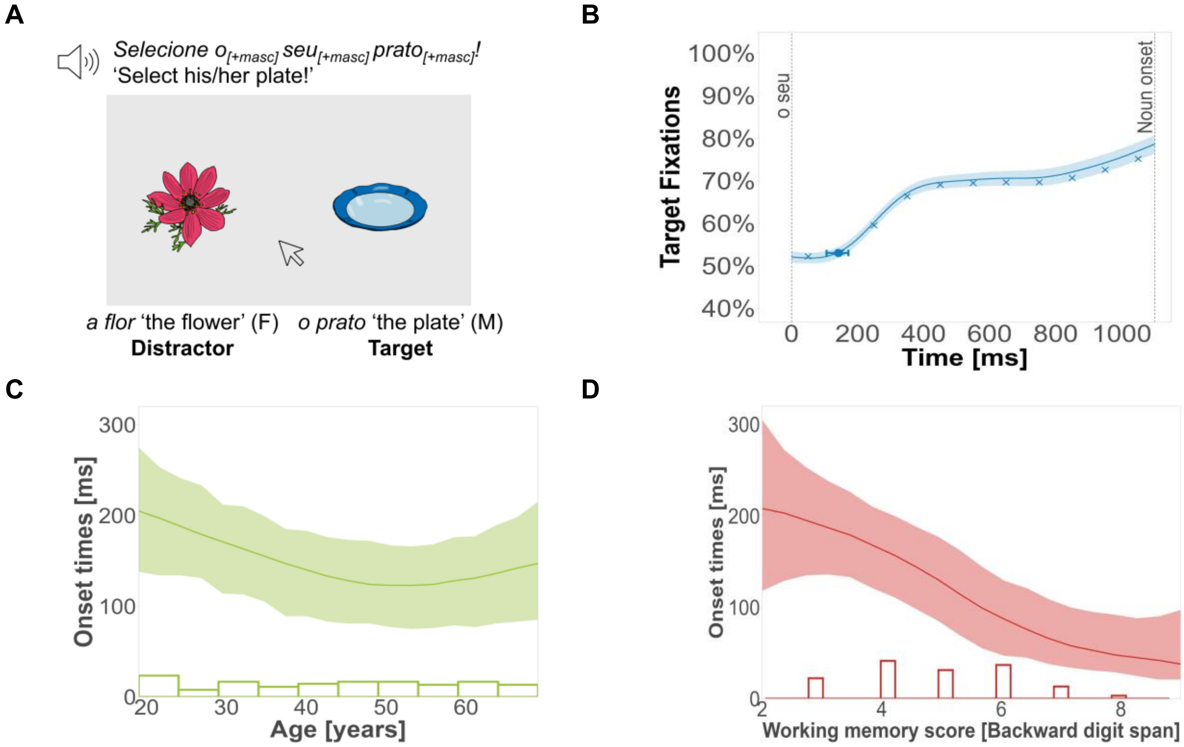
[7] Arciuli, J., McMahon, K., & Zubicaray, G. de. (2012). Probabilistic orthographic cues to grammatical category in the brain. *Brain and Language*, *123*(3), 202–210. https://doi.org/10.1016/j.bandl.2012.09.009

Ageing yields improvements in morphosyntactic prediction

Jéssica Gomes[1], Sol Lago[2], João Veríssimo[1]
[1]Center of Linguistics, School of Arts and Humanities, University of Lisbon; [2]Goethe University Frankfurt
jgomes4@edu.ulisboa.pt

Background: Language prediction is affected by ageing in different ways, as age-related changes in cognitive and linguistic abilities influence individuals' predictive efficiency. For example, ageing is associated with declines in working memory (WM) and processing speed (PS) [1], while syntactic processing tends to be preserved, likely due to compensatory mechanisms that result from the lifelong accumulation of linguistic experience [2,3,4]. A previous study on individual differences in language prediction suggested that increased age was associated with a more efficient use of gender cues to predict upcoming linguistic information [5]. However, the authors only analysed the overall proportions of fixations to the target image, without considering their timecourse. In our study, we used a continuous-age design with a well-distributed sample of participants to examine the use of morphosyntactic cues for language prediction across the lifespan. Additionally, we employed a novel statistical method to estimate the onset of predictive effects and their modulation by age [6]. We also explored how cognitive factors, such as WM and PS, influence predictive efficiency. We addressed two research questions: first, how does age impact the temporal onset of morphosyntactic prediction? Second, do WM and PS modulate predictive processing across the lifespan?
Method: We tested 142 European Portuguese native speakers (age range: 19–69 years) in a visual world eye-tracking experiment. In each trial, participants saw two objects while hearing an auditory instruction that included a determiner and a possessive pronoun (e.g., *o seu* '(the)[+masc] his/her[+masc]'), both expressing gender information that allowed predicting the target noun (e.g., *livro* 'book'[+masc]) (Fig. 1A). The eye-tracking data was analysed with a GAMM-based method [6] that allowed estimating the effects of age on three complementary measures: (i) overall proportion of fixations to the target image during the predictive time-window; (ii) timecourse of fixations over time; and (iii) temporal onsets of predictive effects. Participants also completed WM and PS tasks, which were used as covariates in the statistical models.
Results: At the group level, the onset of the predictive effect was estimated at 141 ms, 95% CI [107, 173] ms (Fig. 1B). At the individual level, we observed (i) a positive effect of age on the proportions of target fixations, (ii) a significant age effect on the timecourse of fixations, and (iii) faster prediction onsets for older than younger adults (at least until the mid-50s) (Fig. 1C). These effects were obtained when controlling for WM capacity and PS. Moreover, sensitivity analysis showed that the effects remained significant even when age was not adjusted for any covariates, contrasting with [5], where age-related effects only emerged after controlling for covariates. We also observed an effect of WM (correlation between age and WM, $r = -0.39$): participants with higher WM capacity exhibited higher proportions of target fixations and shorter predictive onsets (Fig.1D). There was no evidence for an effect of PS (correlation age and PS, $r = 0.40$).
Discussion and conclusion: Our findings suggest that ageing can enhance the use of morphosyntactic cues to predict upcoming linguistic information. These results support the view that aspects of syntactic processing can not only be well-preserved at older ages, but might even improve, likely due to compensatory mechanisms resulting from lifelong linguistic experience [4,5]. In addition, the WM effect is consistent with research suggesting that WM capacity plays a role in vision-language tasks [5], and highlights the importance of considering the role of individual differences and cognitive abilities in predictive processing.

Figure 1: (A) Sample experimental trial. The images depict objects that differ in grammatical gender (e.g., *flor* 'flower'[+fem] vs. *prato* 'plate'[+masc]). Participants were introduced to the gender-neutral character *Glipi*, and asked to find their belongings by selecting the target object mentioned in the auditory instruction. (B). Fitted GAMM curve overlaid onto empirical means of fixation proportions across participants. The point with error bars indicate the median predictive onset with the 95% highest density interval obtained from the GAMM-based method. (C) Age effect and (D) working memory effect on the onset times of prediction. The ribbons show 95% confidence intervals. The histogram on the x-axis shows the number of participants at each age bin and working memory score.



References

[1] Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*(3), 179-183.

[2] Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex, 20*(2), 352-364.

[3] Veríssimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., & Ullman, M. T. (2022). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nature Human Behaviour, 6*(1), 97-110.

[4] Cabeza, R., Albert, M., Belleville, S., Craik, F. I., Duarte, A., Grady, C. L., ... & Rajah, M. N. (2018). Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, *19*(11), 701-710.

[5] Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language Cognition and Neuroscience, 31*(1), 80–93.

[6] Veríssimo, J. & Lago, S. (2025). A novel method for detecting the onset of experimental effects in visual world eye-tracking [Preprint]. https://doi.org/10.17605/OSF.IO/9ETHQ.

Sensitivity to Verb Bias as a Continuous Variable in L1 and L2 Processing

Danny Dixon[1] & Lisa Levinson[2]
[1]University of Kansas; [2]University of Michigan
dannydixon@ku.edu

Background: Verb structural continuation bias, or how often a verb occurs with different complement types, has been shown to influence syntactic processing in L1 [1] and in L2 [2,3], though its role in L2 processing is less well understood. While previous studies have shown that the bias of L2 verbs is used in L2 processing [2,3], the potential influence of L1 verb bias during L2 processing has not been demonstrated conclusively. The existence of such influence would align with prior findings about L1 interference regarding transitivity frequency in verbs [4] and would be expected under theories of bilingualism assuming parallel activation during processing [5]. Prior studies have largely treated verb bias as a categorical variable, with verbs categorized as direct object (DO) biased, sentential complement (SC) biased, or equi-biased. However, these categorizations could obscure fine-grained differences in bias strength that could influence processing in a continuous manner; furthermore, they limit the detail of comparisons of a verb's bias across languages, which is necessary to investigate the role of L1 verb bias in L2 processing.

Current Study: The current study aims to obtain continuous measurements of verb bias and use them in analyzing behavioral data collected by Wolford 2022 [3] to better understand the role of L1 and L2 verb bias in L2 processing. Wolford used a maze task to examine the L2 processing behavior of Spanish-English bilinguals ($n = 60$) in comparison with a group of English monolinguals ($n = 60$); stimuli were adapted from Dussias and Cramer Scaltz [2] and consisted of English sentences for each continuation type (Fig. 1).

Methods: Approximately 280 million words each of English and Spanish plaintext corpora were programmatically parsed, collected from the Corpus of Contemporary English (COCA) and the Corpus del Español respectively. The corpora were tokenized by sentence, and dependency representations of sentence structure were created using UDPipe, a multilingual dependency parser [6]. These dependency parses were used to categorize the type of complement each verb occurred with, and counts for each continuation type were recorded for all verbs used in the experimental stimuli by Wolford [3]; these counts were then used to calculate continuous verb bias probabilities and corresponding syntactic surprisal values.

Results: The calculated continuous verb bias values agreed reasonably well with prior verb bias categorizations for the English verbs by Garnsey et al. [7] and for the Spanish verbs by Dussias et al. [8] (Fig. 2). When used to analyze Wolford's [3] maze task data, the results showed that the continuous English verb bias measurements were a significant predictor of RTs for the monolingual control group ($\beta = 0.13$, $SE = 0.02$, $p < 0.001$; Fig. 3). The analysis of the bilingual participants' RTs indicated that the verb bias values of the equivalent Spanish verbs were a better predictor of RTs ($\beta = 0.03$, $SE = 0.01$, $p < 0.05$) than the verb bias of the English verbs encountered in the stimuli which were not shown to be a significant predictor (Fig. 4); the analysis also found that the range of DO syntactic surprisal values for the Spanish verbs was significantly smaller than for the SC values (Fig 4.)

Conclusion: These results suggest that verb bias influences processing in a gradient manner. Moreover, they indicate that L1 verb bias may influence L2 sentence processing, potentially to a greater extent than L2 verb bias. Future research should seek to replicate these findings with other language pairs and levels of L2 proficiency.

References: [1] Clifton et al., 1984. *Journal of Verbal Learning and Verbal Behavior*. [2] Dussias & Cramer Scaltz, 2008. *Acta Psychologica*. [3] Wolford, 2022. *HSP*. [4] Frenck-Mestre & Pynte, 1997. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*. [5] Kroll & Ma, 2017. [6] Straka & Straková, 2017. *Association for Computational Linguistics*. [7] Garnsey et al., 1997. *Journal of Memory and Language*. [8] Dussias et al., 2010. *Behavior Research Methods*.

Figure 1.

Unambiguous	The ticket agent admitted that the mistake might not have been caught.
DO Continuation (Ambiguous)	The ticket agent admitted the mistake because she had been caught.
SC Continuation (Ambiguous)	The ticket agent admitted the mistake might not have been caught.

Figure 2.

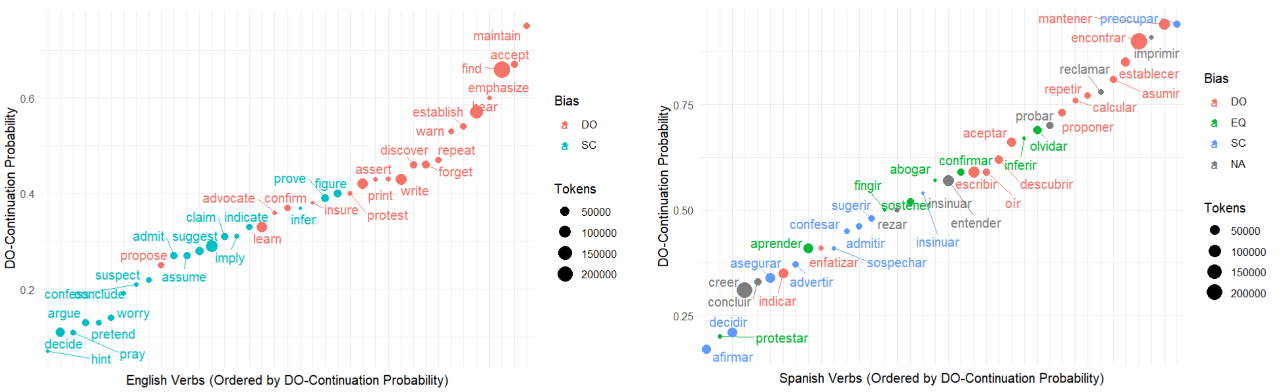


Figure 3.

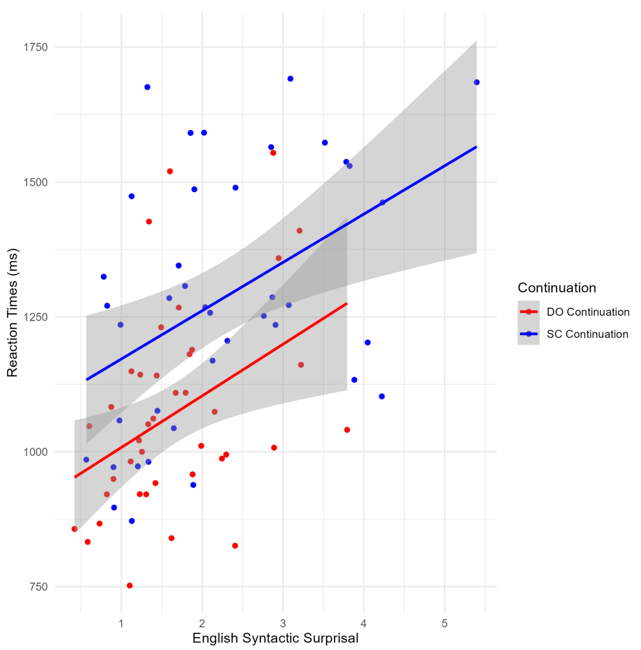
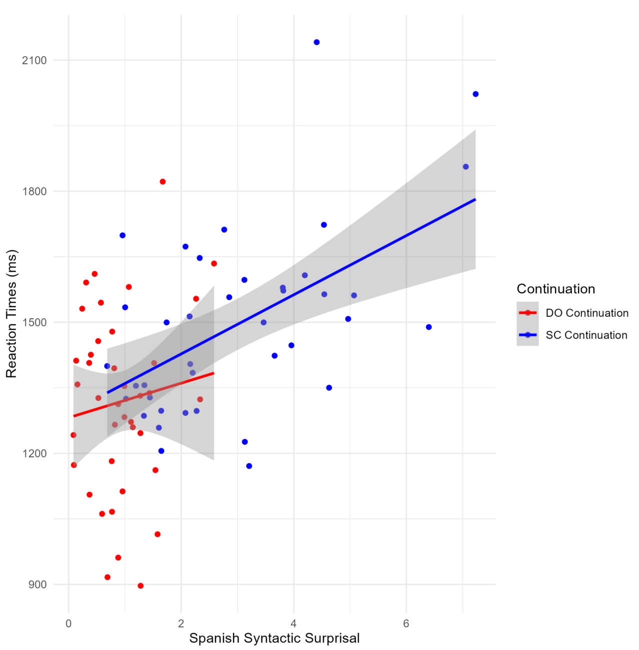


Figure 4.



The task dependence of misinterpretation effects: A comparison of L1 and L2 speakers of German

Anna-Fiona Weiss[1], Markus Bader[2], Michael Meng[3]
[1]Catholic University of Eichstätt-Ingolstadt; [2]Goethe University Frankfurt; [3]Merseburg University
annafiona.weiss@ku.de

Background: The role of heuristic processing routines and shallow representations has been a subject of debate both in research on L1 and L2 sentence processing. In L1 processing, heuristic processing has been argued to lead to so-called misinterpretation effects that several studies observed for unambiguous sentences with non-canonical order of agent and patient ([1, 2], among many others). However, it has also been demonstrated that whether or not misinterpretation effects can be observed depends on the task used to assess comprehension: whereas agent/patient naming and similar tasks lead to significant error rates for noncanonical sentences, tasks such as plausibility judgments can be solved with high accuracy [3]. The task-dependence of such misinterpretation effects may be explained in terms of a post-interpretive account proposed by [2].For L2 processing, some accounts postulate an even stronger impact of heuristic processing than in the L1. According to the Shallow Structure Hypothesis [4], for example, L2 learners rely more heavily on lexical and semantic cues rather than complex syntactic structures to derive meaning, which may lead to nonliteral sentence interpretations, especially for sentences with noncanonical word order. The current study asks whether the task dependence of misinterpretation effects can also be observed with L2 speakers. We report an experiment that compared the performance of L1 and L2 speakers on the agent/patient naming task and the plausibility judgment task following the presentation of unambiguous sentences with canonical or noncanonical word order.
Method: Participants read 32 German main clauses varying according to two factors: word order (subject-before-object/SO or object-before-subject/OS) and plausibility (plausible or implausible; see Table 1). Participants first read a sentence and then had to do one of two tasks: either judging the plausibility of the sentence (plausible versus implausible) or retrieving a phrase from the sentence corresponding to a cue given after sentence presentation ("do-er" or "acted-on"). The task required on each trial was signaled only after sentence presentation was complete. The stimuli were interspersed with 64 filler sentences and the experiment was run online on PC-Ibex. We recruited 48 L1 and L2 German speakers each via Prolific.
Results: For plausibility judgments, we found significant effects of Word Order and Plausibility. In addition, L2 speakers showed higher error rates overall, but had particular difficulty with implausible OS sentences (indicated by a significant interaction of Plausibility x Language). For agent-patient naming, however, we again found higher error rates for L2 speakers, but no between-group differences in the basic pattern, which includes a significant Word Order x Plausibility interaction (see Figure 1 and Table 2).
Discussion: The similar patterns observed for L1 and L2 for agent/patient naming suggests that the strategies and the grammatical information used to solve the task do not fundamentally differ between L1 and L2. However, L1 and L2 speakers apparently differ with respect to how syntactic and semantic information is weighted when giving plausibility judgments. While L1 speakers rely on syntactic information, L2 speakers show an over-reliance on event probability, resulting in high error rates for implausible sentences. This suggests that L1 and L2 speakers differ in how they approach metalinguistic tasks which may reflect differences in the sentence representations L1 and L2 speakers arrive at.

Table 1: A stimulus item in all four versions

plausible	SO	Der Lehrer hat den Fehler korrigiert. the-nom teacher has the-acc mistake corrected 'The teacher corrected the mistake.'
	OS	Den Fehler hat der Lehrer korrigiert. the-acc mistake has the-nom teacher corrected 'The teacher corrected the mistake.'
implausible	SO	Der Fehler hat den Lehrer korrigiert. the-nom mistake has the-acc teacher corrected 'The mistake corrected the teacher.'
	OS	Den Lehrer hat der Fehler korrigiert. the-acc teacher has the-nom mistake corrected 'The mistake corrected the teacher.'

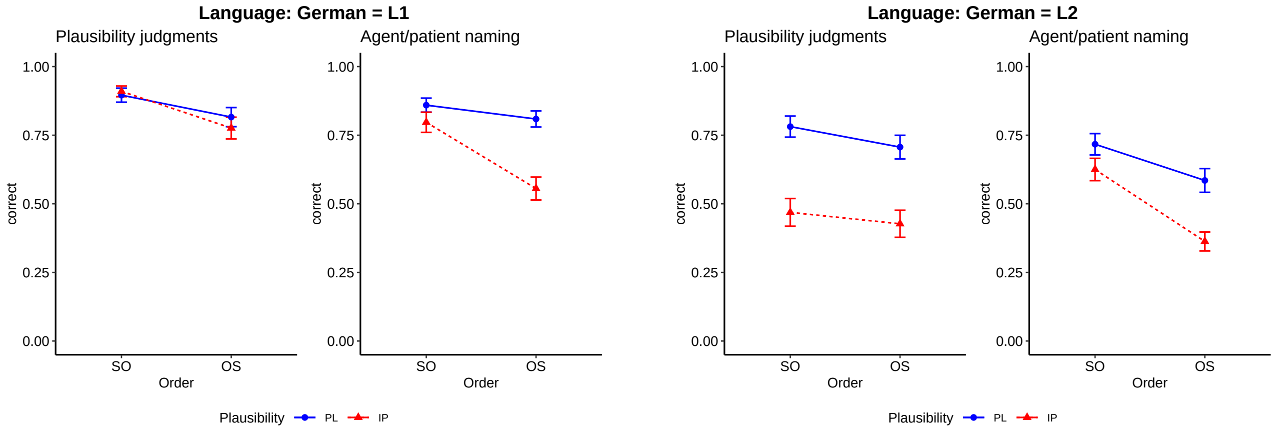


Figure 1: Percentages of correct answers. Error bars show standard errors by participants.

Table 2: Mixed-effects model for correctness data. Only significant effects and interactions are shown.

Contrast	Estimate	SE	z value	p value
Task = Plausibility judgments				
Word Order	0.547	0.198	2.77	< 0.01
Plausibility	0.772	0.231	3.34	< 0.01
Language	1.800	0.263	6.85	< 0.01
Plausibility × Language	-1.675	0.442	-3.79	< 0.01
Task = Agent-patient naming				
Word Order	0.8869	0.1255	7.07	< 0.01
Plausibility	0.8388	0.1255	6.68	< 0.01
Language	0.9909	0.1807	5.48	< 0.01
Word Order × Plausibility	-0.7415	0.2486	-2.98	< 0.01

References

[1] Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

[2] Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

[3] Meng, M., & Bader, M. (2025). Sentences with non-canonical word order – easy to interpret but hard to remember. *Journal of Cognitive Psychology*.

[4] Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics*, *27*(1), 3–42.

Session 5

Predictive processing adapts to prediction error in a non-linear fashion – that's why adaptation effects are so difficult to detect using LMERs

Katja Haeuser
Department of Psychology, Saarland University, Germany
khaeuser@coli.uni-saarland.de

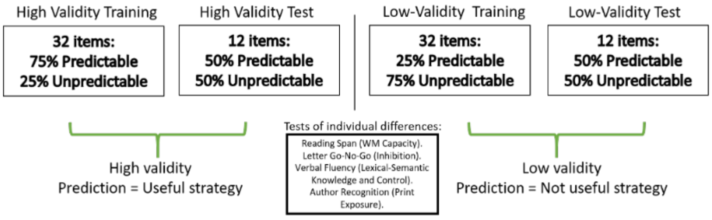
Background: Language processing is predictive in nature [1]. But do readers adapt their predictions flexibly depending on the likelihood of prediction success? Persistently generating predictions may not be rational when these predictions are rarely borne out. Prior findings remain inconsistent on whether linguistic prediction is adaptive [3-6]. Here, we suggest that this may be due to the use of common analysis methods (e.g. LMERs) which fail to capture effects when adaptation unfolds in a non-linear fashion over time.
Method: We investigated adaptiveness of linguistic prediction in a large web-based self-paced reading study (n=323 subjects; for design, see Figure 1). Participants were exposed to two blocks, a high- and low-validity block. Each block consisted of a training phase and a test phase. In the training phase, we trained participants to rely or not rely on linguistic prediction, by presenting participants with a large proportion of prediction-confirming or -disconfirming sentences (75% vs 25%, respectively; e.g., "Ariel was fired and desperately needed a job/loan to survive"; verified by means of prior cloze ratings). In the subsequent test phase, predictability effects were measured for equal proportions of predictable and unpredictable sentences. Here, we focus on the training blocks, in which we expected to find a critical three-way interaction (i.e., predictability * trial * validity), suggesting that predictability effects become larger (or smaller) with repeated exposure depending on validity.
Results: Data were initially analyzed using planned LMER models [7], which assume a linear relationship between predictor and outcome variables. We found main effects of predictability and validity, but no interactions across the board (see Table 1.1). However, to address the possibility that adaptation effects are non-linear over time, we re-analyzed the training data using generalized additive models (i.e., GAMMs, [8]) that specified a non-linear smooth term for trial number and an interaction variable between validity and predictability [9]. The model (see Table 1.2) showed significant non-linear effects of trial number across all four interaction terms. Contour plots showing the reading time difference between unpredictable and predictable sentences suggested that, in low-validity blocks (Figure 2, top right), predictability effects became progressively smaller with greater exposure to training trials. This was likely driven by predictable items, whose reading-time facilitation levelled out after repeated exposure to unpredictable items (see Figure 2, top left). For high-validity training blocks, trial effects were wigglier overall, but were also greater than zero for most of the experiment.[1]
Discussion: Our findings indicate that adaptation to prediction error unfolds non-linearly, by reducing the reading-time advantage that normally occurs for predictable sentences. Thus, readers dynamically regulate predictive processing, tuning their engagement up or down according to prediction reliability. These results support the view that predictive language processing operates under rational principles that flexibly optimize behavior in response to changing conditions.

[1] We followed up on the wiggly RT pattern for predictable items in high-validity blocks by running a model that included not only the predictability of trial-1 as a control predictor, but also the predictability of trial-2 and trial-3. The rationale for this new model was that the wiggly RT pattern could be driven by local context effects, when chunks of unpredictable sentences appeared in a row (note that the order of items was fixed over subjects and not counter-balanced). However, the EDF term associated with the critical smooth term in the new model was unchanged (i.e., the wiggliness remained). In addition, the two new control predictors did not explain much variance.

Table 1. LMER (1) and GAMM (2) output for reading times in the training blocks.

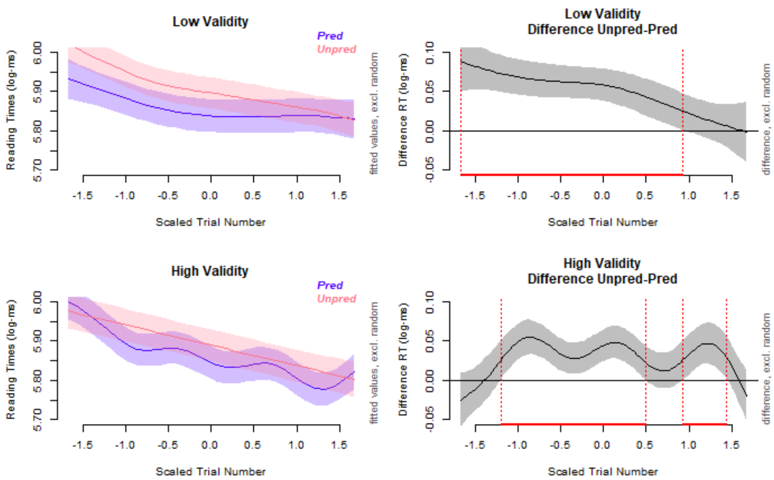
(1)				(2)				
lmer(log(noun_spill1_RT) ~ cond * val * scale(trial) + scale(length) + scale(lgSUBTLEX) + scale(position) + scale(previous_word_RT) + (1+val|subject) + (1|item), data=train)				bam(log(noun_spill1_RT) ~ condval + s(trial, by=condval) + s(length, k=5) + s(lgSUBTLEX) + s(position) + predictability_tial-1 + list + s(subject, bs="re") + s(subject, cond, bs="re") + s(item, bs="re"), data=train, AR.start = AR.start, rho=acf[2])				
Fixed Effects	**Estimate**	**SE**	**p-value**	**A. Parametric Terms**	**Estimate**	**SE**	**t-value**	**p-value**
Intercept	5.79	0.03		(Intercept)	5.90	0.04		
Predictability	0.01	0.002	< .001	Unpredictable-High	0.03	0.01	3.85	< .001
Validity	-0.01	0.005	.08	Predictable-Low	-0.005	0.01	-0.58	.56
Trial	-0.04	0.002	< .001	Unpredictable-Low	0.04	0.01	6.90	< .001
Length	0.02	0.004	< .001	Predictability Trial-1	0.02	0.004	4.98	< .001
lgSUBTLEX	-0.01	0.005	< .01	List	-0.03	0.01	-1.89	.06
Position	0.01	0.002	< .001	**B. Smooth Terms**	**EDF**	**Ref.df**	**F-value**	**p-value**
Previous RT	0.14	0.003	< .001	s(trial) * Predictable-High	7.25	8.23	35.74	< .001
Predictability * Validity	-0.001	0.001	.92	s(trial) * Unpredictable-High	1.00	1.00	86.90	< .001
Predictability * Trial	-0.003	0.002	.28	s(trial) * Predictable-Low	2.792	3.44	8.65	< .001
Validity * Trial	-0.006	0.002	< .05	s(trial) * Unpredictable-Low	3.43	4.23	53.15	< .001
Predictability * Validity * Trial	-0.001	0.002	0.69	s(length)	1.84	1.88	1.18	.23
				s(lgSUBTLEX)	1.00	1.00	0.02	.90
				s(position)	1.00	1.00	19.78	< .001

Figure 1. Experimental design.



Note. The split between training and test blocks was entirely implicit. The order of low- and high validity exposure was counterbalanced over lists.

Figure 2. Fitted smooth curves (left column) and difference smooths (right column) of reading times in low and high-validity blocks for the GAMM model in Table 1.2



References

[1] Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences, 27(11),* 1032-1052. [2] Ness, T., & Meltzer-Asscher, A. (2021). Rational adaptation in lexical prediction: The influence of prediction strength. *Frontiers in Psychology*, *12*, 622873. [3] Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10-20. [4] Baker-Kukona, A., & Hasshim, N. (in press). Mouse cursor trajectories capture the flexible adaptivity of predictive sentence processing. To appear in: *Journal of Experimental Psychology: Learning, Memory, and Cognition*. [5] Nieuwland, M. S. (2021). How 'rational' is semantic prediction? A critique and re-analysis of Delaney-Busch, Morgan, Lau, and Kuperberg (2019). *Cognition*, *215*, 104848. [6] Van Wonderen, E., & Nieuwland, M. S. (2023). Lexical prediction does not rationally adapt to prediction error: ERP evidence from pre-nominal articles. *Journal of Memory and Language*, *132*, 104435. [7] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. [8] Wood, S. N. (2017). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC. [9] Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86-116.

How gender information influences spontaneous speech in context

Amalia Spyromilio, Holly Jenkins, Elizabeth Wonnacott, Michael Ramscar
University Tübingen, University of Oxford
amalia.spyromilio@student.uni-tuebingen.de, holly.jenkins@education.ox.ac.uk,
elizabeth.wonnacott@education.ox.ac.uk, michael.ramscar@uni-tuebingen.de

While historically the functionality of grammatical gender (aka noun class) has often been called into question, from a predictive / information theoretic perspective one possible function of noun class seems obvious: nouns are the least predictable part of speech, and gender marked articles can help make nouns more predictable in context. Similarly, prenominal adjectives can be seen to provide similar information in languages without noun classes, such as English (Dye et al., 2018). Employing a written paradigm, Hoppe et al. (2025) found that when English and German speakers produced noun phrases in context (German has three noun classes), German speakers used fewer prenominal colour adjectives than English speakers when articles helped predict nouns in context (but not when they were uninformative).

To expand on these findings, we examined whether these effects are visible when answers are spoken rather than written, and also whether subsequent mentions affect the amount of information in the noun phrases produced. German speaking participants were presented with four coloured objects and asked to describe their movements so that they could later be reproduced. One target object either had the same noun class as the other three objects (uninformative condition) or a unique noun class (informative condition; Figure 1).

In contrast to previous findings from typed responses, participants speaking spontaneously produced high numbers of adjectives regardless of whether the articles were informative or not. Analysing the articulation of the responses, we found that at first mention they tended to speak slower and louder in the informative condition as compared to the uninformative condition, whereas at second mention they spoke faster and quieter when describing the items in the informative condition, and slower and louder in the uninformative condition. That is, we observed an interaction between the informativity of the context and repeated mentions of the target item, with speakers appearing to adjust to both the information provided by articles and the ambiguity of the context, an effect that was most clearly visible in their entire noun phrase productions. We suggest that this not only reflects speakers' sensitivity to gender information in context, but also that they adjust their later communicative behaviour according to what they learn about how this information helps or fails to help disambiguate the ambiguity of a message in that context.

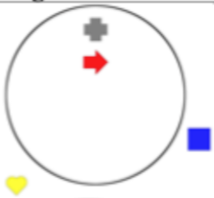
Condition	Image	Target	Distractors
German informative		**der** Pfeil, m.	**das** Kreuz, n. **das** Quadrat, n. **das** Herz, n.
German uninformative		**der** Blitz, m.	**der** Kreis, m. **der** Stern, m. **der** Pfeil, m.

Figure 1. The two different conditions participants encountered in the experiment. In the informative condition (top), the target noun has a distinct noun class. In the uninformative condition (bottom) all nouns share the same noun class. The objects start outside the reference frame (the large circle) and then successively moved into it and around it, such that participants were required to mention each target object at least twice.

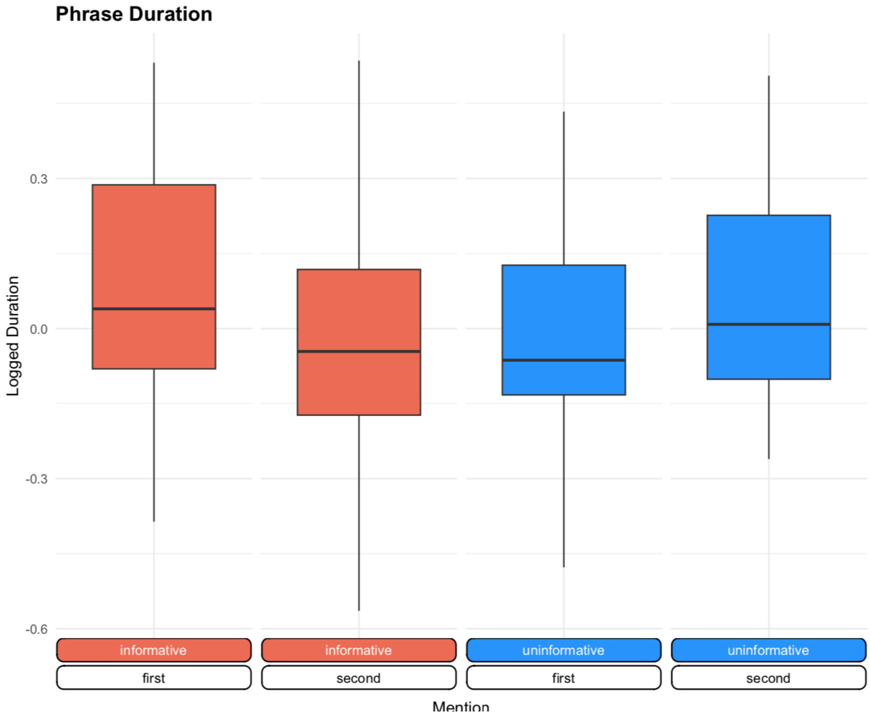


Figure 2. Phrase durations of identical article-adjective-noun phrases for the four possible test conditions: informative article first mention, informative article second mention, uninformative article first mention and uninformative article second mention. Depending on the context (informative or uninformative) of first mention, behaviour on second mention changes.

References

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2018). Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. Topics in cognitive science, 10(1), 209-224.

Hoppe, D., Gibson, Van Rijl, J., Hendriks, P. & Ramscar, M. (2025) How grammatical gender supports efficient communication, Proceedings of the Annual Meeting of the Cognitive Science Society

Linguistic complexity measures (surprisal, entropy, semantic similarity, syntactic node counts) differentially impact initial vs. re-reading

Susanne Eisenhauer[1], David Hernández-Gutiérrez[1,2], Simona Mancini[1,3]
[1]Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain, [2]University of Rochester, Rochester, New York, USA, [3]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
s.eisenhauer@bcbl.eu

More complex texts are harder to read. This complexity may arise at different linguistic levels. For example, at the semantic level, words that are less well aligned with the meaning of the preceding context are expected to increase reading duration [1]. Likewise, higher syntactic complexity increases reading time [1,2]. However, since few eyetracking studies aimed at disentangling contributions from different types of linguistic complexity [1,2], how and when different types of linguistic complexity impact reading behavior is not well understood. In the present study, we used metrics from large language models to simultaneously investigate the impact of semantic and syntactic complexity on eyetracking measures that reflect different types of reading behavior. We used publicly available data from the MECO eyetracking corpus [3], focusing on 49 participants who read twelve Wikipedia-style articles in their native Spanish. Information-theoretic measures, next-word surprisal and entropy [4] based on GPT-2, served as broad measures of linguistic complexity. Semantic complexity was reflected in the semantic similarity of each content word to the three preceding content words, based on word vector representations from the ELMo language model [5]. Syntactic complexity was reflected in two node count measures based on the Stanford constituency parser [6]: Bottom-up node count (the number of nodes that are closed at each word) reflects a parsing strategy in which the constituency structure is only formed once the final word of a constituent is reached, while top-down node count (the number of nodes that are opened at each word) reflects a parsing strategy during which the reader predicts the upcoming constituency structure in advance [7]. Using (generalized) linear mixed models on content words (all variance inflation factors <4), we found that surprisal and entropy enhanced gaze and total reading duration (Table 1), while higher entropy in addition reduced the number of backwards regressions onto a word (Table 2). Higher semantic similarity, contrary to our expectation, increased total reading duration (Table 1) as well as the number of first-pass and total backward regressions on a word (Table 2). Of the syntactic measures, only bottom-up and not top-down node count significantly affected the investigated eye movement parameters, in line with prior evidence that bottom-up node count is a better predictor of brain activity during comprehension [7]. Effects of bottom-up node count were in an unexpected direction, as a higher node count – associated with increased syntactic complexity [7] - reduced total reading time (Table 1) as well as the number of first-pass and total backward regressions on a word (Table 2). Our findings reveal that different measures of linguistic complexity differentially affect reading behaviour. In detail, more surprising as well as high-entropy words are read longer during first-pass reading, while low-entropy words, words that are semantically more similar to the preceding context, as well as words with a less complex bottom-up syntactic structure, are re-read more often. Thus, broader word difficulty enhances initial reading duration, while re-reading focuses on semantically and syntactically ‚easier‘ words, potentially aiding the semantic interpretation of the text as a whole. These findings extend prior evidence that mainly focused on the role of re-reading for integrating complexities and disambiguation [8]. Moreover, they can serve as basis for future studies aimed at identifying impaired subprocesses in reading-impaired individuals, such as people with dyslexia or acquired language disorders.

Table 1. Results from linear mixed models investigating effects of linguistic complexity metrics on gaze and total reading duration.

	Gaze duration				**Total reading duration**			
	Estimate	SE	t	p	Estimate	SE	t	p
Surprisal	6.465	1.500	4.310	1.79e-05	10.287	2.919	3.525	0.0004
Entropy	4.420	1.457	3.033	0.002	6.341	2.831	2.240	0.025
SemSim	-0.734	1.320	-0.556	0.579	5.337	2.553	2.090	0.037
BU node count	1.653	1.634	1.012	0.312	-12.526	3.180	-3.939	8.74e-05
TD node count	2.123	1.530	1.388	0.166	2.664	2.982	0.893	0.372
SE = standard error, SemSim = semantic similarity, BU = bottom-up, TD = top-down.								

Table 2. Results from generalized linear mixed models investigating effects of linguistic complexity metrics on the number of first-pass and total regressions.

	First-pass regressions				**Total regressions**			
	Estimate	SE	z	p	Estimate	SE	z	p
Surprisal	0.114	0.066	1.721	0.085	0.066	0.040	1.634	0.102
Entropy	-0.153	0.063	-2.421	0.015	-0.067	0.039	-1.727	0.084
SemSim	0.129	0.048	2.677	0.007	0.111	0.032	3.476	0.0005
BU node count	-0.402	0.088	-4.582	4.6e-06	-0.399	0.049	-8.217	< 2e-16
TD node count	0.090	0.071	1.275	0.202	-0.006	0.042	-0.148	0.882
SE = standard error, SemSim = semantic similarity, BU = bottom-up, TD = top-down.								

[1] Pynte, J., New, B., & Kennedy, A. (2008). A multiple regression analysis of syntactic and semantic influences in reading normal text. Journal of Eye Movement Research, 2(1), 1–11.

[2] Demberg, V., & Keller, F. (2008). Data from eyetracking corpora as evidence for theories of syntactic processing complexity. Cognition, 109(2), 193–210.

[3] Siegelman, N., Schroeder, S., Acartürk, C., et al. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). Behavior Research Methods, 54, 2843–2863.

[4] Hale, J. (2016). Information-theoretical Complexity Metrics. Language and Linguistics Compass, 10(9), 397–412.

[5] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv, 1802.05365.

[6] Klein, D., & Manning, C. D. (2003). Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, 423–430.

[7] Giglio, L., Ostarek, M., Sharoh, D., & Hagoort, P. (2024). Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. Proceedings of the National Academy of Sciences of the United States of America, 121(11), e2310766121.

[8] Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125-134.

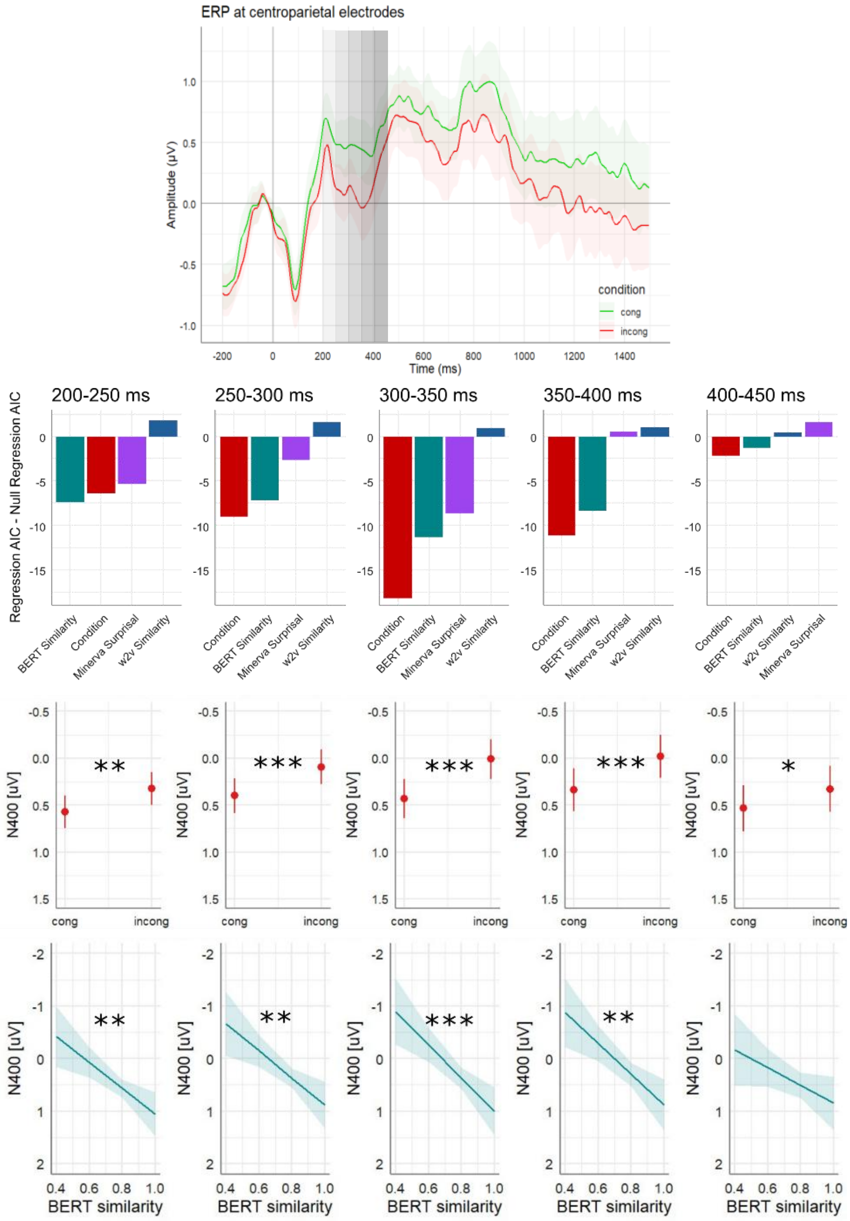
Session 6

Context-based encoding of novel meanings after minimal exposure to natural text: an EEG study on integration of linguistic chimeras

Fabio Marson[1], Giulia Loca[1], Marco Ciapparelli[1], Marco Marelli[1]
[1] Department of Psychology, University of Milano-Bicocca, Milan, Italy
fabio.marson@unimib.it

Background: During communication, meanings expressed through language might be unknown to one of the parts involved. In this case, language works as a source of experience for learning concepts that were never encountered before through other modalities (e.g., sensorimotor; [1]). The distributional properties of language have been found to support the integration of novel meanings during sentence processing (e.g., via sentential context [2]). Data-driven computational measures of such regularities (e.g., surprisal or distributional similarity obtained by language models) can predict event-related potential amplitude elicited by word processing in context [3,4]. Thus, we tested whether minimal exposure to naturalistic text could foster the integration of novel meanings observable in the N400 amplitude, and if such modulation could be predicted by data-driven computational measures. **Method:** In the current study, we used textual stimuli known as linguistic chimeras: language-based representations constructed by merging two compatible concepts. For example, replacing the word *cat* in "*the cat was chasing the mouse*" and *eagle* in "*the sky is crowded by eagles flying*" with a single pseudoword (e.g., *mohalk*) would produce a chimera, which would probably be represented as an animal that eat mice and can fly. We identified a total of 54 different concepts from a feature-based norms database [5], paired in 27 chimeras, and extracted naturalistic sentences related to each concept from the ItWac corpus [6]. 45 participants read sentences on a screen in rapid serial visual presentation in a total of 189 trials. In each trial, two sentences (one for each concept of the chimera) provided the semantic context of the chimera (encoding phase). A third sentence, that could be either congruent (semantically related/same label), incongruent (semantically unrelated/same label), or control (semantically unrelated/different label), was shown after (testing phase). The presence of a congruency effect (i.e., stronger N400 following the novel label in incongruent condition compared to congruent ones) would signal the success of semantic integration. Additionally, three data-driven computational measures were adopted to predict ERPs amplitude: two data-driven measures of cosine similarity (one from word2vec and one from Italian BERT) and surprisal (from Minerva, an Italian autoregressive LLM). Congruency was assessed using a cluster-based permutation analysis approach, while contribution of computational measures in the temporal profile of N400 was assessed by models' AIC comparison. **Results:** We observed a significant Condition effect (incongruent > congruent) at a topography and latency compatible with the N400. At a more fine-grained analysis, Condition predicted amplitude in all the time windows related to N400 (200~450 ms). BERT predicted N400 amplitude in the 200~400 ms time-window, outperforming Condition in the earliest time window (200~250 ms). In the early 200~350 ms time windows, surprisal was also found to significantly predict N400. **Discussion:** These results suggest that minimal exposure to naturalistic written sentences fosters the integration of novel meanings and that such process is observable at the neural level. Computational measures suggest that language distributional properties (only when accounting for sentential context, i.e., BERT & Surprisal > word2vec) play a role in guiding novel meanings processing, especially in the early stages of novel words retrieval. Taken together, these results fit with the idea that the semantic system exploits linguistic experience to facilitate the learning process of novel words meaning.

Figure1. Event-related potential time course related to the presentation of the novel words in the testing phase (top panel). Results of the model comparison analysis in terms of AIC (second row), BERT cosine similarity effect (third row) and condition effect (last row) in the N400 time window for each time bin (50 ms each, ranging from 250 to 450 ms). Grey bars in the top panel represent the 50 ms bins used in the second analysis to characterize the time course of N400.



References

[1] Dove, G. O. (2023). Rethinking the role of language in embodied cognition. Philosophical Transactions of the Royal Society B, 378(1870), 20210375.

[2] Lazaridou, A., Marelli, M., & Baroni, M. (2017). Multimodal word meaning induction from minimal exposure to natural text. Cognitive science, 41, 677-705.

[3] Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. Language, Cognition and Neuroscience, 32(9), 1192-1203.

[4] Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2024). Strong prediction: Language model surprisal explains multiple N400 effects. Neurobiology of language, 5(1), 107-135.

[5] Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: A feature-based analysis and new norms for Italian. Behavior research methods, 45, 440-461.

[6] Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. Language resources and evaluation, 43, 209-226.

This page may be used only for additional information about the generally less-known language you are targeting in the abstract.

Vision Language Model Representations Predict EEG Response to Visual and Auditory Attributes in Property Verification

Harshada Vinaya, Sean Trott, Seana Coulson
University of California San Diego (USA)
Correspondence to: hyadav@ucsd.edu

Background: A major issue in 21st century psycholinguistics has been the extent to which lexical semantics recruits perceptual “grounded” resources versus amodal representations derived from the statistics of language. Here we use advances in NLP to investigate how well measures of both linguistic distribution information and experiential grounding predict brain responses to words in a property verification task. To operationalize distributional contributions to word meaning, we use embeddings from GPT2, a Large Language Model (LLM) trained only on language corpora [5]. To approximate contributions from visual experience, we use embeddings from the vision language model CLIP that links images to text [6]. Though CLIP has a text encoder with a similar architecture to GPT2, its training incorporates information about what things look like - thus, embeddings from CLIP’s text encoder are informed at least partially by a visual grounding process [4]. If human semantic representations primarily reflect distributional information, EEG elicited by words such as “red” should be well-predicted by GPT2. If semantic representations also reflect grounded resources, predictions of the brain response to these words should be improved by CLIP.

Method: We conducted a reanalysis of single trial EEG data collected from 19 participants who performed the property verification task reported in [3]. This reanalysis includes the TRUE response trials (e.g., “APPLE - red”, “DYNAMITE - booms”) with visual (e.g., ”red”, n = 1373) and auditory (e.g., ”loud”, n =1160) property words. As a measure of semantic dissimilarity, we measured the cosine distances between the concept and property vectors obtained from the final layer of both GPT-2 and CLIP. As control predictors, we include logarithmic word frequency [2] and number of letters in each word. All predictors were z-scored. We measured mean voltage at each electrode in successive 100 ms windows, from the onset of the property word until 700 ms, and fit the following five models for each of seven windows. The NULL model included no lexical or semantic predictor, but only interactions of scalp-dimensions [7] and modality type (auditory/visual) as fixed effects, and item and subject-level random effects [voltage ~ (X+Y+Z) * modality + (1|word) + (1|subject)]. The rest of the models progressively include predictors of interest in interaction with the scalp dimensions and modality. The second model, BASE, included word frequency and number of letters. The third model added GPT2 cosine distance, and the fourth model added CLIP cosine distance to the BASE. The final/full model added all the predictors. We used Akaike Information Criterion (AIC) scores for statistical model comparison considering AIC differences of 10 as robust evidence for a more likely model.

Results: Figure 1 compares the ΔAIC (scaled to the null model) values for the Base, Base + GPT2, Base + CLIP, and Base + GPT2 + CLIP models across the seven 100ms windows. Results show that the Base model improves the fit above the null model from 200-700 ms, GPT2 improves over Base in 0-400 and 500-700 ms, and CLIP improves over Base from 200-700 ms. The full model improves over GPT2 and CLIP when included alone across all windows except the first one (i.e. 0-100 ms).

Discussion: After the first 100ms, the best fit to the neural data came from the combined model. This suggests that human semantic representations in this task are informed both by the statistical regularities of a word and by its associated visual properties. Superior performance of GPT2 in the first 100ms may support the proposal that early word processing is sensitive to pure distributional information [1].

This page serves for Tables, Figures, and References.

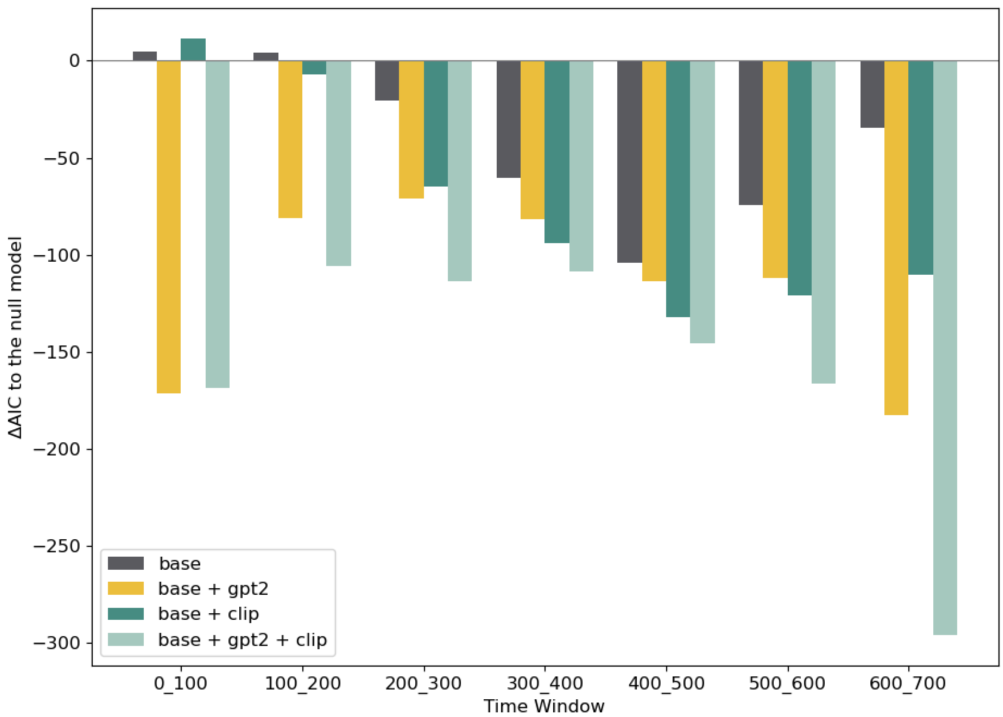


Figure 01: AIC comparisons of the four models of theoretical interest. AIC is scaled by subtracting the null model's AIC. The X-axis marks each successive 100 ms time window. A more negative value indicates a better model fit relative to the null model.

References

[1] Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), Symbols and Embodiment: Debates on meaning and cognition (p. 0). Oxford University Press.

[2] Brysbaert, M., New, B., & Keuleers, E. (2012, December). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44 (4), 991–997. Retrieved 2024-01-14, from https://doi.org/10.3758/s13428-012-0190-4 doi: 10.3758/s13428-012- 0190-4

[3] Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality Switching in a Property Verification Task: An ERP Study of What Happens When Candles Flicker after High Heels Click. *Frontiers in Psychology*, 2. doi: 0.3389/fpsyg.2011.00010

[4] Jones, C. R., Bergen, B., & Trott, S. (2024). Do Multimodal Large Language Models and Humans Ground Language Similarly? Computational Linguistics, 50(4), 1415–1440. doi: 10.1162/coli_a_00531

[5] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI.

[6] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. arXiv. 10.48550/arXiv.2103.00020

[7] Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2018, September). An electrophysiological megastudy of spoken word recognition. Language, Cognition and Neuroscience, 33 (8), 1063–1082. doi: 10.1080/23273798.2018.1455985

This page may be used only for additional information about the generally less-known language you are targeting in the abstract.

Comprehension of Pragmatically Licensed Sentential Negation and Its Influence on Memory Retention

Svetlana Mnogogreshnova[1], Sol Lago[1], Esther Rinke[1], Petra Schulz[1], Clara Vilà Dolado[1]
[1] Goethe University Frankfurt
Mnogogreshnova@em.uni-frankfurt.de

Negation is a universal phenomenon investigated in linguistics and in psychology, but some findings are still controversial. When it comes to the processing of negation, some accounts claim that negation is a two-step process, in which comprehenders first process the information that is being negated and then integrate the negator to derive the factual state of affairs [1, 2]. This two-step scheme is argued to hold even in the face of pragmatic licensing, particularly for postverbal negation [3]. But it also has been postulated that a pragmatically licensing discourse context facilitates comprehension [4] and allows the full meaning of the negative utterance to be processed in one step [5, 6]. As for the influence of negation on memory, findings are mixed: some studies show that negation makes it more difficult to remember the negated information [7, 8], whereas other studies show that the negated items can be falsely remembered as existing [9]. It is yet unclear how pragmatically licensed negated sentences influence processing and memory retention.

Our study tested 40 German native speakers. Participants performed a two-alternative forced choice task (2AFC) followed by a 5-minute Flanker task (as a distractor) and a memory test. In the 2AFC task, participants first read a context that pragmatically licensed negation by introducing a clear question under discussion (QUD): in each context, a character faced a choice between two actions and had to choose one. Each story ended with a target sentence revealing the character's decision either via an affirmative or a negative statement (see Table 1). The target sentence was followed by two pictures representing the two actions mentioned in the story, and participants had to choose the picture matching the target sentence. The results of mixed-effects models revealed that reaction times were significantly longer in the negative than in the affirmative condition (Fig. 1). Consistent with the two-step accounts [1, 2, 3], this pattern indicates higher processing difficulty for negative sentences, despite the presence of a pragmatically licensing context.

In the memory test, participants were presented with probes in the form of verbal phrases (e.g., *eine Birne schälen, "to peel a pear"*). The probes represented either actions that had been encountered in the 2AFC task in an affirmative or a negative target sentence, or filler probes representing actions that had not been mentioned before (e.g., *einen Ball werfen, "to throw a ball"*). Participants had to determine whether the action had taken place in the stories in the 2AFC task. The results showed that the negative condition (i.e., probes mentioned in negative sentences in the 2AFC task) caused significantly longer reaction times and higher error rates compared to both the affirmative and the filler conditions. Importantly, this pattern was not due to the general difficulty in rejecting a probe, since there was no significant difference between the time it took to reject a filler and the time it took to confirm a probe in the affirmative condition (Fig. 2A). Neither was there a general bias towards *yes*-responses, since the number of errors in the affirmative condition was significantly higher than in the filler condition (Fig. 2B). We conclude that negative utterances, although they were pragmatically licensed, are still more difficult to remember than their affirmative counterparts.

Table 1. A sample item set for the 2AFC task.

Context	
Papa macht das Mittagessen und bittet Mama um Hilfe. Mama soll den Käse reiben. Aber vielleicht schält sie lieber die Birne. Mal schauen! “Dad is making lunch and asks Mum for help. Mum has to grate the cheese. But perhaps she would rather peel the pear. Let's see!”	
Target sentences	
affirmative	negative
Sie reibt den Käse schon. “She is already grating the cheese.”	*Sie reibt den Käse nicht.* “She does not grate the cheese.”
Sie schält die Birne schon. “She is already peeling the pear.”	*Sie schält die Birne nicht.* “She does not peel the pear.”

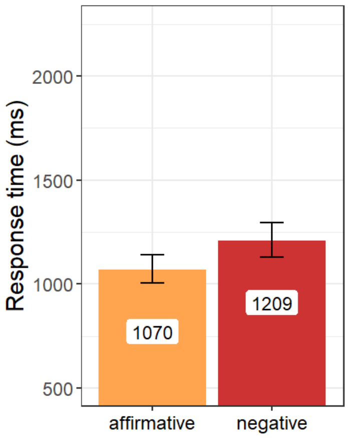


Figure 1. Response times in the 2AFC task.

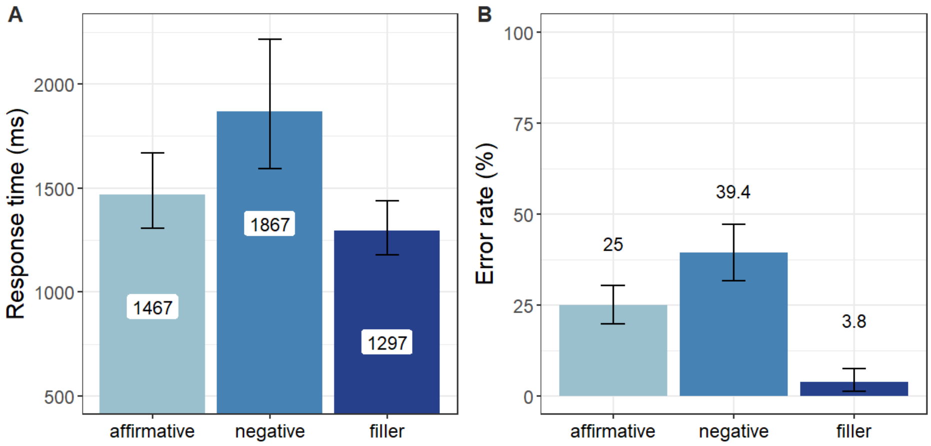


Figure 2. Response times (A) and error rates (B) in the memory test.

References

[1] Kaup, B., Ludtke, J., & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. In Proceedings of the annual meeting of the Cognitive Science Society (Vol. 27, No. 27).

[2] Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed?. Journal of Pragmatics, 38(7), 1033-1050.

[3] Lee, M., Kim, H. S., Lee, G., Noh, Y., & Kim, S. Y. (2025). Incremental processing of postverbal negation: ERP evidence from Korean. Journal of Neurolinguistics, 75, 101259.

[4] Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In Proceedings of the 28th annual conference of the cognitive science society (Vol. 27, pp.1735–1740).

[5] Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why We Simulate Negated Information: A Dynamic Pragmatic Account [PMID: 21058217]. Quarterly Journal of Experimental Psychology, 63 (12), 2305–2312. https://doi.org/10.1080/17470218.2010.525712.

[6] Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. Language, Cognition and Neuroscience, 31(5), 683-698.

[7] Mayo, R., Schul, Y., & Rosenthal, M. (2014). If you negate, you may forget: Negated repetitions impair memory compared with affirmative repetitions. Journal of Experimental Psychology: General, 143(4), 1541.

[8] Zang, A., Beltrán, D., Wang, H., Rolán, K., & de Vega, M. (2023). The negation-induced forgetting effect remains even after reducing associative interference. Cognition, 235, 105412.

[9] Maciuszek, J., & Polczyk, R. (2017). There was not, they did not: May negation cause the negated ideas to be remembered as existing?. PLoS One, 12(4), e0176452.

Where L2 Still Looks L2: Spatial Undershoot and Logographic Saccadic Programming in Advanced L1 Chinese/L2 English readers

Shiyu He[1], Petar Milin[2], Dagmar Divjak[2]

[1]University of Bristol [2]University of Birmingham

celia.shiyu.he@bristol.ac.uk

Background Saccadic targeting during reading integrates visual, linguistic, and motor planning processes [1, 2]. While second language (L2) learners may develop native-like fixation durations in their L2 [3], it is unclear whether spatial adaptations extend to their first language (L1), especially when scripts are vastly different. This study investigates whether advanced English proficiency and exposure reshape saccadic precision and landing stability during L2 English reading, and whether these patterns transfer to L1 Chinese reading.

Method Forty Chinese university students were categorised into Advanced and Upper Intermediate L2 English learner groups using Principal Component Analysis and Cluster Analysis of English vocabulary size, IELTS scores, and years of residence in the UK [4]. Participants read 24 matched Chinese (38–71 words each, M = 55.9, SD = 10.3) and English (44–98 words each, M = 70.7, SD = 12.9) texts in a counterbalanced order while their eye movements were recorded. Relative Landing Positions (RLPs) were computed using the x-coordinate of the first fixation relative to word centre, capturing both central tendency and dispersion.

Results In L1 Chinese reading, both groups demonstrated highly stable, centre-based RLPs that did not vary with word length, aligning with native-like saccadic patterns [5, 6]. In contrast, during L2 English reading, both groups showed increasingly leftward RLPs of the first fixation from word centre as word length increased, consistent with spatial undershoot patterns attributed to oculomotor constraints or motor planning strategies. This spatial undershoot occurred despite the Advanced group showing native-like fixation durations, indicating that temporal and spatial components of eye movement control may be governed by different mechanisms.

Discussion The present findings highlight a dissociation between temporal fluency and spatial targeting in bilingual reading. While L2 proficiency and exposure may support native-like fixations, they do not affect saccadic targeting. The persistence of L1 Chinese saccadic patterns suggests that spatial aspects of eye movement control are script-specific and relatively resistant to change [3], particularly when the scripts of the two languages differ markedly. These findings suggest that L1-based saccadic routines persist in L2 reading, highlighting the need for bilingual models to account for the limited cross-script adaptability of spatial oculomotor control.

References

[1] Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

[2] Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

[3] He, S. (2024). *Learning to optimise reading: An exploratory study of eye movement transfer across languages* [Doctoral dissertation, University of Birmingham].

[4] He, S., Divjak, D., & Milin, P. (2025). Optimising participant grouping methods in bilingualism studies: Insights from eye-tracking data. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.24019.he

[5] Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word N+1 modulates preprocessing effectiveness of word N+2 in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1669–1676. https://doi.org/10.1037/a0019329

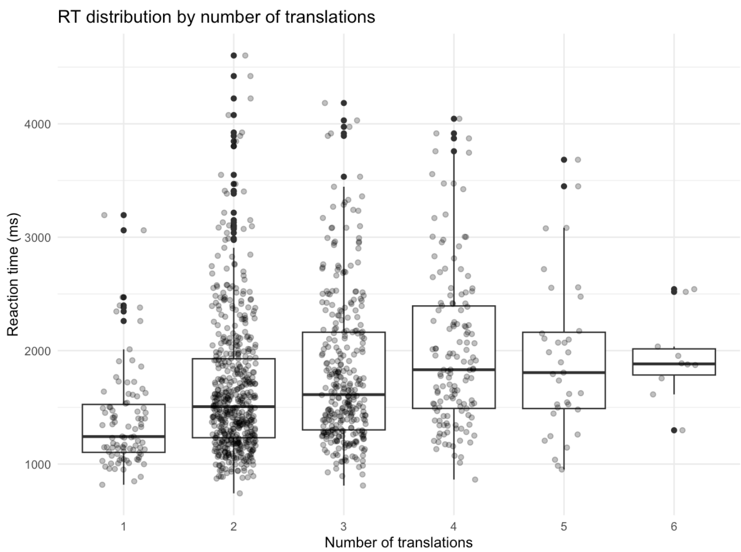
[6] Liu, Y., Reichle, E. D., & Wang, J. (2018). Word predictability affects saccade targeting during Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 756–764. https://doi.org/10.3758/s13423-017-1357-x

Session 7

Language planning and lexical competition: evidence from lexical processing for sociopragmatic differentiation of lexical doublets

Tamara Butigan[1], Norbert Vanek[1] Robert Greenberg[2],
[1] University of Auckland, New Zealand [2] University of Newcastle, Australia
Corresponding author: tbut559@aucklanduni.ac.nz

Background: This study investigates the status of lexical doublets (syntactically and semantically equivalent lexical items associated with either the Serbian or Croatian language norm) in the mental lexicons of Bosnian speakers. It is the first experimental study using time-sensitive measures to examine how planned language change affects lexical representation in Bosnian speakers. While translation ambiguity research [1,2] has generally focused on words with different meanings or forms, our study offers the novel perspective that sociopragmatic associations alone may lead to the differentiation of doublets in the mental lexicon. We draw on the Revised Hierarchical Model with Translation Ambiguity [3, 4] and Blythe and Croft's model of variant competition to investigate whether lexical items that differ only in sociopragmatic aspects are represented as separate competing lexical entries or as single entries with multiple surface forms. Our findings advance the theoretical understanding of how language planning initiatives impact lexical processing and representation.
Method: We recruited 24 native Bosnian–English bilinguals, comprising 12 professional translators (average 23.4 years of experience, *SD* = 6.1) and 12 advanced Bosnian students of English. Each participant completed a web-based, timed translation production task: 51 English target words (all corresponding to Bosnian doublets differing by sociopragmatic norm) and 49 distractors were presented in isolation in randomized order. Participants provided Bosnian translations of stimuli using a microphone; speech-onset latencies were manually determined via waveform analysis in Audacity. To quantify translation ambiguity, we calculated Shannon entropy [5,6] from the distribution of translations produced per stimulus. Continuous predictors included word frequency on the Zipf scale [7], familiarity, concreteness, word length, and entropy. Linear mixed-effects models were fitted in R using the lme4 package [8], with random intercepts for participant and stimulus, and random slopes for entropy where supported, to evaluate effects of number of alternatives (Min(1), Max(9) and entropy on log-RTs for doublet (D-type) items (examples on p. 3).
Results: Linear mixed-effects modeling of D-type (doublet) items revealed that introducing a second translation alternative imposes a substantial processing cost, consistent with two separate lexical entries being co-activated. Specifically, a significant main effect of number of translations ($F(5, 171.43) = 4.86$, $p < .001$) was driven by a 273 ms increase in response time when moving from unambiguous words (one translation) to doublets (two translations), $SE = 92.4$, $t(260.8) = –2.958$, $p = .039$, whereas a further increase in the number of elicited translations (3+) did not yield proportional RT increases (all ps > .32). Crucially, there was no significant difference in RTs between SR and HR variants ($\beta = –0.018$, $p = .554$), and total entropy robustly predicted slower responses ($\beta = 0.115$, $p = .003$), indicating that the core ambiguity cost stems from managing two distinct entries rather than surface variation within a single entry.
Discussion: The introduction of a second translation caused a 273 ms processing cost, while additional alternatives didn't produce proportional delays. This suggests that doublets function as distinct co-activated lexical entries rather than variants of one lemma. Despite identical semantics, social indexing alone drove competition costs and entropy effects on RTs. This extends the RHM-TA model by showing that the bilingual lexicon encodes social meaning as rigorously as semantics, requiring speakers to resolve both lexical ambiguity and embedded social identity cues.



Pairwise Contrasts for Response Times

Contrast	Difference_ms	SE	df	t_ratio	p_value
1 vs 2	-273.4	92.4	260.8	-2.958	0.0392*
2 vs 3	-105.8	52.0	467.4	-2.036	0.3235
3 vs 4	-113.3	81.0	259.6	-1.399	0.7277
4 vs 5	-16.8	184.0	62.0	-0.091	1
5 vs 6	-60.8	274.0	180.5	-0.222	0.9999

[a] Significance threshold: *p < 0.05

Figure 1. Distribution of response times (ms) across items with different numbers of translations (1-6) (left) and Pairwise contrasts for response times (in milliseconds) between stimuli with a different number of translations (right)

References

[1] Schwieter, J. W., & Prior, A. (2020). Translation Ambiguity. In A. B. Cieślicka & R. R. Heredia (Eds.), *Bilingual Lexical Ambiguity Resolution* (pp. 96–125). Cambridge University Press. https://doi.org/10.1017/9781316535967.006

[2] Degani, T., Prior, A., Eddington, C. M., Arêas da Luz Fontes, A. B., & Tokowicz, N. (2016). Determinants of translation ambiguity. *Linguistic Approaches to Bilingualism*, *6*(3), 290–307. https://doi.org/10.1075/lab.14013.deg

[3] Kroll, J. F., Hell, J. G. V., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

[4] Blythe, R. A., & Croft, W. (2012). S-Curves and the Mechanisms of Propagation in Language Change. *Language*, *88*(2), 269–304.

[5] Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., & Carl, M. (2016). Word Translation Entropy: Evidence of Early Target Language Activation During Reading for Translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* (pp. 183–210). Springer International Publishing. https://doi.org/10.1007/978-3-319-20358-4_9

[6] Wei, Y. (2022, September 12). *Entropy as a measurement of cognitive load in translation*.

[7] Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

[8] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Doublets (D-type) in this study refer to Bosnian translations that belong to a *lexical pair,* where one member of the pair corresponds to the Croatian-standard variant ("Hr") and the other to the Serbian-standard variant ("Sr"). For instance, the English stimulus 'translation' can elicit either 'prijevod' or 'prevod'. More examples are provided in the table below. "Elicited Hr" and "Elicited Sr" indicate which variant could be produced by participants; "Other" lists any additional non-doublet responses.

Example of D-type items (doublets) in Bosnian elicited translations

Stimulus	Elicited_Hr	Elicited_Sr	Other
translations	prijevod	prevod	—
employee	zaposlenik	uposlenik	radnik
transportation	prijevoz	prevoz	transport; transportiranje
sponsored	sponzoriran	sponzorisan	podržan
qualified	kvalificiran	kvalifikovan	osposobljen
municipality	općina	opština	—
criticise	kritizirati	kritikovati	—

Augmenting the psycholinguistic toolbox with AI-generated word characteristics

Marc Brysbaert[1], Javier Conde[2], Pedro Reviriego[2], Gonzalo Martínez[3]
[1] Department of Experimental Psychology, Ghent University; [2] ETSI de Telecomunicación, Universidad Politécnica de Madrid, [3] Universidad Carlos III de Madrid
marc.brysbaert@ugent.be

Background: Until recently, there were two main ways of obtaining information about words and expressions. The first was to analyze large text data sets (corpora) and calculate the frequency with which words and phrases occur, as well as the typical contexts in which they occur. The second was to ask participants to provide subjective information about words and phrases, such as the familiarity of the stimuli or the age at which they are typically acquired. The development of large language models has given us a third option. Instead of asking participants for information, we can query large language models. The results show that the information obtained from those models is just as good and often even better than the information obtained from people [1,2,3]. This talk will discuss the recent developments and indicate how results can further be improved by finetuning the model to a few thousand stimuli.

Method: Large language models (mostly GPT4) are queried to obtain information about word familiarity, concreteness, valence, arousal, and age of acquisition. Importantly, the results are compared to human data (e.g., ratings provided by people, predictive value of lexical decision times and megastudies).

Results: Converging data will be discussed for several languages, including English, Spanish, German, and Mandarin Chinese. These show that LLM-generated word information provides an interesting addition to the tools currently available.

In the second part of the talk, we will discuss developments that are not yet available in published form, such as the extension from single words to multiword expressions, direct prediction of lexical decision times obtained in megastudies, and the use of model finetuning to generate a large dataset on the basis of a few thousand seed words. Each time the outcome of LLM-queries is compared to human performance, sometimes newly collected to test specific predictions based on the models.

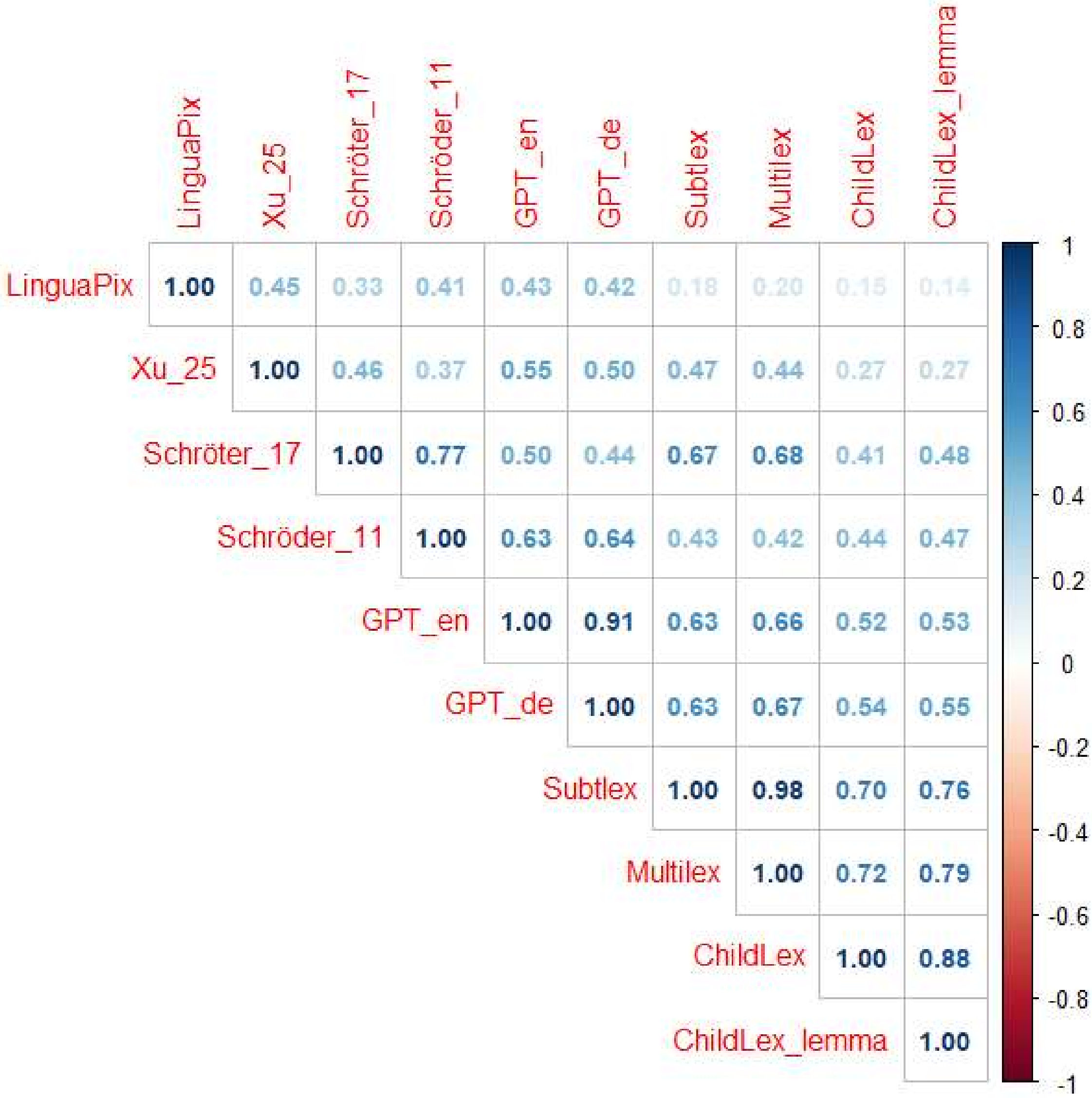
References

[1] Trott, S. (2024a). Can large language models help augment English psycholinguistic datasets? *Behavior Research Methods, 56*, 6082–6100.

[2] Brysbaert, M., Martínez, G., & Reviriego, P. (2025). Moving beyond word frequency based on tally counting: AI-generated familiarity estimates of words and phrases are an interesting additional index of language knowledge. *Behavior Research Methods, 57*, 28.

[3] Martínez, G., Conde, J., Reviriego, P., & Brysbaert, M. (2025). AI-generated estimates of familiarity, concreteness, valence, and arousal for over 100,000 Spanish words. *Quarterly Journal of Experimental Psychology*, 17470218241306694.

Example of a study: A comparison of GPT4 generated familiarity estimates for German, compared to the available datasets of human familiarity ratings and German word frequency estimates. The data show that the correlations of the human ratings with GPT estimates are higher than the correlations between the studies themselves. They also show that very much the same estimates are obtained whether the GPT instructions are given in German (de) or in English (en). Correlations between GPT-generated familiarity estimates and word frequency are between .5 and .7.



Pronoun accentuation produces interference effects in memory for alternatives

Sophie Repp, Heiko Seeliger, Sven David Weber
University of Cologne
sophie.repp@uni-koeln.de

Prosodic prominence (PProm) influences the representation of referents in discourse memory: High PProm of a noun may improve recall of the denoted referent and of information associated with it, as This has been shown for high PProm indicating focus [1]–[5] and contrastive topichood [1][6], which both involve (contrastive) alternatives. Better recall for focus is thought to stem from a richer semantic representation (a more specific memory trace) for the referent itself, and/or the additional representation of the alternatives. Better recall for contrastive topics has been linked to a structured memory representation with topics as anchors [6]. PProm-induced memory effects have not only been shown for the focus/topic itself but also for the alternatives [2]–[4]. The effects can be long-lasting, e.g. a day [1]. So far, memory-enhancing effects of PProm have not been studied for pronouns. When a pronoun is uttered, the pronominal referent is accessed in memory. Accentuation, i.e. high PProm, may trigger a contrastive interpretation, i.e. indicate the presence of alternatives [7]–[9]. Hence, pronoun accentuation might also enhance memory for the referent and the alternatives.

We conducted a recall study where 96 German participants listened to stories (30 items, 30 fillers) containing a clause with a subject *d*-pronoun (1). *D*-pronouns are used in colloquial speech [10] and are accented more naturally than personal pronouns. Each story introduced a set of referents (e.g., animal breeders), then named two of the referents (rabbit/horse breeder), followed by a third. The *d*-pronoun in the target clause referred to the third referent and was unaccented or carried a L+H* accent (Lat. sq.). The story ended with an evaluative statement. In the recall task, which followed directly or with a 24h delay (between-participants), participants verified written statements (Lat sq.): the target clause with (2i) the original subject referent $T_{TrueRef}$, (2ii) one of the mentioned contextual alternatives as subject $T_{ContAlt}$ or (2iii) a plausible unmentioned alternative T_{NewAlt}. Our (G)LMM analysis showed that ***recall accuracy* (RA)** (Fig. 1) was higher for $T_{TrueRef}$ than $T_{ContAlt}$/T_{NewAlt}, and lower for $T_{ContAlt}$ than T_{NewAlt} (but above chance) suggesting that overall, the proposition expressed by $T_{TrueRef}$ is remembered best, and that *contextual alternatives* are also stored in memory turning them into lures for an illusory recall of $T_{ContAlt}$ (esp. after 24h); unmentioned referents are safer rejects. For $T_{ContAlt}$ RA was lower when the *d*-pronoun was accented, suggesting that accentuation raises the discourse prominence of the contextual alternatives, adding to their lure quality. RA dropped after 24h except for $T_{TrueRef}$, indicating a lasting advantage for true propositions. ***Reaction times (RT)*** *for correct answers* were faster for $T_{TrueRef}$ than $T_{ContAlt}$/T_{NewAlt}; the 24h delay increased RTs (Fig. 2). The former effect confirms enhanced memory for true propositions, the latter confirms memory decay with time. *RTs for incorrect answers* to $T_{TrueRef}$ were slower with an accented *d*-pronoun; for $T_{ContAlt}$/T_{NewAlt} they were faster with accentuation (Fig. 3). We propose that PProm makes not only the mentioned alternatives more discourse-prominent but also activates likely alternatives, slowing down the incorrect rejection of $T_{TrueRef,}$ and speeding up the incorrect acceptance of T_{Alt}: participants are tricked into this choice faster with PProm. The effect vanishes after 24h. **Overall**, PProm of a pronoun has similar memory effects for the true proposition as nouns do. For contextual alternatives, pronouns show the opposite effect. Pronouns are interpreted against the previous discourse. So an accented pronoun induces the *re*-activation both of the pronominal referent and the alternatives. Competition from alternatives is greater than for nouns due to reduced cue informativity (nb/gender).

(1) Sample item listening phase (left and right context: English translation only)

The results from the animal breeding contest in Altdorf have just come in. A few of our people were there, remember. For example, our <u>rabbit breeder</u> and our <u>horse breeder</u> took part. Our <u>pigeon breeder</u> had the most beautiful animals.

Am Ende hat {der/DER$_{L+H*}$} echt viele Preise gewonnen.

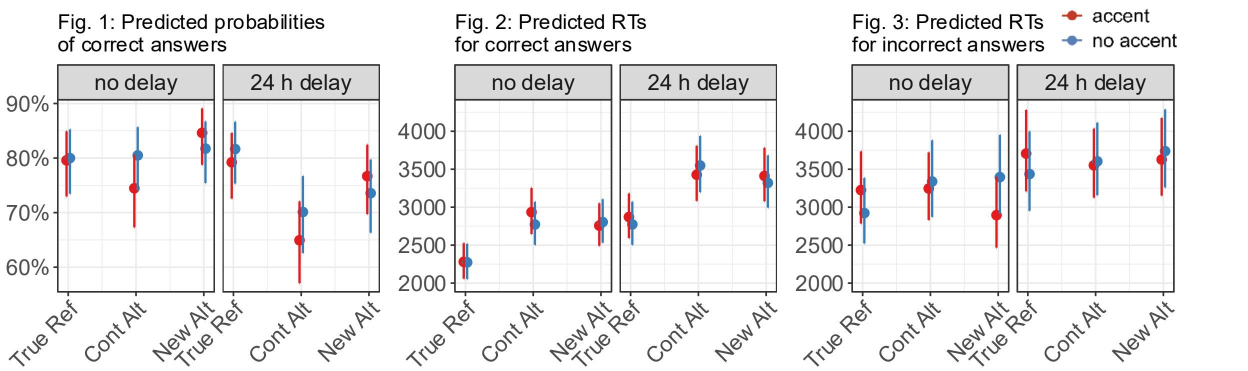
at end has he really many prizes won

'In the end he really won lots of prizes.'

That was quite impressive.

(2) Sample item recall task

i.	**TrueRef**	*Der Taubenzüchter*	*hat echt viele Preise gewonnen.*
ii.	**ContAlt**	*Der Kaninchenzüchter/Pferdezüchter*	*hat echt viele Preise gewonnen.*
iii.	**NewAlt**	*Der Hühnerzüchter*	*hat echt viele Preise gewonnen.*
		the {pigeon//rabbit/horse//chicken} breeder	has really many prizes won



[1] Sanford, A.J.S., Sanford, A.J., Molle, J., & Emmott, C. 2006. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 42(2), 109-130.

[2] Fraundorf, S.H., Watson, D.G., & Benjamin, A.S. 2010. Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *JMemLang* 63(3), 367-386.

[3] Gotzner, N., & Spalek, K. 2019. The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *LangLinguisticCompass* 13(2). e12310.

[4] Yan, M., Calhoun, S., & Warren, P. (2023). The role of prominence in activating focused words and their alternatives in Mandarin: Evidence from lexical priming and recognition memory. *Lang&Speech*, 66(3), 678-705.

[5] Kember, H., Choi, J. Yu, J. & Cutler, A.. 2021. The processing of linguistic prominence. *Lang&Speech* 64(2), 413-436.

[6] Repp, S., & Drenhaus, H. 2015. Intonation influences processing and recall of left-dislocation sentences by indicating topic vs. focus status of dislocated referent. *LangCogn&Neuroscience 30*, 324-346.

[7] Venditti, J., Stone, M., Nanda, P., & Tepper, P. 2002. Discourse constraints on the interpretation of nuclear-accented pronouns. *Speech Prosody* 2002, 675-678.

[8] Brown-Schmidt, S., Byron, D.K., & Tanenhaus, M.K. 2005. Beyond salience: Interpretation of personal and demonstrative pronouns. *JMemLang 53*(2), 292-313.

[9] Kaiser, E. 2011. Salience and contrast effects in reference resolution: The interpretation of Dutch pronouns and demonstratives. *Language and Cognitive Processes 26*(10), 1587-1624.

[10] Hinterwimmer, S. 2019. Prominent protagonists. *Journal of Pragmatics 154*, 79-91.

Building Structures Left to Right *and* Bottom Up: The Production and Perception of Syntactic Branching by L1 and L2 Users of a Tone Language

Chien-Jer Charles Lin, Zeping Liu, Xiao Dong
Indiana University Bloomington
chiclin@iu.edu

Background: Syntactic branching and structure building are fundamental to human sentence processing. While structural and semantic composition proceeds in a bottom-up fashion, language unfolds linearly from left to right in both perception and production. This interplay creates particular challenges when bottom-up processing leads to right-branching (RB) structures, which require additional effort to resolve discontinuities and form syntactic boundaries. For example, the phrase *the purple unicorn hat* is structurally ambiguous, allowing for a left-branching (LB) interpretation (a hat with purple unicorn patterns) or a right-branching (RB) interpretation (e.g., a purple hat designed for a unicorn). This paper examines the incremental processing and production of LB and RB structures in three-word sequences in Mandarin Chinese, a tone language where pitch is phonemic and prosodic cues for disambiguation are limited. We address three central questions: (i) How do LB and RB structures differ in processing? Is one more dominant? (ii) How do speakers signal branching direction in a prosodically constrained (tone) language? (iii) How do L2 learners with varying proficiencies acquire and produce these structures?

Method: We report findings from three experiments: one comprehension study (Exp1) and two production studies (Exps 2&3). Exp 1 used a phrase-picture matching task where participants (N=37) decided whether acoustically neutral ambiguous phrases matched the LB/RB pictures (Fig. 1, left panel). Exp 2 used a sentence reading-out task with an audience design [1-2] to collect how Mandarin speakers use prosody to disambiguate the phrase (N=16), where branching type (LB or RB) and Contrastiveness (No Contrast vs. Contrast, depending on whether the visual scene included only LB or RB images or both) were manipulated (Fig. 1, right panel). Exp 3 assessed productions of trisyllabic Tone 3 sequences (T3T3T3) by L1 speakers (N = 24) and L2 learners at intermediate-low (N = 16) and intermediate-high (N = 21) proficiency levels [3]. These sandhi sequences yield distinct outputs based on branching type due to Tone 3 Sandhi (T3S): T2T2T3 for LB structures and T3T2T3 for RB structures (see page 3 for information about T3S).

Results & Discussion: Results (Linear Mixed-Effects Models) show that LB structures were processed more quickly and accurately than RB structures (Exp 1, Fig. 2), suggesting a general preference for LB. In production (Exp 2), speakers inserted silent pauses & prolonged durations—rather than using pitch or intensity—as cues for branching direction, with pause/lengthening location aligning with expected structural boundaries (Fig. 3). Exp 3 (Fig. 4) revealed that while intermediate-high learners were sensitive to morphosyntactic cues, they struggled more with RB structures compared to L1 speakers. Intermediate-low learners predominantly applied sandhi in a left-to-right manner, favoring LB outputs. These findings suggest that LB is preferred in both comprehension and production due to the alignment of left-to-right and bottom-up processing. In tone languages like Mandarin, *temporal* cues such as pauses and lengthening are employed to mark syntactic boundaries since pitch cues are constrained. Furthermore, the acquisition of RB structures appears to emerge later in L2 development, as learners accumulate more linguistic knowledge and processing experience.

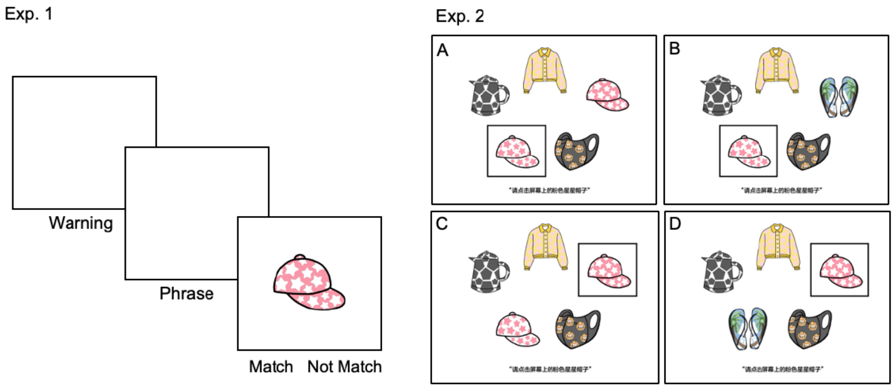


Fig. 1 Design in Exp.1 & 2. In Exp. 2, selected images are Contrast LB (A), NoContrast LB (B), Contrast RB (C), and NoContrast RB (D). Below images is the written sentence to be read out by the speakers.

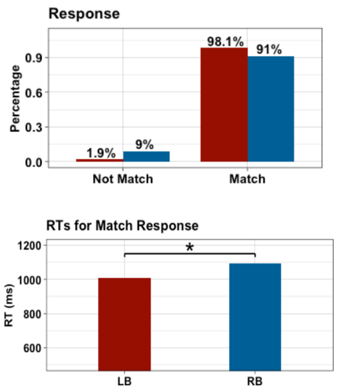


Fig. 2 Percentages & RTs for LB & RB in Exp 1.

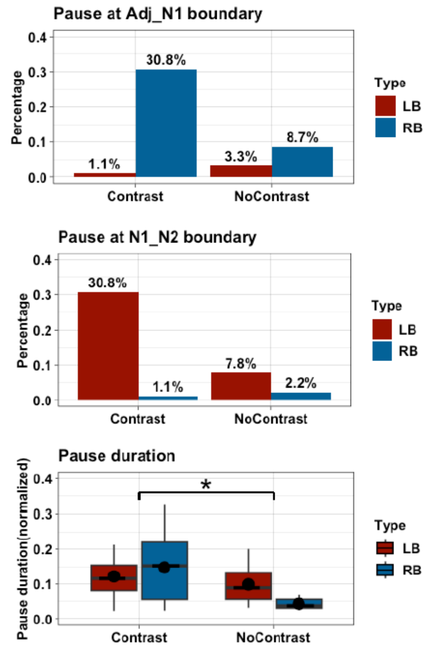


Fig. 3 Perceived pauses at phrasal boundaries in Exp 2.

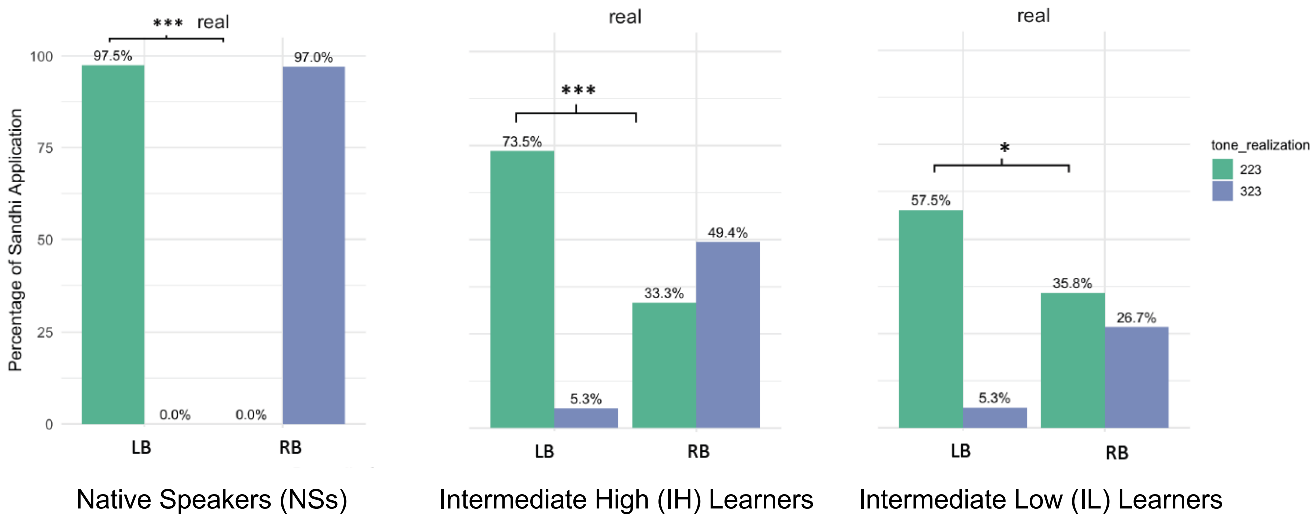


Fig. 4 Trisyllabic sandhi application in Exp 3.

Additional Information on Mandarin Tone 3 Sandhi (T3S)

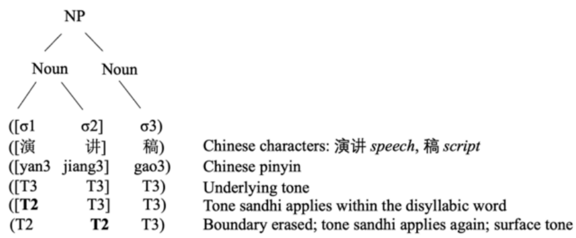
Mandarin Tone 3 Sandhi (T3S) is an obligatory phonological rule in Mandarin Chinese that applies when two adjacent low tones (Tone 3, or T3) occur within a defined prosodic or morphosyntactic domain. In such cases, the first T3 is converted to a rising tone (Tone 2, or T2), resulting in a T3T3 → T2T3 pattern. This tonal alternation is a regular and productive process in Mandarin phonology.

For example, in the disyllabic compound /mei3-xau3/ "wonderful"—composed of /mei3/ "beautiful" and /xau3/ "good"—the first syllable undergoes T3S and is produced as [mei2], yielding the surface form [mei2-xau3]. (The numbers 2 and 3 following the pinyin romanizations indicate the conventional tone categories Tone 2 and Tone 3 in Mandarin.)

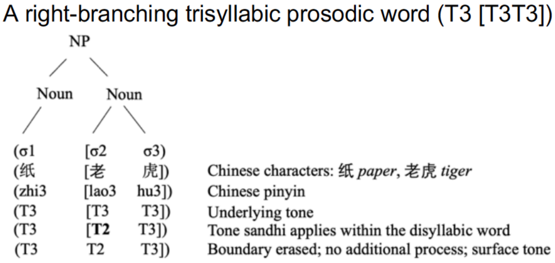
T3S involves two core processes: (a) identifying the appropriate domain for rule application based on morphosyntactic and prosodic structures—typically the bottom-level words or phrases in a syntactic tree—and (b) applying the phonological alternation to change the first of two adjacent T3 tones within that domain, producing the characteristic T2T3 output.

In trisyllabic Tone 3 sequences (T3T3T3), which are the focus of Exp 3 in the current study, T3S interacts with syntactic branching, yielding distinct surface forms that reflect underlying structural configurations. In left-branching (LB) structures like (1) below, T3S applies twice; first on the initial syllable, and again on the second syllable, resulting in T2T2T3, which shows left-to-right application. In contrast, right-branching (RB) structures trigger right-to-left application, yielding T3T2T3. These structurally determined outputs serve as diagnostic cues for distinguishing underlying phrase structure, as illustrated in examples (1–2) below.

(1) A left-branching trisyllabic prosodic word ([T3T3] T3)



(2) A right-branching trisyllabic prosodic word (T3 [T3T3])



This distinctive production makes Mandarin as an ideal case for studying productions that are unambiguously LB or RB phonologically.

References

[1] Lockridge & Brennan, 2002. Psychonomic bulletin & review. [2] Clark & Carlson, 1982. Language. [3] Zhang, J., & Lai, Y. (2010). Phonology, 27(1), 153–201.

Session 8

Redefining psycholinguistic cognates:
Linguistic and historical considerations

Isabella Fritz (isabella.fritz@ling-phil.ox.ac.uk), Joshua Booth (joshua.booth@ling-phil.ox.ac.uk) & Aditi Lahiri
University of Oxford, Language & Brain Laboratory

Much of what we know about the bilingual mental lexicon is based on experiments comparing cognates with non-cognates. In psycholinguistics, the term *cognate* refers to translation equivalents that don't only share meaning but also have a phonological and orthographical (for same script-languages) overlap [1]. As highlighted in a recent review paper [2], the current psycholinguistic cognate definition raises the question of a cut-off point between cognates and non-cognates which is operationalised differently across studies. Concretely, this means that various approaches (algorithms/ratings) exist to calculate phonological and orthographical similarity, often based on the assumption that all differences are either weighted equally (e.g., [3]), or the weighting is not based on empirical evidence (e.g., ALINE [4]). Importantly, current approaches neglect the fact that languages are structured systems and that phonological variation across languages is systematic.

Based on our recent line of research, we propose an alternative approach to investigate the role of the L1 phonology in L2 word recognition that is based on theoretical and historical linguistics. In linguistics, the term *cognate* refers exclusively to inherited words (English: *daughter* [ˈdɔːtə], German: *Tochter* [ˈtɔxtɐ̯]) whilst all borrowed words are considered loans. Loans may come from the same source (e.g., from French: E *reptile* [ˈɹɛpˌtʌɪ̯l], G *Reptil* [ˈʁɛpˈtiːl]) which we refer to as *shared loans*. Alternatively, the first language has borrowed the relevant item from the second (English) as in *chocolate* [tʃɔklət] – Mandarin 巧克力 [t͡ɕʰjaʊ̯³kʰɤ⁴li⁴] which was then adapted to the L1 phonological grammar. Cognates only exist in related languages and they do undergo regular sound changes. For example, following the Second Sound Shift (approximately dated to the sixth century), all initial interdental fricatives [θ] in English became voiced stops in German [d] (e.g., *thick* – *dick*, see Table 1).

Why is this relevant for psycholinguistic studies? For cognates, sound change is regular, systematic, and sound correspondences between classes of words remain. This systematicity holds for shared loans, i.e., the ways in which non-native phonemes and metrical patterns are adapted from one language into another is entirely systematic. However, phonological grammars differ between shared loans and cognates leading to varying degrees of phonological and orthographic overlap and on different phonological levels (see Table 2).

In regard to models of bilingual word recognition, recent accounts supporting an integrated bilingual mental lexicon [5] assume orthographically-driven co-activation based on visual word recognition studies. If, however, we want to advance our understanding of phonology-based co-activation [cf. 6], we need to test step-by-step which types of phonological mismatch between L1 and L2 are permitted so that phonological co-activation still occurs [7]. However, previous studies on L2 processing leave open the question of how this "mismatch" is to be defined or quantified. And this is exactly what our research agenda set out to achieve.

We will draw from a set of ERP cross-modal priming studies with a Lexical Decision Task and with a variety of L1-L2 combinations (i.e., Bengali, Dutch & German L1, English L2). Across multiple studies, we show that bilingual speakers are sensitive to systematic cross-linguistic correspondences and differences in both shared loans and cognates. In our research so far, we investigated systematic cross-linguistic differences/similarities on a featural and prosodic level. However, we will show that such linguistically informed study designs can be used to test for a wide range of phonological phenomena.

/s/–/z/ [OBS], [CONT], [COR], [STRID]				/w/–/v/ [CONT], [LAB], [VOICE]				/θ/–/d/ [OBS], [COR]			
sack	/sæk/	*Sack*	/zak/	*wind*	/wɪnd/	*Wind*	/vɪnd/	*thick*	/θɪk/	*dick*	/dɪk/
sand	/sænd/	*Sand*	/zand/	*wine*	/wʌɪn/	*Wein*	/vaɪ̯n/	*thing*	/θɪŋ/	*Ding*	/dɪŋ/
sea	/siː/	*See*	/zeː/	*wolf*	/wʊlf/	*Wolf*	/vɔlf/	*thorn*	/θɔːn/	*Dorn*	/dɔʁn/
/s/ has no voice feature /z/ = [VOICE]				/w/ = [SONORANT] /v/ = [OBSTRUENT]				/θ/ = [CONTINUANT] /d/ = [PLOSIVE] (underspecified)			

Table 1 Classes of corresponding initial consonants in English–German cognates, with the relevant differentiating phonological feature provided below. Shared features are provided in the first row of the table (OBS = OBSTRUENT, CONT = CONTINUANT, LAB = LABIAL, COR = CORONAL, STRID = STRIDENT).

Status	Related E—G words	Orthography	Phonology
Cognates	nest–Nest [nɛst]	✓	✓
	mild–mild [mɪlt]	✓	×
	mouse–Maus [maʊ̯s]	×	✓
	night–Nacht [naxt]	×	×
Recent loans	chef–Chef [ʃɛf]	✓	✓
	garage–Garage [gaˈʁaːʒə]	✓	×
	Cognac–Kognak [ˈkɔnjak]	×	✓
	bureau–Büro [byˈʁoː]	×	×

Table 2: The ways in which related words can differ across English and German in terms of orthographic and phonological overlap.

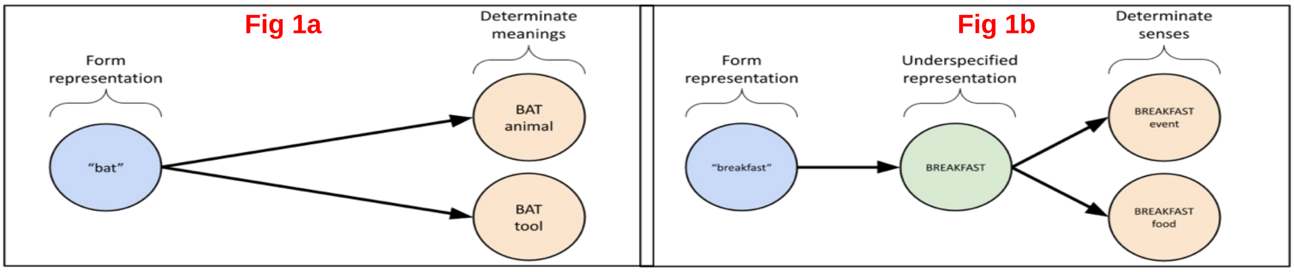
References

[1] Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, *41*, 496-518.
[2] Roembke, T. C., Koch, I., & Philipp, A. M. (2024). What makes a cognate? Implications for research on bilingualism. *Bilingualism: Language and Cognition*, 1-6.
[3] Mulder, K., Wloch, L., Boves, L., ten Bosch, L., & Ernestus, M. (2021). Cognate status modulates the comprehension of isolated reduced forms. *Language, Cognition and Neuroscience, 37*(5),
[4] Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities, 37*(3), 273-291.
[5] Dijkstra T, Wahl A, Buyuenhuijs F., et al. Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*. 2019;22(4):657-679.
[6] Jiang, N. (2019). Phonology-based bilingual activation among different-script bilinguals?. *Bilingualism: Language and Cognition*, *22*(4), 693-694.
[7] Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-197. doi:10.1017/S1366728902003012

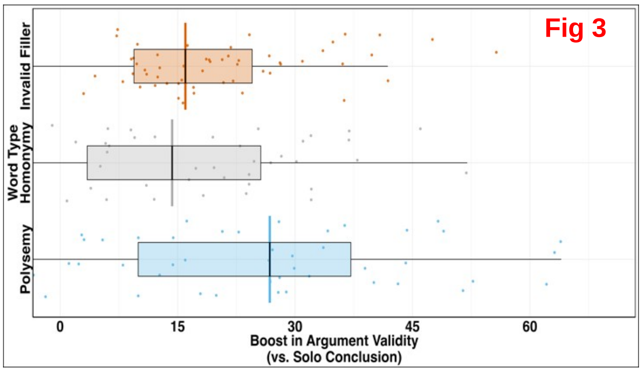
Thinking about nothing: the processing and mental representation of lexical ambiguity

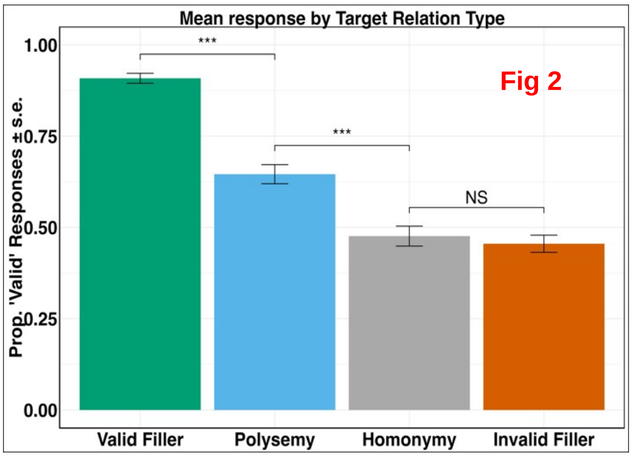
Elliot Schwartz[1], Griffin Pion[1], Jake Quilty-Dunn[2], Eric Mandelbaum[1,3], Spencer Caplan[1]
[1]CUNY Graduate Center, [2]Rutgers University, [3]Baruch College
scaplan@gc.cuny.edu

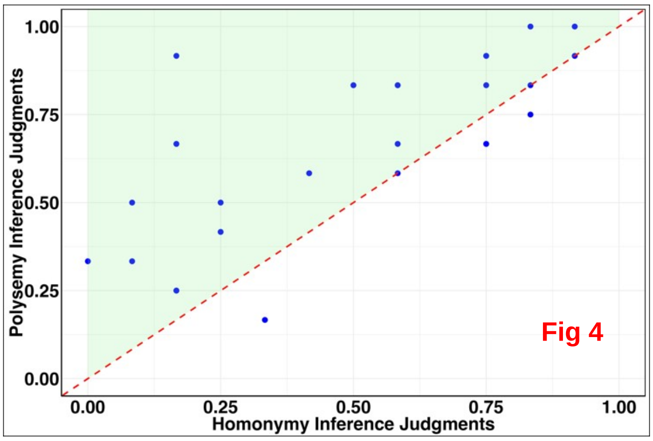
Background: How are words—and their many potential meanings—represented in the mind [1]? The study of ambiguity in the mental lexicon has a long history in psycholinguistics [2] with lexical ambiguity traditionally categorized into two kinds. Homonyms—single word forms that map to multiple unrelated meanings (e.g. "bat" as a flying mammal or a club used in sport)—are generally understood to employ a *one-to-many* mapping [3] (Fig 1a). Polysemes—word forms that map to multiple *related* senses (e.g. "breakfast" refers to both dining events and foods)—conversely, are more debated. The List View [4] holds that polysemes share the same representational structure as homonyms, with the differences between "classes" of ambiguity reduced to graded relatedness. In contrast, the Underspecification View [5] posits that polysemes involve a *one-to-one-many* structure (Fig 1b) where a word form passes through an *underspecified* representation before resolving to a determinate sense. While previous work has found processing time differences between polysemy and homonymy [6], such effects do not directly reveal how these forms are represented.
Methods: We developed a novel inference paradigm to test predictions of the List and Underspecification views of polysemy. Across two experiments (N=148) we tasked participants with evaluating the validity of Aristotelian syllogisms (a total of 224 arguments evenly divided between four forms; Table 1) in which the middle term was either a polyseme or homonym along with valid and invalid fillers. The polysemy trial terms were drawn from [7] and each premise designed to uniquely constrain the intended sense. For example: "*P1: all breakfasts are under three hours. P2: Some smoothies are breakfasts. C: Some smoothies are under three hours*"—since this argument equivocates on the senses of "breakfast" it is formally invalid. The Underspecification view predicts that, if polyseme-arguments involve processing an underspecified representation, participants will draw the conclusion anyway (at higher rates than in the homonymy or invalid filler trials). The List view predicts that if all lexical ambiguity is represented in the same way, then participants should endorse polyseme- and homonym-arguments at similar rates. We also conducted norming studies to gather judgements on the plausibility of the premises and conclusions when presented in isolation (i.e. no context of argument) which were included as additional factors in the mixed-effects models.
Results: Participants endorsed arguments with equivocal polysemes as valid (65%) more often than arguments with equivocal homonyms (48%— Fig 2). Our primary analyses (elided for space) confirm this difference using mixed-effects logistic regression via nested model comparison. Beyond classic belief bias effects, polyseme-based arguments show a validity boost twice as large as those involving homonyms or fillers (Fig 3). Finally, while participants vary in their overall rates of argument acceptance, the boost to polyseme items is stable across participants (Fig 4). These findings provide behavioral evidence that polysemous word forms engage a qualitatively different mental representation than homonyms, i.e. polysemes are not merely stored as lists of related senses but instead include an intermediate, underspecified representation that is active in reasoning. If concepts are elements of inference, these results indicate that humans can reason using "placeholders." Such conceptual structures are meaningful yet not fully specified: we can think without thinking about anything in particular. This raises questions such as how such structures are learned and how long such underspecified states persist during online comprehension.



Dimatis	Disamis	Datisi	Darii
Sm P are M	Sm M are P	All M are P	All M are P
All M are S	All M are S	Sm M are S	Sm S are M
Sm S are P	Sm S are P	Sm S are P	Sm S are P







References

[1] Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*, *15*(2), 411-427.

[2] Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of verbal learning and verbal behavior*, *18*(6), 645-659.

[3] Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of memory and language*, *29*(2), 181-200.

[4] Devitt, M. (2021). Semantic polysemy and psycholinguistics. *Mind & Language*, *36*(1), 134-157.

[5] Frisson, S. (2009). Semantic underspecification in language processing. *Language and linguistics compass*, *3*(1), 111-127.

[6] Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1-24.

[7] Murphy, E. (2021). *Linguistic representation and processing of copredication* (Doctoral dissertation, UCL (University College London)).

Not that *cloze*: Semantic sentence constraint is influenced by language background and dominance

Ana Bautista[1, 2] & Clara Martin[1, 3]

[1] BCBL, Basque Center on Cognition, Brain & Language, Donostia-San Sebastián, Spain

[2] University of the Basque Country (UPV-EHU), Bilbao, Spain

[3] Ikerbasque, Basque foundation for science, Bilbao, Spain

a.bautista@bcbl.eu

Background: Research in language processing and prediction often relies on cloze tests to establish sentence constraint norms. In these tests, participants are presented with incomplete sentences and are asked to write the most likely continuation. The degree of answer consistency across participants then informs about sentence constraint. However, the validity of cloze test results depends heavily on how closely the pre-test population matches the one in the main study, and there is limited evidence on how different sociolinguistic profiles may influence sentence constraint. Given the practical challenges of running cloze pre-tests with every target population, this study explores whether sentence constraint measures are affected by linguistic and cultural background (Experiment 1) and by language dominance (Experiment 2).

Method: We tested different speaker groups on a cloze task hosted on Gorilla, where participants completed 330 sentence contexts with the first word that came to mind. Responses were processed to extract the first noun mentioned, and within-group noun response entropy was calculated as a measure of semantic sentence constraint. In Experiment 1, 60 Spanish-dominant and 60 French-dominant speakers completed the task in Spanish and French, respectively. In Experiment 2, French sentence contexts were presented to three groups: 60 French-dominant, 60 Spanish-dominant, and 60 English-dominant speakers. To assess alignment in semantic constraint across groups, we computed Spearman correlations between group pairs. To further interpret the observed correlations, we developed two additional correlation benchmarks tailored to our data: a distribution of correlations representing null alignment, obtained through permutation tests, and a distribution representing maximum alignment, derived from split-half iterations.

Results: In Experiment 1, response entropy in Spanish-dominant speakers was only mildly aligned with that of French-dominant speakers tested on the same translated sentences (Figure 1). In Experiment 2, dominant and non-dominant French speakers showed only mild alignment, distant from the distribution representing maximum alignment (Figures 2 and 3). However, the two groups of non-dominant speakers of French who differed in their dominant language were strongly aligned, their observed correlation being indistinguishable from the maximum alignment distribution (Figure 4).

Discussion: Since presenting the same translated sentence contexts to dominant speakers of different languages resulted in only mild alignment, our findings suggest that sentence expectations are shaped by language-specific usage and cultural factors associated with high proficiency and dominance. Moreover, differences in constraint scores between dominant and non-dominant speakers highlight the role of language dominance in shaping predictive processing. In contrast, participants' most dominant language did not affect sentence constraint in a shared non-dominant language, suggesting minimal L1 influence on L2 semantic expectations. These findings contribute to theoretical proposals of L2 prediction and support more efficient, informed practices for designing language prediction studies across diverse speaker populations.

Figures

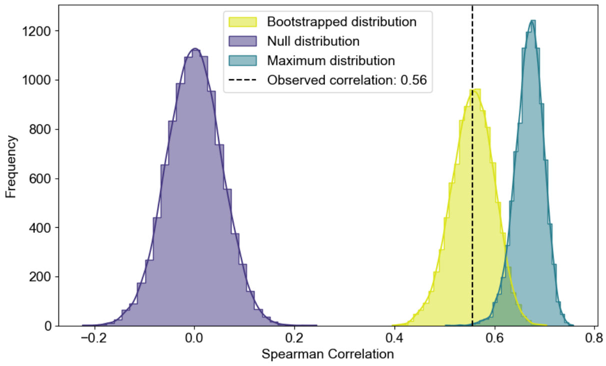


Figure 1. Alignment distributions between Spanish- and French-dominant speakers completing Spanish and French sentences, respectively.

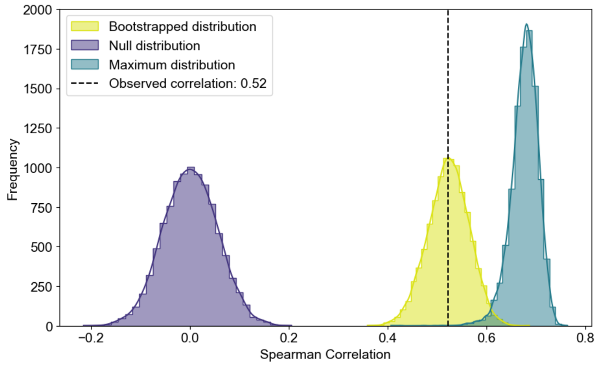


Figure 2. Alignment distributions between French- and Spanish-dominant speakers completing French sentences.

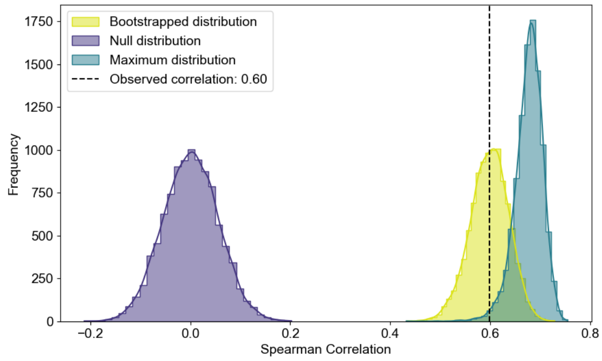


Figure 3. Alignment distributions between French- and English-dominant speakers completing French sentences.

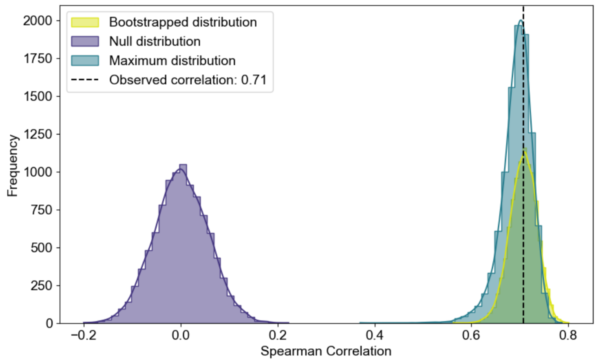


Figure 4. Alignment distributions between Spanish- and English-dominant speakers completing French sentences.

Jointly modeling maze RT and accuracy using diffusion models: A first case study

John Duff
UCLA Linguistics
duff@ucla.edu

Laura Pissani
Saarland University, Language Science & Technology
laura.pissani@uni-saarland.de

In the maze task [1], participants must decide between target continuations and inappropriate foils. Choices and RTs in this task have so far been modeled independently, but we suggest they could be jointly well-described by the **diffusion model** of decision-making [2]. In addition to jointly explaining these components, diffusion modeling can also help distinguish between several sources for differences across conditions in decision tasks, including non-decision processing time (t_0), initial response attractiveness (z), and rate of preference accumulation (v) (Fig 1). Here, for one recent effect of interest, we offer a proof-of-concept that diffusion models can be applied to maze data, and probe the source of the effect.

Effect of interest When a single maze decision followed comprehension of a familiar two-word metaphor embedded in context (Fig 2), [4] observed that target responses were slower, and foil selection was more likely, when the foil was RELATED to the literal interpretation of the metaphor. These effects reduced when the critical maze position was postponed farther away from the metaphor [5]. This pattern, known as the *metaphor awakening effect*, can be related to the general finding that maze decisions are more difficult when concepts related to the foil are activated in preceding context [6-7]. [4-5] consider that the literal meaning may be temporarily activated during metaphor comprehension, and before it rapidly decays, it spreads activation to related concepts. With the diffusion modeling we present here, we find that a joint explanation for the RT and accuracy components of this effect is feasible, and conclude that the principal source for both is a slower accumulation of preference for the target in the RELATED-foil condition.

Data from [4-5] comprises 40 participants in each of SHORT, MEDIUM, and LONG delays between metaphor and maze, each seeing 12 RELATED-foil trials and 12 UNRELATED. Foils were matched in length, frequency, and contextual appropriateness across conditions, differing only in association to the preceding metaphor.

Diffusion models Using `rtdists` [8], we fit max-likelihood parameters to observed RT and response distributions with 10 diffusion model variants, pruning away which parameters could vary by condition. AIC (Table 1) favors model #5, in which only drift rate v varied across all conditions. Fitted v_{REL} towards the target is slower than v_{UN}, less so at larger delays, while z_{REL} is weakly biased against the target, but $t_{0\,REL}$ is slightly facilitated. Simulated RTs using model #5 closely match the observed distributions of RTs for both target and foil selections (Fig 4). Simulated accuracies likewise capture the observed pattern (Fig 3).

Discussion Model performance supports the proposal that when foils in maze decisions come shortly after a related word, they jointly slow RTs and depress accuracy by interfering in the decision process. In particular, we conclude such foils principally slow the rate of preference accumulation towards the target. If this interference comes from residual priming, it suggests that the core process of maze decisions, judging contextual fit, is susceptible to other sources of activation. Future work should pursue this idea further, and perhaps validate the connection to residual priming by comparing stimulus-specific predictors of spreading activation [e.g., 9] to itemwise diffusion model estimates.

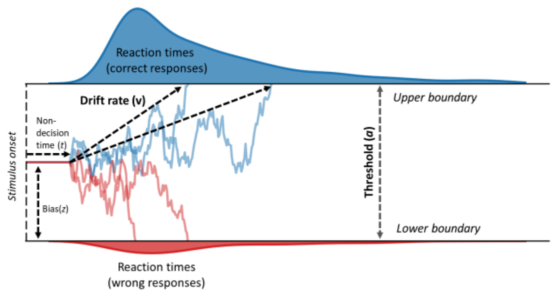


Figure 1. Illustrating the core mechanisms of a diffusion model. Figure from [10]. (NB: $t = t_0$)

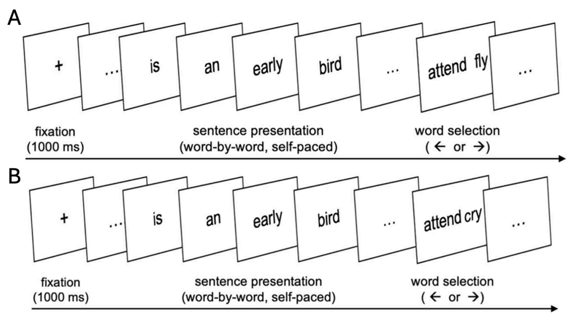


Figure 2. [4-5]'s procedure for the maze task with related (A) and unrelated (B) foils.

Table 2. Parameters from Diffusion Model #5.

$t_{0\,REL} = 0.40$	$t_{0\,UN} = 0.46$	$z_{REL} = 0.43$	$z_{UN} = 0.53$	
$v_{REL,S} = 1.08$	$v_{REL,M} = 1.40$	$v_{REL,L} = 1.53$		
$v_{UN,S} = 1.60$	$v_{UN,M} = 1.81$	$v_{UN,L} = 1.84$		
$a = 1.96$	$s_{t0} = 0.50$	$s_z = 0.30$	$s_v = 0.49$	$s = 1$

Table 1. Model comparisons, best AIC in bold.

$_{\pm REL}$: parameter fit with two values, for related vs. unrelated foils.
$_{\pm REL\times D}$: parameter fit with six values, crossing ±related and {S/M/L} delay.

	Model	∑ LogLik	AIC
1	$t_{0\,\pm REL\times D}$, $z_{\pm REL\times D}$, $v_{\pm REL\times D}$	–2016.04	4076.08
2	$t_{0\,\pm REL}$, $z_{\pm REL\times D}$, $v_{\pm REL\times D}$	–2019.54	4075.07
3	$t_{0\,\pm REL\times D}$, $z_{\pm REL}$, $v_{\pm REL\times D}$	–2017.09	4070.18
4	$t_{0\,\pm REL\times D}$, $z_{\pm REL\times D}$, $v_{\pm REL}$	–2027.52	4091.05
5	$\boldsymbol{t_{0\,\pm REL}}$, $\boldsymbol{z_{\pm REL}}$, $\boldsymbol{v_{\pm REL\times D}}$	**–2020.46**	**4068.92**
6	$t_{0\,\pm REL}$, $z_{\pm REL\times D}$, $v_{\pm REL}$	–2032.00	4092.00
7	$t_{0\,\pm REL\times D}$, $z_{\pm REL}$, $v_{\pm REL}$	–2036.45	4100.90
8	t_0, $z_{\pm REL}$, $v_{\pm REL\times D}$	–2029.35	4084.69
9	$t_{0\,\pm REL}$, z, $v_{\pm REL\times D}$	–2028.01	4082.02
10	t_0, z, $v_{\pm REL\times D}$	–2031.16	4086.32

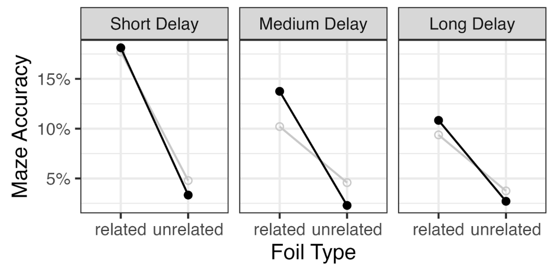


Figure 3. Observed (black) and predicted (grey) accuracy (% "target") from Diffusion Model #5.

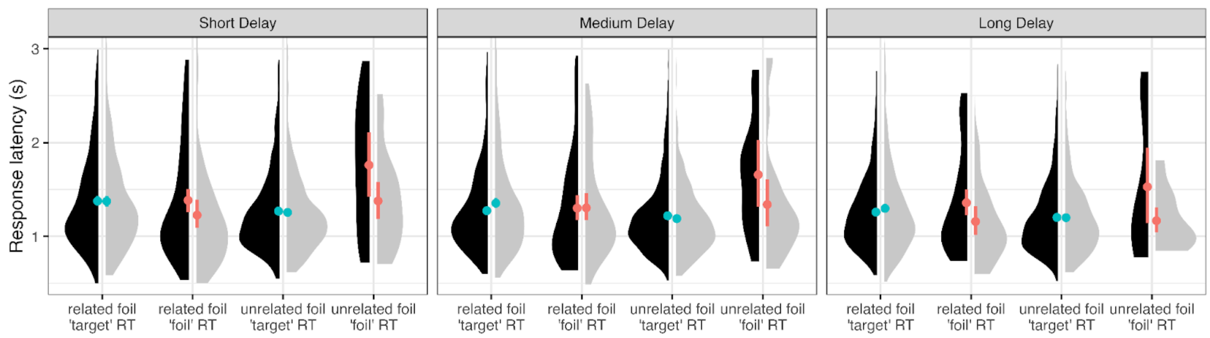


Figure 4. Observed (black) and model-predicted (grey) RT distributions from Diffusion Model #5. Points indicate observed and predicted means with bootstrapped 95% CIs.

References [1] Forster et al. (2009) *Behav Res Meth* [2] Ratcliff & McKoon (2008) *Neural Comput* [3] Myers et al. (2022) *Front Psychol* [4] Pissani & de Almeida (2022) *Psychon Bull Rev* [5] Pissani & de Almeida (2023) *Metaph Symb* [6] Gallant & Libben (2020) *Ment Lex* [7] de Almeida et al. (2025) *JEP: LMC* [8] Singmann et al. (2022) R [9] Mandera et al. (2017) *JML*. [10] Vinding et al. (2021) *Consc Cogn*.

Poster sessions

Thursday Morning Posters

Acquisition of the blocking effect in L2 Chinese by L1 Japanese speakers

Jun Lyu
Peking University
junlyu@pku.edu.cn

Background This study investigates L1 Japanese speakers' processing of reflexive *ziji* ('self') in L2 Chinese in relation to the **blocking effect** (**BE**) phenomenon. In Chinese, the reflexive *ziji* can be long-distance (LD) bound by a non-local subject; however, when the local subject is a 1st-person pronoun, LD binding is not allowed (e.g., 'John$_1$ said I$_2$ like self$_{*1/2}$'s shirt'). This is because (i) LD *ziji* is empathic, requiring the speaker or reader to empathize with the non-local subject or ***empathy locus*** and (ii) the 1st-person pronoun 'I' is a **stronger empathy locus** and is priortized over a 3rd-person empathy locus (e.g., Kuno'87). In contrast, although Japanese bare reflexive *zibun* ('self') can also be bound non-locally, it does not seem to show BE in the absence of empathy-inducing morphemes (see e.g., Oshima'07). Indeed, LD *zibun* tends to be used as a non-empathic, attitudinal reflexive (e.g., Nishigauchi'14), bound by an ***attitude holder***. Thus, Chinese *ziji* and Japanese *zibun* by default are construed differently: *ziji* is necessarily empathic and shows BE, while *zibun* tends to be non-empathic and does not show BE. Therefore, one challenge facing Japanese learners of L2 Chinese is to overcome transfering the properties of LD *zibun* to LD *ziji*. As these two reflexives share similar properties, from a psychotypological perspective (e.g., Kellerman'83), the acquisition of BE in L2 Chinese may be highly challenging for Japanese learners as the perceived linguistic proximity may prompt L1 transfer.

Methods Thirty-two L2 Chinese learners, in addition to 33 L1 Chinese and 27 L1 Japanese control participants, participated in the study. Two factors, ANTECEDENT (local vs. non-local) and BLOCKER (blocker vs. non-blocker) were fully crossed in a 2x2 factorial design. See **Fig.1** for an example target set (16 target items and 24 fillers) and how local/non-local binding is created in blocker and non-blocker conditions. In the experiment, participants first made a forced choice acceptability judgment and then rated the acceptability of a particular reading using a 7-point Likert scale. The L2 participants also completed two cloze tests and a language background survey.

Results Mixed effects logistic and linear regressions were run over forced choice judgments and acceptability ratings, respectively. We mainly report results from the forced choice task as acceptability ratings show similar results. In the **L1 Japanese** experiment (**Fig. 2**), we discovered an ANTECEDENT x BLOCKER interaction ($p < 0.001$): within the local conditions, participants accepted local binding more often in the presence of a blocker; within the non-local conditions, they accepted non-local binding more often in the absence of a blocker. Crucially, in the non-local/blocker condition, L1 Japanese speakers prefer non-local binding, which contrasts sharply with L1 Chinese speakers. In the **L1 Chinese** experiment (**Fig. 3**), in addition to an interaction ($p < 0.001$), we further found a strong BE as Chinese natives strongly reject non-local binding in the presence of a blocker. **L2 participants** (**Fig. 4**) showed interpretation patterns of *ziji* similar to L1 Chinese speakers as they also mostly rejected non-local binding when the local subject is 'I'. Taken together, these findings suggest that Japanese learners of L2 Chinese can acquire BE in the processing of *ziji*. (We omit the discussion of some minor aspects of our findings due to space.)

Discussion The main findings of this study are two-fold. First, consistent with prior theoretical discussions, we found different interpretation patterns of *ziji* and *zibun* (e.g., Oshima'07; Wang & Pan'15). Second, Japanese learners of L2 Chinese can suppress L1 transfer and show BE in L2 Chinese. Furthermore, the weak BE effect in L1 Japanese also provides new data relating to the linguistic properties of *zibun*.

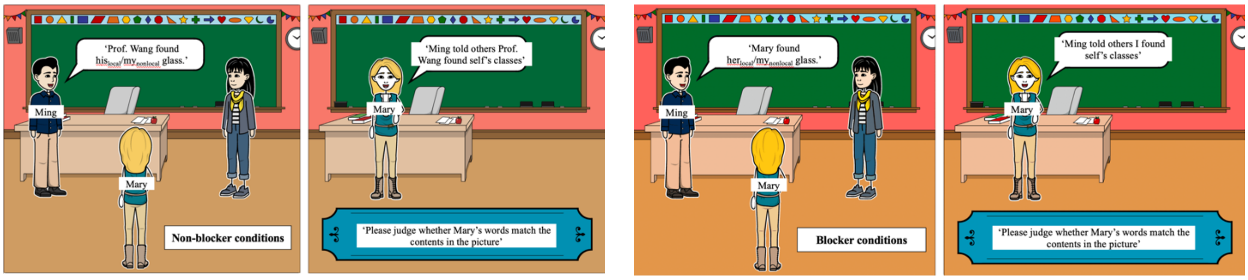


Fig. 1: An example target set with English translations. The original texts were shown in Japanese and Chinese in respective experiments. The left panel shows two **non-blocker** conditions (i.e., the local subject in Mary's speech is **3rd-person** 'Prof. Wang') with local or non-local antecedents. The right panel shows two **blocker** conditions (i.e., the local subject is **1st-person** 'I'). Local or non-local binding readings are contigent upon Ming's speech ('**his**' indicates **local** binding; '**my**' indicates **non-local** binding).

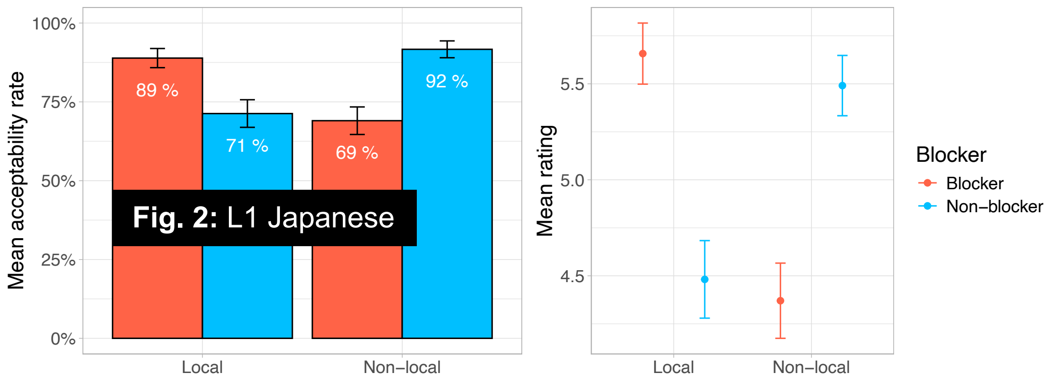


Fig. 2: L1 Japanese

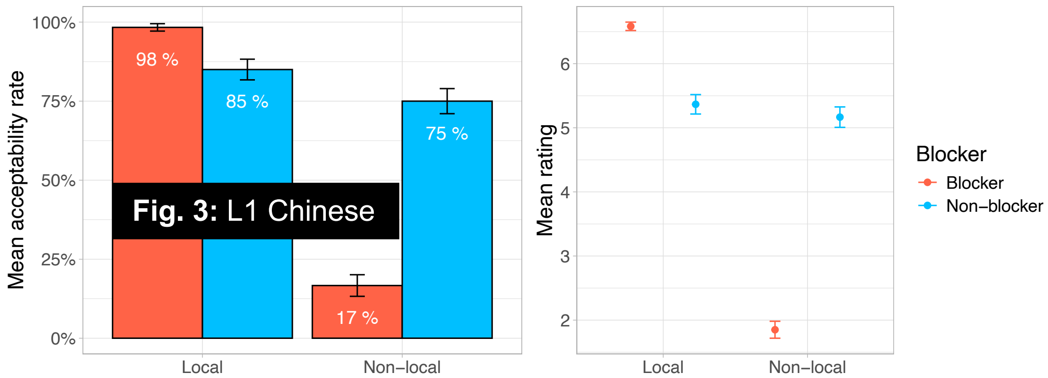


Fig. 3: L1 Chinese

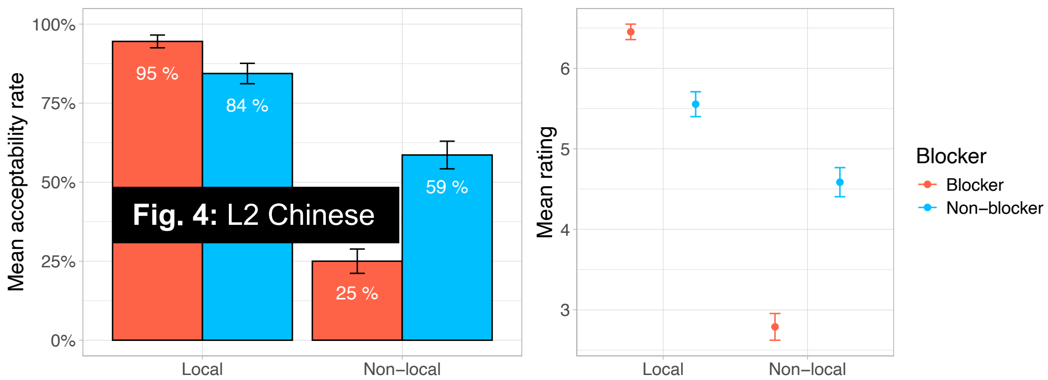


Fig. 4: L2 Chinese

Finding, sharing, and losing words: word associations and the mental lexicon

Tess Fitzpatrick
Swansea University
t.fitzpatrick@swansea.ac.uk

This paper reports outcomes from the "Finding, sharing, and losing words" project concerned with lexical acquisition and attrition through the lifespan, and the degree to which word meaning is shared. Specifically, the project attempts to determine the potential of lexical retrieval research to i) identify patterns of lexical retrieval in healthy ageing; ii) identify language-specific patterns, and the effects of multilingualism on lexical retrieval; and iii) investigate the impact of lived experience, including sensory experience, on lexical organisation.

This ambitious set of objectives is addressed through analysis of word association data. Evidence from over a century of research indicates that word associations – i.e. the first word one thinks of when confronted with a 'cue' word – may carry information about the structure of the mental lexicon, cognitive development, linguistic proficiency, lexical connotation, attitude, and more [1, 2, 3, 4]. It also hints at important differences in the way individuals process language, in that the kinds of association people make (e.g. synonyms, collocations, orthographically/phonologically similar) differ between people, but that individuals' association profiles are stable across time and (for bilinguals) across their languages [5, 6, 7]. This project analyses bespoke word association data from populations ranging in size from n=10 to n=680, with distinct characteristics (dizygotic and monozygotic twins, 16-year-olds and over-65-year-olds, multilingual individuals, people with visual impairment), in order to interrogate those propositions. Additionally, the project has produced an online resource giving access to seven previously unavailable sets of word association data, including the first comprehensive set of word association norms in Welsh (900 cues, 85 respondents).

In the first part of this paper we present project findings that suggest ways in which the architecture of the mental lexicon is affected by factors including age, the interaction of bilinguals' languages, features of one's L1, and sensory impairment. After reporting our headline findings we will introduce the online resource, demonstrating how its databases, norms lists (group responses ranked by frequency), and visualisations can be used by researchers and other users to address a range of research questions. Lastly, we will talk about the real world applications, in education and healthcare, of our research findings, and will consider ways in which word association analyses might be further harnessed to augment our understanding of the mental lexicon.

The research reported here was funded by the Arts and Humanities Research Council (AHRC), part of UK Research and Innovation. (AH/Y003020/1).

References

[1] Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*(3), 305-312.

[2] De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior research methods*, *51*, 987-1006.

[3] De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior research methods*, *40*(1), 213-231.

[4] Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237-274.

[5] Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*(3).

[6] Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In Fitzpatrick, T. and Barfield, A. *Lexical processing in second language learners.* Multilingual Matters. 38-52.

[7] Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. J. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, *36*(1), 23-50.

A Computational Perspective on the Stage of Acquisition of Grammatical Competence: Testing the Growing Tree Approach

Robiatu Al Addawiyah[1], Cristiano Chesi[2], Adriana Belletti[1]

[1]University of Siena, [2] IUSS Pavia

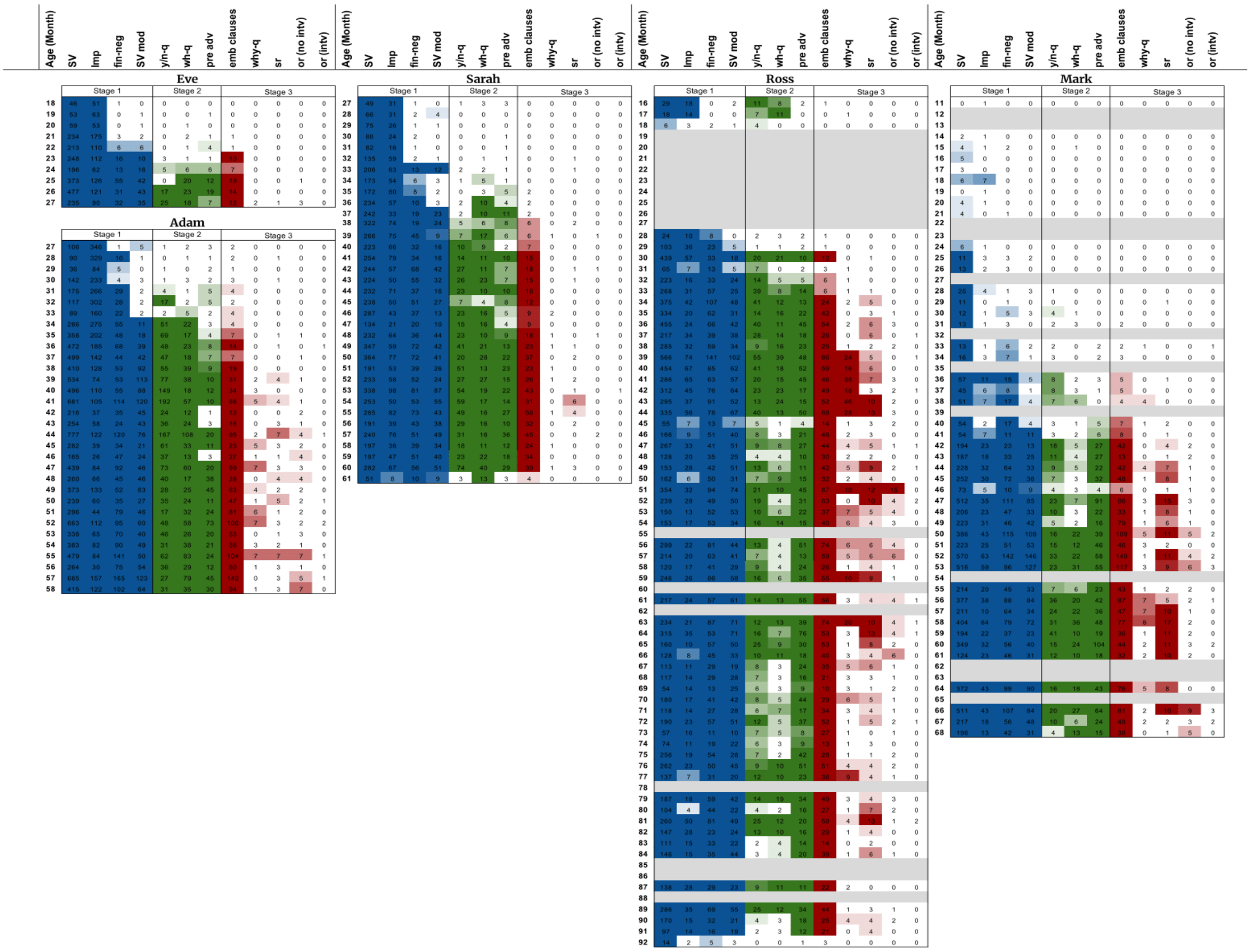
Background. Determining the Age of Acquisition (AoA) of complex morphosyntactic constructions (Guasti 2017) is fundamental not only from a classic psycholinguistic perspective but also for assessing the training regimens and performances of Large Language Models (Charpentier et al. 2025). A novel approach for assessing the Stage of Acquisition (SoA)—a useful proxy for accounting for linguistically non-significant individual variability in AoA—is the 'Growing Tree' model (Friedmann, Belletti, & Rizzi, 2021). The idea that grammatical competence develops in (discrete) stages, 'from bottom to top' following the functional spine above the lexical predicate, has been proposed to account for the order in which L1 children master different complex syntactic constructions at specific SoA. Three clear-cut stages have been identified so far. In the first stage, children produce structures involving vP and IP. The second stage comprises structures involving the first portion of the Left Periphery (LP), up to the Q/Foc head (Rizzi & Bocci, 2017). In the third stage, the syntactic tree becomes fully accessible, with the availability of the highest portion of the LP, up to the ForceP. Studies in Hebrew (Friedmann & Reznick, 2021) and Italian (Casadei, 2023) support this acquisition model.

However, there is no study yet examining the syntactic acquisition order in English L1 children. This work aims to fill this gap by adopting an automated system—a rule-based syntactic structure identifier—specifically developed for this investigation, targeting the Growing Tree's structural predictions.

Method. We used longitudinal naturalistic speech from five children taken from two CHILDES corpora—Brown (1973) and MacWhinney (1991)—as dataset (940,873 tokens in the original corpus, 494,352 tokens after cleaning). Data were analyzed in two phases: 1) pre-processing and 2) structure identification. In the pre-processing phase, we used RegEx to extract the necessary information, remove noise, and store the data into a dataframe before inspecting the structure. In the structure identification phase, we parsed and inspected the structure using predefined rules. The identifier relies mainly on the constituency parse tree and Part of Speech (PoS) tagging. Thus, we used spaCy and integrated Benepar for the parsing process. With multilabel identification, one or more structures can be assigned to a given sentence. Additionally, since the system also detects the presence of non-standard question production, we observed the question formation patterns between standard and non-standard forms. Each rule is duly assessed for accuracy, both precision and recall.

Finding. With 77.5% subset accuracy and a 2.4% Hamming loss, the results showed that the children's syntactic acquisition pattern aligns with the acquisition model proposed by the Growing Tree approach, yielding a Guttman scale. Children who produced sentences with structures reflecting lower CP projections also produced those reflecting vP/IP, and those who produced sentences with structures reflecting higher CP projections also produced those reflecting lower CP and vP/IP projections. The result also showed variability in the age of milestone, supporting the claim that age is not a reliable predictor (Friedmann & Reznick, 2021; Friedmann, Belletti, & Rizzi, 2021). Furthermore, the question productions yielded intermittent result, indicating that children do not always use the correct construction of questions consistently after acquiring it. This finding aligns with Rowland (2000), who noted that the failure of applying grammatical rules governing question formation in children occurs at the same time as they produce correct forms of questions.

Figures. The order of production frequency of various structures in the five children's longitudinal data. Grey-colored cells indicate no data recorded at the corresponding ages.



References

Casadei, E. (2023). *A new sentence repetition task tool to investigate the acquisition of syntactic structures in typical and atypical language development: A view from growing trees and syntactic cartography* (Master thesis). University of Siena.

Charpentier, L., Choshen, L., Cotterell, R., Gul, M. O., Hu, M., Jumelet, J., Linzen, T., Liu, J., Mueller, A., Ross, C., Shah, R. S., Warstadt, A., Wilcox, E., & Williams, A. (2025). BabyLM Turns 3: Call for papers for the 2025 BabyLM workshop (Version 2). arXiv. doi: 10.48550/ARXIV.2502.10645

Friedmann, N. & Reznick, J., (2021) Stages rather than ages in the acquisition of movement structures: Data from sentence repetition and 27696 spontaneous clauses, *Glossa: a journal of general linguistics 39*(1): 143. doi: 10.16995/glossa.5716.

Friedmann, N., Belletti, A. & Rizzi, L., (2021) Growing trees: The acquisition of the left periphery, *Glossa: a journal of general linguistics 6*(1): 131. doi: 10.16995/glossa.5877.

Guasti, M. T. (2017) *Language acquisition: The growth of grammar*. MIT press.

Rizzi, L., & Bocci, G. (2017). Left Periphery of the Clause: Primarily Illustrated for Italian. *In The Wiley Blackwell Companion to Syntax, Second Edition.* (pp. 1-30). 10.1002/9781118358733.wbsyncom104.

Rowland, C. F., & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: 'what children do know?' *Journal of Child Language, 27*(1), 157–181. doi: 10.1017/S0305000999004055.

To see the whole code of the rule-based syntactic structure identifier, visit this github.

Shared syntax in bilingual humans and cognitive models: Code-switching increases cross-language structural priming

Yung Han Khoe[1,2], Gerrit-Jan Kootstra[1], Stefan L. Frank[1], Rob Schoonen[1], Edith Kaan[3]
[1]Radboud University; [2]University of Birmingham; [3]University of Florida
y.h.khoe@bham.ac.uk

Background: After hearing a grammatical structure in one language, bilinguals become more likely to produce that structure in their other language. Such cross-language structural priming is commonly interpreted as evidence for shared syntax in bilinguals [1,2]. However, bilinguals switch languages not only between but also within sentences, unlike the script of typical cross-language structural priming experiments [3]. We hypothesize that increased prediction error caused by a code-switch in the prime sentence increases implicit learning of shared syntax, leading to stronger cross-language priming. We test for this effect in an implicit learning model of priming [4] and in Spanish-English bilinguals.

Method: We conducted four simulated structural priming experiments, using instances of the Spanish-English Bilingual Dual-path model [5], which was previously used to simulate both code-switching [5] and cross-language structural priming [6]. Following [6], we trained model instances on artificial versions of Spanish and English, to be used as simulated participants. In a structural priming experiment, these simulated participants were presented with Spanish active or passive primes before producing English transitives. The primes either had an English (code-switched) determiner and noun (Examples a and b) or noun only (Examples c and d), or were entirely in Spanish (non-code-switched control). Code-switches occurred at the beginning (Examples a and c) or the end (Examples b and d) of sentences.

(a) "**the boy** empuja el juguete"
(b) "el niño empuja **the toy**"
(c) "el **boy** empuja el juguete"
(d) "el niño empuja el **toy**"

We also tested for the interaction between code-switching and prime structure in 190 Spanish-English bilinguals in the US. In a pre-registered online experiment, participants wrote 60 English picture descriptions after hearing active or passive primes that were entirely Spanish or code-switched.

Results: In the model simulations, structural priming was increased compared to entirely Spanish primes only with a code-switched determiner and noun at the beginning (Example a), as evidenced by a weak but significant positive interaction between code-switch condition and priming ($\beta = 0.18$, $p < 0.001$) in our mixed effects analysis and visualized in Fig. 1 (left). We therefore only used code-switched primes of type (a) in our follow-up experiment with human participants. As predicted by the model, the experiment revealed that structural priming in participants was stronger after code-switched compared to entirely Spanish primes as evidenced by a significant positive interaction between code-switch condition and priming ($\beta = 0.15$, $p = 0.002$) in the mixed effects analysis and visualized in Fig. 1 (right).

Discussion: Together, these results suggest that processing code-switches in a prime sentence can result in increased prediction error, which in turn can lead to increased implicit learning of shared syntactic representations, resulting in stronger cross-language structural priming. These results also demonstrate that the Bilingual Dual-path model can be used to predict novel psycholinguistic effects in human participants.

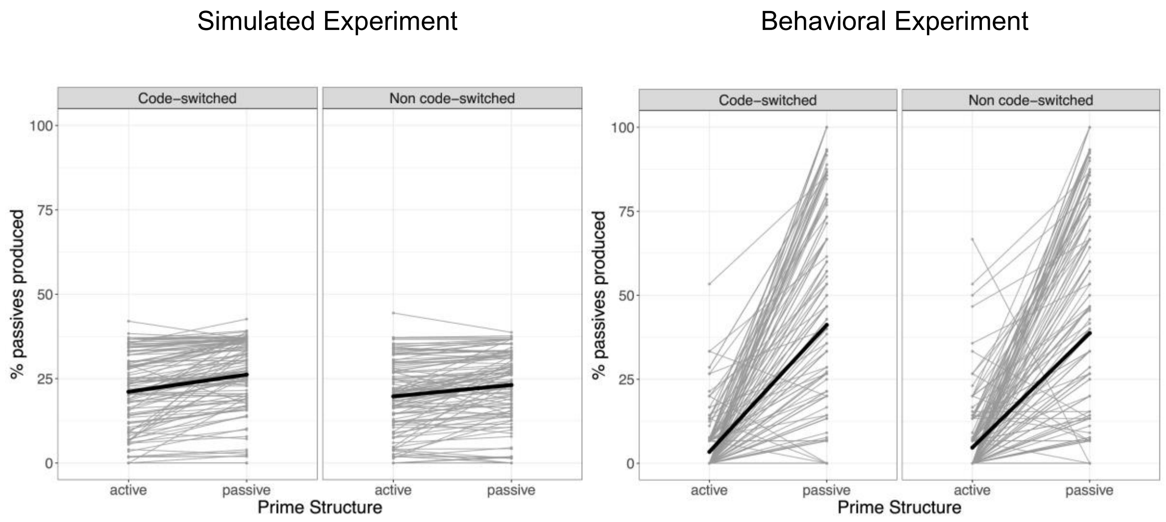


Fig. 1: Results from the simulated experiment (left half) and behavioral experiment (right half) with code-switched determiner and noun at beginning of prime sentence (left panels), see Example (a), compared to non-code-switched control primes (right panels). The thick black lines visualize the priming effect across all analyzed trials by connecting the percentage of passive responses after active primes to the percentage of passive responses after passive primes. The thin grey lines show the same for each individual (simulated) participant.

References

[1] Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414.

[2] Van Gompel, R. P., & Arai, M. (2018). Structural priming in bilinguals. *Bilingualism: Language and Cognition*, *21*(3), 448-455.

[3] Kootstra, G. J., & Rossi, E. (2017). Moving beyond the priming of single-language sentences: A proposal for a comprehensive model to account for linguistic representation in bilinguals. *Behavioral and Brain Sciences*, *40*.

[4] Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234.

[5] Tsoukala, C., Broersma, M., Van den Bosch, A., & Frank, S. L. (2021). Simulating code-switching using a neural network model of bilingual sentence production. *Computational Brain & Behavior*, *4*(1), 87-100.

[6] Khoe, Y. H., Tsoukala, C., Kootstra, G. J., & Frank, S. L. (2023). Is structural priming between different languages a learning effect? Modelling priming as error-driven implicit learning. *Language, Cognition and Neuroscience*, 1-21.

The Causal Role of Supplementary Motor Area (SMA) in Orthographic Retrieval During Chinese Character Handwriting

Yichi ZHANG[1], Xufeng DUAN[1], Zhenguang CAI[1]
[1] The Chinese University of Hong Kong
yichizhang@ling.cuhk.edu.hk

In the digital age, the act of handwriting has become increasingly rare, leading to widespread difficulties in recalling and writing characters for non-alphebetic speakers, a phenomenon known as “character amnesia.“ Writing involves complex cognitive processes, including central processes such as orthographic retrieval from long-term memory (LTM), phonology-orthography conversion (POC) and holding orthographic information in the working memory, together with peripheral processes such as allographic/letter shape selection and graphic-motor planning. While fMRI studies have suggested that the Supplementary Motor Area (SMA) plays a role in motor sequence memory and motor planning during handwriting [1, 2], its involvement in central processes like orthographic retrieval remains unclear. We therefore investigated the causal role of SMA in orthographic retrieval during Chinese character handwriting.

We conducted a between-subject offline transcranial magnetic stimulation (TMS) experiment with 40 healthy adult native speakers of simplified Chinese. Using neuronavigation with coordinates from our previous fMRI study [3], we applied inhibitory offline repetitive TMS (rTMS) to either SMA (experimental group) or the vertex (control group). Each participant underwent two 20-minute rTMS sessions (1 Hz frequency, 1200 pulses per session) organized in the following procedure: first rTMS session, followed by behavioral battery tasks (character generation by pronunciation/radicals, character copying, and writing speed tests), then second rTMS session, followed by the character dictation task. This structure allowed us to maintain the inhibitory effect throughout the critical dictation task while ruling out confounding effects on peripheral processing and POC route in central processing. The character dictation task assessed orthographic retrieval using 60 characters, which have an average amnesia rates with healthy adults of 32.8% in a previous study [4]. Participants wrote the 60 characters one by one according to a dictation prompt (e.g., 灶台的灶, "the character 灶 in the word 灶台"); after the handwriting, they were shown the target character and reported whether they had correctly written the target character, knew what character to write but could not write it, and incorrect handwriting (see also Figure 2). We collected the handwriting latency (reflecting orthographic access), handwriting duration (reflecting motor execution), and handwriting accuracy (reflecting orthographic retrieval failure).

Linear mixed-effects models revealed a significant increase in orthographic retrieval failure in the SMA stimulation group compared to controls (β = 0.90, z = 3.03, p = .002). Importantly, handwriting latency and duration remained unaffected, as did performance on other cognitive tasks, suggesting dissociable neural circuits where retrieval may fail while successful execution proceeds normally. The selective impairment in character dictation, without changes in timing measures, suggests that the SMA specifically contributes to orthographic retrieval, independent of motor planning or execution

These findings demonstrate that the SMA plays a causal role as a critical neural substrate for orthographic retrieval in Chinese handwriting, providing new insights into the neurocognitive mechanisms underlying written language production in logographic writing systems.

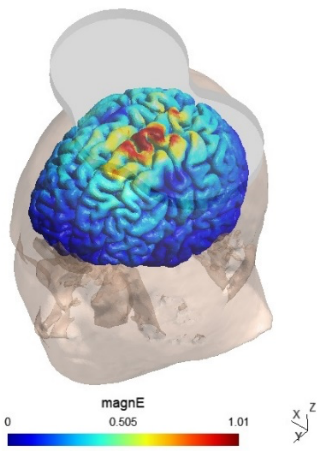


Figure 1. The simulated electric field for TMS targeting the SMA at Montreal Neurological Institute (MNI) coordinates x = -8, y = 14, z = 54. The color bar represents the electric field strength scaled from 0 (blue) to the maximum (red).

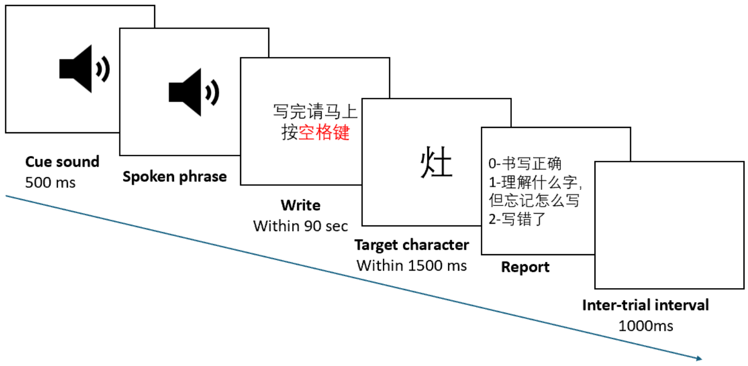


Figure 2. Schematic illustration of the trial structure. On the “Write” screen, the instruction “Please press SPACE upon finishing writing” was presented; on the “Target character” screen, the target character was shown; on the “Report” screen, participants were asked to press a number on the number pad indicating whether their handwriting was accurate (i.e., 0) or not (1 or 2 depending on the reason for the inaccuracy; see the text for more details)

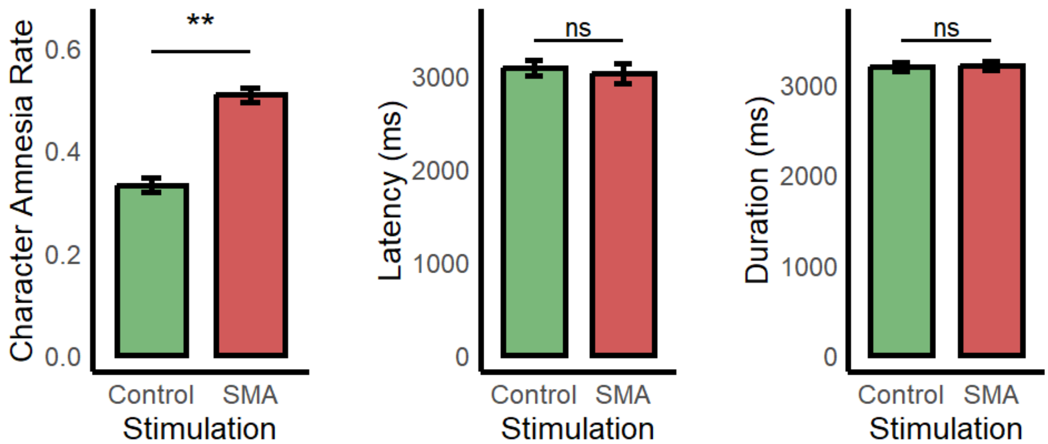


Figure 3. Accuracy rate, latency and duration as a function of stimulation group

References

[1] Planton, S., Jucla, M., Roux, F. E., & Démonet, J. F. (2013). The “handwriting brain”: a meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex*, *49*(10), 2772-2787.

[2] Purcell, J. J., Turkeltaub, P. E., Eden, G. F., & Rapp, B. (2011). Examining the central and peripheral processes of written word production through meta-analysis. *Frontiers in psychology*, *2*, 239.

[3] Duan, X. (2024). The Neural Basis of Decline in Written Production: Evidence from Chinese Handwriting. *The annual meeting of Society for the Neurobiology of Language (SNL)*. Oral presentation. Brisbane, Australia.

[4] Wang, R., Huang, S., Zhou, Y., & Cai, Z. G. (2020). Chinese character handwriting: A large-scale behavioral study and a database. *Behavior Research Methods*, *52*, 82-96.

Emergence of suffixing bias: Affixation patterns in L1 and sequence processing by statistical learning mechanisms

Mikhail Ordin, Leona Polyanskaya
Universidade de Coimbra, Faculdade de Medicina
mikhail.ordin@uc.pt

Background: The suffixing bias (the tendency to use suffixes more often than prefixes for grammatical meanings) in world languages is a typological bias [1-3]. The number of strongly suffixing languages is 4.5 times larger than the number of strongly prefixing languages [4]. This typological bias recapitulates in morphology a strong skew towards right-hand branching observed in syntax [5]. The origin of this typological bias is not clear. Among multiple theories, we focussed on the existence of cognitive bias that could explain the emergence of a typological bias: general cognitive mechanisms shape languages to be more easily processed by available cognitive machinery. Alternatively, the typological bias may be language-specific.
Method: We used statistical learning (SL) experiments to compare processing of suffixed and prefixed sequences on linguistic and non-linguistic material. SL is an evolutionary-ancient ability that was recycled for speech processing, but it is shared by taxonomically different species that do not have linguistic faculties [6-7], and is engaged for non-linguistic purposes by humans [8-9]. At the same time, SL is associated with syntactic and morphological processing [10]. We expected to observe a suffixing bias (the tendency to segment and memorize sequences with variable endings more easily than sequences with variable beginnings) in an artificial language learning paradigm, which would reveal a cognitive suffixing bias. The nonsense words embedded into a continuous stream were two-syllabic, and the sequences were tri-syllabic (the third syllable was added either at the beginning – prefix-modelling – or at the end – suffix-modelling – of the fixed bi-syllabic sequence (3 “prefixes and 3 “suffixes” were used). The recurrent sequences were concatenated into a continuous stream that participants had to listen to. After exposure, participants (1) listened to a legal prefixed sequence and legal suffixed sequence, and had to choose which one stem from the language (both answers are correct, and the choice shows the preference); (2) listened to a sequence with a prefix, suffix, no affix, or foil (random concatenation of the same syllables in the order, in which they never occurred consecutively during exposure) and to say if it is a word from the language or not. Also, the same paradigm was used with non-language stimuli, when non-verbalizable noises were used instead of syllables following the same transitional probabilities as in linguistic stimuli. The experiment was run with Basque speakers who are familiar both with inflectional prefixes and suffixes in their native language) and monolingual Spanish speakers (familiar only with inflectional suffixes) [11].
Results: No cognitive bias was observed. On non-linguistic material, suffixed and prefixed sequences of noises are learnt equally well (or badly), and when prefixed and suffixed sequences are paired, no preference was observed (suffixed sequences are chosen as frequently as prefixed). On linguistic material, the behaviour of the participants is determined by the morphology of the native language: Spanish monolinguals sow clear suffixing preference and learn suffixed sequences better than prefixed. Basque participants show no preference and no learning bias.
Discussion: No cognitive bias – at least in the SL ability – was observed to account for the emergency of the typological bias. Rather, it seems like the typological properties of the native language may influence doamin-general cognitive mechanisms.

References

[1] Sapir, E. (1921). *Language*. NY: Harcourt Brace.

[2] Hawkins, J. A., and Cutler, A. (1988) ‘Psycholinguistic Factors in Morphological Asymmetry’, in Hawkins J. A. (ed.) Explaining Language Universals. NY: Basil Blackwell.

[3] Cutler, A., Hawkins, J. A., and Gilligan, G. (1985) ‘The Suffixing Preference: A Processing Explanation’, Linguistics, 23(5): 723758.

[4] Dryer, M. S. (2005) ‘Prefixing Versus Suffixing in Inflectional Morphology’, in Haspelmath M., Dryer M. S., Gil D., and Comrie B. (eds.) The World Atlas of Language Structures. Oxford, UK: Oxford University Press.

[5] Hawkins, J. A., and Gilligan, G. (1988) ‘Prefixing and Suffixing Universals In Relation To Basic Word Order’. Lingua, 74(2-3), 219–259.

[6] Kikuchi, Y., Sedley, W., Griffiths, T., & Petkov, C. (2018). Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. Current Opinions in Behavioral Sciences 21, 145–153.

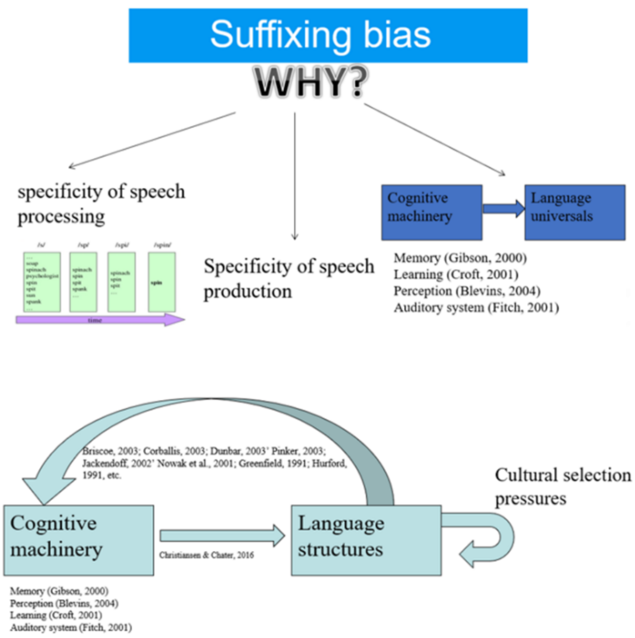
[7] Milne, A.E., Petkov, C., & Wilson, B. (2018). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. Neuroscience 389, 104–117.

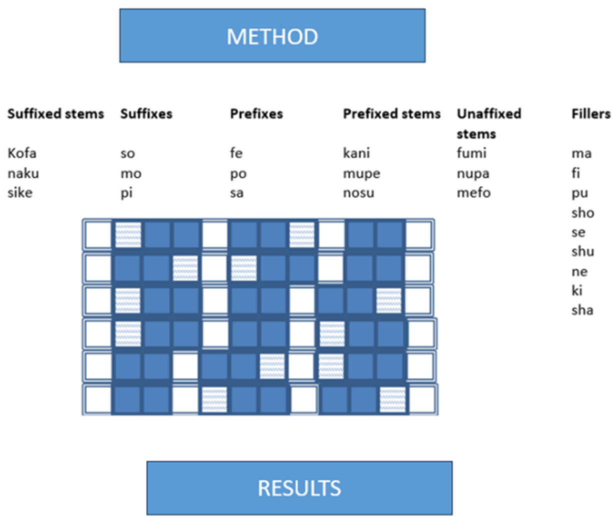
[8] Mitchel, A.D., Weiss, D., Senses, L. (2011). Cross-modal effects in multisensory statistical learning. Journal of Experimental Psychology: Learning, Memory and Cognition 37, 1081–1091.

[9] Conway, C.M. 2020. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. Neuroscience and Biobehavioral Review 112, 279–299.

[10] Kahta, S. & R. Schiff. 2016. Implicit learning deficits among adults with developmental dyslexia. Annals of Dyslexia 66: 235–250.

[11] De la Cruz-Pavía, I., Elordieta, G., Sebastián-Gallés, N., & Laka, I. (2014). On the role of frequency-based cues in the segmentation strategies of adult OV/VO bilinguals. International Journal of Bilingual Education and Bilingualism, 18(2), 225–241.





Grammatical prefixes in native language reduce suffixing preference (i.e., suffixing bias) in statistical artificial language learning on linguistic but not on non-linguistic material. Suffixing bias is probably determined by language experience, not by general cognitive mechanisms.

Does prediction require executive resources?

Leigh B. Fernandez[1], Muzna Shezad[2], & Lauren V. Hadley[2]
[1]University of Kaiserslautern Landau; [2] Hearing Sciences-Scottish Section, University of Nottingham
Corresponding author: leigh.fernandez@rptu.de

Prediction is the preactivation of information before it is directly encountered, and is an important theoretical framework for understanding how the brain processes information. It has been relatively well established that monolinguals actively make use of multiple sources of information to predict upcoming linguistic information before they encounter it. However, there are still several aspects of prediction that are not well understood. One key question focuses on whether prediction is necessary to understand language. While some current frameworks argue that this is the case[1], others argue that while prediction is useful, it is by no means a requirement[2]. A key distinction between these views stems from the proposed contribution of executive resources during prediction, with only the latter inferring their involvement[3].

Compelling evidence that prediction requires investment of executive resources comes from reduced prediction in groups of speakers with fewer available executive resources[2]. However, these groups (e.g., L2, children, and older adults) also often differ in linguistic experience[3], which has been proposed as an alternative basis of such prediction differences. The current study investigates how prediction is affected by executive resource availability when linguistic experience is matched, by investigating older adults with normal hearing (PwNH) vs older adults with hearing loss (PwHL), as well as within these participants through a memory load manipulation. PwNH and PwHL should have similar linguistic experience, but PwHL are faced with degraded auditory signals leading to a greater reliance on the cognitive system to extract meaning[4] and thus potentially reduced executive resources available for prediction[5].

30 PwNH (M_{age}=65.63) and 41 PwHL (M_{age}=71.27), matched in working memory, participated in a VWP study with semantically predictable and unpredictable sentences (see Table 1). The experiment featured two conditions: listening alone (no load) and listening while concurrently retaining a three-item visuo-spatial memory load (load). To investigate prediction, we compared looks to the target (*suit*) vs. a distractor (*tree*) while listening to *The tailor trimmed the suit*. Using divergence point analysis, we found that the timing in PwNH was later in the load relative to the no load condition (263.38ms [CI:180,340]). Timing in PwHL did not differ between the load and no load condition (47.34ms [CI:-180,120]). Additionally, the PwHL were reliably later in the no load condition relative to the PwNH (123.99 [CI:20,220]), but not in the load condition (-183.44 [CI:-280,100]). There were no differences in the unpredictable items. See Figure 1.

Overall, predictions were delayed for PwHL in the no load condition compared to PwNH. However, while PwNH were impaired by load, PwHL showed no such detriment. Specifically, for PwNH, the addition of load delays prediction such that PwNH behave similarly to PwHL. No differences emerged in the unpredictable items. This suggests that both hearing loss and cognitive load impact prediction specifically, which supports literature that prediction requires executive resources[6], and that taxing those resources (be they from more demanding listening or with an additional tasks) delays prediction mechanisms. It is unclear why the PwHL did not show an impact of load, but this may reflect a cognitive bottle neck, or perhaps a resiliency to cognitively demanding tasks (due to the greater experience of listening being challenging). Uncovering the cognitive architecture of linguistic prediction is an imperative step in forming precise theories of prediction and will provide a greater theoretical understanding of human intelligence and learning.

Table 1: Example stimuli

Condition	Example sentence	Corresponding visual array[7]
Predictable	The tailor trimmed the <u>suit</u>.	
Unpredictable	The guardian sells the <u>suit</u>.	

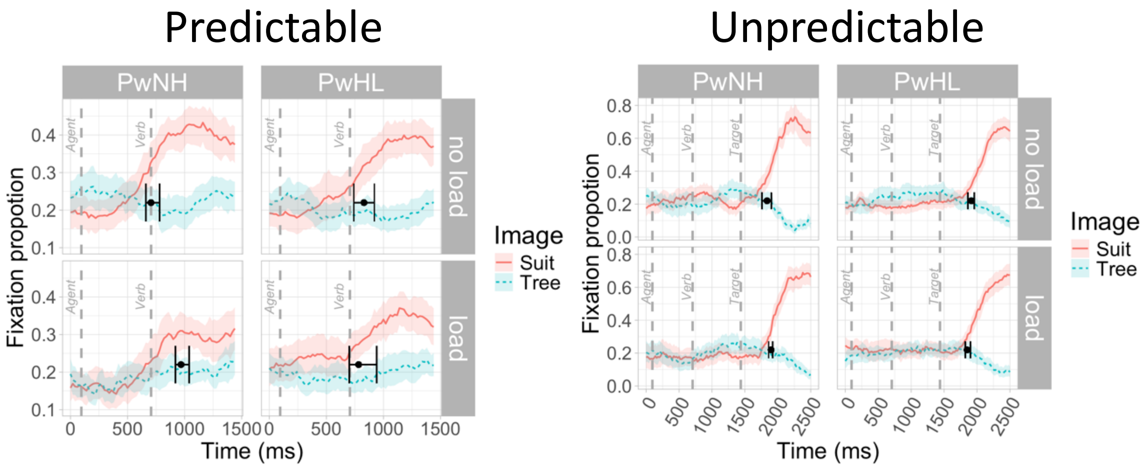


Figure 1. Divergence point and 95% confidence intervals superimposed on the fixation proportion of looks to the predictable items (left) and unpredictable items (right) across load and participant group.

1. Kuperberg, G. R., & and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

2. Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

3. Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258.

4. Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2).

5. Fernandez, L. B., Pickering, M. J., Naylor, G., & Hadley, L. V. (2024). Uses of Linguistic Context in Speech Listening: Does Acquired Hearing Loss Lead to Reduced Engagement of Prediction? *Ear & Hearing*, *45*(5), 1107–1114.

6. Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

7. Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816.

Extraction out of wh-clauses depends on the construction: evidence from French

Anne Abeillé, Emma Kious, Laboratoire de Linguistique Formelle, Université Paris Cité

Theories about locality (« island ») constraints come in three main varieties : for syntactic approaches, they are universal constraints, applying to all long distance dependencies (LDD) alike ; for processing approaches, they come from processing overload (Kluender & Kutas 1993) or low-frequency structures (Dabrowska 2008, Liu et al. 2021) ; for discourse-base approaches, they are backgrounded constituents out of which extraction is infelicitous (« backgrounded constituents are islands » BCI, Goldberg 2006, 2013).

A series of experiments have shown a systematic difference between LDDs for a variety of languages: while wh-questions show island effects for (finite) adjunct clauses and (finite) wh-clauses in Norwegian, relative clauses (RCs) and topicalizations do not (Kush et al. 2017, 2019, Kobzeva et al. 2022) ; in English, adjunct-if clauses show a penalty (compared to complement that-clauses) with wh-questions, but not with RCs (Sprouse et al. 2016, Nyvad et al. 2022); wh-questions and it-clefts show island effects for (nominal) subjects, compared to objects, in French and English, but RCs do not (Abeillé et al. 2020, Winckel et al 2024). This has led to a new discourse-based theory, building on the discourse function of the construction: « it is infelicitous to extract out of a backgrounded constituent with a focalizing LDD » (Focus Background Conflict (FBC) constraint: Abeillé et al. 2020, Winckel et al. 2024). If wh-questions and it-clefts are focalizing constructions (the extracted element is an information or contrastive focus), RCs and topicalizations are not (the extracted element is a topic).

We aim to test this theory on wh-clauses, which have been claimed to be weak islands (Kluender 1998, Szabolcsi, 2006), i.e. to allow for extracting complements but not adjuncts. For French, Godard (1988) claimed only finite wh-clauses to be islands, not infinitival ones (1). In a Google search, we found attested examples of relativizing out a non-finite wh-complement (2), but no examples of wh-questions.

In two on-line experiments, we tested (object) extraction out of non-finite wh-clauses, compared to infinitival complements of the same verbs. In Exp.1, we tested RCs (3), in Exp.2, wh-questions (4). Each has a 2x2 design: $\pm$wh, $\pm$extraction. The -extraction conditions were paraphrases with clausal coordinations in Exp.1, and yes/no questions in Exp.2. We chose d-linked wh-questions (*quel* 'which'+noun) and not bare ones, to make them closer to RCs (which also have a head noun). Each experiment had 16 items with comprehension questions, 16 fillers and 12 grammaticality controls. Both were acceptability judgment tasks on a 1-10 scale (IbexFarm). For Exp.1, we used data from 89 native French speakers from the RISC and social media, and data from 35 native French speakers from Prolific for Exp.2. The predictions of most syntactic theories are a penalty for extraction out wh-clauses, in both experiments, and the same holds for processing-based theories (since wh-clauses are less frequent than infinitival complements) and for the BCI, assuming wh-clauses are more backgrounded than non-wh ones, while the FBC predicts a penalty for Exp.2 only.

Exp.1 shows a preference for extraction and for -wh-complements, with no interaction (Fig.1, 0$\in$CI). Relativizing out of a +wh-clause rated much higher (6.13) than ungrammatical controls (2.75). In Exp.2, -wh-complements were also rated higher than wh-clauses: ordinal logistic Bayesian regression models show a robust interaction (β=3.15, CI=[2.2;4.1], P(β>0)=1) with a superadditive penalty for wh-questions out of a wh-clause.

These results contradict previous claims about non-finite wh-clauses (Godard 1988). They are predicted by the FBC : no wh-penalty with RCs and a penalty only with wh-questions. They are not compatible with frequency-based theories, nor the BCI. They are also compatible with Rizzi (1990)'s relativized minimality: the LDD is blocked if the filler crosses a barrier with similar syntactic features: a +rel filler may move out of a wh-clause in a RC construction, but not a +wh filler out of a wh-clause in a wh-question. We conclude that in French, extraction out of (non finite) wh-clauses depends on the LDD and can be acceptable with a relative clause, as with Norwegian (finite) wh-clauses (Kobzeva et al. 2022).

Selected references

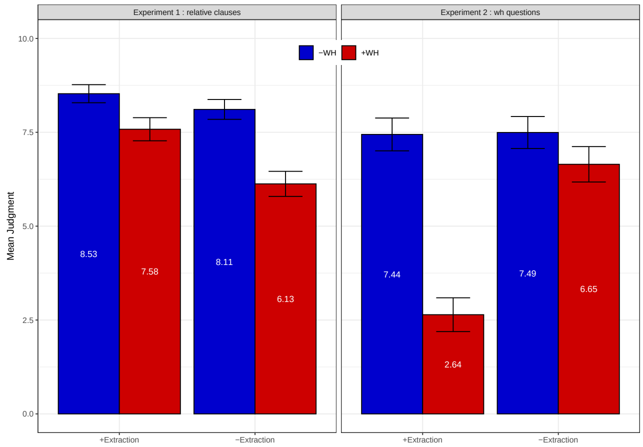


Figure 1: Mean ratings of Experiments 1 and 2

(1) Un enfant que je ne sais pas à qui confier/*je confierai. (Godard 1988)
'A child that I don't know to whom to.give /* I would.give'

(2) Voilà un curieux insecte que je ne sais où placer. (insecte.org, nov. 2016) 'This a curious bug that I don't know where to put'
Vous avez abordé la réponse à une question que je ne sais pas comment poser (Citemusique-marseille.com, may 2013) 'You have adressed the answer to a question that I don't know how to raise'

(3) Experiment 1 (16 items, 89 participants)
-wh,+extr : Mon frère a trouvé un appartement que je lui dis de rénover.
'My brother has found a flat that I tell him to renovate'
+wh,+extr : Mon frère a trouvé un appartement que je lui dis comment rénover.
'My brother has found a flat that I tell him how to renovate'
-wh,-extr : Mon frère a trouvé un appartement et je lui dis de le rénover.
'My brother has found a flat and I tell him to renovate it'
+wh,-extr : Mon frère a trouvé un appartement et je lui dis comment le rénover.
'My brother has found a flat and I tell him how to renovate it'

(4) Experiment 2 (16 items, 35 participants)
-wh,+extr : Quel appartement lui as-tu dit de rénover ?
'Which flat did you tell him to renovate?'
+wh,+extr : Quel appartement lui as-tu dit comment rénover ?
'Which flat did you tell him how to renovate?'
-wh,-extr : Lui as-tu dit de rénover l'appartement ?
'Did you tell him to renovate the flat?'
+wh,-extr : Lui as-tu dit comment rénover l'appartement ?
'Did you tell him how to renovate the flat?'

Novel word learning over different time scales: A comparison of paired-associate and cross-situational paradigms

Anne Neveu[1], Emma Libersky[1], Margarita Kaushanskaya[2]
[1] Kean University; [2]University of Wisconsin – Madison
aneveu@kean.edu

Word learning can take place through non-ambiguous word-referent pairings, as in paired-associate word learning (PAL), or in ambiguous contexts, as in cross-situational word learning (CSWL). Little is known about retention of novel words beyond immediate testing in PAL, and even less so in CSWL. Moreover, work in PAL has drawn on Baddeley's working memory model but the role of phonological working memory in supporting CSWL remains unclear.

We assigned 378 adult participants to PAL or CSWL and tested retention either immediately, after a 6-minute delay or a 24-hour delay. In both the PAL and the CSWL conditions, participants were taught twelve novel word-object pairings. In the PAL condition, participants viewed a single object and heard its name on each trial. In the CSWL condition, participants viewed two objects on each trial and heard them labeled in a random order. Participants then advanced to the same four-alternative forced choice recognition task. We examined the role of learning condition (PAL vs. CSWL), delay (immediate vs. 6-minute delay – Exp. 1; and immediate vs. 24-hour delay – Exp 2), and phonological working memory (measured with a backward digit-span task) in learning performance.

Experiment 1 results showed no difference in performance at immediate and 6-minute delayed testing, and better performance in PAL than in CSWL. Backward digit-span performance positively associated with word learning performance, over and above word learning paradigm and testing session, suggesting that phonological working memory plays a similar role in supporting word learning in PAL and in CSWL, at immediate and delayed testing.

Experiment 2 results indicate better performance in PAL than CSWL and at immediate than at 24-hour delayed testing. Moreover, there was a sharper drop in performance with the 24-hour delay testing in PAL than in CSWL. Phonological working memory tended to be positively associated with learning. Better phonological working memory was more robustly associated with higher performance at delayed testing in CSWL than in PAL suggesting that phonological working memory capacity may be especially critical to support longer-term retention in the more demanding CSWL paradigm.

These results indicate that word learning can take place in a non-ambiguous context, akin to a classroom, or in a more ambiguous immersion context. Our study shows that while we retain a higher proportion of new words when learning in a non-ambiguous context, memory for words learned in a more ambiguous context decays at a slower pace over time. Practically, this suggests that the two types of learning complement each other in supporting the learning of new vocabulary.

Figure 1

Probability of learning novel words as a function of word learning paradigm and timing condition, with standard error bars at immediate, 6-minute delay (Experiment 1) and 24-hour delay (Experiment 2).

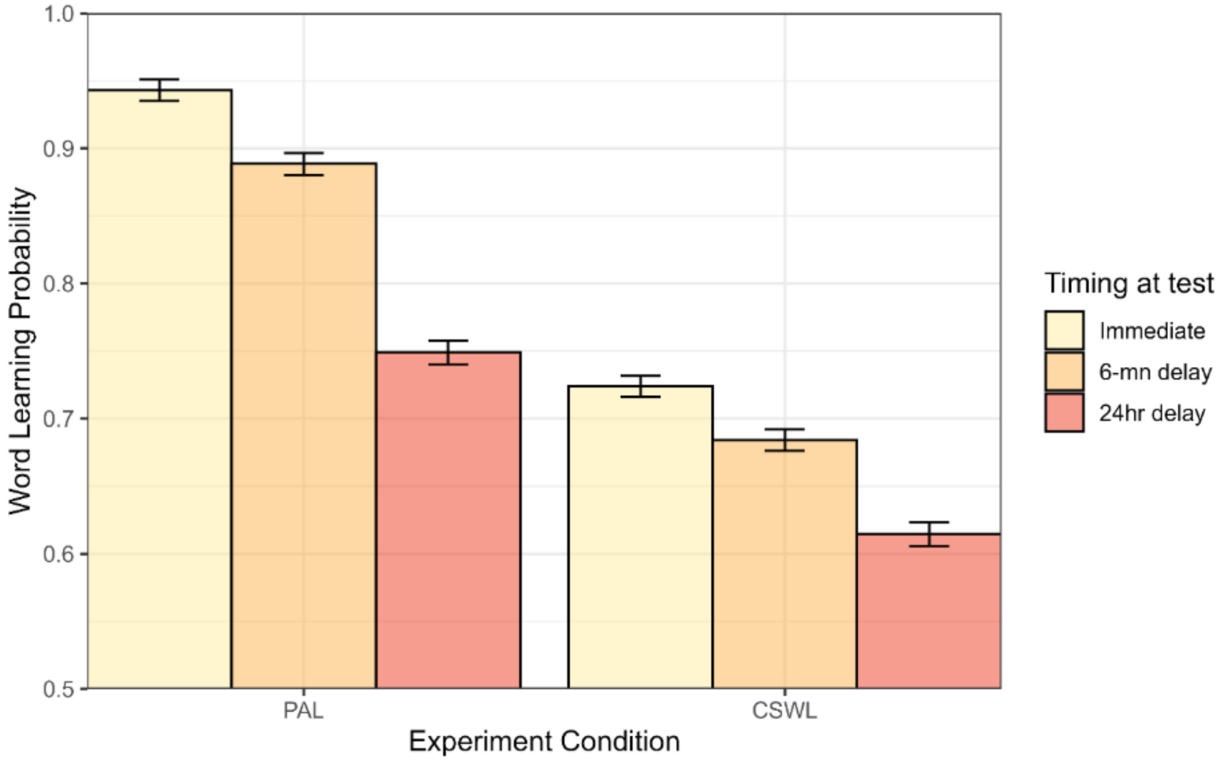
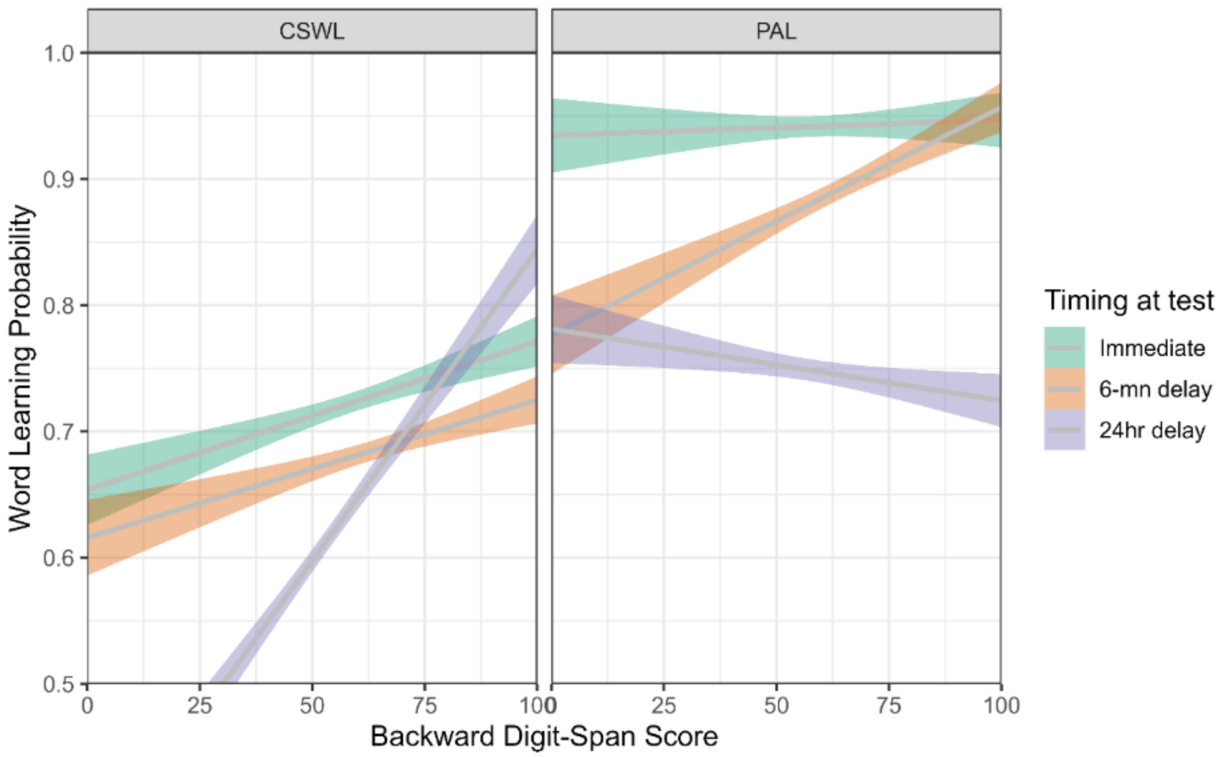


Figure 2

Probability of learning novel words as a function of word learning paradigm, phonological working memory score and timing condition, with standard error bars.



The processing costs of generic and specific singular *they*: A self-paced reading study

Dominic Schmitz
English Language and Linguistics, Heinrich Heine University Düsseldorf, Germany
dominic.schmitz@uni-duesseldorf.de

Background: The present study investigates the online processing of *he*, *she*, generic definite (GD) and specific ungendered (SU) singular *they* [cf. 1] by speakers of British English. Following the study by [2], it aims to show whether almost thirty years of language use lead to a change in the processing of singular *they*. That is, does GD *they* still come with shorter reading times than a stereotypically mismatched *he* or *she*, and does SU *they* still show the opposite effect?

Method: A self-paced reading task was conducted, following the design of [2]. Eighty native speakers of British English (mean age: 39 years, range 18-84) were recruited via Prolific. Their task was reading sentences consisting of three clauses, where each clause was given on a separate slide and disappeared when proceeded to the next (see Example 1 and Figure 1).

1. A / The **magician** has to perform flawlessly,
   even if PRONOUN may be nervous,
   because illusions rely on confidence and precision.

The first clause contained a stereotypically male or female role noun (bold in Example 1) based on the stereotypicality data from [3]. The introduced referent was either generic definite (the sentence started with an indefinite article) or specific ungendered (the sentence started with a definite article). The second clause contained one of the pronouns under investigation (*he*, *she*, *they*) referring to the referent. The third clause justified the claim made by the first two clauses. For each sentence, participants were asked whether they agreed with its statement to ensure parsing of the content. After the task, participants were asked their preferred pronouns as a more relevant alternative control variable for gender, as genders and pronouns do not show a one-to-one correspondence. Participants' response is henceforth used to differentiate between *participant type*s: HE and SHE participants. Reading time data normalised by number of syllables for the second clause were analysed in linear mixed-effects models following standard procedures [4], including variables such as *pronoun*, *stereotypicality*, *participant type*, and *age group*.

Results: With stereotypically male generic definite referents, the reading times of *they* are significantly longer than those of *he*, but not different to those of *she*. Figure 2 illustrates these findings. With stereotypically female generic definite referents, no differences between the pronouns were found. With specific ungendered referents, an effect of *pronoun* does not emerge.

Discussion: The present results indicate two main findings. First, the two types of singular *they* are processed differently. Second, for GD *they*, an influence of participant stereotypicality is found. In sum, it appears that the processing of GD *they* remains influenced by stereotypicality, whereas the processing of SU *they* does not. This finding indicates that SU *they* became more accepted over time, suggesting a change in its processing.

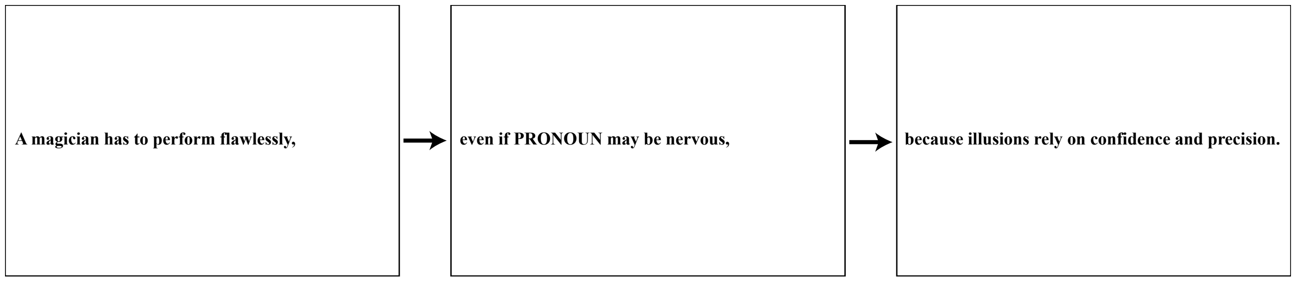


Figure 1: Trial structure: With each key press, the next clause of a trial appears.

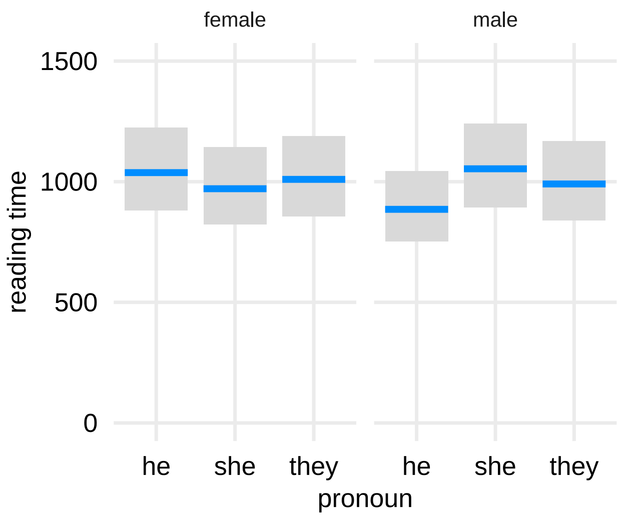


Figure 2: Effect of *pronoun* and *stereotypicality* (columns) on reading times for generic definite referents.

References

[1] Conrod, K. (2020). Pronouns and gender in language. In *The Oxford Handbook of Language and Sexuality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190212926.013.63

[2] Foertsch, J., & Gernsbacher, M. A. (1997). In search of gender neutrality: Is singular they a cognitively efficient substitute for generic he? *Psychological Science*, *8*(2), 106–111. https://doi.org/10.1111/j.1467-9280.1997.tb00691.x

[3] Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulikova, A., Öttl, A., Valdrova, J., Von Stockhausen, L., & Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871. https://doi.org/10.3758/S13428-013-0409-Z

[4] Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

The success of Neural Language Models on syntactic island effects is not universal: strong *wh*-island sensitivity in English but not in Dutch

Michelle Suijkerbuijk, Naomi Shapiro, Peter de Swart, and Stefan Frank
Centre for Language Studies, Radboud University

Background: A much-debated question in linguistics is how humans acquire grammatical knowledge: are we born with a language-specific learning capacity, or can we learn language from input alone? The recent introduction of neural language models (NLMs) is greatly influencing this debate. NLMs learn solely from their input in combination with their inductive biases, and thus without any built-in linguistic representations. If these networks can learn specific grammatical phenomena similarly to humans, this suggests that these phenomena can, in principle, be learned based on input alone. Recent research has looked at the learnability of one of the most studied phenomena in experimental syntax, syntactic island effects (see example in Table 1), to investigate whether NLMs are sensitive to island violations (e.g., [1]). Syntactic islands are an ideal test bed because they rarely occur in training data and NLMs do not have built-in linguistic knowledge to fall back on. Research has mostly shown successful results: NLMs seem able to model island effects in English. Yet, the behaviors or NLMs are almost never compared to human data and are almost exclusively researched in English. This makes it difficult to claim that NLMs can model island effects in ways that are comparable to humans. The present study addresses these gaps by incorporating data from human experiments and by looking beyond English.

Method: We make two important improvements on earlier work. First, we present an NLM and human participants with the same sentences (Table 1).[1] By collecting both model-assigned sentence probabilities and participant acceptability judgments, we directly compare whether the model represents island sensitivity similarly to humans. Second, we take this approach beyond English and compare NLM and human behavior in both English and Dutch, since the languages, though related, differ in their word order (SVO vs. SOV). Moreover, we are currently running similar experiments in Turkish, a language that is more morphologically complex than English and Dutch and has a flexible word order (we expect to have these results at the conference).

Results: The results for English and Dutch are shown in Figure 1. Figure 1 (top two plots) shows that the strong *wh*-island sensitivity of NLMs in English is replicated and that this sensitivity is comparable to that of English participants: the NLM and the participants show the same patterns in their results. The same cannot be said for Dutch, however (bottom two plots of Figure 1). While the Dutch participants showed a strong sensitivity to *wh*-island violations, with patterns comparable to English participants, the sensitivity of the Dutch model was not statistically significant (although in the right direction).

Conclusion: NLMs are not successful in all languages (yet) (e.g., Dutch), so more cross-linguistic research is necessary before NLMs can be claimed to bear on the human capacity for grammar learning.

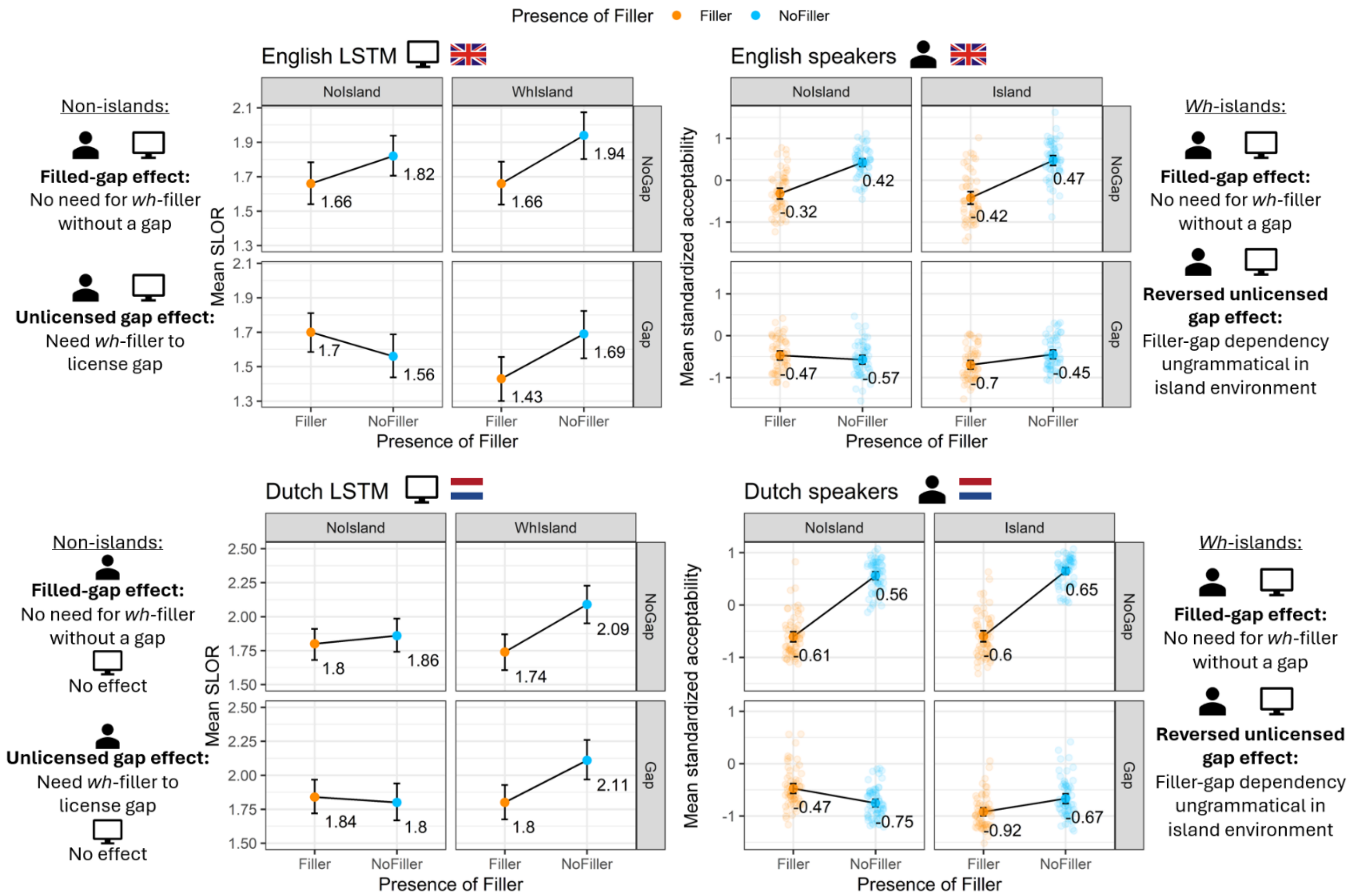
[1] We trained one Long Short-Term Memory network with one recurrent layer per language. The Dutch model was trained on 8,940,314 sentences from the NLCOW corpus (vocabulary: 20,194 tokens) and the English model on 10,904,314 sentences from the ENCOW corpus (vocabulary: 21,004 tokens), consisting of individual sentences collected from the World Wide Web.

Table 1. *Example sentences used in the Dutch (NL) and English (EN) experiments, crossing the factors* PRESENCE OF GAP *('cookies' vs. _ (a gap); indicated in orange),* PRESENCE OF FILLER *('that' vs. 'what'; indicated in blue) in non-islands and* wh*-islands (indicated in red between curly brackets).*

Gap?	Filler?		Example sentence in non-island and *wh*-island configuration
No	No	NL	*Ik weet dat jij {denkt dat/betwijfelt of} de bakker koekjes maakt in de bakkerij.* I know that you {think that/doubt if} the baker cookies makes in the bakery
		EN	I know that you {think that/doubt whether} the baker makes cookies in the bakery.
No	Yes	NL	**Ik weet wat jij {denkt dat/betwijfelt of} de bakker koekjes maakt in de bakkerij.* I know what you {think that/doubt if} the baker cookies makes in the bakery
		EN	*I know what you {think that/doubt whether} the baker makes cookies in the bakery.
Yes	No	NL	**Ik weet dat jij {denkt dat/betwijfelt of} de bakker _ maakt in de bakkerij.* I know that you {think that/doubt if} the baker GAP makes in the bakery
		EN	*I know that you {think that/doubt whether} the baker makes _ in the bakery.
Yes	Yes	NL	*Ik weet wat jij {denkt dat/*betwijfelt of} de bakker _ maakt in de bakkerij.* I know what you {think that/doubt if} the baker GAP makes in the bakery
		EN	I know what you {think that/*doubt whether} the baker makes _ in the bakery.

Note. The Dutch and English sentences only differ in the object-verb order in the embedded sentence (*koekjes maakt* vs. 'makes cookies').

Figure 1. *Mean standardized acceptability judgements (right plot) and mean Syntactic Log-Odds Ratio value (i.e., average, frequency- and length-corrected surprisal; left plot) for every combination of* PRESENCE OF GAP *and* PRESENCE OF FILLER *within non-islands (top and bottom left) and* wh*-islands (top and bottom right) for English (top plots) and Dutch (bottom plots). Error bars represent 95% confidence intervals.*



References

[1] Wilcox, E. G., Futrell, R., & Levy, R. P. (2024). Using Computational Models to Test Syntactic Learnability. *Linguistic Inquiry*, *55*(4), 805-848.

Utterances with Decreasing Entropy Facilitate Speech Comprehension and Concurrent Planning

Cecília Hustá[1], James Trujillo[1, 2], Judith Holler[1, 3], Linda Drijvers[1, 3], Antje Meyer [1, 3]

[1] Max Planck Institute for Psycholinguistics; [2] University of Amsterdam; [3] Donders Institute for Brain, Cognition and Behaviour

cecilia.husta@mpi.nl

Background: Effective dialogue relies on the seamless coordination of speech planning and comprehension, facilitating rapid turn-taking in conversation. Previous research has shown that Dutch utterances that require a response tend to have a decreasing entropy pattern [1], meaning that the uncertainty in upcoming words decreases as the utterance progresses. For next-turn speakers, lower entropy near the end of an utterance may reduce cognitive demands on comprehension, allowing more resources for speech planning.

Method: In this study, we investigated whether and how the degree of entropy decrease in utterances affects concurrent speech planning. Using a subset of utterances from casual Dutch face-to-face conversations [2, 3, 4], we examined whether comprehending utterances with more pronounced entropy decrease (MPED) versus less pronounced entropy decrease (LPED) affected concurrent speech planning of unrelated sentences. Participants planned sentences (e.g., De man hoort dat de vrouw klapt. [The man hears that the woman claps.]) based on two pictures and two verbs that appeared on the screen [5] while they listened to utterances (for the trial sequence see Figure 1). They held their response until the comprehension utterance ended and a response cue appeared on the screen, and then produced the sentence as quickly as possible. We included comprehension questions to measure attention to comprehension.

Results: Our results indicated that participants exhibited shorter speech onset latencies and produced fewer errors in the MPED condition compared to the LPED condition (for the speech onset latency results see Figure 2A), suggesting that a more predictable linguistic structure facilitated speech planning. Additionally, participants responded faster to the comprehension questions in the MPED condition than in the LPED condition (Figure 2B).

Discussion: These findings support the hypothesis that entropy reduction facilitates conversational turn-taking by optimizing resource allocation between comprehension and planning. Our study provides insights into how linguistic strategies naturally employed by speakers may contribute to communicative efficiency.

Figure 1. *Trial sequence of the experiment.*

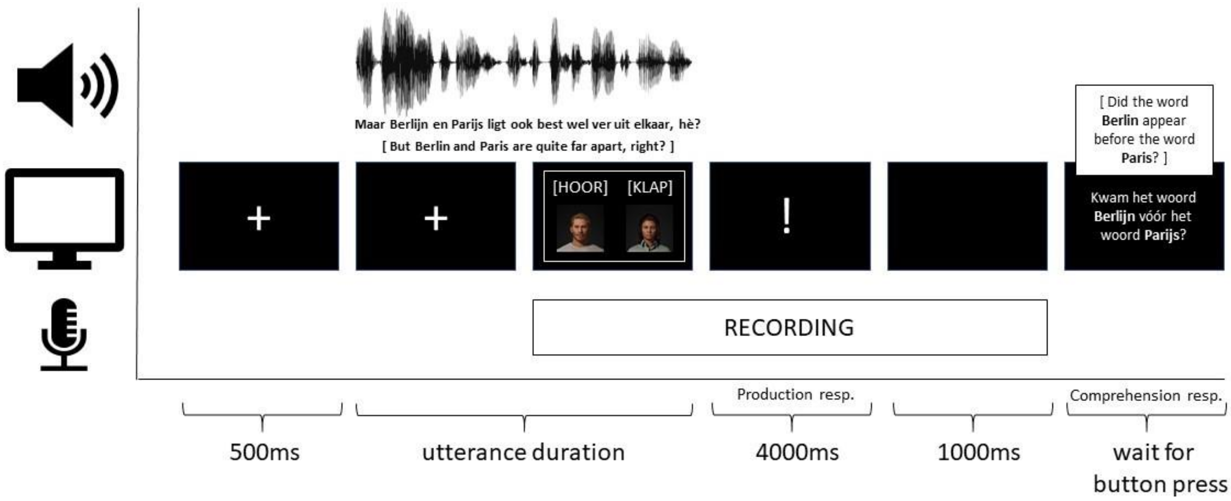
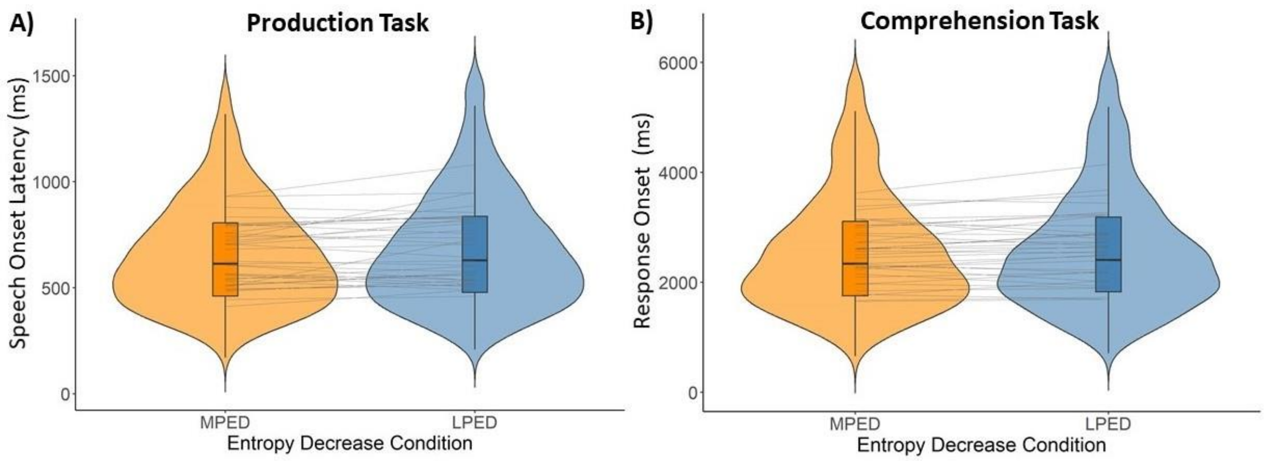


Figure 2. *Results. A) Shows the speech onset for the production task; B) shows the response onset for the comprehension task.*



References

[1] Trujillo, J. P., & Holler, J. (2025). Multimodal information density is highest in question beginnings, and early entropy is associated with fewer but longer visual signals. Discourse Processes, 0(0), 1–20. https://doi.org/10.1080/0163853X.2024.2413314

[2] Nota, N., Trujillo, J. P., & Holler, J. (2021). Facial Signals and Social Actions in Multimodal Face-to-Face Interaction. *Brain Sciences*, *11*(8), Article 8. https://doi.org/10.3390/brainsci11081017

[3] ter Bekke, M., Drijvers, L., & Holler, J. (2024). Hand Gestures Have Predictive Potential During Conversation: An Investigation of the Timing of Gestures in Relation to Speech. *Cognitive Science*, *48*(1), e13407. https://doi.org/10.1111/cogs.13407

[4] Trujillo, J. P., & Holler, J. (2021). The Kinematics of Social Action: Visual Signals Provide Cues for What Interlocutors Do in Conversation. *Brain Sciences*, *11*(8), Article 8. https://doi.org/10.3390/brainsci11080996

[5] Giglio, L., Ostarek, M., Weber, K., & Hagoort, P. (2022). Commonalities and Asymmetries in the Neurobiological Infrastructure for Language Production and Comprehension. *Cerebral Cortex*, *32*(7), 1405–1418. https://doi.org/10.1093/cercor/bhab287

It's not all Chinese to them: Differential heritage bilingual processing and rating of classifiers in Cantonese and Mandarin

Hannah Bou-Lai Lam, Johanne Paradis
University of Alberta
hblam@ualberta.ca

Background: Producing appropriate sortal classifiers in Chinese languages is challenging for heritage bilinguals, who tend to overuse the general classifier [1, 2, 3] rather than semantically appropriate classifiers. However, research on bilingual classifier comprehension is scarce, and scarcer still for online processing. Furthermore, there has been no systematic research examining whether variation in classifier grammar [4] and usage [5, 6] in different Chinese languages like Cantonese and Mandarin impacts bilingual acquisition.
Research questions: 1) How does classifier usage affect processing and acceptability ratings for heritage Chinese-English bilinguals, and 2) do different Chinese languages and participant individual difference factors modulate classifier processing and acceptability ratings?
Method: Using self-paced listening and acceptability ratings, we studied processing of classifiers by Cantonese- (N = 44) and Mandarin-heritage bilinguals (N = 25) in Canada. In the critical segment, participants were presented with either matching, mismatching, or the general classifier followed by nouns. To measure language proficiency, we administered a sentence repetition task in Cantonese/Mandarin. We also used the LEAP-Q [7] to collect information on language experience factors.
Results: To address RQ #1, we first ran linear mixed-effects models with the output variable of either reaction time at the critical segment or acceptability rating and the fixed effect being classifier condition. Random effects were participant and item. Linear mixed-effects models showed that mismatching classifiers were processed slowest (mismatch-match: p=.002; mismatch-general: p=.032) and rated worst (mismatch-match: p<.001; mismatch-general: p<.001). There were no processing differences between matching and general classifiers (p=.665).
Individual differences were added as additional fixed effects in linear mixed-effect models to answer RQ #2. Processing was modulated by an interaction between Chinese proficiency and exposure to Chinese from friends (Figure 1), where participants with greater Chinese proficiency were much faster in processing the critical segment if they had more exposure to Chinese-speaking friends. Chinese proficiency interacted with exposure to English from family (Figure 2): participants with the lowest amount of Chinese proficiency were slower at processing classifiers if they were exposed to more English at home, but higher Chinese proficiency coupled with more English exposure at home led to faster processing.
With acceptability ratings, language-specific effects emerged—only Cantonese-heritage bilinguals rated general classifiers worse than matching classifiers (p=.001) while Mandarin-heritage bilinguals rated general classifiers as being as acceptable as matching classifiers (p=.960) (Figure 3).
Discussion: Due to the heterogeneous linguistic experiences of these heritage bilinguals in our study, we found that processing of classifiers was influenced by language experience factors as well as proficiency in Chinese. The general classifier was processed similarly to matching classifiers, lending some insight into why the general classifier is overproduced [1, 2, 3]. However, differential rates of acceptability between the two languages may be attributable to Cantonese classifiers being more frequently used than Mandarin ones.

Figure 1

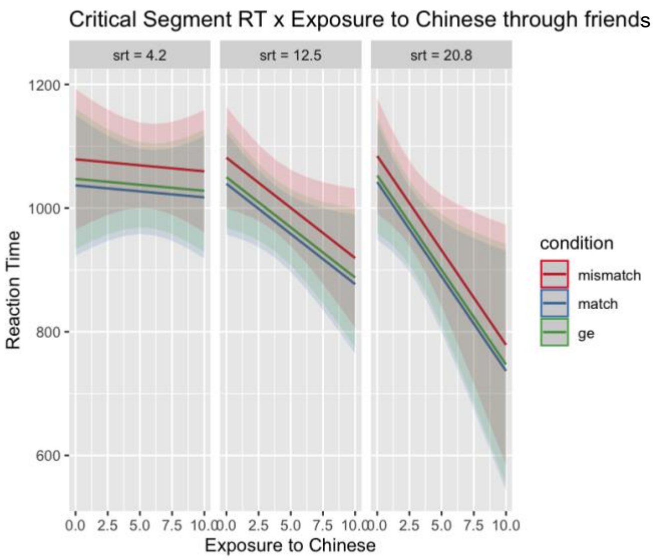


Figure 2

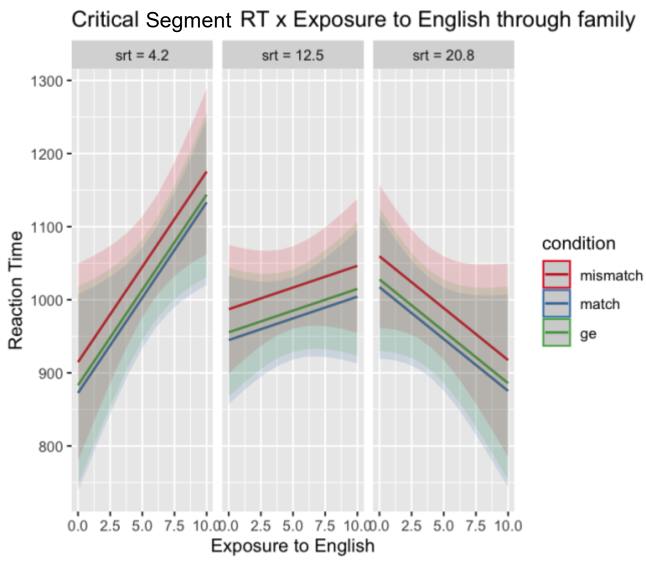
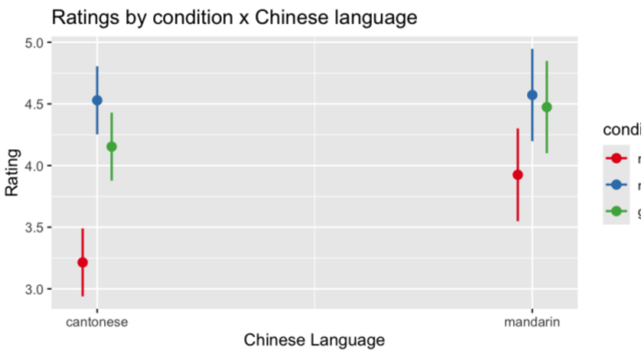


Figure 3



References

[1] Li, W., & Lee, S. (2001). L1 development in an L2 environment: The use of Cantonese classifiers and quantifiers by young British-born Chinese in Tyneside*. International Journal of Bilingual Education and Bilingualism*, *4*(6), 359-382.

[2] Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, *18*(4), 737-752.

[3] Kan, R. T. (2019). Production of Cantonese classifiers in young heritage speakers and majority language speakers. *International Journal of Bilingualism*, *23*(6), 1531-1548.

[4] Cheng, L. L. S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. *Handbook of Comparative Syntax*, 259-292.

[5] Erbaugh, M. S. (2002). Classifiers are for specification: Complementary functions for sortal and general classifiers in Cantonese and Mandarin. *Cahiers de linguistique Asie orientale*, *31*(1), 33-69.

[6] Erbaugh, M. (2006). Chinese classifiers: Their use and acquisition. In Li, P., Tan, L. H., Bates, E. & Tzeng, O. J. L. (eds), *Handbook of East Asian Psycholingusitics. Volume 1: Chinese*, 39–51. Cambridge: Cambridge University.

[7] Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940-967.

Sortal classifiers are matched with specific noun classes based on properties such as shape, natural kind, and function. The semantically correct classifier for (1a) is *zek* in Cantonese and *zhi* in Mandarin, which is reserved for non-human animate entities, and for (1b) the classifier *bun* in Cantonese and *ben* in Mandarin, which is reserved for books.

1. Cantonese

a.	*Jat/go*	*zek*	*gaau*
	One/that	CL	dog
	'One/that dog'		
b.	*Jat/go*	*bun*	*syu*
	One/that	CL	book
	'One/that book'		

2. Mandarin

a.	*Yi/zhe*	*zhi*	*gou*
	One/that	CL	dog
	'One/that dog'		
b.	*Yi/zhe*	*ben*	*shu*
	One/this	CL	book
	'One/this book'		

Table 1. Classifiers that were tested in Cantonese and Mandarin, each with two semantically-matching nouns.

Cantonese		**Mandarin**	
1. 本(書/雜誌)	'bun' – books	1. 本(書/雜誌)	'ben' – books
2. 條(魚/繩)	'tiu' – long objects	2. 條(魚/繩子)	'tiao' – long objects
3. 隻(貓咪/手)	'zek' – animate, non-human	3. 隻(貓咪/手)	'zhi' – animate, non-human
4. 架(火車/的士)	'gaa' – machines	4. 輛(汽車/自行車)	'liang' – vehicles
5. 幅(畫/相)	'fuk' – pictures/photos	5. 幅(畫兒/照片)	'fu' – pictures/photos
6. 棵(松樹/大白菜)	'po' – plants/trees	6. 棵(松樹/大白菜)	'ke' – plants/trees
7. 首(歌/唐詩)	'saau' – poems/songs	7. 首(歌/唐詩)	'shou' – poems/songs
8. 把(刀/尺)	'ba' – tools	8. 把(刀/鑰匙)	'ba' – tools
9. 座(山/大厦)	'zo' – mountain/large building	9. 座(山/樓)	'zuo' – mountain/large building
10. 粒(糖/核)	'lap' – small object	10. 粒(米/沙子)	'li' – small object
11. 件(事/大褸)	'gin' – clothing/situation	11. 件(事情/大衣)	'jian' – clothing/situation
12. 張(紙/凳)	'zoeng' – paper/furniture	12. 張(紙/桌子)	'zhang' – paper/furniture
13. 枝(牙膏/鉛筆)	'zi' – long, rigid object	13. 枝(牙膏/鉛筆)	'zhi' – long, rigid object
14. 塊(餅乾/檯布)	'faai' – cracker/soap	14. 朵(花/雲)	'duo' – flower/cloud
15. 部(電影/車)	'bou' – movie/car	15. 片(樹葉/沙灘)	'pian' – leaf/beach
16. 嚿(番梘/石)	'gaau' – blockish objects	16. 根(手指/黃瓜)	'gen' – long, pointy object
17. 間(舖頭/屋)	'gaan' – smaller building	17. 匹(馬/布)	'pi' – horse/cloth
18. 啖(味/風)	'zaam' – fragrance/wind	18. 頭(大蒜/騾子)	'tou' – garlic/donkey

Beyond Familiar Verbs: Czech-learning Children's Comprehension of Noncanonical OVS Word Order

Jolana Treichelová[1,2], Anna Chromá[1,2], Filip Smolík[2]
[1]Faculty of Arts, Charles University; [2]Institute of Psychology, Czech Academy of Sciences
jolana.treichelova@ff.cuni.cz

Children acquiring different languages respond differently to canonical and non-canonical word orders, depending on the cues for semantic-role interpretation such as word order and case marking. Some researchers have argued that using familiar verbs may bias sentence interpretation, as children might rely on verb-specific patterns rather than abstract syntactic knowledge (the verb island hypothesis) [1]. To control for these effects, some studies on English and German have used novel verbs to investigate how children assign thematic roles based on (morpho)syntactic cues alone, independent of lexical-semantic knowledge [2–4]. In languages with case morphology like Czech, an effect of verb-specific knowledge might be less expected if the role interpretation is driven by the case marker. However, this has never been studied before.

A prior study on Czech reported above-chance performance in OVS-structure comprehension in children before the age of four but used only known verbs [5]. The present study uses prior Czech data with familiar verbs [6] and compares them with novel-verb data obtained in a parallelly designed procedure.

The comprehension of OVS structures with known verbs [6] was tested among 30 children between 3;6 and 5;6 (mean = 51.5 months, SD = 4.8). In each trial, children were presented with two pictures showing reversed agent–patient roles and a recorded sentence describing one of them. Children's gaze was tracked while listening, and they were asked to point at the corresponding picture afterwards. Each child saw the same picture pair four times, with both SVO and OVS sentences for each of the pictures. There were eight items (32 stimuli) in total. Mixed-effect models were fitted on the data: binomial models of SVO-pointing, and linear models of SVO-looking proportion. The results supported findings from a previous study [5], showing that children correctly interpreted the case marker in both SVO and OVS conditions. The post-hoc division of the sample into a younger and an older group (by age-median split) revealed no striking change in development: the difference between SVO and OVS condition was significant at all sentence segments from the verb onwards in both groups (Figure 1). However, on the verb segment, the difference was much less pronounced in the younger group which suggests that younger children process the case-marking slower. Nevertheless, given the use of familiar verbs, the role of verb-specific expectations could not be excluded as support for the general high performance.

To address this limitation, we are currently conducting a follow-up study using pictures of unfamiliar transitive actions performed by familiar characters (see example in Figure 2), with sentence stimuli created using novel verbs and familiar nouns (see example in Table 1). The design and procedure of the study are otherwise the same as before. The study is planned to be carried out with Czech-speaking children aged 2;6 to 5;0 years.

We expect that children will correctly interpret both SVO and OVS sentences based on case marking, even with novel verbs. However, OVS structures may be more demanding to process, resulting in slower or less confident responses. This would reflect general processing difficulty rather than reliance on verb-specific knowledge.

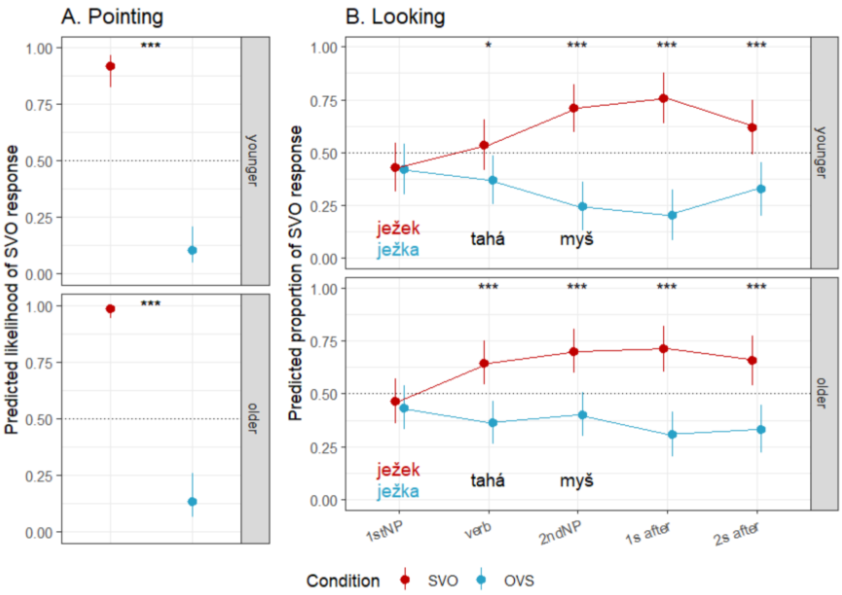


Figure 1 Values predicted by the mixed-effect regression: likelihood of the SVO response in pointing (panel A) and proportion of SVO response in gaze behavior (panel B) in the experiment with familiar verbs. Error bars represent 95% confidence intervals. Asterisks mark significance level of the word-order effect in the particular condition tested by the mixed-effect model (*** p<.001, * p<.05).

Figure 2 Example of an unfamiliar transitive action performed by familiar characters (mirror-reversed)

Condition	Czech	English translation
SVO	Tučňák tebá vlka.	*The penguin tebs the wolf.*
	Vlk tebá tučňáka.	*The wolf tebs the penguin.*
OVS	Vlka tebá tučňák.	*The penguin tebs the wolf.*
	Tučňáka tebá vlk.	*The wolf tebs the penguin.*

Table 1 Example of one novel verb item across SVO and OVS conditions

Bibliography

[1] Tomasello, M. (1992). First verbs: a case study of early grammatical development. Cambridge: Cambridge University Press.

[2] Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. Journal of Child Language, 26, 339-356.

[3] Chan, A. Meints, K., Lieven, E., & Tomasello, M. (2010). Young children's comprehension of English SVO word order revisited: Testing the same children in act-out and intermodal preferential looking tasks. Cognitive Development 25(1), 30-45.

[4] Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. Child development, 79(4), 1152-1167.

[5] Smolík, F. (2015). Word order and information structure in Czech 3-and 4-year-olds' comprehension. First language, 35(3), 237-25.

[6] Chromá, A., Smolík, F., Treichelová, J., & Friedrich, C. (in prep). Cross-linguistic comparison of Czech- and German-learning four-year-olds' comprehension of noncanonical word order.

Word order and case marking in Czech

Czech is a language with flexible word order, where grammatical relations between sentence constituents are primarily marked by case marking rather than word order. For example, the sentences "Pes$_{nom}$ honí kočku$_{acc}$" and "Kočku$_{acc}$ honí pes$_{nom}$" are interpreted identically as *The dog chases the cat*, because the noun endings indicate which noun is the subject and which is the object, regardless of their position in the sentence. In nouns, a single suffix can simultaneously encode gender, number, and case. Moreover, the morphemes marking different case and number combinations can be homophonous across words; for example, the suffix -a in the feminine noun "sov**a**" (owl) marks nominative singular, while -a in the masculine noun "kluk**a**" (boy) marks accusative or genitive singular. Additionally, some case forms can be syncretic and do not differ from each other — for example, in transitive sentences, the nominative and accusative forms can be identical (e.g., "kuře" nom/acc).

Reliable measures of orthographic statistical learning predict spelling but not reading skill

Haoyu Zhou[1], Fabienne Chetail[2], Louisa Bogaerts[1]
[1] Department of Experimental Psychology, Ghent University ; [2] Laboratoire Cognition Langage Développement (LCLD), Research Centre in Cognition & Neuroscience (CRCN), Université Libre de Bruxelles
haoyu.zhou@ugent.be

Background: Our language environment is rich in statistical structure. It is therefore natural to assume that statistical learning (SL) —the ability to detect regularities in the environmental input—plays a role in processing written text and supports efficient literacy skills. Small to moderate correlations between SL and reading have been reported [1], but other studies reported null findings [2]. We propose that this link can be better revealed by measuring individual sensitivity to orthographic regularities, a type of regularity more directly relevant to literacy. We focused on bigram frequency as a specific orthographic regularity, since the bigram processing was proposed as a key step toward whole-word processing [3]. The current study aimed to clarify how sensitivity to bigram frequency in natural language and learning of novel bigrams (i.e., orthographic learning) relate to literacy outcomes in French, using reliable tasks.

Method: Sensitivity to bigram frequency was measured using a word-likeness task in which participants chose between pseudoword pairs differing in bigram frequency. Orthographic learning was assessed using an incidental exposure paradigm in which participants were introduced to four illegal French bigrams embedded in consonant strings and then tested on recognition of new strings containing the trained bigrams. Literacy skills were evaluated using standardized word spelling and reading tests (BOQS, [4]; ECLA 16+, [5]), with transcription accuracy and read-aloud accuracy as key outcomes. Note that the read-aloud accuracy is a ratio of total reading time over correctly read words (seconds) and thus negative correlations between reading and the other measures are expected. Data from 113 French-speaking participants were analyzed. Preprint of the study is available at https://osf.io/preprints/psyarxiv/6kf4u_v1.

Results: Both sensitivity to bigram frequency and orthographic learning tasks showed moderate-to-high split-half reliability (Table 1), yet they were not significantly correlated with each other (Figure 1). Crucially, only spelling scores correlated significantly with these measures (Figure 1), while reading accuracy did not.

Discussion: Our findings suggest that reliable measures of orthographic SL predict individual differences in spelling but not reading ability in French. This may reflect the greater demand for precise orthographic representations in spelling than in reading, and is consistent with prior work linking orthographic learning to spelling even in more transparent languages such as German [6]. In contrast, these results may not generalize to languages like Arabic and Hebrew, where readers are more sensitive to letter positions. The lack of correlation between the two orthographic SL tasks suggests that learning regularities from scratch in an aritifical task does not reflect the same individual differences as sensitivity to regularities in one's native language. However, the fact that both measures relate to spelling underscores the complementary contributions of artificial and naturalistic SL to literacy outcomes. Future work could further explore the relation between SL and reading in a child sample where there might be more variability in reading skills, or by including other reading components such as silent visual word recognition or text-level reading.

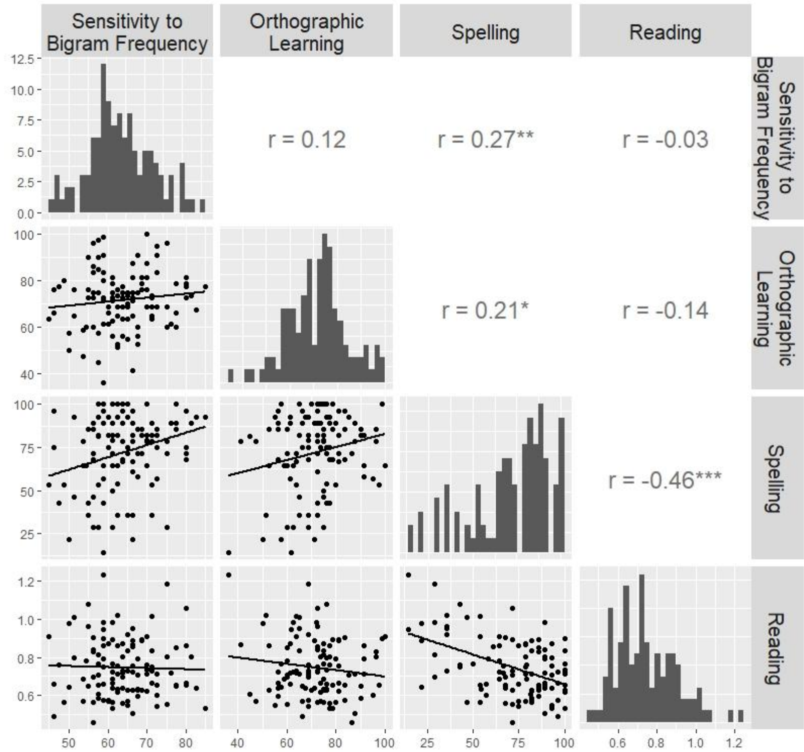
Table 1.

Summary of results for the two orthographic SL tasks

Task measures	Mean	SD	Split-half reliability
Sensitivity to bigram frequency (% accuracy)	63.86	8.36	r = 0.60; 95CI (0.48; 0.70)
Orthographic learning (% accuracy)	71.62	12.36	r = 0.84; 95CI (0.80; 0.88)

Figure 1.

Correlation results matrix



Note. Effect sizes are labeled in the corresponding grids. Significance: *, p <.05; **, p <.01; ***, p < .001 (not corrected for multiple comparisons).

References

[1] Ren, J., Wang, M., & Arciuli, J. (2023). A meta-analysis on the correlations between statistical learning, language, and reading outcomes. *Developmental psychology, 59*(9), 1626.

[2] van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2021). The contribution of individual differences in statistical learning to reading and spelling performance in children with and without dyslexia. *Dyslexia*, *27*(2), 168-186.

[3] Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe?. *Trends in cognitive sciences*, *8*(2), 58-59.

[4] Chetail, F., Porteous, F., Patout, P.-A., Content, A., & Collette, E. (2019). BOQS: An orthographic quality scale for adults*. Revue de Neuropsychologie, 11*, 214–226.

[5] Gola-Asmussen C., Lequette C., Pouget G. *ECLA16 + : évaluation de compétences de lecture chez l'adulte de plus de 16 ans*. CeFoCOP/Université de Provence Aix-Marseille I: Cognisciences LSE Université Pierre Mendès, France, Grenoble; 2010.

[6] Schmalz, X., Mehlhase, H., Schulte-Körne, G., Moll, K., & Wang, H.-C. (2024). *Do faster learners know more? Orthographic learning and knowledge, sensitivity to graphotactic regularities, spelling accuracy, and reading ability in German primary school children.* PsyArXiv. https://doi.org/10.31234/osf.io/ywjrb.

When cues collide: The role of contextual and classifier-based prediction in Mandarin comprehension

Bernard A.J. Jap[1] & Yu-Yin Hsu[2]
[1]School of Arts and Social Sciences, Hong Kong Metropolitan University
[2]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
bajjap@hkmu.edu.hk & yu-yin.hsu@polyu.edu.hk

Background: Predictive processing in language comprehension integrates cues across different levels [5, 6]. In Mandarin Chinese, classifiers must precede and agree (somewhat semantically) with their associated nouns, which provide local morphosyntactic cues to upcoming nouns. While previous research [1] reported N400 effects at the classifier position when it mismatched an upcoming predictable noun, suggesting an effect of pre-activation of semantic features, other studies [2, 3, 4] found ERP primarily at the noun position rather than at the classifier itself. We investigated how hierarchical prediction operates when global contextual cues and local classifier cues provide conflicting information about an upcoming noun.

Method: We recorded EEG data from 50 native Mandarin speakers while they read 160 sentences word-by-word (except the context and final part). Stimuli (see below) contained a high-constraint context (e.g., "woodcutter chop") followed by either an appropriate (*ke*) or inappropriate (*feng*) classifier for the expected noun ("tree"), which was then followed by either the expected noun ("tree") or an unexpected noun ("grass"). This design has four conditions: (1) congruent classifier + expected noun, (2) congruent classifier + unexpected noun, (3) incongruent classifier + expected noun, and (4) incongruent classifier + unexpected noun. EEG data were preprocessed using EEGLAB (0.1Hz high-pass, 40Hz low-pass filters, ICA, & artifact rejection) and analyzed with FieldTrip using cluster-based permutation test: 5000 permutations, whole epoch, one-tailed for negative effects.

Example stimuli (Context: 樵夫砍伐的 'The woodcutter chopped...')

Condition	Det	Classifier	Noun	Gloss (Classifier - Noun)	Congruency (-/+)
1	这	棵 (*kē*)	树木 (*shùmù*)	this CL:tree - tree	+C +N
2	这	棵 (*kē*)	小草 (*xiǎocǎo*)	this CL:tree - grass	+C -N
3	这	封 (*fēng*)	树木 (*shùmù*)	this CL:letter - tree	-C +N
4	这	封 (*fēng*)	小草 (*xiǎocǎo*)	this CL:letter - grass	-C -N

(Note: The final part of the sentence, e.g., 非常大 'very large', followed the noun)

Results: Figures 1-2 show grand average ERPs and topographical plots of the N400 effects. Unlike [1], we found no significant effects at the classifier position. At the noun position, unexpected nouns elicited significantly larger N400s than expected nouns following both appropriate classifiers (p=0.0008, cluster: 297-588ms) and inappropriate classifiers (p=0.0234, cluster: 268-451ms), with typical centro-parietal distributions. Visual inspection suggested a potential reduction in N400 effect for inappropriate versus appropriate classifier conditions; however, this interaction did not reach significance (p=0.0604, d=-0.11).

Discussion: Our findings reveal a prediction system where strong contextual cues remain effective despite conflicting local morphosyntactic information [7, 8]. Under high contextual constraint, the comprehension system prioritizes global contextual predictions over classifier-noun agreement—perhaps putting the cue from the classifier 'on-hold'. This aligns with predictive coding frameworks [6], where prediction errors are weighted by precision, as well as noisy channel models [9], where conflicting cues may be treated as "noise" when they contradict reliable contextual information. These results support a cue integration system [10] where highly constraining context may contribute more to predictions than classifier information. However, one caveat remains: the interaction results (p=0.6) may suggest that there is *still* an effect of local morphosyntactic violations, where readers have to resolve the tension between the context cue and the cue from the classifier.

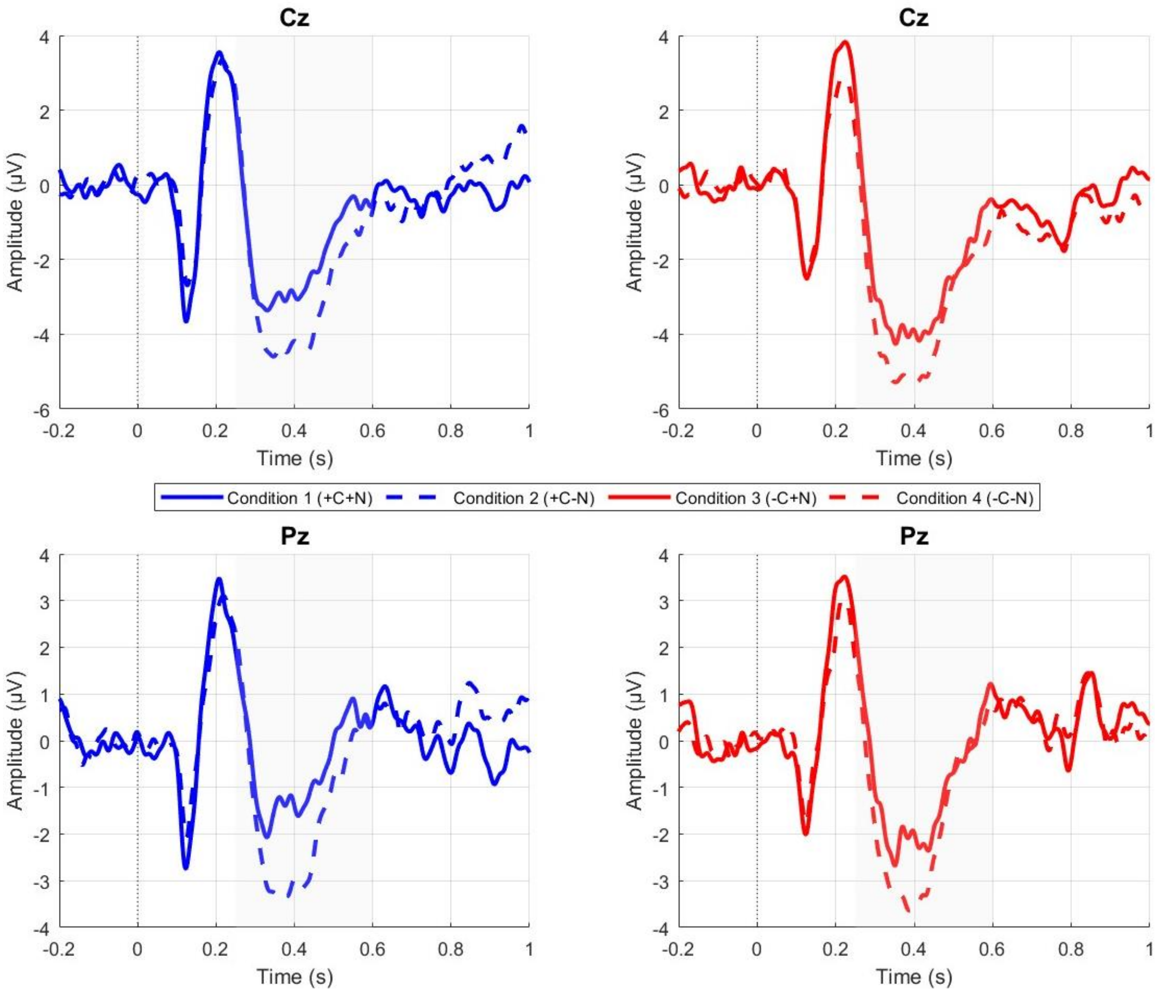


Figure 1. Grand average ERPs at noun position for electrode Cz (top) and Pz (bottom).

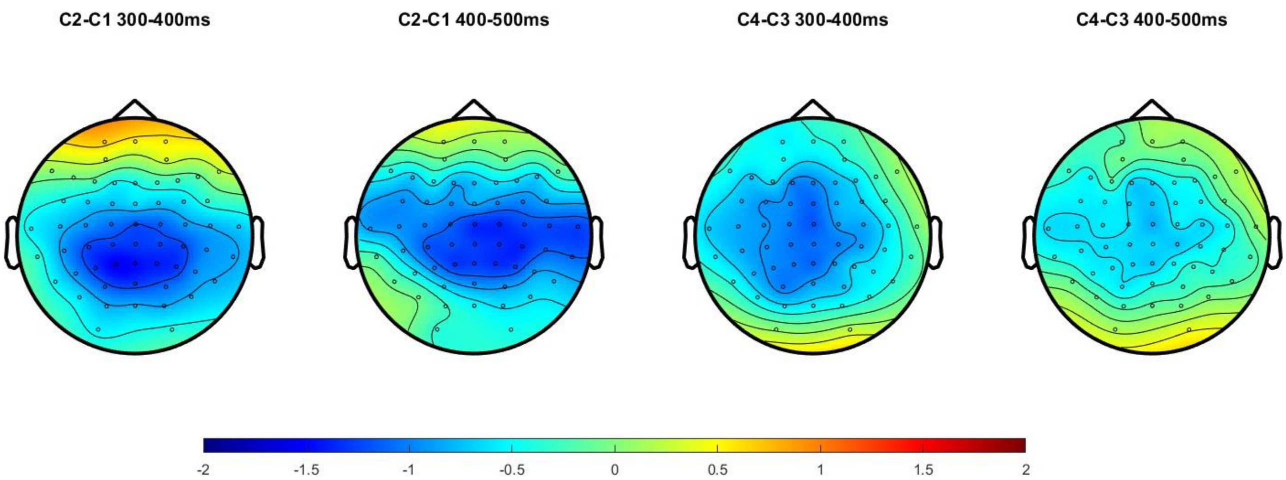


Figure 2. Topographic distribution of N400 effects (unexpected minus expected noun) for two time windows.

References

[1] Kwon et al., 2017, Cognition
[2] Zhang et al., 2012, Brain Lang.
[3] Zhou et al., 2010, Neuropsychologia
[4] Hsu et al., 2014, Brain Lang.
[5] Kuperberg et al., 2016, Lang. Cogn. Neurosci.
[6] Rabagliati et al., 2019, Psychol. Med.
[7] Ferreira et al., 2018, Curr. Dir. Psychol. Sci.
[8] DeLong et al., 2005, Nat. Neurosci.
[9] Gibson et al., 2013, PNAS
[10] MacDonald et al., 1994, Psychol. Rev.

On the lexical representation(s) of compounds: Evidence from continuous naming in young and older healthy speakers

Antje Lorenz[1], Anna-Lisa Döring[2], Lara Mundt[2], Pienie Zwitserlood[3], Rasha Abdel Rahman[2]
[1]Faculty of Linguistics and Literary Studies, Bielefeld University; [2]Department of Psychology, Humboldt-Universität zu Berlin; [3]Department of Psychology, University of Münster
antje.lorenz@uni-bielefeld.de

Background: Competing accounts exist on how compound nouns (e.g., Teekanne [teapot]) are lexically stored and processed in speech production ([1] vs. [2]). Previous data from the continuous naming paradigm with young adult speakers suggest a multiple-lemma representation of compounds, consisting of a compound lemma that is directly linked to the compound's constituents at the lemma level ([3])? Whether the lexical representation of compounds changes with age is also a matter of debate. Older neurotypical speakers are generally slower and have more word-finding difficulties than young speakers. Previous data from the picture-word interference task with compound targets suggests stable lexical representations of compounds with age (e.g., [4]). The current study aims to complement earlier findings and to replicate the cumulative semantic effects in compound production.

Method: In the continuous picture naming task, members of different semantic categories, such as *dog*, *fox*, *horse*, *zebra*, and *donkey* for the category ANIMALS are named in a seemingly random order, separated by 2 to 8 unrelated objects from other categories. Participants' naming latencies increase with each additional member of a semantic category, that is interference accumulates within categories. This cumulative semantic interference reflects lexical-semantic processes in speech production ([5, 6]). The present study used the continuous naming paradigm with noun-noun compound targets ([3]). Effects of healthy aging were assessed in a group design with 32 participants each (young: M=27 years, SD = 4.59; Min. = 19, Max. = 39 years; older: M=70 years, SD = 5.36, min. = 60, max. = 83 years). To test whether the constituents of compounds are stored and retrieved separately, category membership was established through the compounds' first constituents, their modifiers, whereas the compounds (and their heads) were not categorically related (e.g., for the category animals: Hundeleine (dog lead), Fuchsbau (fox's burrow), Pferdekutsche (horse-drawn carriage), Zebrastreifen (zebra crossing), Eselsohr (dog-ear (in a book), lit. donkey's ear)). In a control condition, pictures depicting the compounds' modifiers (e.g., dog, fox, horse, zebra, donkey) were presented to assess whether the paradigm works.

Results: Separate linear-mixed effects models were used to assess age-related effects in simple-noun and compound production. In line with our predictions, older speakers needed more time and produced more picture-naming errors than young speakers. Furthermore, cumulative semantic interference was confirmed for both age groups, both with simple nouns and with the first constituents of compound targets (for compounds, see Figure 1).

Discussion: Our data suggest a multiple-lemma representation of noun-noun compounds in German, replicating data from a previous study with young adult speakers ([3]). While the older participants were generally slower than the young, no age-related effects were observed in the strength of cumulative semantic interference, in line with stable lexical-semantic processes in the elderly ([4]).
Potential effects of the semantic transparency of compounds will also be presented.

References

[1] Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1- 38.

[2] Marelli, M., Aggujaro, S., Molteni, F., & Luzzatti, C. (2012). The multiple-lemma representation of Italian compound nouns: A single case study of deep dyslexia. Neuropsychologia, 50(5), 852–861.

[3] Döring, A.-L., Abdel Rahman, R., Zwitserlood, P., & Lorenz, A. (2022). On the lexical representation(s) of compounds: A continuous picture naming study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 48(1), 43–59.

[4] Lorenz, A., Regel, S., Zwitserlood, P., & Abdel Rahman, R. (2018). Age-related effects in compound production: Intact lexical representations but more effortful encoding. Acta Psychologica, 191, 289–309.

[5] Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. Cognition, 100, 464–482.

[6] Abdel Rahman, R. & Melinger, A. (2019). Semantic processing during language production: an update of the swinging lexical network. Language, Cognition and Neuroscience, 34, 1176–1192.

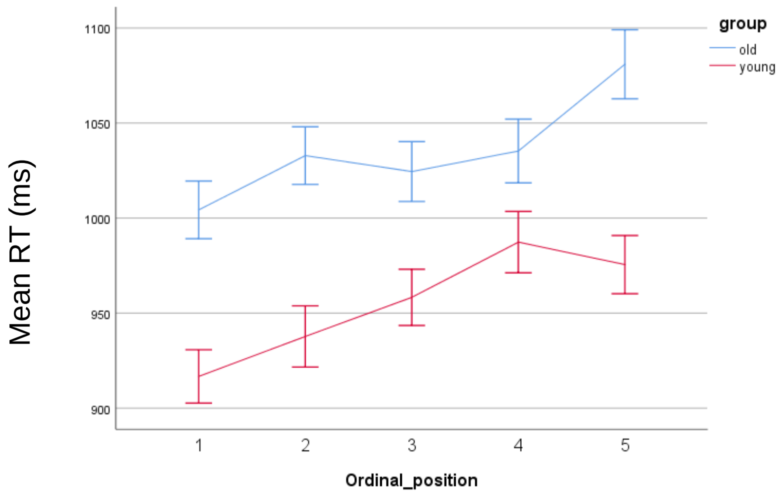


Figure 1: Mean picture-naming latencies in the continuous naming task with noun-noun compound targets named by young vs. older healthy speakers (n=32, each; run 1). Ordinal position refers to the sequence of the five category members.

Note: Cumulative semantic interference is significant in both age groups (nested linear mixed model: older speakers: ordinal position-effect (linear): $t = 2.205$, $p = 0.028$; young speakers: ordinal position-effect (linear): $t = 3.725$, $p < 0.001$)

Spatial Order and Cognition Difficulties: An Eye Tracking Study of Comic Panel Layouts

Koyel Mukherjee[1], Bidisha Som[2], Abhishek Shrivastava[3]

1. Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati
2. Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati
3. Department of Design, Indian Institute of Technology Guwahati

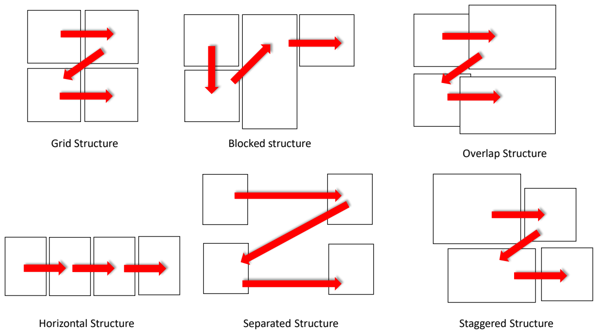
Corresponding email: koyel.mukherjee@iitg.ac.in

The Scene Perception and Event Comprehension Theory (Loschky et al., 2020) posits that visual fixations across a scene adapts according to back-end processing that connects information into a continuous event. In case of comics, the continuity is guided by its spatial framework in the form of panel layout structure. Panel layouts in comics provide aesthetic as well as cognitive functions. Prior studies on comics have suggested an External Composition Structure (Cohn, 2018), and the placement of text influences reading order especially in critical panel structures such as 'Blocked' and 'Overlap' (Kirtley et al., 2023). To explore further into how different layouts contribute to visual information processing and their perceived complexities, an experiment with 9 comic panels (Sources: Peanuts, Calvin & Hobbes) was devised. The comic strips (3 worded, 3 wordless, and 3 with only onomatopic and emotive expressions) were re-arranged into 6 panel types: 'Grid', 'Blocked', 'Overlap', 'Horizontal', 'Separated' and 'Staggered'; giving a stimuli size of 54. This experiment considers quantitative data such as fixation, saccadic movements, pupil sizes, and Response Times (RT) as well as qualitative data that includes reader's perceived levels of 'Understandability', 'Enjoyability', and 'Navigational difficulties. The data was collected using E-Prime 3.0, Tobii Pro Fusion (250 Hz), and E-Prime Extension for Tobii Pro 3.2.

100 respondents (77 male, 23 female) between age group of 18-35 (M=22, SD=2.4) participated in the experiment. They were presented a randomized stimuli set (one comic strip at a time) on the screen, which would advance on button pressing. It was followed by 3 questions (Based on 'Understandability', 'Enjoyability', and 'Navigational difficulties') on the screen that required rating on a Likert scale. The results are congruent with the current literature on Visual Language Processing where critical panel structure (where the navigation of readers is not intuitive and flouts from the Z-path) such as 'blocked' showed the highest response time (M=5389 ms) and eye fixation duration (M= 5376 ms), followed by 'staggered' with RT (M=4989.7 ms) and eye fixation duration (M=4913 ms), and Overlap with RT (M= 4706.9 ms) and fixation duration (M=4692). A Mann Whitney U test reveals 'Blocked' layouts having statistically significant difference from all the other layouts (Refer to Fig. 5 and 6). Moreover, highest saccade regression (reverting to previous fixation) counts have also been recorded on 'Blocked' (M=13,523) and 'Overlap' (M=13,196). These data suggest higher cognitive load for both these layouts. However, the qualitative data shows that 'Overlap' has been considered as 'easily understood' (M=1.86 in a 1-7 Likert scale, 1= easy), and with less navigation difficulty (M= 5.08 in a 1-7 Likert scale, 1= difficult). Furthermore, pupillometry data suggests that both 'Blocked' and 'Overlap' have smaller pupil sizes (M= 3.4 mm), while 'Horizontal' and 'Separated' have larger pupil sizes (M= 3.5 mm). Mann-Whitney U test shows a statistically significant difference between 'Overlap' and 'Horizontal' (p=<0.001) and 'Separated' (p=<0.001).

Since panel layouts are indicative of implicit navigational schema in comics, the 'Overlap' layout becomes interesting as it is similar to 'Blocked' in its demand for Visio-Spatial reasoning, while maintaining its efficiency in delivering semantic information. This difference in qualitative (subjective understanding) and quantitative reports (RT, fixations) offers a dissociation between conscious understanding and automated eye-movements.

Figures:



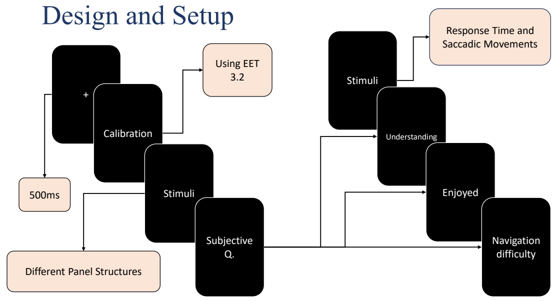
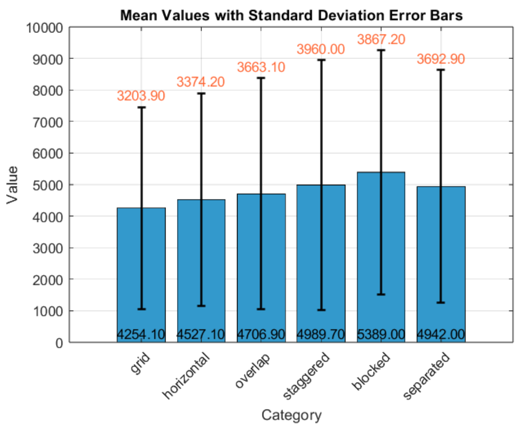


Figure 1 Stimuli: Layout Structures

Figure 2: Design of the Experiment



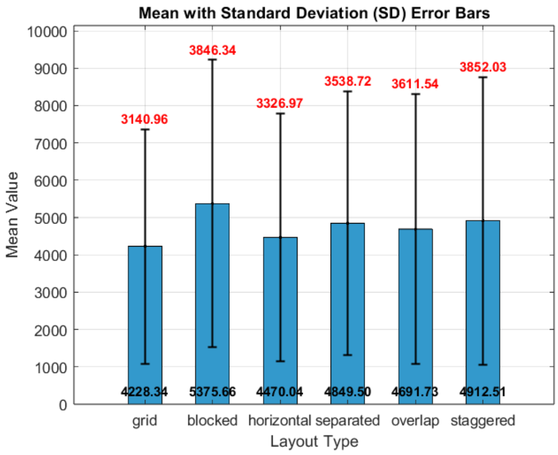
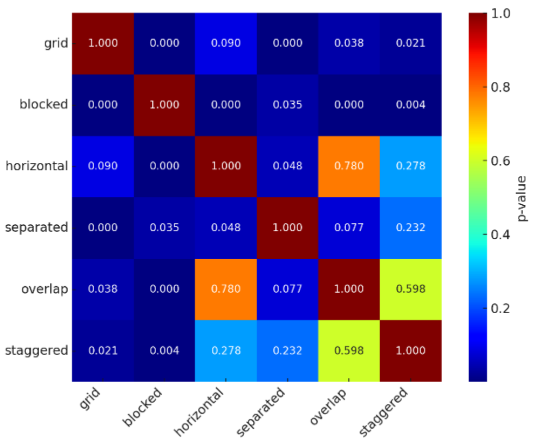


Figure 3. Mean and SD of Response Times

Figure 4: Mean and SD of Fixation Duration



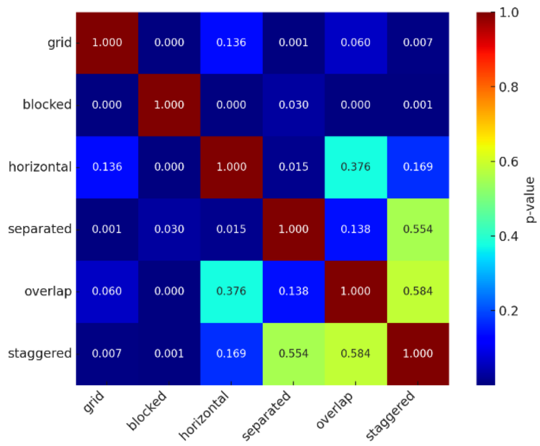


Figure 5. Mann-Whitney results-RT

Figure 6. Mann-Whitney results fixation duration

References:

Cohn, N. (2018). *Visual Language Theory and the scientific study of comics*.

Kirtley, C., Murray, C., Vaughan, P. B., & Tatler, B. W. (2023). Navigating the narrative: An eye-tracking study of readers' strategies when Reading comic page layouts. *Applied Cognitive Psychology*, *37*(1), 52–70. https://doi.org/10.1002/acp.4018

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The Scene Perception & Event Comprehension Theory (SPECT) Applied to Visual Narratives. *Topics in Cognitive Science*, *12*(1), 311–351. https://doi.org/10.1111/tops.12455

Online Generalization of Speaker-Specific Lexical Preferences

Nitzan Trainin[1,2], Einat Shetreet[1,2] & Aya Meltzer - Asscher[1,2]

[1] Sagol School of Neuroscience, Tel Aviv University; [2] Department of Linguistics, Tel Aviv University

nitzant@mail.tau.ac.il

Background: It has long been observed that listeners adapt to their interlocutor's language use, when salient social/perceptual cues are present [e.g., 1-4]. In addition, when interacting with multiple people, listeners expect references to be consistent within-speaker [e.g., 5-6]. Recently [7], it was observed that when a speaker repeatedly used a small set of repeated disfavored words, listeners expected him/her to generally use disfavored words. Thus, at least when a strong signal is presented, individuals not only expect lexical consistency, but a more general, stylistic one. In this study, we investigate whether such generalization happens based on a weaker signal (without same-word repetitions; Experiment 1), and whether expectations are adapted online in a speaker-specific manner (Experiment 2).

Methods: Experiment 1. In the *Exposure phase*, on each trial, participants (N = 32) were presented with an image, followed by a recording of a word that matched the image. Then, participants were instructed to type the word they heard in a designated textbox. The words were either in a feminine or a masculine voice and were presented with an avatar of a woman or a man (respectively). One of the speakers (counterbalanced for gender) consistently used disfavored words (the uncommon speaker), and the other one consistently used favored words (the common speaker). For example, for a picture of a car, the uncommon speaker would use 'automobile' and the common speaker would use 'car'. Each speaker produced 16 different words. Crucially, each word repeated only once. In the *Test phase*, participants were presented with an image and a written word, along with an image of one of the avatars, and were asked if they expected the presented speaker to use this word for the image (by responding yes/no). Crucially, all the images and words in the test phase were previously unmentioned, such that positive responses are the result of generalization, and not simple memory.
Experiment 2 (N = 32) followed *Experiment 1*, with the addition of measuring the time it took to type the first (appropriate) letter in the textbox, to evaluate online adaptation.

Results and Discussion: Experiment 1. We fitted a mixed-effects logistic regression predicting the log-odds of a positive response to the coupling of a disfavored word with each speaker (common/uncommon). This model revealed that participants successfully generalized speaker-specific naming preferences to previously unmentioned words. That is, they were far more likely to expect a disfavored word from the uncommon speaker than from the common one (b = 8.085, p <.001; Fig. 1). Experiment 2 replicated this finding (b = 3.079, p < 0.001; Fig. 1). In addition, to assess online adaptation, we fitted a mixed-effects linear regression model to predict (log) Reaction Times (log RT) by trial number and speaker status. This model revealed that trial number (b = -0.007, p < 0.001) and speaker status (b = -0.061, p < 0.001) both reliably predicted log RT, such that log RT decreased with each trial and were lower for the common speaker than for the uncommon speaker. No trial number X speaker status interaction was observed. To conclude, listeners generalize speaker-specific naming preferences based on a small set of non-repeating items. The shortening of RTs further suggests that participants adapt to their interlocutors. Still, the absence of an interaction between trial number and speaker status could also mean that this adaptation comes from general improvement in the task and not from speaker-specific adaptation. It therefore remains to be (and will soon be) investigated using an online measure (ERP) whether listeners adapt their expectations in a speaker-specific manner.

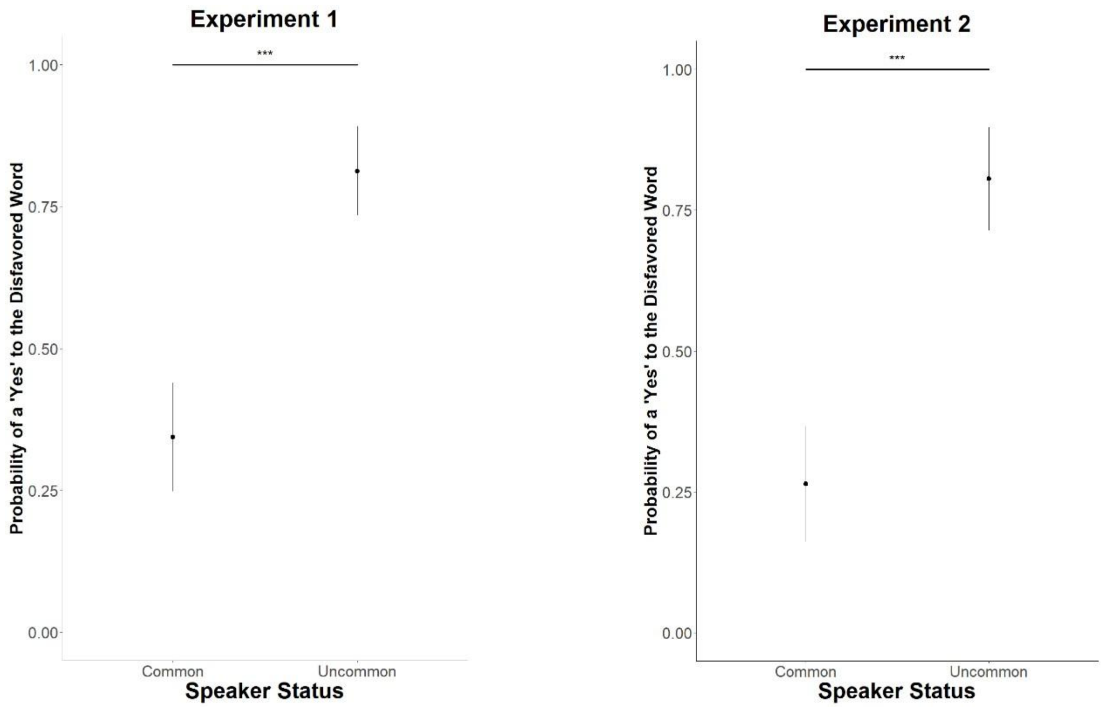


Figure 1. Probability of responding 'yes' to a pairing between a speaker and a disfavored word, by speaker status, in Experiment 1 (left) and in Experiment 2 (right).

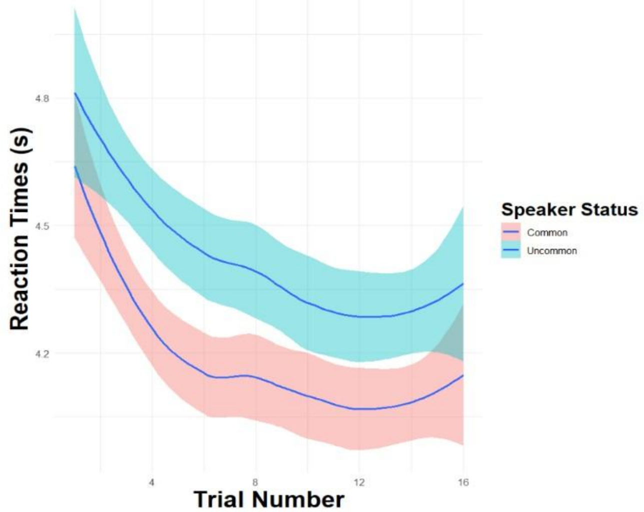


Figure 2. The time it took participants to type in the first letter in Experiment 2, by trial number and speaker status.

References

[1] Trude, A. M., & Brown-Schmidt, S. (2012). *Language and Cognitive Processes* **[2]** Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). *Cognition*. **[3]** Kroczek, L., & Gunter, T. (2017). *Scientific Reports.* **[4]** Schuster, S., & Degen, J. (2019). Speaker specific adaptation to variable use of uncertainty expressions. *CogSci* **[5]** Metzing, C., & Brennan, S. E. (2003). *Journal of Memory and Language.* **[6]** Shintel, H., & Keysar, B. (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition.* **[7]** The authors (2024). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Tracing the development of German number cues: A case study from the LEO corpus

Dinah Baer-Henney[1], Alexander Clemen[2]
[1]Ruhr University Bochum, [2]Heinrich Heine University Düsseldorf
dinah.baerhenney@googlemail.com

Background: In German, no single plural cue is predictable and thus there is a debate about how grammatical number is acquired in the nominal paradigm. A rule-based learning mechanism seems unlikely. Among the alternative, i.e., usage-based accounts, there is discussed which underlying mechanism links a form with its number meaning. The present case study examines the development of grammatical number cues using the most comprehensive spontaneous speech data corpus of a German-speaking preschool child available [1,2]. We researched the development of selected suffixal cues and umlaut in the input as well as the child's productions. We analyzed the input over the time course, investigating the predictions derived by two possible mechanisms underlying the development of the German plural system – a frequency-based [3] and the naive discriminative learning [4] account. Accounts differ with respect to the underlying association mechanism and we determined the importance of the selected number cues in the input for the child. In addition, we traced the plural marker production errors of the child. Our aim is to trace the child's production errors per number cue along the expected developmental path based on the input based on [3,4].

Method: We present a case study of the LEO corpus from the CHILDES database [1,2]. The longitudinal corpus with recordings of only one child comprises the largest collection of spontaneous speech in German with a child aged three to five. We filtered all nouns uttered towards Leo analyzed the cue availability under a frequency-driven [3] and discriminative-strength driven [4] usage-based account to generate theory-dependent hypotheses on the expected learning trajectory for plural cues for this particular child. We thus investigated the expected learning trajectory of the learning child based on the varying accessibility of number cues in the input during the ages of three to five years. We then identified suffix and umlaut errors Leo produced over time. We aimed at analyzing how Leo oriented himself given the input he received.

Results: We estimated the mechanism-specific predictions for the plural markers /-e/, /-er/, /-s/, /-en/ and umlaut. A frequency-driven account assumes that all markings except /-en/ are constantly relatively bad plural markers. Discriminative strength predicts /-e/, /-er/ and /-s/ to be a rather bad plural cue, while it becomes better with time only for /-s/. Both /-en/ and umlaut tend to be good plural cues but this weakens over time. Turning to the output of Leo we observe few errors in general. As expected, there is a decreasing error rate over time with a notable previous increase at the age of 4 years for suffixes containing a schwa. Comparing the error rates across cue types we observe in line with [3,4] that /-er/ cue has the highest error rates. We find relatively few umlaut errors as expected by [4].

Discussion: Overall, the analyses of the input show that – based on the same data – learning models predict different developmental trajectories of different cues depending on distinct underlying mechanisms. The cue type with the highest error rate /-er/ is rated as a bad cue under both theoretical assumptions [3,4] and is therefore not useful to disentangle between theories. Relatively few umlaut errors could be attributed to the fact that initially this was a rather strong plural cue in Leo's input according to its discriminative strength. Higher error rates at the age of four as compared to the age of three may indicate a developmental step from memorized learning to the application of a developing underlying mechanism.

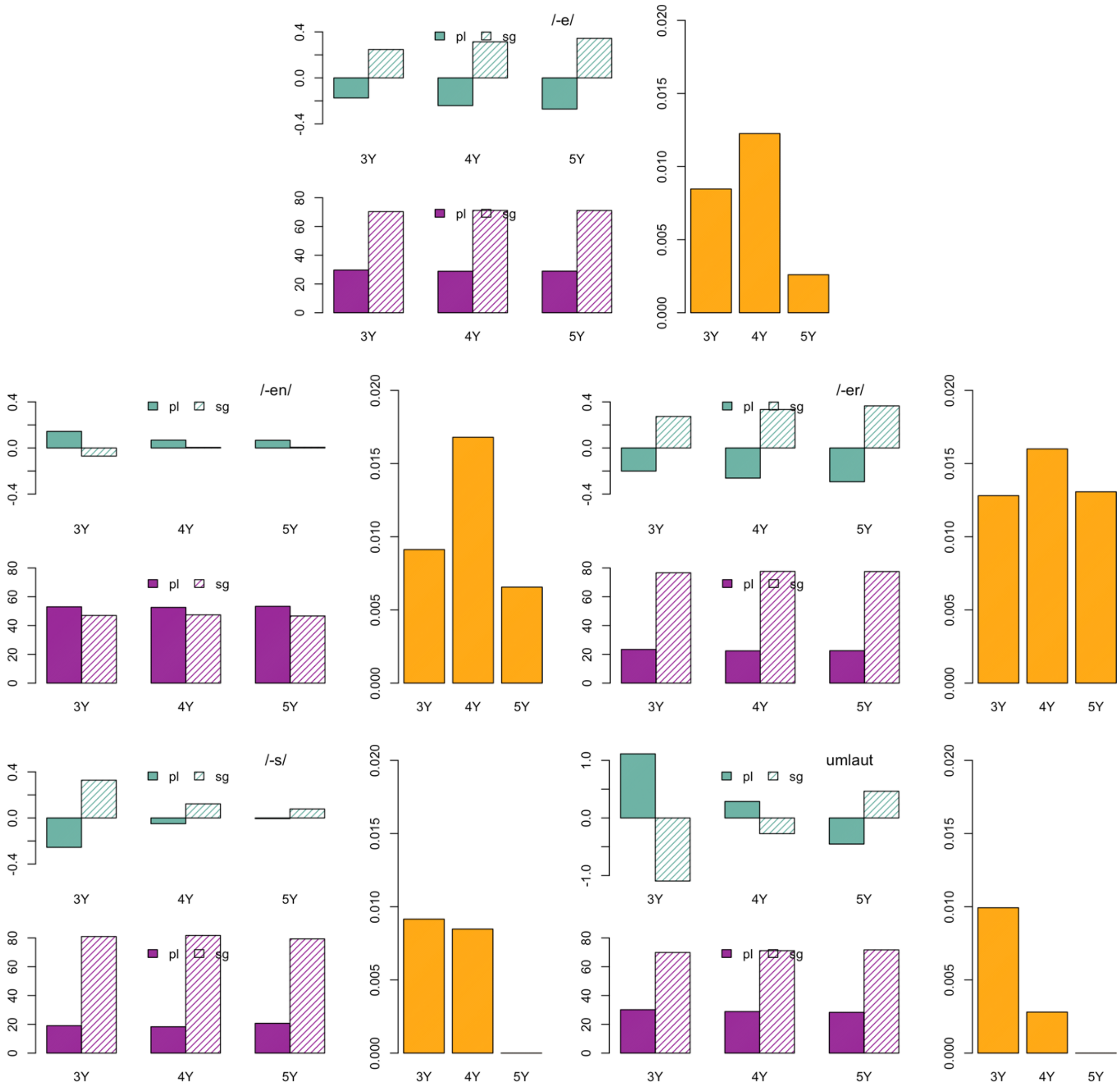


Fig. 1: For each plural marker, we show the predicted developmental trajectories resulting from LEO's input driven by discriminative strength (green) and frequency (purple). The yellow bars show the error rates in LEO's productions within the plural markers over the years.

References

[1] Behrens, H. (2006). The input–output relationship in first language acquisition. Language and Cognitive Processes, 21(1): 2–24.

[2] MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs. New York & London: Psychology Press.

[3] Köpcke, K.-M., Schimke, S. & V. Wecker (2021). Processing of German noun plurals: Evidence for first-and second-order schemata. Word Structure, 14(1): 1–24.

[4] Baayen, R. H., Milin, P., Filipović, D., Durdević, D., Hendrix P. & Marelli, M. (2011) An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. Psychological Review 118(3). 438–481.

The Role of Heritage and Societal Languages in L3 Aspect Processing: Insights from Eye-Tracking

Natalia Mitrofanova[1], Serge Minor[1], Christina Athanasiadi[1], Nadine Kolb[2], Marit Westergaard[3,1]
[1]*UiT-The Arctic University of Norway,* [2]*University of Stavanger,* [3]*Norwegian Academy of Science and Letters*
natalia.mitrofanova@uit.no

Using Visual World eye-tracking, we examine how bilingual German-Russian and Greek-German children process English aspectual forms, comparing them to age- and proficiency-matched monolinguals. The languages involved exhibit distinct aspectual systems: Russian and Greek differentiate between perfective and imperfective aspects, associating them with completed and ongoing events, respectively. German lacks grammaticalized viewpoint aspect, while English uses past progressive form for ongoing events in the past, with simple past forms allowing for both ongoing and completed event interpretations [1]. Present perfect forms of accomplishment verbs in English are strongly associated with completed event interpretations [2], while in German (given the absence of grammaticalized aspect) the semantics of the Present Perfect has been argued to be broader allowing for both perfective and imperfective readings [3]. We adapted the paradigm from [1] to test whether the aspectual systems of the participants' heritage languages (HLs) influence their processing of L3 English. Our study involved 117 children (age 8-13 y., mean age 10 y.): 52 German-Russian bilinguals and 24 German monolinguals tested in Germany, and 41 Greek-German bilinguals tested in Athens, all matched for English exposure and lexical proficiency. Results showed that German-Russian and Greek-German adolescents were sensitive to aspect in both offline judgments and online gaze preferences, unlike the monolingual German controls, who showed no sensitivity to aspect manipulation.

Interestingly, the Greek-German bilinguals in Athens exhibited a stronger and earlier aspectual effect than the German-Russian bilinguals in Germany. These findings are discussed in relation to contemporary L3 acquisition models [4, 5] emphasizing the importance of structural similarity (see [6, 7]) and language dominance [8]. The study highlights the role of grammatical representations from previously acquired languages in shaping L3 processing, particularly in the context of aspectual distinctions.

Examples of experimental stimuli:

a) Grandma **was knitting** a new jumper. b) Grandma **knitted** a new jumper. c) Grandma **has knitted** a new jumper.

Ongoing event (OE)

Completed event (CE)

Ex.1 An experimental trial included an audio preamble which located the narrative in the past (e.g. *It was a rainy day*), followed by a sentence-picture matching task where the participants were presented with a pair of pictures on a screen: one representing an Ongoing Event (OE), i.e. an action in progress, and one representing a Completed Event (CE), see examples (a-b) above from the English task. Each experiment included 24 fillers and 24 test trials (12 Past Progressive and 12 Simple Past) involving 48 verbs/event types and visual stimuli. Eye-movements were recorded.

Ex. 2 involved two conditions: Present Perfect (see c above) vs sentences involving phasal verbs (*started, finished, continued* doing something). No preambles were included.

Figures:

Exp.1 Picture selection task:
Proportion of trials with Ongoing Event choice

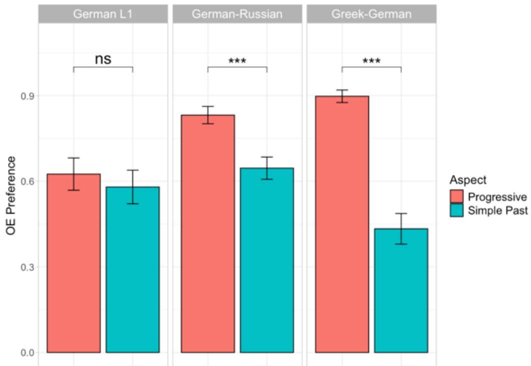


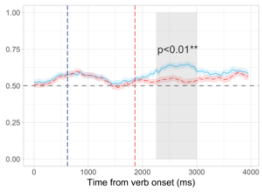
Fig. 1. Offline results

Significant effect of aspect in the bilingual groups
No effect of aspect in the L1 German group
Significant interaction between Group and Aspect (Greek-Ger > Ger-Rus > Ger L1)

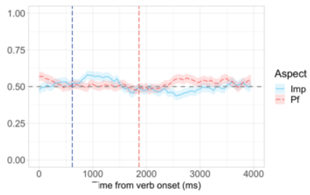
*Mixed effects logistic regression. Group, Aspect, English Proficiency and their interactions as predictors, random intercepts and random slopes for Aspect by Participant and by Item.

Looks to Ongoing Event picture by Aspect

German-Russian



German



Greek-German

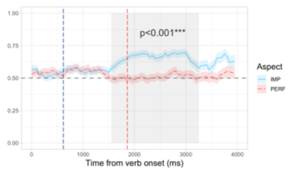


Fig. 2. Online results

Significant effect of aspect in bilingual groups
No effect of aspect in the L1 German group
Significant difference between the groups

*Cluster based permutation analysis + Bayesian mixed effects zero one inflated beta regression (Group, Aspect and their interaction as predictors, English proficiency included as covariate, random slopes for Group*Aspect by Participants and Items)

References:

[1] Minor et al. (2023). Aspect processing across languages: A visual world eye-tracking study. Front.in Lang.Sci. [2] Vos et al. (2025). The Simple Past vs. Perfect in English: Evidence from Visual World Eye Tracking. [3] Zwart (2008). Almost perfect: some notes on the present perfect puzzle. Germania Latina VI. [4] Westergaard et al. (2023). Full Transfer Potential in L3/Ln acquisition: Crosslinguistic influence as a property-by-property process. CUP. [5] Rothman et al. (2019) *Third Language Acquisition and Linguistic Transfer.* CUP. [6] Jensen et al. (2023). Crosslinguistic influence in L3 acquisition across linguistic modules, *IJM*. [7] Kolb et al. (2022). Crosslinguistic influence in child L3 English: An empirical study on Russian-German heritage bilinguals. *IJB*. [8] Athanasiadi (2024) An investigation into L3 acquisition and processing of Grammatical Aspect in English by bilingual Greek-German children. MA thesis, EMCL++.

Sight translation of non-canonical structures: Eye movement patterns and individual differences

In real-time language tasks such as sight translation (STR), translators must rapidly decode written input and produce a spoken equivalent. This dual processing makes STR an ideal paradigm for studying how linguistic structure and referential properties, such as syntax and animacy, interact with cognitive load. Prior research shows that non-canonical structures, particularly object-relative clauses, are more cognitively demanding due to their departure from the preferred subject–verb–object order and because they often reverse animacy hierarchies by placing animate entities in object position. These configurations may interfere with syntactic expectations and processing routines, increasing cognitive load [1–3]. This study examines how such structures affect real-time processing in STR and whether individual differences in working memory and cognitive control modulate these effects.

Twenty-one Spanish-speaking participants (B1+ English, LexTALE-verified) translated 28 English sentences while their eye movements and speech production were recorded in a 2×2 within-subjects design, with the following experimental conditions: relative clause type (subject vs. object) and animacy (animate vs. inanimate). Eye movements were captured with the EyeLink system, allowing for the simultaneous recording of gaze and speech production. Key metrics included total trial fixation duration, mean fixation duration, number of fixations, and eye-voice span (EVS). Behavioural measures included Reading Span Test, Digit Span Test, and Flanker Task. No significant effects were observed across these measures, suggesting that the observed effects are robust across participants.

Statistical analyses were performed using RStudio. Mixed-effects regression models (generalised linear mixed models, GLMM) were applied with random intercepts for participants and items. Predictors were z-standardised. A Gamma distribution with log link was used for continuous time-based measures (fixation durations, EVS), and a Poisson distribution with log link was used for count data (number of fixations). Thus, effects are expressed through estimates (b or IRR), z-values, and p-values. Results showed that animate objects in object-relative clauses triggered significantly higher cognitive load compared to canonical animate-subject structures. The animate object condition showed significant differences in total trial fixation duration, ($b = 0.31$, $SE = 0.08$, $z = 3.88$, $p < .001$); eye-voice span (EVS), ($b = 0.43$, $SE = 0.09$, $z = 4.71$, $p < .001$); and number of fixations, where the incidence rate ratio (IRR) was 1.43 ($SE = 0.06$, $z = 6.24$, $p < .001$). While overall fixation duration and EVS increased in animate ORCs, mean fixation duration decreased, indicating a compensatory pattern of faster but more numerous fixations. This fixation behaviour may reflect an adaptive mechanism for local reanalysis, previously observed in tasks involving syntactic ambiguity, disfluency repair, or self-monitoring during reading and interpreting [4]. When animacy expectations are violated—i.e., when an animate noun appears in a structurally unexpected object position—the parser may engage in faster, repeated fixations to resolve conflict between word order and thematic role assignment.

In sum, non-canonical syntax and animacy reversal jointly increase processing demands in STR, yet participants appear to deploy visual strategies to maintain fluency under pressure. As sight translation engages additional syntactic and semantic processes, the observed effects cannot be attributed solely to comprehension, but rather to the combined demands of bilingual production and processing. These findings seek to inform both models of incremental language production and translator training, highlighting the role of structural expectations and adaptive control in bilingual processing.

References

1. Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*(1), 50–68.
2. Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55*, 37–76.
3. Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*(2), 204–224.
4. Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production. *Journal of Memory and Language, 42*(2), 168–182.

Table 1. Stimuli examples per condition

Condition	Sentence example
Animate subject	The musician that witnessed the accident angered the policeman a lot.
Animate object	The musician that the accident terrified angered the policeman a lot.
Inanimate subject	The accident that terrified the musician angered the policeman a lot.
Inanimate object	The accident that the musician witnessed angered the policeman a lot.

Table 2. Model outputs for predictors across four eye-tracking measures

	Total trial fixation				**Fixation duration**				**Number of fixations**				**Eye-voice span**			
Predictors	*Estimates*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*	*Incidence Rate Ratios*	*std. Error*	*CI*	*p*	*Estimates*	*std. Error*	*CI*	*p*
Animate object	1.37	0.10	1.18 – 1.59	**<0.001**	1.00	0.02	0.97 – 1.04	0.981	1.43	0.08	1.28 – 1.61	**<0.001**	1.54	0.15	1.27 – 1.86	**<0.001**
Inanimate object	1.02	0.08	0.88 – 1.19	0.755	1.02	0.02	0.99 – 1.06	0.251	1.03	0.06	0.92 – 1.15	0.594	0.92	0.09	0.76 – 1.11	0.372
Inanimate subject	1.08	0.08	0.93 – 1.25	0.324	1.03	0.02	0.99 – 1.06	0.142	1.07	0.06	0.96 – 1.20	0.240	0.99	0.10	0.82 – 1.19	0.903
Random Effects																
σ^2	0.08				0.01				0.02				0.20			
τ_{00}	0.03_{item}				0.00_{item}				0.04_{item}				0.05_{item}			
	0.03_{id}				0.00_{id}				0.16_{id}				0.12_{id}			
ICC	0.41				0.40				0.92				0.46			
N	21_{id}				21_{id}				21_{id}				19_{id}			
	112_{item}				112_{item}				112_{item}				112_{item}			
Observations	585				585				585				447			
Marginal R^2 / Conditional R^2	0.102 / 0.475				0.009 / 0.405				0.086 / 0.929				0.102 / 0.515			

(Dis)agreement across languages: Cues to control in English and Romanian

Juliana Gerard[a] and Camelia Bleotu[b,c]

[a]Ulster University

[b]University of Vienna

[c]University of Bucharest

j.gerard@ulster.ac.uk

Background: Subject control in temporal adjuncts is observed in sentences like (1), in which the adjunct subject, notated as PRO, is interpreted as the main clause subject "John." Subject control in non-finite adjuncts is reported across languages, and argued to be a universal property [1-5]. However, Romanian has been argued not to be a control language [6-12], with a *pro* element in non-finite clauses, just as in finite clauses (e.g., unlike PRO, null subjects alternate with lexical subjects in non-finite adjuncts (2)). We test the predictions of a *pro* element in Romanian, compared with PRO in English with an acceptability judgement task based on gender agreement between the antecedent and elements in the non-finite adjunct. **While English speakers only accept matching agreement with the subject, Romanian speakers accept agreement with the subject or the object**. We consider the implications for acquisition theory and relevant linguistic input.

Method: We tested native Romanian speaking adults (n=48) and native English adults (n=18) with an acceptability judgment task, using a slider scale (Figure 1). For English, adjunct clauses included a reflexive action and the reflexive *himself* or *herself* which matched or matched with the adjunct subject or object; Romanian had the same manipulation except that adjunct clauses included an adjective with gender agreement (Table 1, see (3) and (4)). Subject and object gender always mismatched, and subject gender was counterbalanced across items. Fillers were also included (5), with a reflexive (English) or adjective (Romanian) matching (50%) or mismatching (50%) the referent in gender,

Results: We conducted a linear regression analysis with z-scored ratings as dependent measure, fixed effects of subject-match/object and language, and random slopes and intercepts for subject and item. Both main effects were significant (p<.001); importantly, there was a significant interaction between subject-match/object-match and language (t=10.7, p<.001) because English speakers rated object-match sentences as ungrammatical, while Romanian speakers did not (Figure 2) - as predicted with a *pro* element. Notably, these higher ratings cannot be due to inattention to gender marking, as Romanian speakers gave low ratings to ungrammatical fillers with a gender-mismatch (Figure 3).

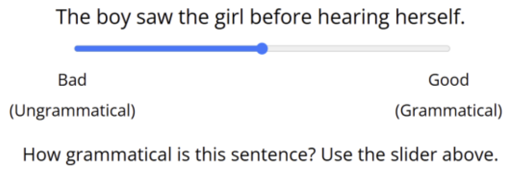
Discussion: While English-speakers' ratings indicated **strict subject control**, Romanian speakers' ratings were indicative of a **discourse-based element** that can take any antecedent in line with the predictions of *pro.* Thus, the results highlight a key difference between English and Romanian in non-finite adjuncts, with English showing strict subject control and Romanian allowing for a more flexible interpretation with *pro*. These findings challenge the universality of subject control. Additionally, the variation has significant implications for acquisition and for the evidence needed for control: if Romanian children's interpretations indicate a subject control grammar, then positive evidence is needed for object control; otherwise, if Romanian children pattern like Romanian adults, this suggests that evidence in the input for control may be more specific to a given language. We are currently testing these predictions with Romanian children.

(1) John$_1$ phoned Bill$_1$ before PRO$_{1/*2/*3}$ running outside.

(2) a. Ion a mâncat înainte de a pleca **pro**.
John has eaten before DE to leave **pro**
'John ate before *pro* leaving.'

b. Ion a mâncat înainte de a pleca **mama**.
John has eaten before DE to leave **mother**
'John ate before mother left.'

Figure 1. Acceptability slider



(3) English: The boy washed the girl before hearing {**himself/herself**}

(4) Romanian: Băiatul a spălat-o pe fată înainte de a deveni {**trist** / **tristă**}
The boy washed the girl before DE to become **sad-M / sad-F**

(5) Filler: *The boy tripped in the hallway and embarrassed herself**.**

Figure 2. Test sentences

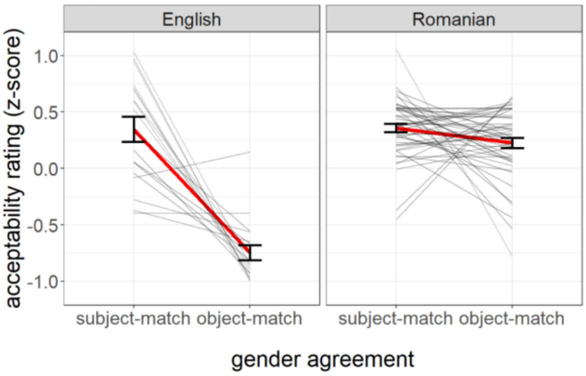


Figure 3. Filler sentences

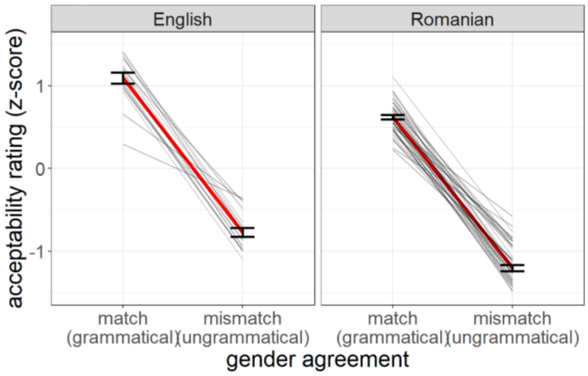


Table 1. Experiment design

Languages	subject-match	object-match
English: The boy washed the girl before…	…hearing **himself**	…hearing **herself**
Romanian: Băiatul a spălat-o pe fată înainte de… The boy washed the girl before DE…	…a deveni **trist** …to become **sad-M**	…a deveni **tristă** …to become **sad-F**

References

[1] Chomsky, N. (1981) *Lectures on Government and Binding.* Foris Publications.

[2] Landau, I. A (2021) Selectional Theory of Adjunct Control. MIT Press.

[3] Ariel, M. (1990) *Accessing noun-phrase antecedents*. Lond. N. Y. Routledge.

[4] Kehler, A. & Rohde, H. (2013) A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39, 1–37 (2013).

[5] Orita, N., Vornov, E. & Feldman, N. H. (2021) Informativity, topicality, and speech cost: comparing models of speakers' choices of referring expressions. *Discourse Process.* 58, 743–765.

[6] Cornilescu, A. (2000) The double subject construction in Romanian. *Comp. Stud. Rom. Syntax* 83–134.

[7] Dindelegan, G. P. (2013) *The Grammar of Romanian.* (OUP Oxford).

[8] Avram, L. & Coene, M. M. R. (2010) Early subjects in a null subject language. in *New Dir. in Language Acquisition: Romance Languages in the Generative Perspective* 189– 218 (Cambridge Scholars Publishing).

[9] Dobrovie-Sorin, C. (1994) *The Syntax of Romanian: Comparative Studies in Romance.* (Walter de Gruyter).

[10] Alboiu, G. (2007) Moving Forward with Romanian Backward Control and Raising. in *New Horizons in the Analysis of Control and Raising* (Springer Netherlands, Dordrecht, 2007).

[11] Cotfas, M. A. (2016) *On the Syntax of the Romanian Subjunctive: Control & Obviation.* (Univ din Bucureşti).

[12] Cotfas, M. A. (2016) Subjunctive complements in Romanian: The view from control. *Gener. Gramm. Geneva* 9, 77–98.

Word-order or Truth-value? Dominant cues during Korean incremental processing with the picture-sentence verification task

Yunju Nam[1], Sun-Young Lee[2], Juno-Baik[3], Hyeonjeong Jeong[4], Youngjoo Kim[5]
[1] Hanyang University; [2] Cyber Hankuk University of Foreign Studies; [3] Brigham Young University; [4] Tohoku University; [5]Kyunghee University
yjnam05@hanyang.ac.kr

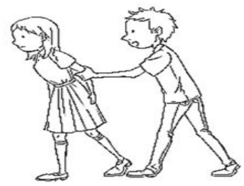
Background: During real-time language comprehension, sentence processing is guided by both structural expectations and contextual plausibility. One well-established cue is the subject-first strategy, whereby parsers tend to construct interpretations that adhere to canonical word order—subject-object-verb (SOV) in Korean. However, this strategy likely reflects the convergence of multiple linguistic and non-linguistic factors, including first-mention bias, animacy hierarchies, and case-marking patterns, rather than purely structural preferences. In Korean, the prominent case-marking system (-ka for nominative, -lul for accusative) provides crucial information for syntactic role assignment, which may interact with word-order expectations. While structural biases can interact with semantic cues such as truth-value in visual contexts, it remains unclear which cue plays a more dominant role during incremental sentence processing in Korean.

Method: We investigated how early structural expectations influence comprehension dynamics with conflicting visual information using self-paced reading (SPR, N=30) and event-related potentials (ERP, N=32) in separate participant groups. Sentences varied by Word Order (SO vs. OS) and Truth Value (True vs. False), forming four conditions. Each sentence included three regions (NP1-ka/lul, NP2-lul/ka, Verb). Linear mixed-effects models examined reaction times and accuracy; repeated-measures ANOVAs analyzed N400 amplitudes at the Verb region. We acknowledge that the absence of pre-stimulus and spillover regions may conflate verb processing with sentence wrap-up effects. However, previous studies using the same structure have confirmed that this is not a problem [1].

Results: Behavioral data revealed significant main effects of Word Order and Truth Value across all regions, with OS word order and False truth-value independently increasing reaction times. A significant interaction emerged only at NP1, indicating cumulative difficulty when OS word order and False truth-value co-occurred. High accuracy rates (95.8-99%) may have created ceiling effects that masked potential interactions in later regions. ERP results showed a significant N400 effect at the Verb region driven by Word Order, with OS structures eliciting larger negativities than SO structures, while truth-value effects were not significant in neural responses.

Discussion: The findings highlight the dominant role of word-order expectations over semantic plausibility during Korean incremental processing. The cumulative difficulty at NP1 and clear N400 effects at the Verb suggest that word-order violations are detected more immediately than truth-value inconsistencies, indicating that early structural expectations guide parsing under ambiguity. However, high behavioral accuracy and absence of additional experimental regions limit our ability to fully characterize these effects' time-course. The data suggest that structural expectations and semantic role interpretation jointly contribute to incremental parsing, though it remains inconclusive whether Korean comprehenders systematically prioritize subject-first or agent-first strategies. Future research should examine whether truth-value cues can override word-order biases in pragmatically natural OS contexts and investigate how case-marking reliability modulates these processing strategies.

[Table 1] Experimental conditions and materials



Picture	Conditions	Examples	
	SO-T	남자-가 여자-를 잡-아요. *Namca-**ka*** *Yeoca-**lul*** *cap-ayo*.	"A man catches a woman."
	OS-T	여자-를 남자-가 잡-아요. *Yeoca-**lul*** *Namca-**ka*** *cap-ayo*.	
	SO-F	여자-가 남자-를 잡-아요. *Yeoca- **ka*** *Namca-**lul*** *cap-ayo*.	"A woman catches a man"
	OS-F	남자-를 여자-가 잡-아요. *Namca-**lul*** *Yeoca-**ka*** *cap-ayo*.	

[Table 2] Descriptive Statistics

Conditions	Self-paced reading				Behavioral results during ERP Exp.	
	Reading Time(Mean ms, (SD))			Accuracy	Delayed Response Time (Mean ms, (SD))	Accuracy
	NP1	NP2	Verb			
SO-T	658.80 (75.64)	561.21 (84.40)	874.67 (139.18)	98.3%	465.39 (281.14)	99%
OS-T	773.40 (126.74)	692.52 (99.96)	1012.47 (183.27)	95.8%	526.51 (348.66)	96%
SO-F	743.01 (102.70)	665.39 (94.04)	1012.71 (263.84)	98.1%	483.76 (291.94)	98%
OS-F	746.44 (89.44)	799.32 (128.00)	1255.13 (448.26)	96.1%	527.16 (335.14)	97%

[Table 3] Statistical Analysis: LMER for reading time in SPR

		Estimate	S.E	t value	Pr(>|t|)	
NP1	(Intercept)	658.8	42.72	15.422	0.000	***
	Word_orderOS	114.6	24.07	4.761	0.000	***
	Truth_valueFALSE	84.21	24.07	3.498	0.000	***
	Word_orderOS:Truth_valueFALSE	-111.17	34.04	-3.266	0.001	**
NP2		Estimate	S.E	t value	Pr(>|t|)	
	(Intercept)	561.21	44.416	12.635	0.000	***
	Word_orderOS	131.308	24.591	5.340	0.000	***
	Truth_valueFALSE	104.182	24.591	4.237	0.000	***
	Word_orderOS:Truth_valueFALSE	2.626	34.777	0.076	0.940	
Verb		Estimate	S.E	t value	Pr(>|t|)	
	(Intercept)	874.67	99.93	8.752	0.000	***
	Word_orderOS	137.81	70.27	1.961	0.049	*
	Truth_valueFALSE	138.05	70.27	1.965	0.049	*
	Word_orderOS:Truth_valueFALSE	104.61	99.37	1.053	0.293	

Figure 1.

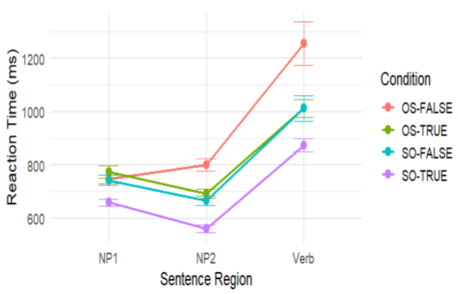
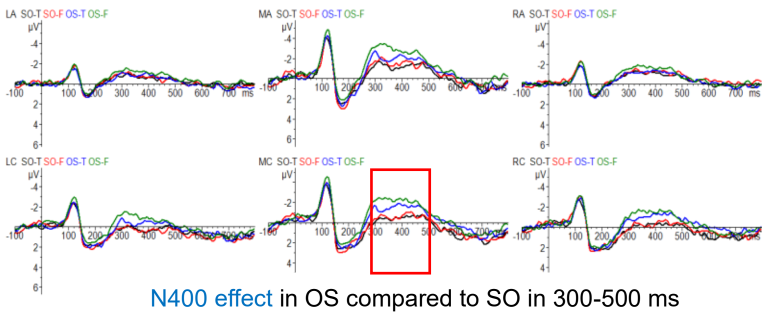


Figure 2. Brain response at verb



N400 effect in OS compared to SO in 300-500 ms
(F(1,27) = 22.864***)

References

[1] Lee SY, Nam Y. (2022) Electrophysiological evidence for a subject-first strategy in visually situated auditory sentence processing in Korean. Acta Psychologica. 231:103799.

Additional Information about Korean Language

Word Order and Flexibility: Korean is a head-final language with canonical Subject-Object-Verb (SOV) word order. However, it allows considerable word order flexibility through scrambling, particularly Object-Subject-Verb (OSV) constructions, without changing the core meaning of sentences.

Case Marking System: Korean employs a robust case marking system that explicitly marks grammatical roles:

- Nominative case: *-ka/-i* (subject marker)
- Accusative case: *-lul/-ul* (object marker)
- Other cases include dative (*-hante/-ege*), genitive (*-uy*), and locative markers

The prominence of case marking enables word order flexibility while maintaining grammatical clarity, making Korean an ideal language for investigating the interaction between structural and morphological cues in sentence processing.

Agglutinative Structure: Korean is highly agglutinative, with extensive use of suffixes and particles. Verbs carry rich morphological information including tense, aspect, mood, and honorifics, while nouns are followed by various particles indicating grammatical and semantic roles.

Pro-drop Characteristics: Korean allows extensive ellipsis of arguments when contextually recoverable. Both subjects and objects can be dropped, making case marking and verb agreement crucial for successful sentence interpretation.

Processing Implications: These linguistic features create a unique processing environment where:

1. Early case marking provides immediate grammatical role information
2. Word order violations can be detected quickly due to strong canonical expectations
3. The interaction between morphological markers and word order creates complex parsing decisions
4. Visual context becomes particularly important given the language's elliptical nature

Relevance to Current Study: The combination of flexible word order, explicit case marking, and strong canonical SOV preference makes Korean particularly suitable for examining the relative contributions of structural expectations versus semantic plausibility in real-time sentence processing.

A cute horgous meets a scary timfil: how do we interpret novel words in context?

Ngoc-Anh Tran[1], Kazimierz Garstecki[2], Giovanni Cassani[2]
[1] Dep. of Methodology and Statistics [2] Dep. of Cognitive Science & Artificial Intelligence (Tilburg University)
g.cassani@tilburguniversity.edu

Background. Studies on form-meaning mappings in the lexicon have highlighted a rich set of systematic relations between the surface form of words and what they mean [1], with some mappings applying cross-linguistically [2, 3, 4]. More recent studies further highlighted how semantic effects such as prime-target similarity and semantic neighborhood density influence lexical decision latencies also for pseudowords [5, 6], blurring the boundary between words and pseudowords. From a language learning perspective such a blurred boundary makes sense: every word a person knows has been a pseudoword during development and many valid words are pseudowords to many speakers, which might eventually encounter them and have to quickly shape a semantic representation. Our work examines the interplay between semantic connotations conveyed by the word form itself [7] and by the sentence context in which the word form is first introduced [8, 9]. We focus on valence, i.e., how positive or negative a stimulus is perceived to be, building on previous work showing people have consistent intuitions about the valence of isolated pseudowords [10]. Ours is a work in progress: this abstract details the methodology and analysis plan we will follow.

Materials and Methods. Each trial of our behavioral task consists of a free association (FA) task and a self-paced reading (SPR) task (order counterbalanced across subjects). The key manipulation involves the valence of the target (pseudo)word, as gauged from word [11] and pseudoword [10] valence ratings, and the valence of the sentence in which they will appear (see Figure 1). We sampled 40 pseudowords from [10] and 40 words from [11] (13 negative, 14 neutral, and 13 positive for both), and created 40 sentences (13 negative, 14 neutral, 13 positive), estimating sentence valence using word valence ratings from [11]. Then, we embedded each target (pseudo)word in 4 sentences from each valence category. In order to control for Part-of-speech effects, all target (pseudo)words are used as nouns. We plan to collect 10 responses per trial.

Hypotheses. The key hypothesis concerns the role of pseudoword valence and its possible interaction with sentence valence. We hypothesize that the connotations evoked by a pseudoword will be strong enough to resist, at least partially, the influence of a sentence context of opposite valence: **horgous** should elicit more negative associates than **timfil**, even when described as *cute*, albeit possibly less negative than when presented in isolation or described as *scary*. In the same vein, we expect increased reading times when pseudowords are embedded in valence-incongruent sentence context.

Analysis plan. We will use thesauri and distributional semantic models to estimate the semantic coherence of associates produced for a (pseudo)word when the FA task was administered first. Then, we will quantify the coherence between associates produced for a (pseudo)word by participants who completed the FA task before and participants who completed it after the SPR task. Moreover, we will analyse reading times on the word immediately following the target (pseudo)word. Associates' coherence and RTs will be analysed using mixed models to capture by-item and by-subject variability. The key predictors will be (pseudo)word valence and sentence valence. We will further control for (pseudo)word length, plausibility, orthographic neighborhood density, and orthographic overlap with the likeliest word(s) to occur in the sentence context.

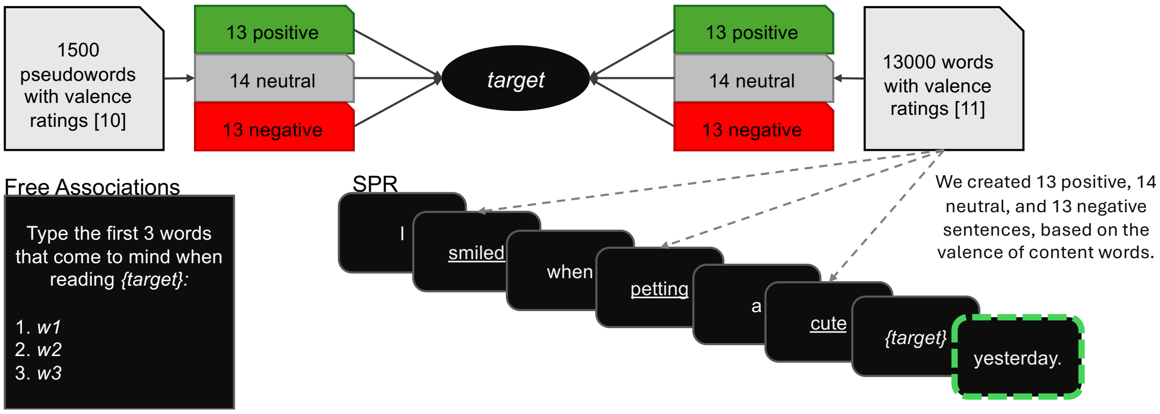


Figure 1: Schematic representation of the experimental design. Half the participants will perform the Free Associations task before the Self-Paced Reading (SPR) task, half after, so that we can compare associates produced for isolated targets with associates produced for targets seen in a context. We manipulate the valence of targets (words or pseudowords) and of the sentence context. We analyse the reading time at the first post-target word (dashed green contour) and the semantic coherence of associates.

References

[1] Sidhu, D. M. (2025). Sound symbolism in the lexicon: A review of iconic-systematicity. *Lang. Linguist. Compass*, *19*(1), e70006.

[2] Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., et al. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philos. Trans. R. Soc. B*, *377*(1841), 20200390.

[3] Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *PNAS*, *113*(39), 10818–10823.

[4] Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled/r/is associated with roughness, linking sound and touch across spoken languages. *Sci. Rep.*, *12*(1), 1035.

[5] Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *J. Exp. Psychol. Learn. Mem. Cogn.*, *47*(1), 157.

[6] Gatti, D., Marelli, M., & Rinaldi, L. (2023). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *J. Exp. Psychol. Gen.*, *152*(3), 851.

[7] Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans'(explicit) intuitions about the meaning of novel words. *Cogn.*, *251*, 105882.

[8] Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *J. Exp. Psychol. Learn. Mem. Cogn.*, *27*(1), 225.

[9] Savic, O., Unger, L., & Sloutsky, V. M. (2022). Exposure to co-occurrence regularities in language drives semantic integration of new words. *J. Exp. Psychol. Learn. Mem. Cogn.*, *48*(7), 1064.

[10] Gatti, D., Raveling, L., Petrenco, A., & Günther, F. (2024). Valence without meaning: Investigating form and semantic components in pseudowords valence. *Psychon. B. Rev.*, *31*(5), 2357–2369.

[11] Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav. Res. Methods*, *45*, 1191–1207.

Can two words mean exactly the same?
Insights from a distributional semantics approach.

Raya Mezeklieva[1], Peter Hendrix[2]
[1]University of Amsterdam, [2]Tilburg University
raya.mezeklieva@student.uva.nl, P.H.G.Hendrix@tilburguniversity.edu

Background: Synonymy—the notion that two words share the same meaning—is a central concept in lexical semantics [1], but what qualifies as "the same meaning" is debated. Linguistic theory distinguishes between near-synonymy, where meanings are similar but not identical, and absolute synonymy, where two words are perfectly interchangeable in all contexts without differences in meaning, style, or connotation [1]. While near-synonymy is uncontroversial, the existence of absolute synonymy remains contested. Critics argue it conflicts with principles of language economy [2], semantic theory [1], and empirical observations [3]. Although some studies have examined limited sets of near-synonyms [4], none have tested the existence of absolute synonymy lexicon-wide. We address this gap using a corpus-based distributional semantics approach [5] to assess whether any words in contemporary English are truly identical in meaning.
Methods: Under absolute synonymy, two words should be as similar to each other as a word is to itself. To test this, we constructed a 14.5 billion word corpus of English and trained distributional semantic models on it using *word2vec* [6]. We randomly tagged each word token with a "*_1*" or "*_2*" suffix, constructing independent distributional representations of the same word (e.g., "*distinguish_1"* vs. "*distinguish_2*"). For each word, we then computed the cosine similarity between its two tagged variants (self-similarity) as well as between the word and all words in a list of approximately 41,000 words (neighbor-similarity). This list consisted of words that appeared either in the English Lexicon Project [7] or in WordNet [8] and that had a minimum frequency of 0.1 per million words. To ensure robustness, we performed stochastic validation by repeating the random tagging and model training procedure 40 times. Paired-sample t-tests compared each word's self-similarity to its highest neighbor-similarity, with Bonferroni correction for multiple comparisons.
Results: Self-similarity was higher than neighbor-similarity for all pairwise comparisons among our list of 41,000 words (i.e., for over 870 million word pairs). Figure 1 shows a density plot of self-similarity versus closest-neighbor-similarity. The average self-similarity was 0.986 (SD = 0.013), compared to an average closest-neighbor-similarity of 0.791 (SD = 0.084). For every word, self-similarity was significantly higher than closest-neighbor-similarity. Further analyses revealed that closest-neighbor-similarity scores were higher for low-frequency words compared to high-frequency words, and for nouns compared to verbs, adjectives, or adverbs (see Figure 2).
Discussion: Our results provide empirical evidence against the existence of absolute synonymy in contemporary English. They show that even similar words retain differences in meaning, however subtle, ensuring that each word occupies a unique role in the language [1,3]. Since absolute synonymy requires full contextual interchangeability, modeling contextual usage patterns through distributional semantics provides a natural basis for empirical evaluation. By leveraging distributional representations over a large, genre-balanced corpus, thus our method offers a powerful and scalable approach to investigating semantic structure and lays the groundwork for exploring whether the absence of absolute synonymy is a universal feature of human language.

Figures

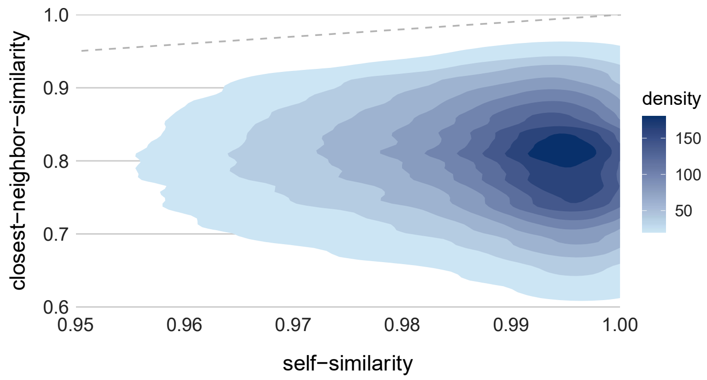


Figure 1. Density plot of the cosine similarity between a word and itself (self-similarity) and between a word and it's closest semantic neighbor (closest neighbor-similarity). The dashed line is the line on which we would expect absolute synonyms to fall.

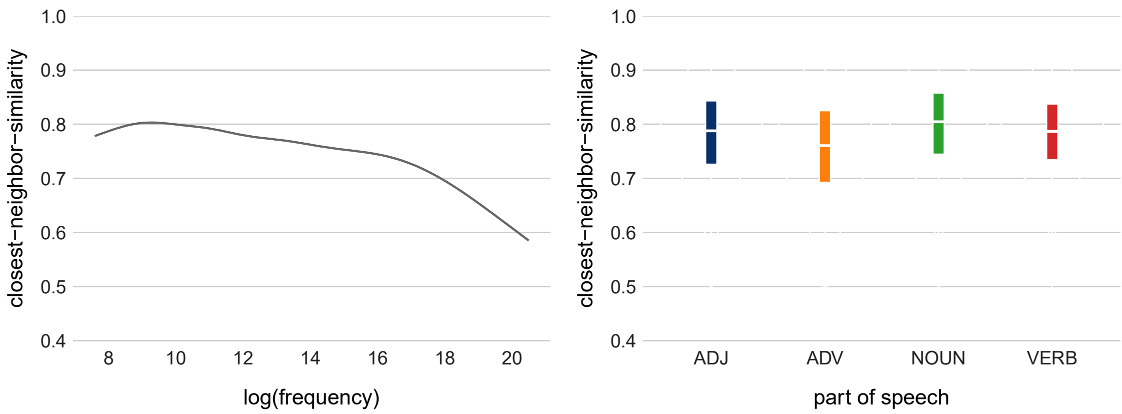


Figure 2. Plots of the effect of word frequency (left panel) and part-of-speech (right panel) on the cosine similarity between a word and its closest neighbor (closest-neighbor-similarity).

References

[1] Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.

[2] Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.

[3] Murphy, M. L. (2003). *Semantic Relations and the Lexicon: Antonymy, Synonymy, and Other Paradigms*. Cambridge University Press.

[4] Abdumanapovna, S. A. (2021). Using corpus-based data to linguistically investigate absolute and loose synonyms. *Int. J. of Aquatic Science, 12* (3), 2886–2899.

[5] Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis* (pp. 1–32). Oxford: Philological Society.

[6] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[7] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

[8] Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41.

The role of contextual alignment in artificial grammar learning

Holly Jenkins[1], Elizabeth Wonnacott[1], Michael Ramscar[2]

[1] University of Oxford, [2] University Tübingen

holly.jenkins@education.ox.ac.uk, elizabeth.wonnacott@education.ox.ac.uk, michael.ramscar@uni-tuebingen.de

Learners normally acquire their first languages in context. While this poses problems – e.g., referential ambiguity – it might also provide benefits, such as ensuring that appropriate patterns of relations between words are acquired. Another benefit that could arise from learning in context is that while obvious similarities among referents might be learnable in isolation, some subtle differences could be better discerned in contexts that allow referents to be aligned (Markman & Gentner, 1997). Since second language learning often occurs in sparse contexts, this may explain why L2 learners struggle to acquire aspects of grammar encoding differences. To explore the impact of learning in and out of context, we trained participants on an artificial language containing a grammatical gender system – which L2 learners typically struggle to learn – that systematically matched a set of morphological markers to differences among the features of its nouns. A between-participants experiment contrasted a traditional "presentation training paradigm," in which referential phrases and corresponding images were introduced individually, with an "alignment error paradigm," where referential phrases were presented alongside two images, and learners had to guess the correct match. Our results show that although noun-phrases were learned well in both paradigms, morphological generalisation only occurred when training allowed for contextual alignment, suggesting that this may be critical for learning many aspects of grammatical structure. Moreover, analysis of eye-movements during training provide evidence that participants used the information in gender-marked determiners in real-time processing to anticipate upcoming nouns: they fixated referents faster in trials where foil and target came from different gender-classes rather than the same noun class (cf Lew-Williams, & Fernald, (2010)). These results suggest that learners can use the prediction error inherent in ambiguity to discriminate the relevant cues to referents and support generalisation.

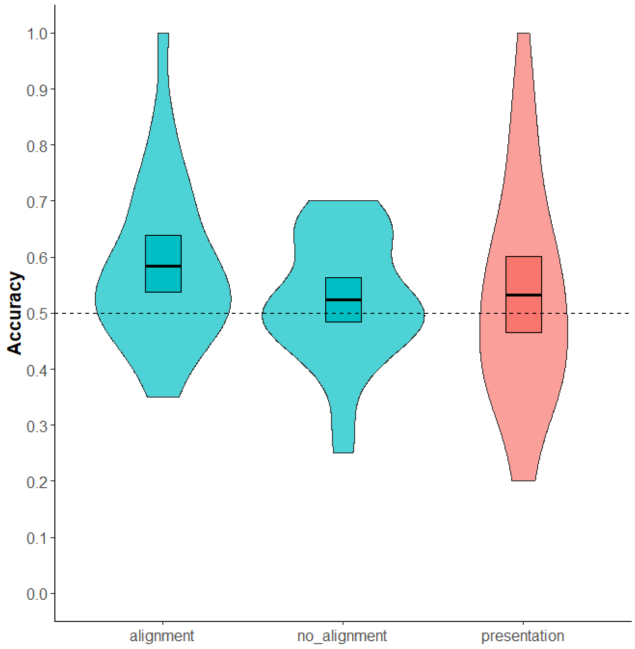


Figure 1. Accuracy in judging the grammaticality of gender-marked novel nouns based semantic cues in the artificial language experiment. Boxes show means and 95% confidence intervals.

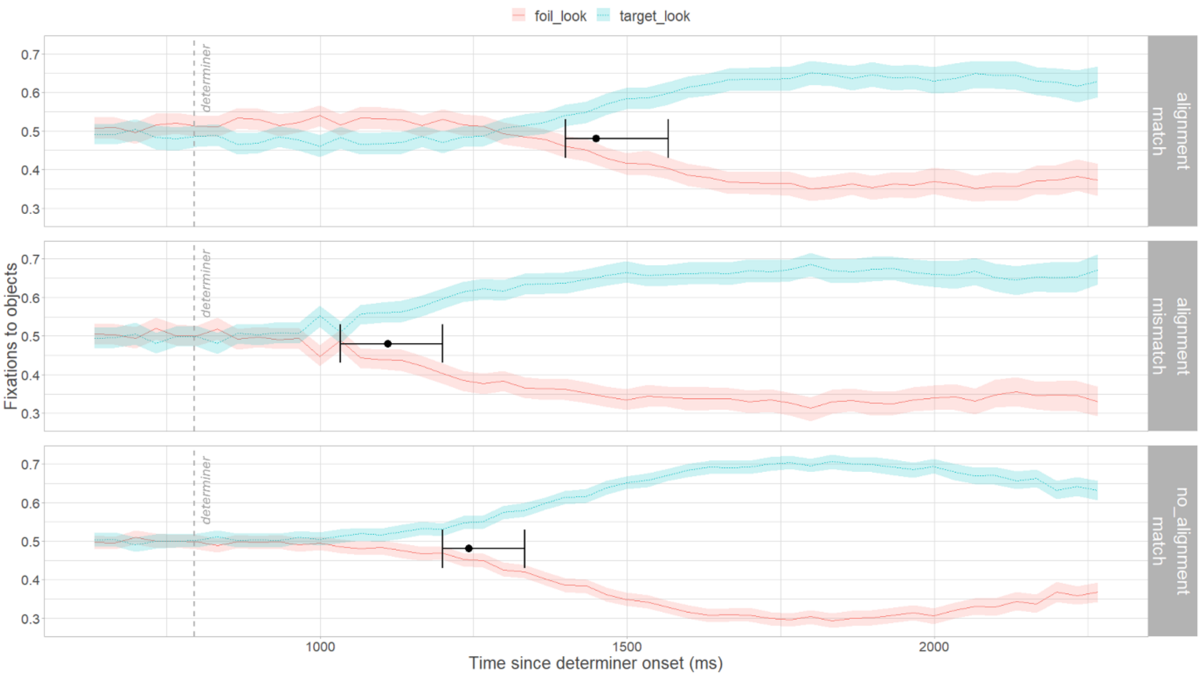


Figure 2. Eye-tracking data from the training phase. Plot shows mean fixations to target and foil (with 95% confidence intervals). Divergence points in black (with confidence intervals) show where looks become consistently greater to the target than the foil (computed using non-parametric bootstrapping). Greater looks to the target prior to the onset of the noun (vertical dashed lines) are based on information afforded by the determiner.

The relationship between perceptual abilities and speech-to-speech synchronization: A Bayesian mixture modeling approach

Andrea Hofmann[a], João Veríssimo[b], Isabell Wartenburger[a],

[a] Cognitive Sciences, Dept. Linguistics, University of Potsdam; [b] Center of Linguistics, School of Arts and Humanities, University of Lisbon

andrea.hofmann@uni-potsdam.de

Background: While the link between perception and production has been shown at the segmental level [1] less is known about its presence at the suprasegmental, prosodic level. Prosody is vital for structuring speech, enhancing auditory sequence memory, and integrating phonological, syntactic, and semantic units [2,3]. Pitch, as a critical prosodic cue, requires sophisticated auditory skills including frequency discrimination and temporal prediction [4]. Neural entrainment - the alignment of neural oscillations with acoustic signals - underlies auditory processing, especially prominent at syllabic rates, and can be assessed through speech-to-speech synchronization (S-to-S sync), where individuals align vocal output with external speech [5,6]. Individual differences in neural entrainment correlate with variations in S-to-S sync ability, which shows a bimodal distribution distinguishing high and low synchronizers [5,6,8]. Our study examines whether individual variability in perceptual acuity for the three primary prosodic cues in German syntactic disambiguation [7] is related to S-to-S sync abilities to test whether established perception-production links extend to prosodic features.

Method: We measured S-to-S sync in 60 native German speakers using the explicit accelerated version the task [8], where participants matched their whispered *'tah'* syllables to an incrementally accelerating auditory stream of synthesized syllables from 4.3 to 4.7 Hz. Phase-locking values (PLVs) were calculated and averaged across two runs. We also assessed participants' perceptual acuity for three prosodic features using adaptive staircase just-noticeable difference tasks: pitch contour, pause duration, and final lengthening discrimination. Using Bayesian mixture modeling, we identified distinct synchronizer groups and examined how the perceptual abilities predicted group membership, employing sensitivity analyses with varying priors to ensure robust conclusions.

Results: Mixture modeling confirmed a bimodal distribution of synchronization abilities with distinct component means (Figure 1) (μ_1 (low) = 0.31 PLV, μ_2 (high) = 0.69 PLV) closely aligning with [8] (μ_1 (low) = 0.26 PLV, μ_2 (high) = 0.63 PLV). Both in separate and combined models (controlling for other abilities), better pitch discrimination emerged as the strongest predictor of being a high synchronizer (Figure 2) (sep.: θ_{high} = 1.70 log-odds, CI [0.45, 3.77], $\ln BF_{10}$ = 2.75; comb.: θ_{high} = 2.14 log-odds, 95% CI [0.54, 4.49], $\ln BF_{10}$ = 2.10), providing reliable evidence. Final lengthening discrimination showed positive but uncertain effects (sep.: θ_{high} = 0.88 log-odds, CI [0.10, 1.82], $\ln BF_{10}$ = 0.54; comb.: θ_{high} = 0.85 log-odds, 95% CI [-0.42, 2.52], $\ln BF_{10}$ = -0.57, with inconclusive evidence (tending toward null when controlled). Pause discrimination showed no reliable contribution (sep.: θ_{high} = 0.74 log-odds, CI [0.02, 1.57], $\ln BF_{10}$ = 0.04; comb.: θ_{high} = -0.58 log-odds, 95% CI [-2.16, 1.08], $\ln BF_{10}$ = -0.95), with evidence tending to favor the null in both models**.**

Discussion: Our findings indicate that pitch discrimination ability is, among those abilities tested, the primary prosodic predictor of synchronization performance: individuals with better pitch perception show better S-to-S synch abilities. This may reflect the complex, multidimensional nature of pitch processing, which integrates frequency detection, contour tracking, and temporal prediction—skills essential for both prosodic boundary perception and synchronized timing. Our findings indicate that the perception-production link is also present at the level of suprasegmental prosodic units.

Acknowledgements: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project ID 317633480, SFB1287.

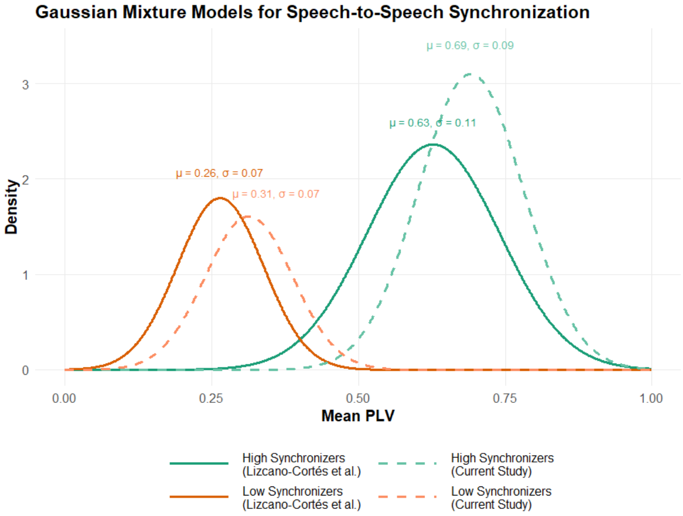


Figure 1: Comparison of Gaussian mixture distributions for the bimodal distribution of speech-to-speech synchronization between the study of Lizcano-Cortés et al. (2022) (solid lines) and the current study (dashed lines).

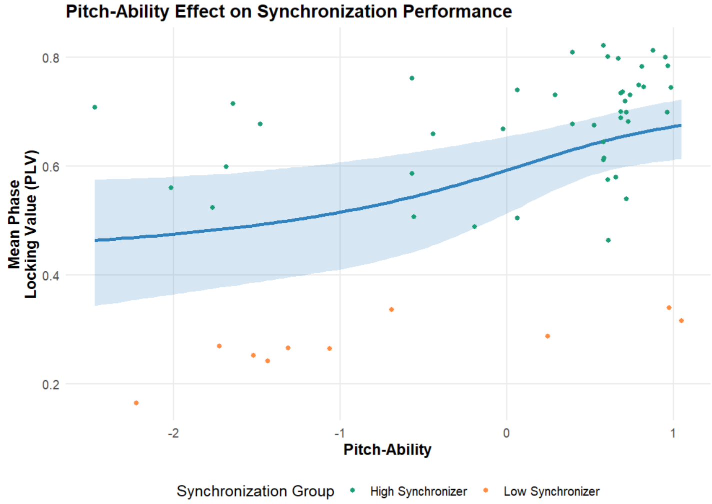


Figure 2: Relationship between pitch discrimination ability and speech synchronization. Better pitch ability (higher values) predicts higher synchronization performance (PLV). Blue curve shows model predictions with 95% credible intervals (shaded areas). Points represent participants colored by synchronizer group from the original S-to-S sync task classification: green = high synchronizers, orange = low synchronizers.

References

[1] Perkell, Joseph S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., … Guenther, F. H. (2004). The distinctness of speakers' /s/-/s/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of Speech, Language, and Hearing Research, 47(6), 1259–1269. https://doi.org/10.1044/1092-4388(2004/095).

[2] Frazier, L., Carlson, K., & Clifton, C., JR. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002.

[3] Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. The European Journal of Neuroscience, 48(7), 2609–2621. https://doi.org/10.1111/ejn.13748.

[4] Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161.

[5] Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nature Neuroscience, 22(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z.

[6] Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. Nature Reviews Neuroscience, 21(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4.

[7] Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in german: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, *61*, 71–92. https://doi.org/10.1016/j.wocn.2017.01.002.

[8] Lizcano-Cortés, F., Gómez-Varela, I., Mares, C., Wallisch, P., Orpella, J., Poeppel, D., Ripollés, P., & Assaneo, M. F. (2022). Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers. STAR Protocols, 3(2), 101248. https://doi.org/10.1016/j.xpro.2022.101248.

Declarative memory effects in L2 morphology learning reflect explicit rule awareness

Giulia Bovolenta[1] and John N. Williams[2]

[1]Bangor University, [2]University of Cambridge

g.bovolenta@bangor.ac.uk

Background: Declarative memory (DM) is a predictor of L2 learning, particularly in its earlier stages [1]. However, the exact way(s) in which DM helps L2 learning is not yet clear. Due to its flexibility, DM may help to learn explicit rules, memorize items as chunks (e.g., "kicked"), or generalise to new forms by analogy (e.g., "blicked") [2]. We carried out a study to tease apart the relative contribution of these learning mechanisms to the effect of verbal DM on L2 acquisition. 80 native English speakers took part in an artificial language learning study where they learned novel morphology and were then tested on both trained and new (untrained) items. Our research questions were:

1. Explicit rule learning: Is the effect of DM modulated by a participant's level of explicit rule knowledge?
2. Chunk memorisation: Is the effect of DM greater for trained than for new items?
3. Analogical processing: Is the effect of DM for for new items modulated by similarity to trained items?

Method: Participants were taught a mini-inflectional system with 48 disyllabic stems and two plural suffixes (-ak, -ek). Suffixation was determined by vowel harmony (back vs. front) between the stem's final syllable and the suffix (e.g., 'nifos' → 'nifosak', 'posim' → 'posimek').

Participants heard stems and their inflected plural forms with pictures depicting their meaning, receiving no explicit instruction. They then gave grammaticality judgments on inflected plural forms of the 48 trained stems and 48 new ones (evenly split between grammatical and ungrammatical).

Verbal declarative memory was measured by MLAT V [3]. Explicit rule knowledge at the end of the test was measured by questionnaire, scored as: 0 = no awareness, 1 = form-aware (aware of suffix alternation, but not tied to stem sound), 2 = partial rule-aware (aware of suffix alternation based on stem sound), 3 = full rule-aware (aware of suffix alternation based on stem's final syllable). Item similarity was calculated for each untrained stem as the average of its Levenshtein distance to the trained stems. Grammaticality judgment data was analysed with logistic mixed-effects modelling; continuous predictors were scaled and centered and categorical ones were sum-coded.

Results: We observed an interaction between MLAT V and Awareness scores (OR = 1.27 [1.04 – 1.55], p = .019): the effect of MLAT V on test accuracy increased with explicit rule awareness. There was no interaction between MLAT V and Item type (OR = 1.04 [0.93 – 1.17], p = .450) or between MLAT V and Similarity scores for untrained items (OR = 1.00 [0.97 – 1.03], p = .915).

Discussion: We did not observe effects attributable to chunk memory (greater effects of DM for trained items) or analogical processing (accuracy on untrained items modulated by similarity to trained ones). But the effect of DM on test scores was modulated by explicit knowledge, increasing across awareness levels. This suggests that one way in which DM supports L2 morphology learning (as measured by grammaticality judgment) is by helping learners to capitalise on their explicit rule knowledge. Further research should look at more implicit language tests, too.

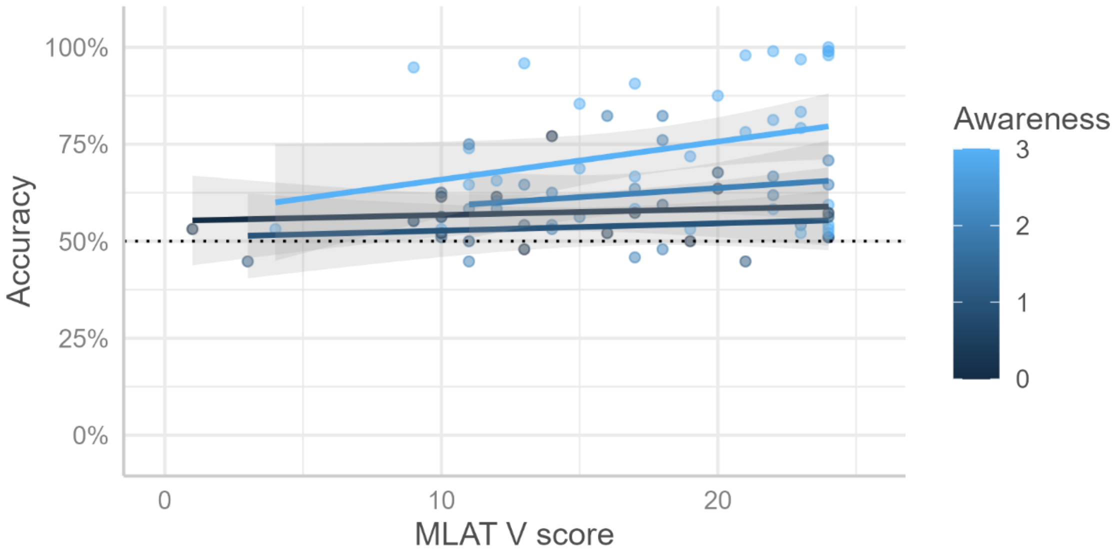


Figure 1. Mean accuracy by MLAT V and rule awareness score

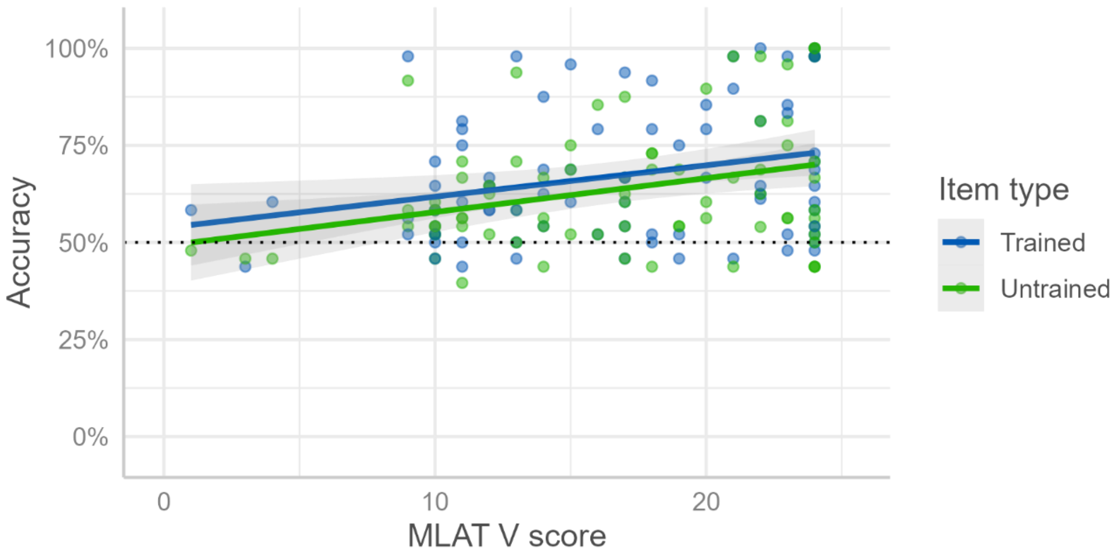


Figure 2. Mean accuracy by MLAT V and item type (trained vs. untrained stems)

References

[1] Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America, 115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

[2] Morgan-Short, K., Hamrick, P., & Ullman, M. T. (2022). Declarative and Procedural Memory as Predictors of Second Language Development. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 67–81). London : Routledge.

[3] Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test. 27.* https://psycnet.apa.org/fulltext/1960-03629-000.pdf

When *yes* sounds like *maybe*: Inferences about respondents in offers and requests

Stefan Blohm[1], Mathias Barthel[1]
[1] Leibniz-Institute for the German Language, Mannheim, Germany
blohm@ids-mannheim; barthel@ids-mannheim.de

Background: In fluent conversation, interlocutors achieve average transition times of 200-300 ms between turns [1]. Longer gaps often indicate 'trouble', e.g., disagreement with a prior assessment [2], and may thus trigger corresponding inferences in comprehenders [3]. Prior studies of such delay-driven inferences have revealed a 'temporal threshold of tolerance' (~700 ms) for inter-turn silence after requests: Longer gaps trigger inferences about respondents' unwillingness to comply [4]. In two experiments in German, we used intuitive judgments to replicate this gap effect on attributions of (un)willingness following requests/offers, to assess whether it extends to interpersonal stances, and to test whether comprehenders adjust their tolerance threshold to speaker characteristics (language proficiency: native vs. non-native). We hypothesized that dis-fluent turn taking triggers inferences about respondents' stance towards both request (Exp. 1: "the respondent is unwilling") and requester (Exp. 2: "the respondent is cold and distant"), but that inferencing is delayed/attenuated if contextual factors plausibly account for the gap. We predicted asymmetrical gap effects for native (=early gap effect) and non-native respondents, whose long silence may be plausibly attributed to increased processing- and planning time (=delayed/reduced gap effect) [5].

Method: We constructed 240 short dialogues modeled as telephone conversations between friends; dialogues consisted of a brief context and a critical question-response pair (=request/offer + yes-response, see Table 2). Questions were spoken by four native speakers of German and responses by four native and four non-native (L1 French) speakers; a pretest ensured that responses were identifiable as (non-)native. We varied the gap length in critical turn transitions (200/720/1200 ms) and respondents' language proficiency (native/non-native) in a fully crossed 3x2 Latin square design. Participants (*n*=42 per experiment) read the context of each dialogue on a screen, then listened to the critical question-response pair via headphones, before they intuitively judged the response on a 5-point rating scale (Exp. 1 "How willingly does she accept?"; Exp. 2: "How cold and distant does she react?"). Bayesian ordinal mixed-effects regression tested for main and interaction effects of gap duration (200/720/1200) and nativeness (native/non-native) on ratings; we used forward-fitting to further test whether model performance improved when phonetic variables known to affect attributional inferences (i.e., response duration and -pitch) [6] were included as covariates.

Results (see Table 1)**:** <u>Exp. 1</u>: We observed main effects of gap duration (200 > 720 > 1200) and language proficiency (native > non-native) as well as the expected interaction (native: 200 >> 720 > 1200; non-native: 200 > 720 >> 1200). <u>Exp. 2</u>: We observed a main effect of gap duration (200 < 720 < 1200) and main effects of language proficiency (native < non-native) and response duration (short > long).

Discussion: **(1)** Delayed responses in recruitments triggered inferences about respondents' stance towards the request and the requester. **(2)** Non-native speakers generally appeared less willing than native speakers and more cold and distant; only judgments of interpersonal distance were sensitive to phonetic response parameters (i.e., duration); presumably because short responses convey curtness and thus increase interpersonal distance. **(3)** Partly supporting our delayed-inference hypothesis, speakers' language proficiency modulates gap-driven attributions of unwillingness (but not interpersonal stance): we observed a strong *early* penalty for native respondents but a strong *late* penalty for non-native ones.

Table 1: Hypothesis tests of Bayesian ordinal mixed-effects regression models.

		Willingness judgments (n=42)		**Distance judgments (n=42)**	
		estimate	$CI_{95\%}$	estimate	$CI_{95\%}$
3 x 2 design	gap720 - gap200	-0.28	[-0.36, -0.20]	0.16	[0.09, 0.23]
	gap1200 - gap720	-0.27	[-0.35, -0.19]	0.15	[0.07, 0.23]
	non-native - native	-0.76	[-0.99, -0.53]	0.46	[0.15, 0.76]
phonetic covariates [6]	response $duration_{+100ms}$	*discarded / no improvement*		-0.21	[-0.28, -0.14]
	response $pitch_{+10Hz}$ (mean)	*discarded / no improvement*		*discarded / no improvement*	
	gap:resp. duration	*discarded / no improvement*		*discarded / no improvement*	
post-hoc tests for gap:native interaction	native: gap720 - gap200	-0.36	[-0.45, -0.26]	–	
	native: gap1200 - gap720	-0.22	[-0.32, -0.13]	–	
	non-native: gap200 - gap720	-0.20	[-0.30, -0.09]	–	
	non-native: gap1200 - gap720	-0.31	[-0.42, -0.21]	–	

References

[1] Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

[2] Pomeranz, A., Atkinson, J., & Heritage, J. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In *Structures of Social Action*, 53–101. Cambridge University Press

[3] Henetz, T. (2017). Don't hesitate! The length of inter-turn gaps influences observers' interactional attributions. PhD Thesis, Stanford University, Stanford. url: http://purl.stanford.edu/zk615rk5483

[4] Roberts, F., & Francis, A.L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America* 133, EL471–EL477. doi: 10.1121/1.4802900

[5] Matzinger, T., Pleyer, M., & Żywiczyński, P. (2023). Pause length and differences in cognitive state attribution in native and non-native speakers. *Languages*, 8(1). doi: 10.3390/languages8010026

[6] Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, 48(9), 1079–1093. doi: 10.1016/j.specom.2006.02.001

Table 2: Example dialogue in German and its English translation.

Speaker A	**Speaker B**	**English translation**	
Hi, was geht bei dir?		*Hi, what's up?*	
	Hi, ich bin fertig mit dem Training.	*Hi, I've finished training.*	context
	Kannst du mich abholen?	*Can you pick me up?*	question
(0.20/0.72/1.20)		(0.20/0.72/1.20)	
Ja.		*Yes*	response after variable silence

Conditioned Delusions: Belief Updating During Naturalistic Reading is Modulated by Individual Cognitive Profile

P. Alex[1], M. Brzeska[1], J. Schwarz[2], B. W. Tatler[1], A. E. Konopka [1*], A. Klimovich-Gray[1*]
[1] University of Aberdeen; [2] University of Cambridge; * joint last authorship
p.alex.24@abdn.ac.uk

Background: We continuously generate and update predictions for incoming input, allowing adaptation and learning from past experiences [1]. In reading studies, predictability effects are well-documented at the word and sentence level [2]: highly predictable words are read faster, while unexpected content generates prediction errors. In theory, prediction errors could be resolved through either (1) belief updating or (2) behavioural adaptation (e.g., changes in reading strategies) without updating beliefs [3]. However, little research has examined how such error resolution operates in broader story-level contexts that reflect real-world reading, or how individual differences shape this process. In an eye-tracking study, we investigated how providing feedback that conflicted with narrative-level context dynamically modulates belief updating during reading, and whether this is modulated by individual differences in cognitive profiles. **Method:** Using a novel paradigm, we investigated trial-by-trial updating of prior beliefs using a False Feedback manipulation [4, 5] in whole-sentence self-paced reading. In a within-subjects design, 52 native English speakers listened to audio narratives about two fictional characters that set up strong expectations (‘lazy Jane’, ’amiable Sophie’). They then read and rated the accuracy of 144 new sentences describing these characters in two counterbalanced blocks, each consisting of three stages. In the first block, participants received 100% True Feedback (TF; i.e., correct feedback regarding response accuracy), whereas in the second block, they received False Feedback (FF; i.e., incorrect feedback) on 17% (Stage 1) and 42% (Stage 2) of trials. In Stage 3 (Test), participants rated sentences without receiving feedback. We hypothesised that receiving FF would generate prediction errors, prompting participants to re-weigh their priors against the new information. To assess individual differences in belief updating (quantified by changes in accuracy and eye-gaze patterns on trials following FF), we assessed phonological working memory (PWM), non-verbal intelligence (NVI) [6], and metacognitive awareness (MAI) [7]. **Results and Discussion:** As expected, accuracy dropped across Blocks (TF: 94.6%; FF: 89.8%) and Stages (94.8% to 89.1%) with progressively more FF. Hierarchical Bayesian modelling revealed that participants gave more incorrect responses, had longer first-pass and total reading times, and made more regressions to sentences read immediately following FF compared to TF trials (Fig 1a-d). Whilst this confirmed that participants changed reading strategies in immediate response to FF, we next asked which adaptations during learning (i.e., FF-induced changes in accuracy, confidence, and/or reading metrics) led to long-term belief updating at Test. Only increases in total reading times and regressions, but not first-pass reading times, predicted higher Test accuracy, suggesting that greater attention to text following FF was linked to less narrative-level belief updating (i.e., maintaining prior beliefs). Individual differences modulated these responses: readers with both high NVI and PWM had the longest reading times and highest frequency of regressions -- a pattern consistent with verification of the FF against previously established priors rather than with immediate integration of FF. These individuals were also least likely to show drops in Test accuracy, suggesting strategic behavioural adaptation without belief change. We argue that PWM supports flexible integration of local context with prior beliefs, while NVI is linked to evaluating the reliability of incoming input.

Figure 1. Reading performance: participants were credibly (a) less accurate, (b) showed longer first-pass fixations and (c) longer total fixations across the entire sentence, and (d) made more regressions when reading sentences immediately following a False Feedback trial compared to True Feedback. Figures show posterior distributions for the group-level effect of False Feedback on each metric: dashed vertical lines indicate the Region of Practical Equivalence (ROPE) and the horizontal bar indicates the 89% Highest Density Interval (HDI).

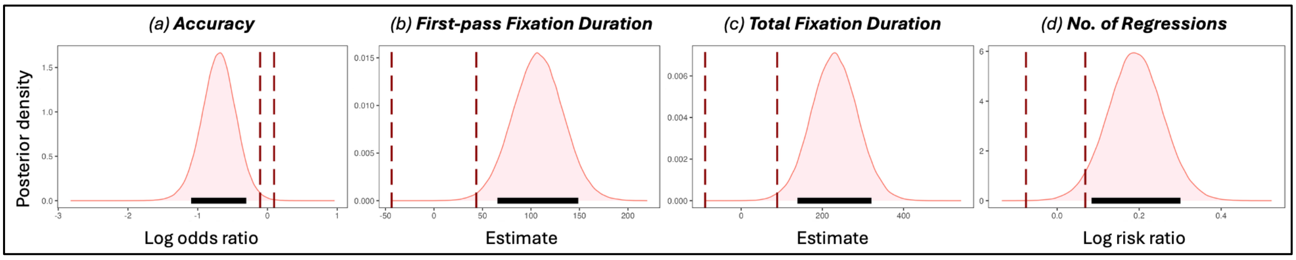
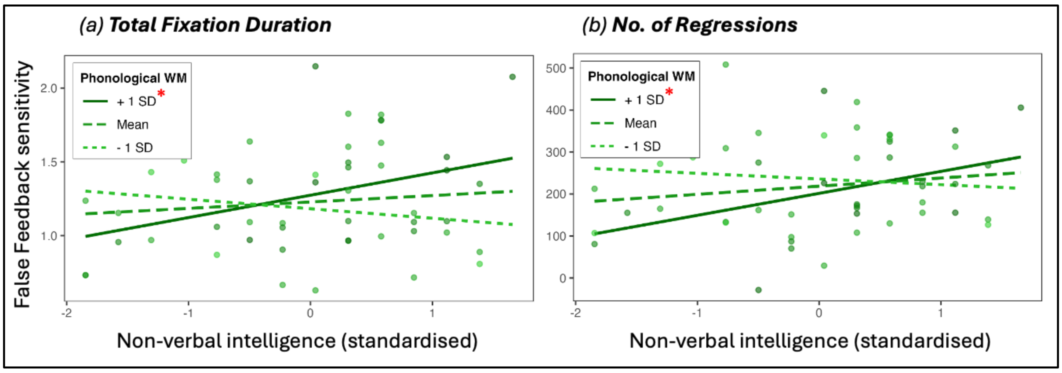


Figure 2. Individual differences: False Feedback sensitivity (i.e., individual-level effects of False Feedback estimated via hierarchical Bayesian models) for (a) total fixation durations and (b) regressions was predicted by an interaction between phonological working memory (PWM) and non-verbal intelligence (NVI). Participants with high PWM showed greater adaptability in reading strategies with increasing NVI.



References

[1] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36(3)*, 181–204.

[2] Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9(8)*, 311-327.

[3] Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation, 29(1)*, 1–49.

[4] Kusev, P., van Schaik, P., Teal, J., Martin, R., Hall, L., & Johansson, P. (2022). How false feedback influences decision-makers' risk preferences. *Journal of Behavioral Decision Making, 35(5)*, e2278.

[5] Wang, J., Otgaar, H., Smeets, T., Howe, M. L., & Zhou, C. (2019). Manipulating memory associations changes decision-making preferences in a preconditioning task. *Consciousness and Cognition, 69*, 103–112.

[6] Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S. J. (2019). The matrix reasoning item bank (MaRs-IB): novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science, 6(10)*, 190232.

[7] Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19(4)*, 460–475.

Context Effects on the Interpretation of Bare Numerals: Evidence from Event Uncertainty and Roundness

Marie Christin Walch
German Department, Leibniz University Hannover
marie.walch@germanistik.uni-hannover.de

Motivation: Theoretical accounts of the interpretation of (bare) numerals dispute whether they receive a punctual (i.e., ‚exactly n‘) or lower-bound (i.e., ‚at least n‘) interpretation [1]. A bare numeral can be interpreted as a lower bound through pragmatic inferences. Experimental findings [2] show that 50% of participants accepted the lower-bound interpretation of bare numerals, supporting their ambiguity. However, these studies used past-tense scenarios, which presuppose event completion. Employing *at least* gives rise to an ignorance inference, suggesting an absence of precise knowledge of the speaker [3]. Consequently, a bare numeral *n* receiving a lower-bound interpretation should trigger the same ignorance inference. Based on this, I hypothesize that future-tense contexts, where outcomes are uncertain, may amplify lower-bound interpretations due to speaker ignorance. This is supported by [4], who, though focusing on modified numerals and question types, examine speaker ignorance in numeral interpretation. As a secondary factor, the characteristic of *n*, its roundness, may influence its interpretation as well by creating a pragmatic halo around the actual numeral, leading to an approximate reading [5] and a reduction in the speaker's commitment to the accuracy of their utterance [6].

Method: A pilot study was conducted as an online experiment (Prolific) in German. It incorporated a 3x3 between-subject design, with the factors EPISTEMIC (competence_certainty vs. ignorance_certainty vs. ignorance_uncertainty) x NUMERAL ($n < m$ vs. $\tilde{n} < m$ vs. $\tilde{n} > m$), where n is the number claimed by a speaker A, round ($\tilde{n}$) or non-round (n), and m the actual number. I distinguish between the conditions ignorance_certainty and ignorance_uncertainty to examine whether there is a difference between speaker ignorance and uncertainty about situations where different outcomes are equally possible. The nine conditions resulting from the 3×3 design are listed representatively in Table 1. 318 native German speakers each completed one condition to avoid carry-over effects. The participants were asked to judge whether a bet had been won or lost in the following context: Two befriended teachers made a bet on the number of attendees in their class. The factor EPISTEMIC was observed in a way so that the class in question was either previously taught (competence_certainty), had already been booked (ignorance_certainty), or was in the process of being booked (ignorance_uncertainty). The teacher A makes a claim about the number of attendees (n), which is refuted by B. Subsequently, both determined that the actual number of attendees *m* was either lower ($n > m$) or larger ($n < m$) than the stated number *n* by A. Regarding the influence of numeral roundness, both round ($\tilde{n}$=30) and non-round numerals (n=31) were tested. Participants responded to a polar question and were required to make a forced-choice judgment between two response options: [*Yes,* A*; No,* A *doesn't*] *won the bet.*.

Discussion & Results: A binomial logistic regression (glm with logit link) assessed the influence of speaker knowledge under event certainty (EPISTEMIC) and numeral roundness (NUMERAL) on the likelihood of a positive response for lower-bound interpretations. The model showed a significant main effect of the ignorance_uncertainty condition ($\beta = 1.64$, $p = .008$, $OR = 5.14$), suggesting that participants were substantially more likely to accept lower-bound interpretations when the speaker lacked knowledge and the event had not yet occurred (see Figure 1). This effect was more pronounced for non-round numerals, indicating both speaker knowledge and numeral roundness influence interpretation, with participants being more likely to interpret round numerals ($\beta = 1.30$, $p = .034$, $OR = 3.67$) as lower bounds. Importantly, speaker ignorance under uncertainty significantly increased acceptance rates ($p < .01$,$OR = 5.14$). In contrast, ignorance under event certainty did not significantly differ from the baseline ($p = .52$). The lack of interaction between speaker competence and ignorance might point to the limited contextual salience in the current study design. Future studies should explicitly manipulate this factor to observe potentially stronger effects. Finally, the upper-bound reading ($\tilde{n} > m = 29$) was significantly less accepted than the lower-bound reading ($\beta = 1.54$, $p < .001$, $OR = 4.67$) across all epistemic conditions, suggesting that results for round numerals cannot be entirely attributed to the pragmatic halo.

		EPISTEMIC		
		competence_certainty	ignorance_certainty	ignorance_uncertainty
NUMERAL	$\tilde{n} < m = 31$	30 did attend	30 did attend	30 will attend
	$\tilde{n} > m = 29$	30 did attend	30 did attend	30 will attend
	$n < m = 32$	31 did attend	31 did attend	31 will attend

Table 1: Standardized format of the A's claim with the claimed number of attendees *n* by A and the actual number m across all conditions

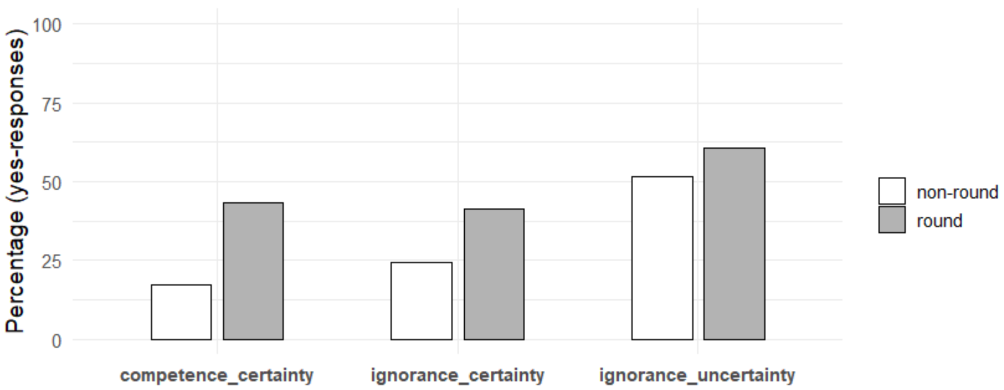


Figure 1: Acceptance rates for lower-bound reading of round and non-round *n* across all EPISTEMIC conditions

References

[1] Spector, B. (2013). Bare numerals and scalar implicatures. *Language and Linguistics Compass*, *7*.

[2] Geurts, B., Katsos, N., Cummins, C., Moons, J., & Noordman, L. (2010). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, *25*, 130–148.

[3] Cummins, C., & Katsos, N. (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics*, *27*(3), 271–305.

[4] Cremers, A., Coppock, L., Dotlacil, J., & Roelofsen, F. (2022). Ignorance implicatures of modified numerals. *Linguistics and Philosophy*, *45*.

[5] Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and evidence in semantics* (pp. 109–132). CSLI Publications.

[6] Van der Henst, J.-B., & Sperber, D. (2004). Testing the cognitive and communicative principles of relevance. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 141–171). Palgrave Macmillan UK.

Testing the Specificity of Human Parser Predictions during ‘Hyper-Active’ Gap Filling

Liliana Nentcheva and Andrea Santi

Department of Linguistics, University College London (UCL)

lily.nentcheva.15@ucl.ac.uk

A key goal of human parsers during filler-gap dependency comprehension is to reduce memory costs by predicting early resolution of the dependency [1]. This is facilitated by strategies such as ‘hyper-active’ gap filling, whereby the parser predicts that the complement domain of the upcoming verb will have an available gap position for the filler to be offloaded [2]. Evidence for this in English filler-gap dependencies is provided by a reading slowdown at an intransitive verb (1a) compared to a verb which subcategorises for a direct object (DO) (1b), relative to the same verbs in syntactic islands (1c-d) [2], where gap integration of the filler is blocked [3]. We present work-in progress to explore the specificity of the gap prediction alongside addressing concerns in this previous work.

We ask whether parsers specifically predict DO gaps in the structure of the upcoming verb, consistent with theories emphasising minimal syntactic structure (e.g. in terms of fewer syntactic nodes [4]), or if a more general gap is predicted within the verb’s complement domain. Additionally, the first-pass reading time (RT) measure, where the slowdown is observed in [2], has largely been associated with Surprisal rather than revision processes [5]. This could be consistent with the interaction effect in [2] being driven by a difference in semantic predictability of the verb based on the filler (*studio-designed* vs. *studio-remained)* that the syntactic island baseline may not effectively control for. RT measures which are more typically linked to reanalysis, namely regression path duration [5], are elevated in the baseline sentences due to the complexity of the island structures. This obscures the effect of interest, as regression path RTs are equivalent across the gap and island conditions in the intransitive (1a, 1c) sentences, but not transitive (1b, 1d) sentences, thus creating an interaction in the direction opposite to that expected.

To test the specificity of the prediction, we include a third verb subcategorsation condition: a clausal complement (CC) where a DO gap is not permitted (2c,f) (supported by corpus and acceptability data (**Table 1**)). We also simplify the baseline condition by utilising a non-island structure (2d-f) for a more straight-forward interpretation of the effect at the verb. The study will utilise the eye-tracking while reading paradigm and data collection is predicted to be complete by July 2025.

If parsers ‘hyper-actively’ predict a specific DO gap, we expect increased RTs in both the CC and intransitive conditions compared to DO conditions, relative to baseline. If a more general gap is predicted in the verbal complement, the DO and CC conditions should pattern similarly relative to baseline, with a slowdown observed in the intransitive condition only. The study’s results will provide insight into the nature of predictive processing in filler-gap dependencies, with broader implications for the relationship between syntactic representations and memory constraints.

1. Example items from [2] demonstrating a manipulation for verb type (intransitive/transitive) and structure type (gap, island)

Gap: The studio that the students **(a) remained**/**(b) designed** peacefully in while the professors conferred was small and ugly.

Island: The studio that the students who **(c) remained**/**(d) designed** peacefully rested in while the professors conferred was small and ugly.

2. Proposed experimental items showing a simplified baseline without a dependency and an additional CC level in verb transitivity manipulation

Gap: The accountant that the investor **(a) remained** reluctantly with/**(b) heard** reluctantly about/**(c) agreed** reluctantly with wanted to implement an alternative financial strategy.

No Gap: The accountant knew that the investor **(d) remained** reluctantly with/**(e) heard** reluctantly about/**(f) agreed** reluctantly with the new CEO who wanted to implement an alternative financial strategy.

Table 1: Corpus and acceptability data demonstrating that CC verbs subcategorise for DO complements less frequently than DO verbs taken from the stimuli in [2]

Mean number of total DO complements from 100 unique hits in the Corpus of Contemporary American English (COCA) for each CC verb in the same form as that used in our stimuli	Acceptability data demonstrating a significant difference between the acceptability of subcategorisations for DO complements between DO and CC verbs (n = 60) from human ratings on a scale of 1 to 7
2.6% (range: 0-10%). Mean number of clausal subcategorisations was higher: 55.2% (range: 18- 87%)	β = -5.3925, p < .001 From the model: *clmm(acceptability_rating ~ Verb_Type + (1+Verb_Type| Participant) + (1| Item), items)*

References

[1] Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium,* 94–126. The MIT Press.

[2] Omaki, A., Lau, E. F., White, I. D., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology,* 6

[3] Phillips, C. (2006). The Real-Time Status of Island Phenomena. *Language 82*(4), 795-823

[4] Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing strategies*. University of Connecticut

[5] Yan, S., & Jaeger, T. F. (2020). Expectation adaptation during natural reading. *Language, Cognition and Neuroscience*, *35*(10), 1394–1422.

Modeling Acceptability in Free Word Order Languages: The Role of Dependency Distance and Projectivity

Kirill Chuprinko, Artem Novozhilov, Artur Stepanov
Center for Cognitive Science of Language, University of Nova Gorica
kirill.chuprinko@ung.si

Background: This study examines whether sentence acceptability in Russian and Serbo-Croatian can be modeled using two factors: Mean Dependency Distance (MDD)[1, 2] and projectivity violations. MDD measures the average syntactic distance between words, while projectivity violations involve crossing dependency arcs or arcs passing the root node (see Fig.1–2) [3, 4]. Although previous research extensively studied MDD as a source of processing difficulty, the role of projectivity violations in acceptability judgments remains largely unexplored. We investigate whether projectivity violations independently increase processing load and evaluate their predictive power relative to MDD.

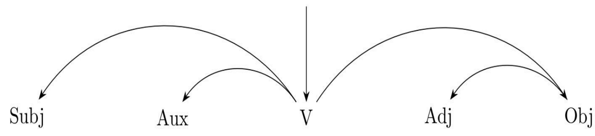
Method: We constructed 5-word kernel sentences in Russian and Serbo-Croatian, two free word order languages, following the template "Subj Aux V Adj Obj" (e.g., Rus: *Ivan budet lovitj boljshuju rybu* "Ivan will catch big fish") and generated a full set of 120 permutations of each kernel to serve as stimuli in an acceptability judgment task. Native speakers of Russian (N=79) and Serbo-Croatian (N=118) rated the resulting sentences on a 5-point Likert scale. Collected acceptability data were modeled using mixed-effects ordinal regressions without interactions. Models were compared using Akaike Information Criterion (AIC).

Results: Results in both languages revealed similar trends, though the rating distribution in SC data differs from the Russian data due to the clitic-second constraint. Both MDD and projectivity violations were significant independent predictors of acceptability ratings, with higher values leading to lower ratings. In particular, our models predict that word orders differing with projectivity violations but with equal MDDs will differ in acceptability. For example, orders Adj Aux Subj Obj V (2 violations) and V Aux Adj Obj Subj (0 violations) both have MDD=2,25. In Russian, the former order is less acceptable than the latter one (mean ratings 2,34 vs 2,97, $t=-2,8$, $df=143$, $p < 0,01$). Furthermore, models based solely on projectivity violations outperformed those based on MDD (see Tables 1–2). A series of likelihood ratio tests showed that combined models, accounting for both predictors, achieved a better fit than the baseline and simpler models, while simpler models achieved a better fit than the baseline (all comparisons statistically significant, $p<0,01$ with Holm-Bonferroni correction where applicable). The baseline models consisted of overall intercepts plus random intercepts for participants and the lexical content of sentences.

Discussion: Our findings show that the projectivity violations metric models the acceptability more accurately than MDD alone, and the combination of both metrics yields better acceptability models than either metric alone. We suggest that MDD and projectivity violations engage distinct processing mechanisms during sentence evaluation. MDD likely reflects working memory constraints: greater syntactic distances increase the burden of maintaining and integrating dependencies, leading to reduced acceptability (cf.[3]). In contrast, projectivity violations may disrupt the incremental parsing process that favors immediate attachment of elements as they are encountered. When dependencies cross or circumscribe the root (V), parsing becomes less predictable and more costly, even if MDD-related memory demands remain low. This supports the conjecture that both memory load and incrementality of parsing play a key role in acceptability evaluations.

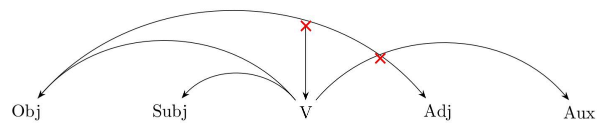
Appendix

Fig. 1. Sentence without projectivity violations. (MDD = 1,5)



Ivan *budet* *lovit'* *boljsh-uju* *ryb-u*
I. will catch.INF big-ACC fish-ACC

Fig. 2. Sentence with one violation of each type. (violations marked with red crosses, MDD=2)



Ryb-u *Ivan* *lovit'* *boljsh-uju* *budet*
fish-ACC I. catch.INF big-ACC will

Table 1. AICs of models with Russian data.

Model	AIC
MDD + Projectivity	24330
Projectivity	24420
MDD	24739
Baseline model	24952

Table 2. AICs of models with SC data.

Model	AIC
MDD + Projectivity	32434
Projectivity	32486
MDD	32700
Baseline model	32708

References

[1] Ferrer i Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E, 70*(5), 056135.

[2] Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences of the United States of America, 112*(33), 10336–10341.

[3] Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. Physics of Life Reviews, 21, 171–193.

[4] Lu, Q., Xu, C., & Liu, H. (2016). Can chunking reduce syntactic complexity of natural languages? Complexity, 21, 33–41.

[5] Kallestinova, E. D.. 2007. Aspects of Word Order in Russian. Ph.D. dissertation, University of Iowa, Iowa City.

[6] Stjepanović, S. 1999. What do second-position cliticization, scrambling and multiple wh-fronting have in common? Ph.D. dissertation, University of Connecticut, Storrs.

Additional information about Russian and Serbo-Croatian

Russian and Serbo-Croatian (SC) both belong to the Slavic language family, with Russian classified as East Slavic and SC as South Slavic. Although both languages canonically follow an SVO word order, they permit considerable word order flexibility compared to strict word order languages like English. Core syntactic units—subject, verb, object, and nominal modifiers—can appear in varying linear positions without rendering the sentence ungrammatical. Deviations from canonical order may however increase processing difficulty, though contextualized information structure and appropriate prosody can mitigate these effects.
A key difference between Russian and SC lies in the auxiliary verb. In SC, the auxiliary used in past and future tenses is a clitic that follows the Wackernagel's Law, requiring it to immediately follow the first word or phrase of the clause. In contrast, Russian uses the auxiliary *byt'* ('to be') in future constructions with imperfective verbs, but it behaves like a regular word without positional constraints. As a result, SC imposes a strict grammatical constraint on clitic placement, whereas Russian does not (compare examples (1) and (2)). Beyond this clitic behavior, we are not aware of other grammatical differences that would systematically affect word order preferences in these languages. In fact, according to syntactic literature, any permutation of word order in both Russian and Serbo-Croatian simple clauses is grammatical, modulo clitic-second requirement in Serbo Croatian [5-6].

(1) SC examples with different auxiliary verb positions

a.	Andrej	je	ulovio	veliku	ribu.
	A.	AUX	caught	big	fish
b.	*Andrej	ulovio	veliku	ribu	je.
	A.	caught	big	fish	AUX

"Andrej caught a big fish."

(2) Russian examples with different auxiliary verb positions

a.	Andrej	budet	lovit'	bol'šuju	rybu.
	A.	AUX	catch	big	fish
b.	Andrej	lovit'	bol'šuju	rybu	budet.
	A.	catch	big	fish	AUX

"Andrej will catch a big fish."

Conceptual similarity, but not informativeness, shapes evidential systems during learning

Benedek Bartha[1], Eva Wittenberg[1], Christophe Heintz[1], Jennifer Culbertson[2]
CEU[1], University of Edinburgh[2]
bartha_benedek@phd.ceu.edu

The evolution of language systems may be explained in terms of a drive for simplicity, and another (often competing) drive for informativeness [1]; but whether both drives are present during language learning is controversial [2]. Evidential marking falls on a continuum of directness, ranging from direct (perceptual, generally visual) evidence type to indirect (reportative) evidence type, with inferential evidence type in between. Of the languages of the world that have grammatical evidential marking, all explicitly mark reported evidence, and the majority mark *only* that type of evidence [3]. Furthermore, no language has marking for visual and reported evidence without also marking inferred evidence. This suggests that languages are sensitive to the continuum of directness: more indirect sources of evidence are more likely to be marked. One explanation for this pattern is that grammatical evidentiality systems are shaped by a pressure to be informative (i.e., to mark less expected information, e.g., [4]). However, evidential marking systems may instead be shaped by a simplicity drive favoring conceptual similarity [5]. Recent work on the learning of evidential systems [6] does not de-confound these two potential drives, so our experiments aim to do so.

We use artificial language learning to explore which of the two drives is active during learning. In an extrapolation paradigm, English-speaking participants are trained on sentences for two evidence types and then extrapolate to the third held-out evidence type (Fig. 1). We measure both accuracy on the two trained evidence types, and extrapolation choices for the unfamiliar evidence type. Both measures can provide evidence about the potential influence of conceptual similarity on semantic partitioning, and informativeness on marking.

Our findings of learning outcomes (Fig. 2) show no learning preference for marking indirect evidence; however, a binomial logistic regression model reveals greater learning difficulty for the system that disrupts the conceptual grouping of visual and inferred evidence (Condition 4: β =-0.34±0.08, p < .001), suggesting an influence from conceptual similarity. In Conditions 1 (N=60), 2 (N=61) and 3 (N=60), we also find clear extrapolation evidence (Fig. 3) for a preference to group visual and inferential evidence types, separately from reportative evidence type. Binomial tests reveal that participants consistently extrapolate a learned marker from visual to inferential (Condition 1: p < .001, 69.2%, 95% CI [65.5%, 72.7%]; Condition 3: p < .001, 62.7%, 95% CI [58.8%, 66.5%]), but not from reportative to inferential (Condition 2: p = .119, 46.9%, 95% CI [43.0%, 50.8%]) evidence type. The extrapolation finding in Condition 4 (N=59) shows that participants prefer a system that preserves the continuum of evidential directness: grouping reportative with inferential, rather than with visual evidence type (p < .001, 66.8%, 95% CI [63.2%, 70.3%]), despite the observed overall grouping preference.

Overall, our results suggest that conceptual similarity shapes the semantic partitioning of evidential systems during learning, but provide no evidence for an influence from informativeness on marking. These findings are in line with previous research showing that language is shaped by distinct pressures during learning and during communication: learning favors simplicity or category naturalness; communication favors informativeness [7].

[1] Kemp & Regier, *Annual Review of Linguistics,* 2018. [2] Carr et al, *Cognition,* 2020. [3] Aikhenvald, *The Oxford Handbook of Evidentiality,* 2018. [4] Kurumada & Grimm, *Cognition*, 2019. [5] Maldonado & Culbertson, *Linguistic Inquiry,* 2022. [6] Saratsli et al., *Cognition,* 2020. [7] Kirby et al., *Cognition,* 2015.

Fig. 1: Examples of training and extrapolation test trials. Participants learn a visual ('bo', left: the boy witnesses the whole event) and a reportative ('di', middle: the boy doesn't witness the event, only reports it later to his friend) evidential marker. They then choose to extrapolate one of them to inferential evidence (right: the boy only witnesses the aftermath of the event).

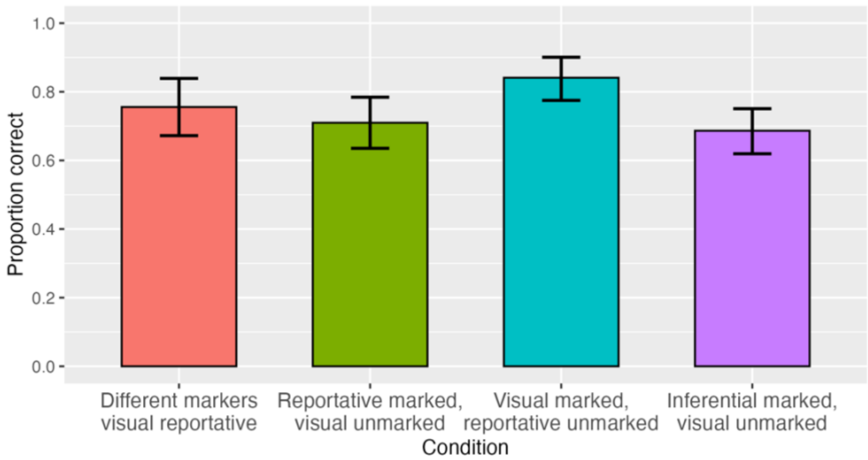


Fig. 2: Learning results by condition with mean accuracy on marking choices for the two familiar evidence types, and 95% confidence intervals.

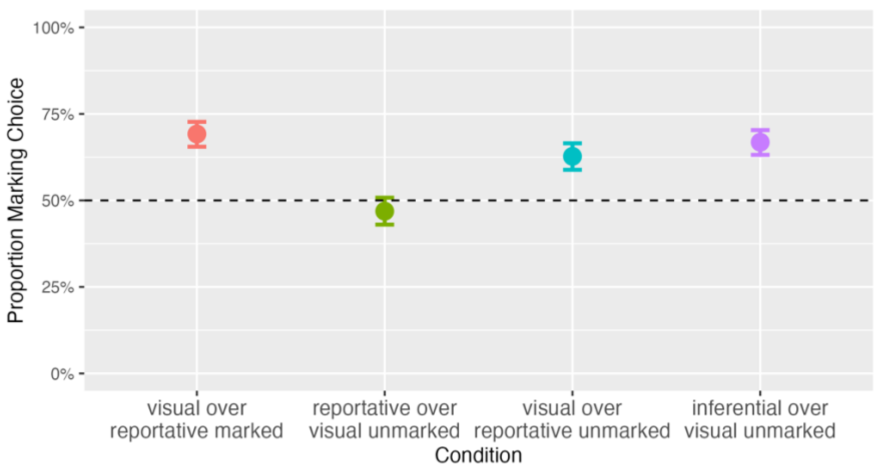


Fig. 3: Extrapolation results by condition with mean proportions and 95% binomial confidence intervals.

Neural speech tracking in a bilingual cocktail party: Does language identity matter?

Maroš Filip[1,2], Kateřina Chládková[1,2]
[1]Institute of Psychology, Czech Academy of Sciences; [2]Faculty of Arts, Charles University
filip@praha.psu.cas.cz

Background: What enables listeners to selectively attend to and comprehend a single talker in a multitalker environment? Research suggests that the brain tracks most robustly and accurately the attended speech stream (target) while attenuating its tracking of the other, unattended, speech streams (maskers; Kaufman & Golumbic, 2023). The robustness and accuracy of neural speech tracking can be measured with the EEG (electroencephalography) as the brain's oscillatory activity that aligns with the attended speech stream.
The aim of this study is to examine whether language similarity and its familiarity to the listener influence the accuracy of neural speech tracking in a multilingual cocktail party environment. We consider the two types of masking proposed in the literature: (1) acoustic masking, which occurs when the acoustic properties of the attended and unattended speech streams overlap, and (2) informational masking, which arises when the messages conveyed by the attended and unattended streams interfere. Acoustic masking should be stronger when the target and masker are identical or similar languages. Informational masking should be stronger when both the target and masker languages are familiar to the listener and reduced when the masker language is unfamiliar.
Objectives: The premise is that the brain will track the attended speech stream more robustly than the unattended speech stream. The hypothesis addressing acoustic masking is that the more dissimilar the masker language is from the target language, the better is the tracking of the target speech. The hypothesis addressing information masking is that tracking of the target will be better in cases of unfamiliar masker. To test these hypotheses, we employ various language combinations Czech (L1, target and masker), English (proficient L2, target or masker), and Dutch (unfamiliar to listeners, masker); see Table 1. Note that English and Dutch were selected because they are rhythmical very similar (i.e. have same status of acoustic masking, White et al., 2012) but differ information-wise (familiar vs. unfamiliar to listeners).
Method: The materials are 25-s podcast clips, narrated by male speakers. Participants (intended n = 40) are young, normal-hearing adults, native speakers of Czech with good proficiency in English (self-rated as B2 or better, according to the CEFR). Before the experiment, participants' L2 proficiency is assessed using LEXtale (Lemhöfer & Broersma, 2012). During the experiment, participants listen dichotically to two different podcasts and will be instructed to focus their attention only on one. Each participant will be exposed to four different language combinations (Table 1). EEG will be recorded throughout the experiment. MATLAB and its mTRF toolbox (Crosse et al., 2016) will be used to fit the temporal response function (TRF), a linear model that describes how stimulus features influence the neural response.
Predictions: We predict that the L1 target will be tracked more robustly when the masker is a foreign language compared to when the masker is L1 (stronger acoustic and informational masking). In a masker environment of L2 (English), participants with lower L2 proficiency, experiencing less informational masking, should track their L1 more effectively than those with higher L2 proficiency. If L1 is the masker, neural tracking of target will be more robust if target is an L2 (less informational masking) than if target is an L1 (more informational masking). Data collection is ongoing; results will be presented at the conference.

TABLE 1. *Language Combinations in the Experiment*

Target Language	Masker Language	Masker Type
Czech (L1)	Czech (L1)	Acoustic and informational
Czech (L1)	English (L2)	Informational (proficiency depending)
Czech (L1)	Dutch (Unknown)	None
English (L2)	Czech (L1)	Informational (proficiency depending)

*Notes***:** Each participant will listen to these podcast pairs in a different order, with the target ear also counterbalanced across participants.

References

Crosse, M. J., G. M. Di Liberto, A. Bednar, and E. C. Lalor. 2016. "The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli." *Frontiers in Human Neuroscience* 10: 604.

Kaufman, M., and E. Zion Golumbic. 2023. "Listening to Two Speakers: Capacity and Tradeoffs in Neural Speech Tracking during Selective and Distributed Attention." *NeuroImage* 270: 119984. https://doi.org/10.1016/j.neuroimage.2023.119984.

Lemhöfer, K., and M. Broersma. 2012. "Introducing LexTALE: A Quick and Valid Lexical Test for Advanced Learners of English." *Behavior Research Methods* 44: 325–43.

White, L., Mattys, S. L., and L. Wiget. 2012. "Language categorization by adults is based on sensitivity to durational cues, not rhythm class." *Journal of Memory and Language* 66(4): 665-679.

Headnods don’t always mean ‘yes’: Ambiguity in gestural responses to negative questions

Sebastian Walter, Lennart Fritzsche
Goethe University Frankfurt
s.walter@em.uni-frankfurt.de

Background: German exhibits a threefold response particle system: *ja* (‘yes’) and *nein* (‘no’) can be used to affirm negative antecedents, and the specialized particle *doch* must be used to reject negative antecedents [1, 2]. In addition to verbal response particles, head gestures, i.e., nods and shakes, can accompany (= co-speech) or substitute them entirely (= pro-speech) [3]. This paper exclusively focuses on pro-speech uses of gestural response elements. Previous research suggests that headshakes serve the same purpose as verbal *nein*, while headnods correspond to *ja* [4]. However, in co-speech cases *doch* often occur with headnods [4], suggesting that headnods may serve a (secondary) *doch*-like function, as well.

Method: We conducted two single-trial forced choice experiments in German testing to which extent headnods can take on a *doch*-like function. After being presented a written context licensing a negative polar question, participants viewed video clips where speaker A uttered the negative polar question *Sind meine Schuhe nicht in deinem Kleiderschrank?* (‘Are my shoes not in your closet?’), followed by a response of speaker B realized either through spoken or gestural response elements. To indicate how they interpeted the response element, they selected one of two images representing either affirmation or rejection of the negative antecedent (cf. Fig. 1). Experiment I (participants: n = 592) tested four conditions: *ja*, *nein*, and pro-speech headnods and headshakes (each repeated three times). The gestures were kept low in amplitude (HEADNOD-small and HEADSHAKE-small) to represent neutral, typical instances. Experiment II (participants: n = 750) investigated whether modifying headnods would boost a *doch*-like interpretation. It introduced five response conditions: spoken *ja* and *doch*, a standard headnod (HEADNOD-small), a larger amplitude headnod (HEADNOD-large), and a nod accompanied by brow furrowing (HEADNOD-furrowing). Brow furrowing was selected based on prior findings associating it with rejection and increased gestural emphasis [5].

Results: In Experiment I, headshakes patterned with spoken *nein*, while headnods significantly diverged from spoken *ja* (cf. choice proportions in Fig. 2), suggesting headnods do not exclusively convey affirmation. They can also reject a negative antecedent and thus take on a function akin to *doch*, in line with findings from its co-speech uses [4]. Experiment II replicated the intermediate status of headnods, with HEADNOD-small differing significantly from both *ja* and *doch* (cf. choice proportions in Fig. 2). Contrary to expectations, neither HEADNOD-large nor HEADNOD-furrowing differed significantly from HEADNOD-small, although the furrowing condition showed a descriptive trend in the predicted direction. This suggests that gesture enhancements may have been too subtle or insufficiently salient to override the default ambiguity of the nod gesture.

Discussion: Our findings support the idea that pro-speech head gestures can reflect verbal response particles, with headnods displaying an ambiguous status between *ja* and *doch* in German. The failure to significantly strengthen the rejecting interpretation through gesture modification points to a high threshold for visual marking when rejecting negative antecedents, possibly due to the backgrounded status of gestures in spoken language (e.g., [6]). Future research will explore whether combining multiple visual cues (e.g., frequency, amplitude, brow furrowing) or using more naturalistic gesture production can more reliably elicit a *doch*-like interpretation.

References

[1] Blühdorn, H. (2012). *Negation im Deutschen: Syntax, Informationsstruktur, Semantik* (Vol. 48). Narr Francke Attempto Verlag.

[2] Claus, B., Meijer, A. M., Repp, S., & Krifka, M. (2017). Puzzling response particles: An experimental study on the german answering system. *Semantics and Pragmatics*, *10*(19), 1. https://doi.org/10.3765/sp.10.19

[3] Kendon, A. (2002). Some uses of the head shake. *Gesture*, *2*(2), 147–182.

[4] Loos, C., & Repp, S. (to appear). Beyond 'yes' and 'no': Multimodal multilexical polar responses in German. *Proceedings of Sinn und Bedeutung 29*.

[5] González-Fuente, S., Tubau, S., Espinal, M. T., & Prieto, P. (2015). Is there a universal answering strategy for rejecting negative propositions? Typological evidence on the use of prosody and gesture. *Frontiers in Psychology*, *6*, 899.

[6] Ebert, C., Ebert, C., & Hörnig, R. (2020). Demonstratives as dimension shifters. *Proceedings of Sinn und Bedeutung 24*, 161–178. https://doi.org/10.18148/sub/2020.v24i1.859

Figures

Figure 1: Forced choice options (left: affirmation of negative antecedent; right: rejection of negative antecedent).

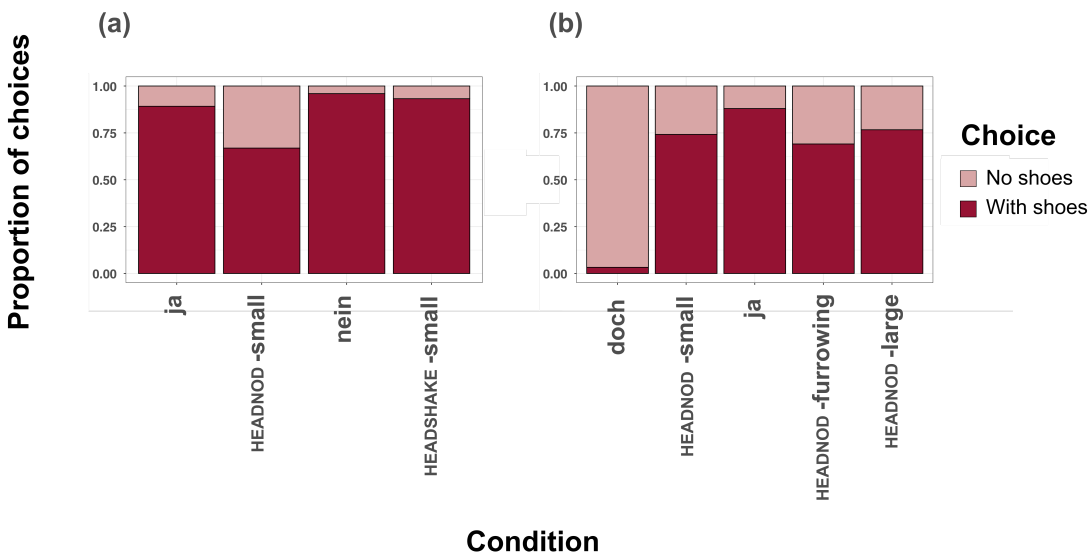


Figure 2: Proportion of choices for each condition in Experiment I (a) and Experiment II (b).

Cross-linguistic structural priming of reciprocal innovations in French-English bilinguals

Foteini Karkaletsou[1], Gunnar Jacob[2], Shanley E. M. Allen[1]
[1] University of Kaiserslautern-Landau; [2] University of Mannheim
karkalet@rptu.de

Cross-linguistic structural priming in bilinguals (i.e., the observation that exposure to a syntactic structure in one language facilitates subsequent processing or production of a similar structure in another language) has been suggested as a key psycholinguistic mechanism underlying contact-induced language change in language-contact settings [1,3]. Specifically, for grammatical structures which exist in only one language of a bilingual, cross-linguistic priming may lead to the activation (and eventually also the production) of innovative structures in the other language. Thus, cross-linguistic innovation priming may constitute a potential pathway which allows L1-specific structures to enter a bilingual's L2.

The present study investigates cross-linguistic priming in the processing and production of French structural innovations in Canadian French-English bilinguals. Unlike previous work, our study focuses on a single contact setting, namely Canada. The innovations tested were French reciprocal structures in which the obligatory reflexive *s(e)* was missing *(*Ils (s') embrassent dans la gare* 'They kiss in the train station'), a structure which is ungrammatical in French, but possible in English. This structure was selected because Canadian bilinguals have been shown to accept such French innovations more upon repeated exposure [2].

In a self-paced reading experiment in Gorilla, seventy-three French-English Canadian bilingual adults read innovative French target sentences such as (2), which were preceded by either a structurally similar, grammatical English prime (1a), a structurally different but otherwise similar grammatical control prime (1b), or an ungrammatical control prime (1c). To control for adaptation to the innovative French structure during the experimental session, stimuli were presented across two blocks, with effects determined separately for each block. As a measure of whether repeated exposure to the innovation during the experiment leads to increased production of such innovations, participants also completed written production tasks before (pretest) and after (posttest) the experiment. If it is possible to prime innovations across languages, English primes such as (1a) should overall facilitate the processing of innovative French structures (2) compared to English primes such as (1b) and (1c).

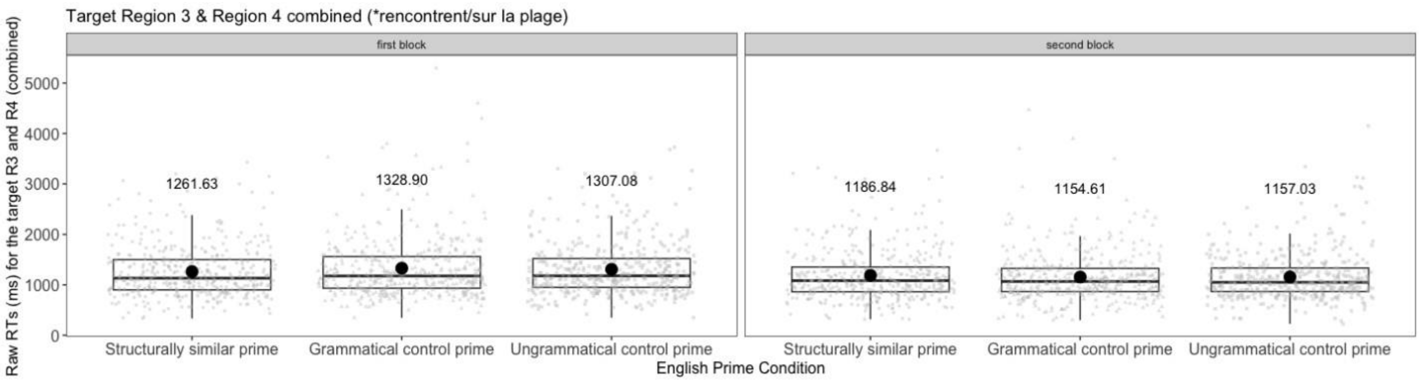
Reading times for the critical regions 3 and 4 (*rencontrent* / *sur la plage*) revealed significant interactions between 'Prime Type' and 'Block', with significant innovation priming in Block 1, but no priming in Block 2 (Figure 1). In addition, the results also showed a significant main effect of 'Block', with faster reading times in Block 2 than Block 1, suggesting that participants adapted to the innovative structure during the experiment. Moreover, although the participants produced numerically more innovations in the posttest (57) than in the pretest (36), this difference was not significant.

Overall, our findings suggest that cross-linguistic priming can also occur for innovative structures which do not exist in the target language yet. The fact that priming took place only in self-paced reading indicates that speakers might rely on similar input from the other language to process innovations, but probably only at initial stages of encountering such structures. Innovation priming in production may have to overcome a participant's reluctance to produce ungrammatical sentences, which may require extensive cumulative exposure over an extended time period.

Table 1. Set of conditions for the English primes and the French targets in the self-paced reading task. The slashes indicate regions of presentation.

English primes	***Condition***
(1a) The writer / and the editor / meet / at the cafe / in the city.	Bare main verb (structurally similar prime)
(1b) The writer / and the editor / meet each other / at the cafe / in the city.	Complementation *each other* on the main verb (grammatical control prime)
(1c) The writer / and the editor / *each other meet / at the cafe / in the city.	Complementation *each other* on the main verb (ungrammatical control prime)
French target	***Condition***
(2) Le sculpteur / et le sorcier / *rencontrent / sur la plage / dans la nuit. 'The sculptor / and the sorcerer / meet / at the square / in the night.'	Bare main verb (no preverbal *se,* ungrammatical)

Figure 1. Distribution of raw reading times (ms) for target regions 3 and 4 combined per block of presentation (first vs. second) and English prime condition. The black dot within each box indicates the mean reading time per condition per block (also provided as text in the plot).



References

[1] Fernández, E. M., De Souza, R. A., & Carando, A. (2017). Bilingual innovations: Experimental evidence offers clues regarding the psycholinguistics of language change. *Bilingualism: Language and Cognition*, *20*(2), 251-268.

[2] Karkaletsou, F., Kholodova, A., & Allen, Shanley E. M. (2024). Exploring bilingual adaptation to structural innovations: Evidence from Canadian French. *Languages, 9*(12), 375.

[3] Kootstra, G. J., & Şahin, H. (2018). Crosslinguistic structural priming as a mechanism of contact-induced language change: Evidence from Papiamento-Dutch bilinguals in Aruba and the Netherlands. *Language*, *94*(4), 902-930.

Between conflict and causality: the connective “but” in discourse processing and recall

Clara Seyfried & Yuki Kamide
University of Dundee
c.seyfried@dundee.ac.uk

Background: Discourse connectives (e.g., “and” (additive), “because” (causal), “then” (temporal)) have long been considered the most distinctive markers of discourse coherence [1]. Indeed, children’s understanding of connectives appears to be indicative of linguistic and non-linguistic academic abilities [2]. However, past research has frequently shown mixed results as to whether the presence of connectives influences discourse comprehension [3, 4, 5]. A possible explanation [6, 7] could be that effects of semantic relationships between discourse units are much larger than the effects of discourse connectives, as found for discourse processing and recall in our study presented at AMLaP 2024. For example, in “I got wet because it was raining”, the strength of the relationship between rain and getting wet likely has a much greater impact than the presence of “because”. Since *causality* has been found to be less often explicitly marked with a connective than other relations [8], suggesting that it is the default continuation relationship (“causality-as-default” [9]), our previous study measured the effect of the causal connective “because” at three levels of causal relatedness. Our findings demonstrated that “because” sped up discourse processing across all levels of causality, but particularly for strongly related items. In the present study, we now explore whether the effect of the contrastive discourse marker “but” mirrors that of the causal “because”, hypothesising that “but” might only facilitate discourse processing for relations that are very weakly causally related (i.e., standing in opposition/conflict). **Method:** Our online study systematically manipulates Connective Presence (presence / absence of “but”) and Causal Relatedness (strongly / moderately / weakly related) between propositions (see (a)). We use LME (reading times) and GLME (delayed recall rates) models to investigate the interplay between online processing and subsequent memory. Our 432 target items (see (a)) and 36 filler items (see (b)) were adapted from our previous study and further validated. Items were distributed across lists (72 targets per participant). Participants (final sample $N = 113$) completed a proposition-by-proposition self-paced reading task on *Gorilla* [10] wherein the first two displays contained the first and second proposition respectively, followed by both together in the third display in each trial. Every 18 trials, participants completed discourse fragments containing one complete and one incomplete proposition with a cue (see (c)). **Results:** Residual reading times on the second display indicate a main effect of Connective ($p = .03$), suggesting that “but” slowed down processing. On the third display, we observe a significant interaction between Causality and Connective ($p = .01$). Specifically, in the presence of “but”, weakly related items were read faster, whereas strongly causally related items were read more slowly (see (d)). In the absence of “but”, strongly causally related items were read the fastest. As in our previous study, we observe that strongly causally related and weakly related items were remembered better than moderately related ones (see (e), $p < .01$). **Discussion:** It has been previously observed that causal relations are processed easier than non-causal ones [11]. Our results suggest that while the possibility of causal interpretations might facilitate processing, congruence of discourse connective and relation type also matters. We find that the presence of the contrastive marker “but” can facilitate or inhibit processing depending on the relation it appears in. However, effects on memory recall continue to be driven by relation type only. Based on these results, we argue that future research should place greater emphasis on exploring the effects of different relation types.

(a) Example Target Item Set (**HC**: Highly Causal, **MC**: Moderately Causal, **WC**: Weakly Causal; **CP**: Connective Present, **CA**: Connective Absent)

(1) **HC / CP**: The teenager had drunk too much last night, but they had a hangover.

(2) **HC / CA:** The teenager had drunk too much last night. They had a hangover.

(3) **MC / CP**: The teenager had drunk only little last night, but they had a hangover.

(4) **MC / CA**: The teenager had drunk only little last night. They had a hangover.

(5) **WC / CP**: The teenager had drunk almost nothing last night, but they had a hangover.

(6) **WC / CA**: The teenager had drunk almost nothing last night. They had a hangover.

(b) Example Filler Item Set (Temporal Structure)

The position of the discourse connective "after" (first or second proposition) was balanced (18 each).

(1) **Connective Position 1:** After all the snow had melted, the week ended.

(2) **Connective Position 2:** Chao walked past Westminster Abbey after taking a selfie.

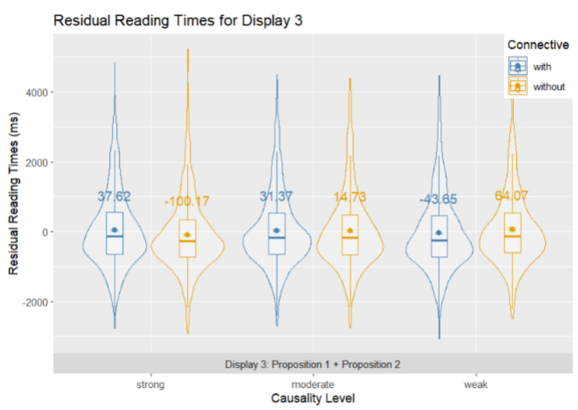
(c) Example Recall Tests

Out of 108 item sets, half had a gap to fill in the first proposition (e.g., (1)), and half in the second proposition (e.g., (2)). The complete proposition visible and the position of the gap were held constant across all 6 conditions for each item. The subject or first word of each clause served as a recall cue.

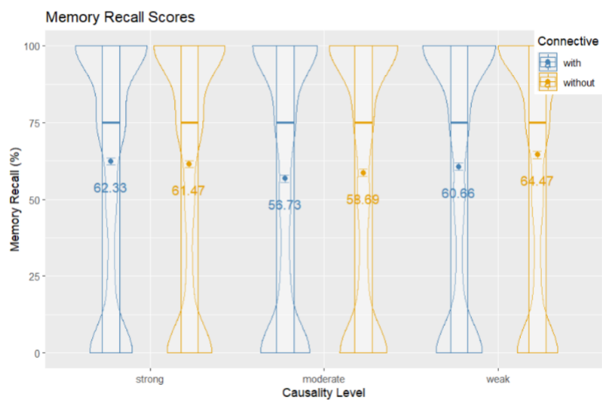
(1) **Gap Position 1:** The teenager ..., but they had a hangover.

(2) **Gap Position 2:** After all the snow had already melted, the week...

(d) Results Discourse Processing



(e) Results Memory Recall



References:

[1] Degand, L., N. and Bestgen, Y. (1999). 'The impact of connectives and anaphoric expressions on expository discourse comprehension'. *Document Design*, 1(1), pp. 39-51. **[2]** Volodina, A., Heppt, B. and Weinert, S. (2021). 'Relations between the comprehension of connectives and school performance in primary school'. *Learning and Instruction*, 74, 101430. **[3]** Degand, L., N. and Bestgen, Y. (1999). 'The impact of connectives and anaphoric expressions on expository discourse comprehension'. *Document Design*, 1(1), pp. 39-51. **[4]** Koda, N. (2008). Connective interference and facilitation: Do connectives really facilitate the understandig of discourse. *The Annual Reports of Graduate School of Arts and Letters*, *56*, 29-42. **[5]** Kleijn, S., Pander Maat, H. L. and Sanders, T. J. (2019). 'Comprehension effects of connectives across texts, readers, and coherence relations.' *Discourse Processes*, 56(5-6), pp. 447-464. **[6]** Kuperberg, G. R., Paczynski, M. and Ditman, T. (2011). 'Establishing causal coherence across sentences: An ERP study'. *Journal of Cognitive Neuroscience*, 23(5), pp. 1230-1246. **[7]** Myers, J. L., Shinjo, M. and Duffy, S. A. (1987). 'Degree of causal relatedness and memory. *Journal of Memory and Language*', 26(4), pp. 453-465. **[8]** Asr, F. T. and Demberg, V. 'Implicitness of discourse relations'. *Proceedings of COLING 2012.* Mumbai, pp. 2669-2684. **[9]** Sanders, T. (2005). 'Coherence, causality and cognitive complexity in discourse'. *Proceedings/Actes SEM-05, First International Symposium on the exploration and modelling of meaning* (pp. 105-114). Toulouse, pp. 105- 114 **[10]** Anwyl-Irvine, A.L., Dalmaijer, E.S., Hodges, N., Evershed, J.K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods.* **[11]** Köhne-Fuetterer, J., Drenhaus, H., Delogu, F., & Demberg, V. (2021). The online processing of causal and concessive discourse connectives. *Linguistics*, *59*(2), 417-448.

Visuospatial cognitive load disrupts predictive gaze behavior but not prediction

Christopher Allison[1,*], Falk Huettig[1,2], Thomas Lachmann[1]
[1]University of Kaiserslautern-Landau, [2]Max Planck Institute for Psycholinguistics
*Corresponding author: Christopher Allison, allison@rptu.de

Background: Simple semantic predictions are often assumed to be driven by more automatic processes such as spreading semantic activation. However, both visuospatial [1] and phonological [2] cognitive load manipulations have been shown to reduce predictive gaze behavior when processing simple semantic sentences in Visual World eye-tracking studies. This has led to questions about the automaticity of predictive language processing. It is unclear at what level (e.g., linguistic, visual, or decision) predictive gaze behavior is affected by increases in visuospatial or phonological cognitive load and whether these effects are mechanistically similar, domain-general effects of cognitive load or mechanistically distinct, domain-specific interferences.

Method: We conducted an EEG (n = 33) and a visual world eye-tracking experiment (n = 39) with a diverse group of moderate to highly proficient L2 English speakers. In both studies, participants heard semantically predictable or unpredictable sentences such as *The waiter/runner brings/remembers the plate*.

EEG: Participants completed a concurrent visuospatial working memory task (modified Corsi block task; remember 4 indicated squares out of a 9 square array). Participants saw 9 squares and in half of the trials 4 of them would be indicated. Participants then heard the sentence, selected the best fitting sentence, and, if previously indicated, recalled the squares in the order and location they were previously shown.

Eye-tracking: Participants completed blocks of Visual World trials in each of three conditions: Control, concurrent visuospatial (remember 4 squares as in the EEG experiment), or concurrent phonological (remember a list of 5 semantically/phonologically unrelated 2-syllable words).

Results: In the EEG study, we used a Bayesian ANOVA to analyze the N400 over the region of interest from 300-500 ms. We found overwhelming evidence ***for*** the effect of predictability (BF_{10}>1000) and moderate evidence for an ***absence*** of an effect of visuospatial cognitive load (BF_1 = 7.5782). In the eye-tracking study, a linear mixed effect model shows significant effects of predictability, visuospatial load, and phonological load, with both load conditions leading to reduced predictive gaze behavior. A divergence point analysis (Figure 2) suggests that the onset of prediction was delayed only in the concurrent phonological condition.

Discussion: In experiment 1, we found evidence against the effect of visuospatial cognitive load on N400 amplitude. In experiment 2, we found that the same visuospatial manipulation significantly reduced predictive gaze behavior. However, the *onset* of predictive gaze behavior is similar to the control condition. These findings suggest that visuospatial load does not affect the preactivation of a target word, as the N400 is not impacted (Exp1) and listeners can make predictive looks at the same time as a control condition (Exp2). Instead, the disruptions seen in predictive gaze behavior are more likely to be occurring at the integrative decision level, with increased visuospatial cognitive load reducing the strength of the spatial representation of the objects in the visual array. Furthermore, the results of experiment 2 suggest that visuospatial cognitive load is likely to be mechanistically distinct from phonological cognitive load, as evidenced by the much later onset of predictions in the phonological load manipulation. Together, these findings suggest specific interference mechanisms at multiple levels as opposed to a more domain-general effect of cognitive load and that disruptions in predictive gaze behavior may be disruptions in *gaze behavior* but not *prediction*.

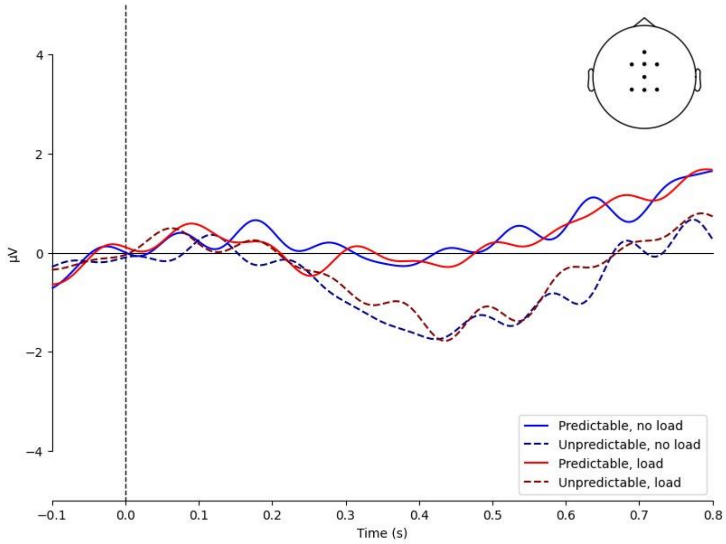


Figure 1. N400 event related potentials evoked upon hearing semantically predictable (solid lines) or unpredictable (dotted lines) sentences either with (red lines) or without (blue lines) a concurrent visuospatial cognitive load. In the visuospatial cognitive load condition, participants had to memorize the order and location of four squares before hearing the sentence.

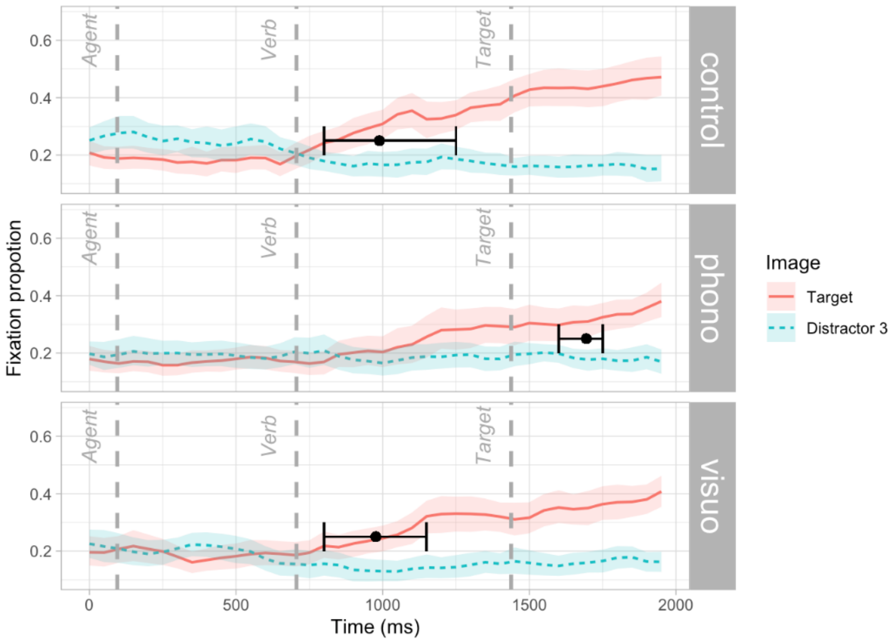


Figure 2. Divergence point analysis and fixation proportions for participants listening to semantically predictable sentences. Participants completed each of three experimental blocks, a control block consisting of standard visual world trials, a phonological load block in which participants had to remember a list of five words, and a visuospatial load block in which participants had to remember the order and location of four squares.

References

[1] Allison, C., Huettig, F., Fernandez, L., & Lachmann, T. (2025). Visuospatial Working Memory Load Reduces Semantic Prediction in the Visual World. Language, Cognition, and Neuroscience, https://doi.org/10.1080/23273798.2025.2522272

[2] Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. Bilingualism: Language and Cognition, 21(2), 251–264. https://doi.org/10.1017/s1366728917000050

Integrating language model embeddings into ACT-R

Maryam Meghdadi, John Duff, Vera Demberg
Department of Language Science and Technology, Saarland University
maryam.meghdadi97@gmail.com

Lexical decision tasks exhibit **associative priming** [1]: exposure to a word *p* leads to shorter response times (RTs) on a subsequent target word *t* if *p* and *t* have associated meanings. This is often explained via automatic spreading activation from *p*'s representation to *t*'s, as a function of their similarity. One principled way to define gradient similarity uses the cosine between multi-dimensional embeddings learned in language modeling tasks (CoSim) [2-5]. E.g. [2] show that a slope for CoSim improves a regression's predictions for target RTs in [6]'s large dataset. Still, these regressions do not provide a model of priming, or capture more complicated patterns, e.g. that similarity has a larger effect for low-frequency words [7]. We go further here, showing that we can incorporate CoSim within a simple mechanistic model of lexical decision in the ACT-R framework [8-9]. ACT-R's relationships between CoSim-based similarity, past use, and retrieval speed can then successfully predict observed RT patterns.

ACT-R models simulate processing as a sequence of symbolic operations in pre-defined modules: e.g. [9-10] simulate lexical decision as a sequence of visual uptake, retrieval of a chunk from memory (the target's lexical representation), and motor execution. The ACT-R retrieval time of chunk *i* is an inverse exponential function of *i*'s activation (Eq 1), itself determined (Eq 2a) by *base-level activation* B_i reflecting frequency and recency of past use, plus *spreading activation* reflecting similarity S_{ji} to chunks *j* in current attention. This latter component can drive priming effects, but to model them in detail, we require a scalable, gradient function for similarity in meaning, instead of ACT-R's native similarity function relying on discrete matching between hand-coded features (animacy, number, etc.); this is where we use CoSim.

Method. To [9]'s model, we add S_{ji} values within the range $[0, \varphi]$, where φ is a free scaling factor, and S_{ji} approaches φ as CoSim increases (Eq 2b). We conducted initial comparisons to identify the best word embeddings suitable for our task, observing that priming effects are best correlated with CoSim computed over word2vec [11] and layer-0 BERT [12] embeddings extracted in isolation, consistent with [5].

Simulations. We use RTs from a primed lexical decision task with 200ms SOA (Fig 1), taken from [6], including 6331 unique prime-target pairs across a variety of relations. We conduct simulations with `pyactr` [9] and evaluate the Spearman rank correlation (ρ) between predicted and observed RTs (Table 1). Models 1-2 simulated RTs from activation values based on only CoSim/S_{ji}, M3 based on only frequency/B_i, and M4-5 based on both components in Eq 2A. The latter models correlate the best with human RTs, and produce the similarity/frequency interaction observed in the human data (Fig 2).

Discussion. Incorporating language model embeddings into ACT-R, while treating spreading activation and frequency as joint components of activation, allows us to model associative priming in a fixed, scalable, and interpretable way. We intend this as a proof of concept, and a first step. Future work may extend this model for other similar tasks [3], or expand it to model potential directional asymmetries in spreading activation via fan effects, or participant-specific associations using fine-tuned language models.

1. $T_i = Fe^{-fA_i}$

Equation 1. Retrieval time of chunk i given activation A_i. In our simulations, we set free parameters F and f to approx 0.36 following fits in [9].

2A. $A_i = B_i + \sum_{j=1}^{m} S_{ji} + \epsilon$

Equation 2A. B_i is the base-level activation of chunk i, estimated from frequency following [9]. S_{ji} is the strength of association between i and some chunk j, which is summed across all m chunks in attention. ϵ is noise (here, set to zero). In our model, i is the target, and we consider only the single prime as j.

2B. $S_{ji} = \varphi * max\{0, cos(Vi, Vj)\}$

Equation 2B. Our adapted equation for S_{ji}, which is sensitive to the CoSim of prime and target word embeddings (V_i, V_j). φ was fit by grid search. Note: for pairs with CoSim < 0 (very few), $S_{ji} = 0$.

Table 1. Spearman correlation between observed RTs and ACT-R model predictions based on CoSim from word2vec (w2v) and BERT, and frequency. Frequency was estimated using `wordfreq` [13]. Applying William's T2, we note that **M4** significantly outperforms **M3** ($p < 0.001$).

ACT-R Model	**ρ**
M1. Cosine (w2v, φ=*1.0*)	*0.20*
M2. Cosine (BERT, φ=*1.0*)	*0.21*
M3. Freq	*0.41*
M4. Freq + Cosine (w2v, φ=*4.0*)	***0.45***
M5. Freq + Cosine (BERT, φ=*3.0*)	*0.43*

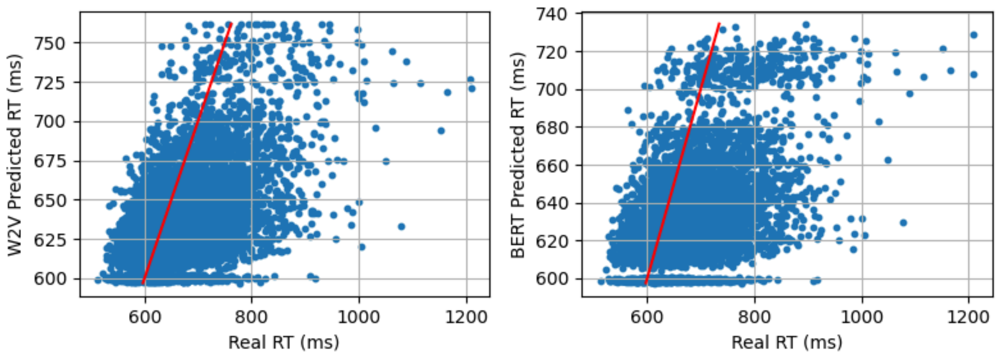


Figure 1. Observed vs. ACT-R-Predicted RTs using w2v (M4) and BERT (M5) CoSims. Unit slope in red.

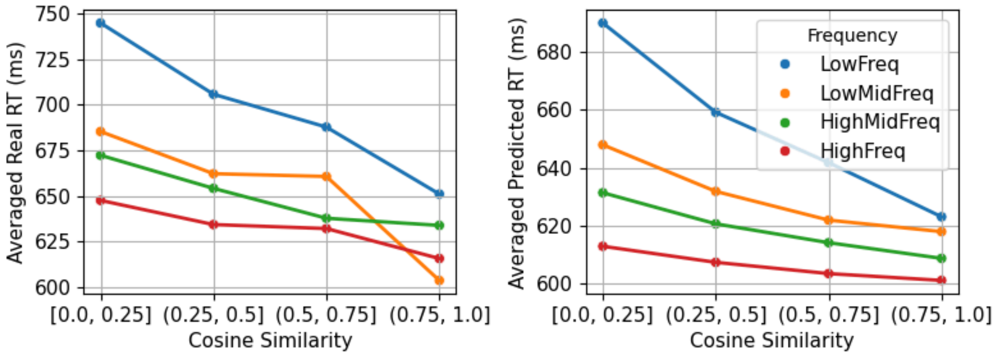


Figure 2. Observed and ACT-R-predicted RTs using w2v (M4), averaged over CoSim and freq. bins.

References. [1] Traxler & Tooley (2012) *Psychology of priming*, [2] Mandera et al. (2017) *JML*, [3] Smith and Vasishth (2020) *Cogn Sci*, [4] Vulić et al. (2020) *Proc EMNLP*, [5] Cassani et al. (2023) PsyArXiv b45ys, [6] Hutchison et al. (2013) *Behav Res Meth*, [7] Becker (1979) *JEP: HPP*, [8] Anderson et al. (2004) *Psychol Rev*, [9] Brasoveanu & Dotlačil (2020) Springer, [10] van Rijn & Anderson (2003) In *Proc ICCM 5*, [11] Mikolov et al. (2013) arXiv, [12] Devlin et al. (2018) arXiv, [13] Speer (2022) Zenodo.

Contrastive Prosody and Pragmatic Meaning: Evidence from Korean L2 Speakers of English

Suzy Park[1]
[1] Yonsei University
suzypark@yonsei.ac.kr

Background: In L2 processing, the interface between linguistic and language-external domains is known to be inherently challenging, due to incomplete representations or cognitive burdens in integrating various sources of information [1]. Despite ample evidence from L1 literature suggesting that prosodic cues in English can be used to derive pragmatic meaning [2,3,4], whether L2 speakers are also able to map prosodic information onto pragmatic inferences remains an open question. In a series of two experiments, this study examines whether L1-Korean learners of English can select the intended referent more effectively by utilizing the contrastive pitch accent (Experiment 1), or by combining the contrastive pitch accent and the rising boundary tone as a composite prosodic cue (Experiment 2).

Experiment 1: Experiment 1 tested whether L2 speakers were sensitive to the function of the L+H* accent in evoking a set of contrastive alternatives in the given discourse context. Fifty-five L1-Korean learners and 32 L1-English speakers participated in a picture selection task with pairs of instructions (*Click on the silver curtains. Now click on the CHERRY curtains*). There were four experimental conditions, which varied by location of the pitch accent (noun/adjective) and whether the pitch accent fell on contrastive or repeated information (felicitous/infelicitous). In contrast to the patterns of Mandarin L1 speakers found in previous literature [5], even highly proficient Korean speakers did not exhibit any facilitation effect on RTs for felicitous conditions, unlike L1 speakers who responded more quickly when the pitch accent was used felicitously. This supports the idea that the mapping between pitch accents and pragmatic processing remains to be a challenge for L2 speakers, especially when the prosodic element is absent in their L1.

Experiment 2: Experiment 2 investigated whether L2 speakers can infer that the intended meaning is the negated form of the utterance when the L+H* accent and a rising boundary tone are used together as a composite prosodic cue. For example, upon hearing "It LOOKS like a puppy," with a rising boundary tone at the end, listeners assume that *it resembles a puppy, but it isn't actually one*. Using *It looks like X* constructions from [6], 55 Korean learners and 32 L1 speakers performed a binary picture task where they heard sentences either with a Noun Focus (NF) or Verb Focus (VF) and were asked to select the picture that corresponds to the sentence. Contrary to previous findings, L1 speakers did not demonstrate any difference between focus types, while the proportion of non-prototypical image selection was higher in the VF condition for L2 speakers. Possible reasons for such unexpected patterns may be due to L2 speakers' heavy reliance on boundary tone due to the prosodic system in their L1, and differences in task design.

Discussion: The current findings suggest that although the prosody-pragmatics interface is difficult for L2 speakers, their sensitivity may be affected by various factors, such as type and location of the prosodic cue. Although more research is needed to determine the cause for L1 speakers' lack of reliance on the composite prosodic cue, such patterns indicate the need to consider the complexities of prosodic elements and pragmatic factors in L2 processing.

References

[1] Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic approaches to bilingualism*, *1*(1), 1-33.

[2] Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language,* 47(2), 292-314.

[3] Dennison, H., & Schafer, A. J. (2017). Processing intonationally implicated contrast versus negation in American English. *Language and Speech*, 60(2), 174-199.

[4] Tomlinson Jr, J. M., Gotzner, N., & Bott, L. (2017). Intonation and pragmatic enrichment: How intonation constrains ad hoc scalar inferences. *Language and Speech*, 60(2), 200-223.

[5] Takahashi, C., Kao, S., Baek, H., Yeung, A. H., Hwang, J., & Broselow, E. (2018). Native and non-native speaker processing and production of contrastive focus prosody. *Proceedings of the Linguistic Society of America*, 3(1), 35:1–13. https://doi.org/10.3765/plsa.v3i1.4323

[6] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn’t it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335-342.

Language and Line Dancing: the Role of Linguistic Labels in Action Learning

Rosa Zaaijer[1], Monique Flecken[1], Marieke Schouwstra[1], Caitlin Meyer[2].
University of Amsterdam, Linguistics[1], Dutch Linguistics[2].
r.p.zaaijer@uva.nl; m.e.p.flecken@uva.nl; m.schouwstra@uva.nl; c.m.meyer@uva.nl.

When learning something new, the presence of an accompanying linguistic label can facilitate forming categories and activating category knowledge. This 'label advantage' effect [1-3] occurs when labels (novel words) accompany novel objects [4], tactile patterns [5], or odor categories [6], during the learning process. However, the effect of novel linguistic labels on how we learn entire events or actions remains unexplored. No previous research exists on the potential effects of novel labels on learning how to *carry out*, and not just recognize, novel actions. This study therefore investigates whether the presence of novel linguistic labels during learning facilitates both the memorization and performance of novel actions.

We conducted a large-scale learning experiment at Lowlands music festival in the Netherlands in which participants (n=297, ages 18 - 55) tried to learn four novel line dance move sequences, after which they did a performance and a memory test. All participants were native Dutch speakers, and as requested by the festival organization, harsh exclusion criteria were avoided so that the majority of festival attendees could participate in this study. They partook in one of three conditions: a novel word label condition (n=105), a hand gesture label condition (n=97) or a no-label control condition (n=95). By comparing novel word labels and (non-linguistic, visual only) gesture labels, we aimed to test whether specifically the linguistic nature of the novel word labels causes the label advantage effect, or whether the process of labelling itself is sufficient. First, participants went through a three-stage learning phase (exposure, practice & final rehearsal) involving an instruction video in which a dance instructor demonstrated the dance sequences and gave instructions to practice. Participants in the novel word label condition received a novel linguistic label to accompany each dance move sequence (*GAM, ZOL, GOL* and *ZAM*, controlled for phonetic frequency). Participants in the gesture label condition received pictures of different non-iconic hand gestures (e.g., a balled fist with a pointed thumb), and the no-label condition participants received no labels whatsoever. After the learning phase, participants completed a performance test where they acted out each dance sequence and were scored live on their performance. They were scored on whether they performed the correct movement, and on how well they performed the movement. After the performance test, participants completed a memory test. The memory test asked questions about the order of individual movements within the dance sequences, and questions that tapped into participants' memory of the accompanying labels. Only participants in the novel word label and hand gesture label conditions were asked to fill out this second part of the memory test.

Participants reached an average accuracy of 64.2% in the performance test and 61.3% in the memory test, which is quite high. Contrary to our expectations, however, we found that the participants who received no accompanying labels at all scored significantly *higher* than participants who received novel word labels or gesture labels on both the performance test (Figure 1) (β=0.94, SE=0.21, z=4.49, $p<.001$) and the memory test (Figure 2) (β=0.97, SE=0.30, z=3.28, $p<.05$). There were no differences between the novel word and the gesture label conditions. We reconcile these findings with those in previous studies by reasoning that (i) learning to categorize static objects is different from learning dynamic action sequences, and hence the presence of (noun-like) artificial labels also functions differently. Moreover, for dynamic concepts, different types of labels may be more appropriate, such as novel labels that are clearly verbal in nature (e.g., *ZAMMING* rather than *ZAM*). Second, learning how to actually *carry out* novel action sequences may require a different learning strategy than just learning to perceptually categorize or discriminate novel actions. These findings contribute to our understanding of novel action learning, allowing us to further investigate what it is about the nature of actions that causes this unexpected lack of the label advantage effect.

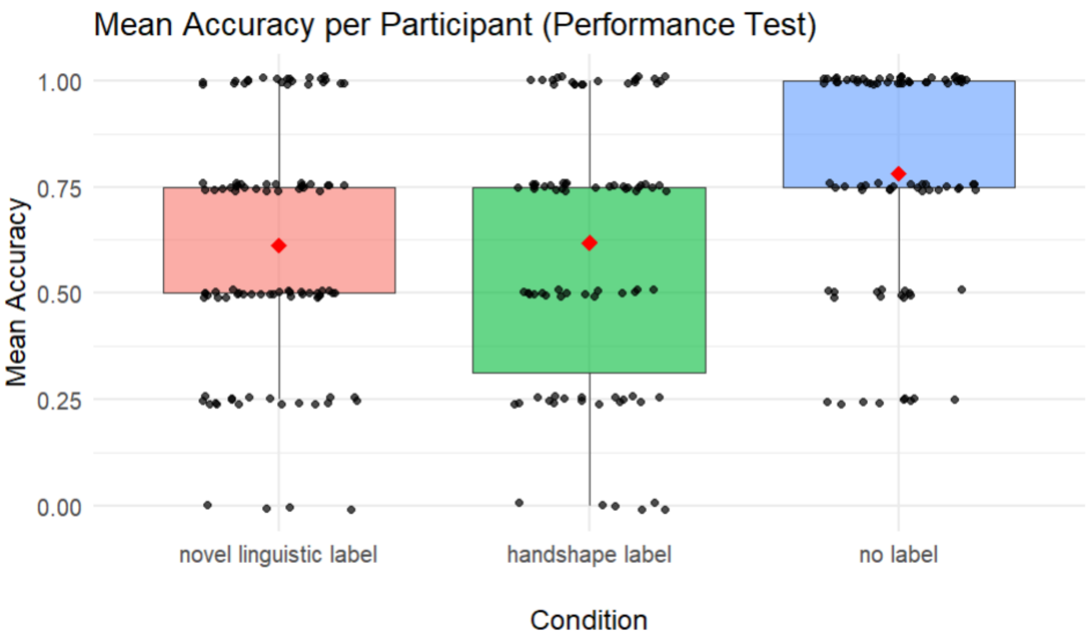


Figure 1: Accuracy in performance of dance sequences

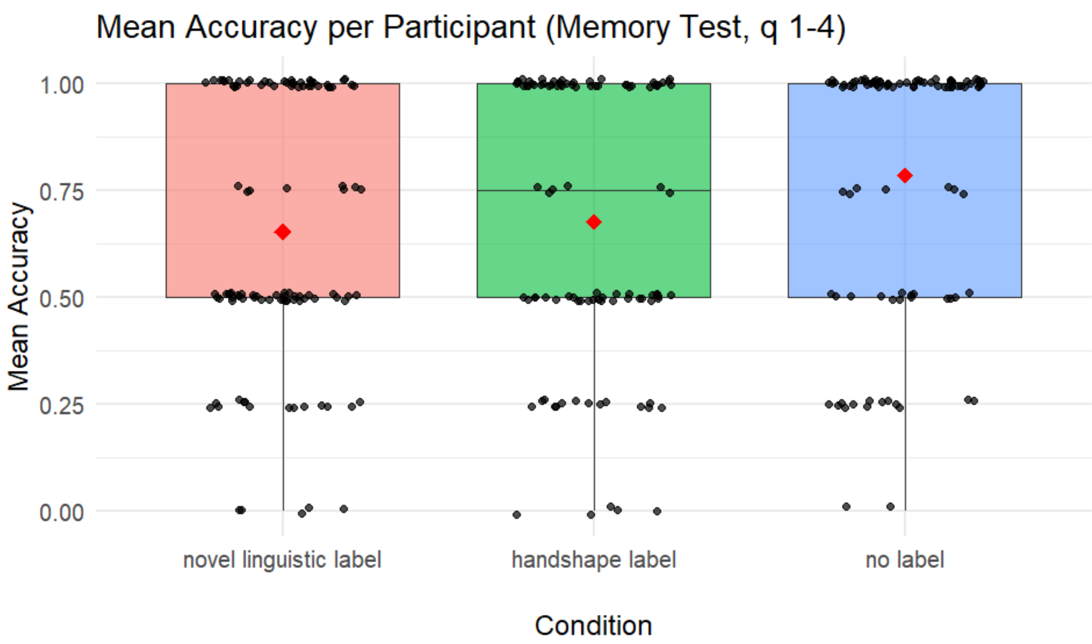


Figure 2: Accuracy in memory of dance sequences

References
[1] Rissman, L., & Lupyan, G. (2024). Words do not just label concepts: activating superordinate categories through labels, lists, and definitions. *Language, Cognition and Neuroscience*. **[2] Boutonnet, B., & Lupyan, G. (2015)**. Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*.**[3] Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007).** Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories. *Psychological Science*.**[4] LaTourrette, A. S., & Waxman, S. R. (2020).** Naming guides how 12-month-old infants encode and remember objects. *P Proc. Natl. Acad. Sci. U.S.A.* **[5] Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018).** Verbal labels facilitate tactile perception. *Cognition*. **[6] Vanek, N., Sóskuthy, M., & Majid, A. (2021).** Consistent verbal labels promote odor category learning. *Cognition*.

Parsing Effect of Structural Prediction in Sentences with Code-Switching

Marina Sokolova
University of Lisbon
marinasokolova@edu.ulisboa.pt, msokolopez@gmail.com

Background: The project investigates how a structural prediction based on the selectional properties of the matrix verb influences subsequent parsing of ambiguous relative clauses (RCs) with code-switching. The design and RQs are based on experimental evidence that a perception verb favors the eventive complement (or pseudo-relative) over the entity one (DP+RC) [1], [3], [4]. It triggers a structural prediction for the clause complement, where the complex DP is the grammatical subject (1b). In this modification, the lower NP cannot be the doer of the activity expressed by the VP. As a result, people prefer HA in sentences with a perception verb even when they are presented with restrictive RCs (1a). Alternatively, CS can impose a prosodic break at the place of language change (2), and RC parsing can be performed by the prosodic structure of the sentence with CS: a prosodic break before 'that' favors HA, a break after the first head noun favors LA [2].

Method: Adult Norwegian-English, Russian-Norwegian, Portuguese-English, Spanish-English and Catalan-English speakers, advanced in their L2, participated in three SPR experiments reading sentences (one word at a time) and comprehension questions (two answer choices) on a computer screen. Exp.1 tested the parsing effect of a structural prediction in English (L1/L2) in sentences with full and reduced RCs (p.3). Exp.2 elicited a parsing effect of CS. Exp.3 checked what has a stronger effect on RC parsing, a structural prediction or a place of CS. CS in Exp.1 and 3 was kept constant L2 -> L1: English->Catalan, English->Portuguese, English->Norwegian, Norwegian->Russian.

Results: Exp.1: Reduced RCs, globally ambiguous to the eventive SC (p.3) return HA and longer response time. Full RCs demonstrate LA and a tendency to slow down the RT at the embedded verb, a place where the prediction for the eventive complement can be discarded. Exp.2: Only L2ers speaking HA+LA languages show light sensitivity to the place of CS in RC resolution. CS does not increase the processing load during sentence parsing. Exp.3: Full RCs with CS are processed like unilingual English full RCs – no effect of a perception verb on RC resolution, but increased in RTs at head nouns and the embedded verb.

Discussion: A structural prediction triggered by a perception matrix verb shapes RC processing, irrespective of the place of CS. A possible prosodic break forced by CS is not strong enough to shape RC resolution. In general, two languages integrate seamlessly and follow the structural prompt available.

(1a) Guille vio [DP al amigo del vecino [RC que hablaba de fútbol en el patio]]

Guille saw the friend-ACC of the neighbor-Gen who talked about football on the patio

Spanish, restrictive RC-reading: *Guille saw the friend of the neighbor who talked about football…*

(1b) Guille vio [CP [DP al amigo del vecino] [CP que hablaba de fútbol en el patio]]]

Guille saw the friend-ACC of the neighbor-Gen who talked about football…

Spanish, eventive reading: *Guille saw the talking about football done by the friend of the neighbor*

(2a) HA: Bill called the friend of the neighbor [break] [RC that was talking about football]

(2b) LA: Bill called the friend [break] of the neighbor [RC that was talking about football]

Design: a. Bill <u>saw/called</u> the friend of the neighbor <u>que hablaba de fútbol en el patio</u>

b. Bill <u>saw/called</u> the friend <u>del vecino que hablaba de fútbol en el patio</u>

Participants: ***Experiment 1***

Group	Part #	Age	C-test Eng, % correct
English	30	28.6 (19-53)	n/a
Portuguese-English	30	23.4 (18-38)	85.7 (75-98)
Norwegian-English	34	25.5 (19-54)	90.5 (72-100)
SpanishCatalan-English	30	26.5 (18-38)	83.3 (72-95)
Armenian-English	30	20.6 (18-28)	83.6 (72-92)

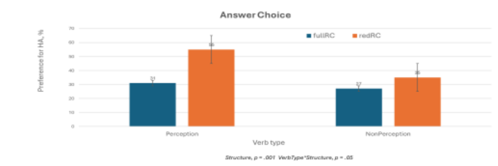
Experiments 2 and 3

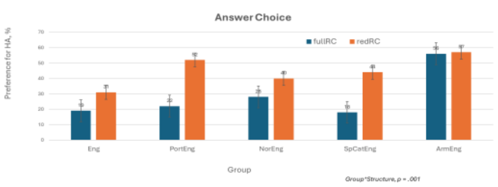
Group	Part #	Age	C-test Eng, % correct
Norwegian-English	34	25.5 (19-54)	90.5 (72-100)
Russian-Norwegian	34	29.5 (19-56)	80.5 (60-100)
Portuguese-English	30	23.4 (18-38)	85.7 (75-98)
HeritageSpanish-English	20	22.5 (18-31)	Spanish: 95 (88-97)
Catalan(Sp)-English	31	26.5 (18-38)	93 (87-100)

Results: Answer choice

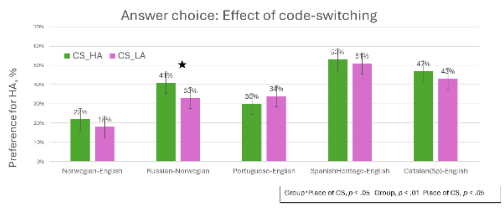
Due to space limitations, only the results of answer choice (proxy for RC resolution) are presented here. At the conference, reading and response time would be reported.

<u>Experiment 1:</u> full RC vs reduced RC in English

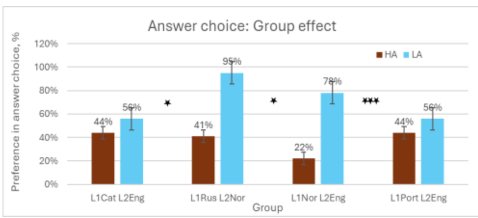


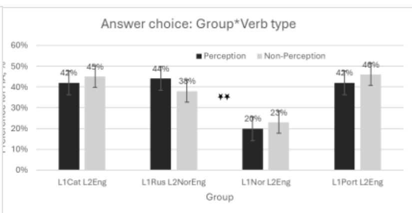


<u>Experiment 2:</u> parsing effect of place of CS



<u>Experiment 3:</u> eventive prediction vs. placeCS





References

[1] Aguilar, M., Ferré, P., Gavilán, J., Hinojosa, J., & Josep Demestre. (2021). The actress was on the balcony, after all: Eye-tracking locality and PR-availability effects in Spanish. *Cognition*. https://doi.org/10.1016/j.cognition.2021.104624

[2] Fodor, J.t D. (2002). Prosodic disambiguation in silent reading. *In Proceedings of the North East Linguistics Society*. Edited by M. Hirotani: 32, vol. 1., pp. 113–132.

[3] Grillo, N., & João Costa. (2014). A novel argument for the universality of parsing principles. *Cognition 133*, (156–187).

[4] Pozniak, C., Hemforth, B., Haendler, Y., Santi, A., & N. Grillo. (2019). Seeing events vs. entities: The processing advantage of pseudo relatives over relative clauses. *Journal of Memory and Language 107*, (128–151)

Selected examples of global ambiguity between full RC and eventive CP (pseudo-relative, Grillo & Costa, 2014)

Ambiguous in Spanish:

(1a) Guille vio [DP al amigo del vecino [RC que hablaba de fútbol en el patio]]

Guille saw the friend-ACC of the neighbor-Gen who **talked** about football on the patio

Spanish, restrictive RC-reading: Guille saw the friend of the neighbor who talked about football on the patio

(1b) Guille vio [CP [DP al amigo del vecino] [CP que hablaba de fútbol en el patio]]]

Guille saw the friend-ACC of the neighbor-Gen who **talked** about football on the patio

Spanish, eventive reading: Bill saw the talking about football performed by the friend of the neighbor

Non-ambiguous in English:

(1c) Bill saw [CP (that) [DP the friend of the neighbor] [VP was drinking coffee]]

(1d) Bill saw [DP the fired of the neighbor [RC that was drinking coffee]]

Selected examples of global ambiguity between reduced RC and eventive SC (reduced pseudo-relative, if following the terminology of Grillo & Costa, 2014)

Ambiguous in English:

(2a) Bill saw [SC [DP the friend of the neighbor] [VP drinking coffee]]

(2b) Bill saw [DP the friend of the neighbor [RC drinking coffee]]

Non-ambiguous in Norwegian:

(2c) Maria så [SC [DP vennen til naboen [VP snakke om fotball på telefonen]]

Maria saw the friend-ACC of the neighbor-Gen **talk-INF** about football on the phone

Target languages: Ambiguity between the restrictive RC (DP+RC) and the eventive complement (CP or SC)

Language & language family	RC attachment	full RC / full CP	reduced RC / SC
English (Germanic)	LA	non-ambiguous	ambiguous
Portuguese (Romance)	LA	non-ambiguous	non-ambiguous
Norwegian (Germanic)	LA	non-ambiguous	non-ambiguous
Spanish (Romance)	HA	ambiguous	ambiguous
Catalan (Romance)	LA!	? non-ambiguous	ambiguous
Armenian (Armenian branch)	HA	non-ambiguous	(!)ambiguous

3

No Evidence for Syntactic or Semantic Interference in Hindi Subject-Verb Processing

Kanika Sachdeva[1], Himanshu Yadav[1]
[1] Department of Cognitive Science, Indian Institute of Technology Kanpur
kanika25@iitk.ac.in, himanshu@iitk.ac.in

Comprehending a sentence requires us to identify and connect together linguistically related pairs of words. For example, we must identify the correct subject of the verb to complete the subject-verb dependency. Cue-based retrieval theories [1, 2, 3], propose that identifying a target chunk is driven by a content-addressable search in memory, i.e., using retrieval cues such as subject, number, etc. A key prediction of cue-based retrieval theories is similarity-based interference: When multiple items in memory share similar features to that of the target item, it causes difficulty in retrieving the target. For example, in processing a subject-verb dependency, a search is triggered at the verb to identify the target subject noun using retrieval cues such as [subject]. If other nouns in memory match the [subject] cue, a difficulty arises in retrieving the target subject noun. Consequently, the retrieval times at the verb will be slower when multiple nouns match the [subject] cue, e.g., in condition (a), compared to condition (b). This predicted effect is called syntactic interference [3]. Similarly, when multiple nouns match the verb's semantic cues, such as the animacy cue, e.g., in (c) versus (d) it is predicted to cause semantic interference [4]. Syntactic and semantic interference effects have been consistently observed in English [3, 4, 5, 6], providing strong support for cue-based retrieval theories. Due to their broad empirical coverage, these theories have been highly influential in sentence processing literature. However, recent self-paced reading and EEG data from German unveil a huge empirical challenge: No syntactic interference is observed in German, and there is weak evidence for semantic interference [7]. Data from German questions the cross-linguistic validity of similarity-based interference and the cue-based retrieval assumption. Given the theoretical importance of similarity-based interference and the open question of its generalizability to non-English languages, we attempt to replicate syntactic and semantic interference effects in Hindi, a language with rich case-marking system and verb-final typology similar to German. Experiment (N_{items}= 40, $N_{subjects}$= 98) consisted of a centered self-paced reading task with a factorial design of 2 x 2. Syntactic and semantic interferences were introduced by manipulating a distractor noun such that it matched the target noun in *syntactic* feature (+nominative) and *semantic* feature (+animate) respectively, resulting in 4 conditions for each item (see sample stimuli 1a-1d). The cue-based retrieval theories [2] predict syntactic interference - a slowdown in reading times at the verb in conditions (1a) and (1b) compared to (1c) and (1d), and semantic interference — a slowdown in (1a) and (1c) compared to (1b) and (1d).

Reading times at the verb were analysed using Bayesian linear mixed models with syntactic interference, semantic interference and their interaction as fixed effects and by-subject and by-item random intercepts and slopes. The estimated semantic interference is largely in the positive range [95% CrI: 0,45]. While the estimated syntactic interference [95% CrI: -38,6] goes against the prediction of the cue-based retrieval. Bayes factors provided anecdotal evidence for semantic interference and no evidence for syntactic interference (see Figure 1). Results contribute to a growing body of evidence that working-memory effects are weak or absent in verb-final languages like Hindi and German. Our findings cast doubts on the cross-linguistic generalizability of cue-based retrieval, demanding a new theory that can explain data across typologically different languages.

(a) The **attorney**$^{+subject}$ whose secretary had forgotten that the **visitor**$^{+subject}$ was important frequently **complained**^{subject} about ...

(b) The **attorney**$^{+subject}$ whose secretary had forgotten about the important **visitor**$^{-subject}$ frequently **complained**^{subject} about ...

(c) The **attorney**$_{+animate}$ whose secretary had forgotten that the **visitor**$_{+animate}$ was important frequently **complained**^{animate} about ...

(d) The **attorney**$_{+animate}$ whose secretary had forgotten that the **meeting**$_{-animate}$ was important frequently **complained**^{animate} about ...

Table 1: Sentence pairs demonstrating predicted interference effects due to syntactic and semantic distractors. In sentence (a) vs (b), the presence of syntactic distractor noun *visitor* is assumed to interfere in retrieval of the target subject *attorney* at the verb *complained* because the distractor *visitor* matches the retrieval cue [subject] at the verb, in contrast to condition (b) where no other noun matches the subject cue.

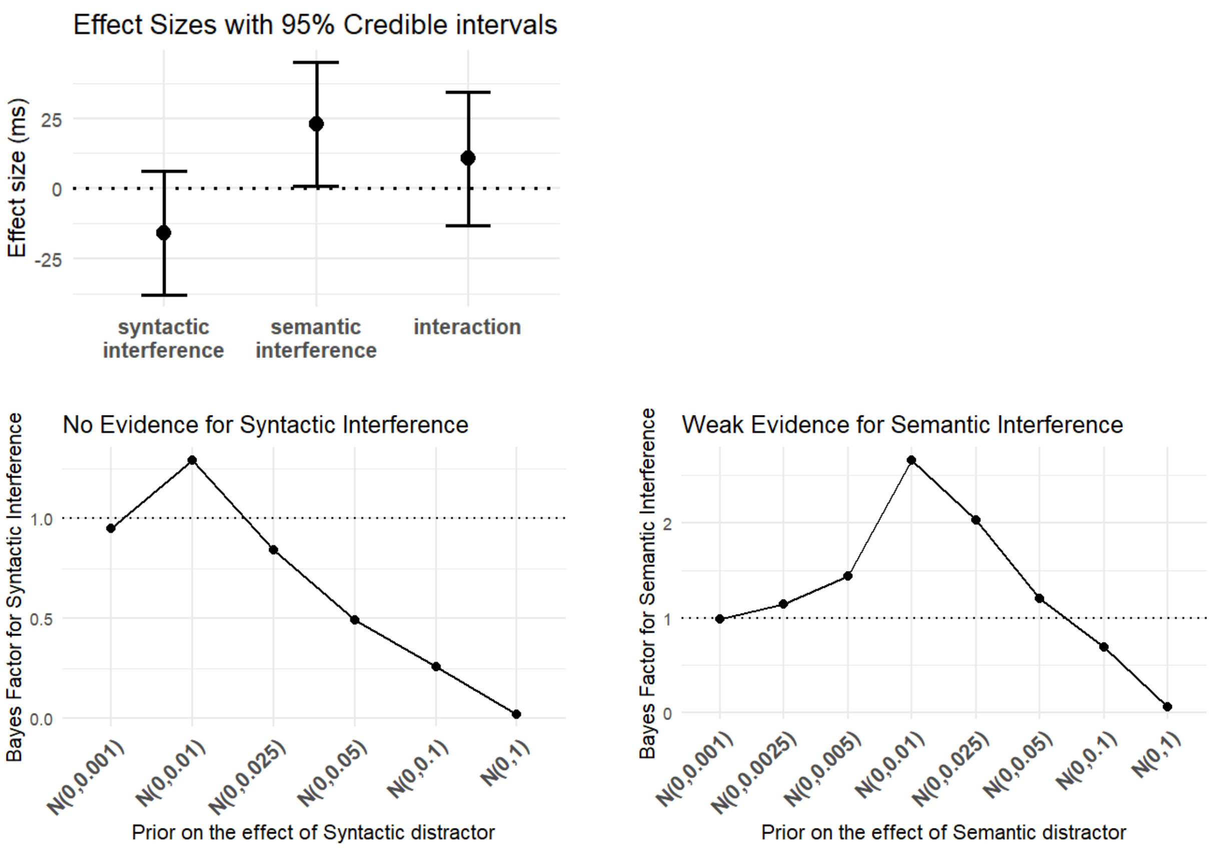


Figure 1: Posterior distributions showing the estimated effect sizes for syntactic and semantic interference and Bayes factors showing weak evidence in favor of semantic interference and no evidence for syntactic interference.

References

[1] McElree, B. (2000). *J Psyhcolinguist Res*.

[2] Lewis, R. L., & Vasishth, S. (2005). *Cognitive science*.

[3] Van Dyke, J. A. (2007). *J Exp Psychol Learn Mem Cogn*.

[4] Mertzen, D., Paape, D., Dillon, B., Engbert, R., & Vasishth, S. (2023). *Glossa Psycholinguistics*.

[5] Arnett, N., & Wagers, M. (2017). *JML*.

[6] Van Dyke, J. A., & McElree, B. (2011). *JML*.

[7] Schoknecht, P., Yadav, H., & Vasishth, S. (2025). *JML*.

A sample experimental stimulus from Hindi consisting of four conditions.

1a **Syntactic and semantic interference condition**

ek **makaanmalik**$^{+nominative}_{+animate}$ jiska **kirayedaar**$^{+nominative}_{+animate}$ kal scooter udhaar le-gaya-tha aaj paidal office ja-rha-hai$^{nominative}_{animate}$, ...

A landlord.*Nominative* whose tenant.*Nominative* yesterday scooter borrow take-go-Past today by-foot office go-PROG-Present, ...

1b **Only syntactic interference condition**

ek **makaanmalik**$^{+nominative}_{+animate}$ jiska **scooter**$^{+nominative}_{-animate}$ kal ped-ke-pas chhoot-gaya-tha aaj paidal office ja-rha-hai$^{nominative}_{animate}$, ...

A landlord.*Nominative* whose scooter.*Nominative* yesterday tree-*Locative* left-go-Past today by-foot office go-PROG-Present, ...

1c **Only semantic interference condition**

ek **makaanmalik**$^{+nominative}_{+animate}$ jisne **kirayedaar-ko**$^{-nominative}_{+animate}$ kal scooter udhaar de-diya-tha aaj paidal office ja-rha-hai$^{nominative}_{animate}$, ...

A landlord.*Nominative* who-*Ergative* tenant-*Dative* yesterday scooter.*Accusative* lend give-give.Perf-Past today by-foot office go-PROG-Present, ...

1d **No interference condition**

ek **makaanmalik**$^{+nominative}_{+animate}$ jisne **scooter-ko**$^{-nominative}_{-animate}$ kal ped-ke-pas chhod-diya-tha aaj paidal office ja-rha-hai$^{nominative}_{animate}$, ...

A landlord.*Nominative* who-*Ergative* scooter-*Accusative* yesterday tree-*Locative* left give-Past today by-foot office go-PROG-Present, ...

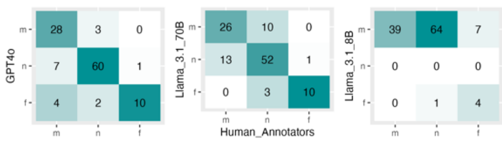
On the suitability of LLM output as an experimental data source in German: Evidence from GPT-4o, LLaMa 3.1 70B and LLaMa 3.1 8B

Franziska Kretzschmar[1], Sandra Hansen[1], Christian Lang[1]

[1] Department of Grammar, Leibniz Institute for the German Language

kretzschmar@ids-mannheim.de

Background: The human-like output of large language models (LLMs) has initiated investigations on whether LLMs may be a valid experimental data source in addition to humans [1-3]. LLMs have been tested on different tasks (e.g., ratings on lexical-semantic dimensions, sentiment analysis, comprehension [1-2,4-5]). The results suggest that despite the qualitative similarity of LLM-based and human-based results, LLM-generated data often show a tendency to over- or undershoot human data. As argued in [1], this may lead to differences in effect sizes between LLMs and humans, and further research is needed to analyze the robustness of this pattern across LLMs, tasks and languages. Here, we present a study in which we compare LLMs and humans presented with a semantic association task, asking to what extent the data sets diverge at trial level. Our study adds three novel aspects to the existing literature: 1) we use an association task in which NPs are to be rated according to their association with a social gender. Social gender, i.e. the association of a word's lexical meaning with a person of female or male gender, establishes an associative relation [6]. So far, it has been found that LLM priming effects are smaller for associative (vs. semantic) relations [4]. 2) we use German as a target language to test LLMs mainly trained on English. 3) we compare the performance of three LLMs of varying size. **Method:** We selected 116 German NPs to be labelled for their association with a social gender (female, male or none) and rated for association strength. This task was used to prompt three LLMs (ChatGPT4o, Llama 3.1 70B Instruct, Llama 3.1 8B Instruct), each instructed to generate 33 iterations. Additionally, 132 human participants from Prolific performed the same task, each with a subset of the NP list so that the number of trials per NP was identical for humans and LLMs. We use majority votes and compare by-item agreement across trials. **Results:** The confusion matrix (Table 1) shows that GPT4o generates the most human-like responses (labels), which is supported by the highest precision, recall and F1 scores. Llama 3.1. 8B performs worst: Most NPs are falsely associated/labelled with a male social gender. Figures 1 and 2 show that the distribution of associative gender labels across trials differs between LLMs and humans. LLMs show more items with perfect agreement of 1 across trials (i.e. the same label for all trials), while there are only two items with an agreement of 1 for humans (Fig. 1). Also, GPT4o shows more items with an agreement < 1 than the other models, Llama 3.1 8B exhibits an agreement < 1 for only one item. Finally, there is divergence for NPs with agreement < 1 in that humans use all three labels (male, female, none) on trials for a single NP item (Fig. 2b), whereas LLMs generate either male+none or female+none, but not male+female associations for an NP (Fig. 2a). **Discussion:** Our study is among the first to compare the behavior of 3 LLMs with humans in an association task. While, superficially, there is overlap in labels and association strength (Fig. 3), the distribution of responses across trials for individual items differs noticeably. This suggests that LLM task performance is qualitatively different at the level of the single item response, confirming that associative relations may pose a challenge for LLMs to perform human-like. Our data also show that LLMs differ vastly from one another on this task in German and that LLMs trained on English are still able to produce interpretable, though not fully human-like, German output. Overall, this implicates limited potential of LLMs as an experimental data source to save research resources.



GPT4o				Llama 3.1 70B				Llama 3.1 8B			
Class	Precision	Recall	F1	Class	Precision	Recall	F1	Class	Precision	Recall	F1
m	0.903	0.718	0.800	m	0.722	0.667	0.693	m	0.355	1.000	0.523
n	0.882	0.923	0.902	n	0.788	0.800	0.794	n	NA	0.000	NA
w	0.625	0.909	0.741	w	0.769	0.909	0.833	w	0.800	0.364	0.500

Table 1. Top panel: Confusion matrix with overlap of gender labels per category. Y-axis: LLM labels, X-axis: human labels. For example, for all instances of male-gender association by humans, GPT-4o correctly labelled 28 items, and incorrectly labelled 11 items (7 neutral, 4 female). Bottom panel: precision, recall and F1 scores per LLM and social-gender label category.

Figure 1.
Majority-vote agreement per item, separated by humans and LLMs. NP names shown: humans agreement = 1, LLMs agreement < 1. (Agreement = 1: the same label for an NP in all trials.)

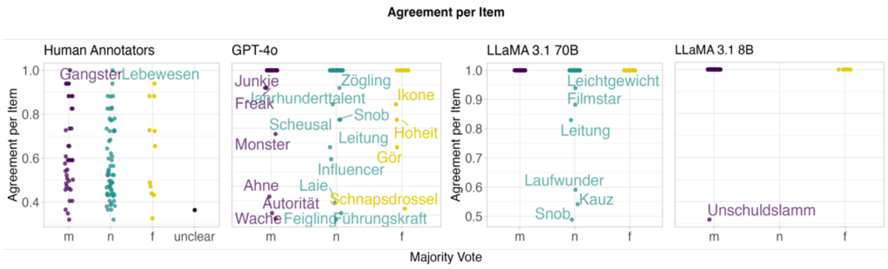


Figure 2a (left). Distribution of LLM-generated labels for items with an agreement < 1.
Figure 2b (right). Distribution of human-generated labels for items with an agreement < 1.

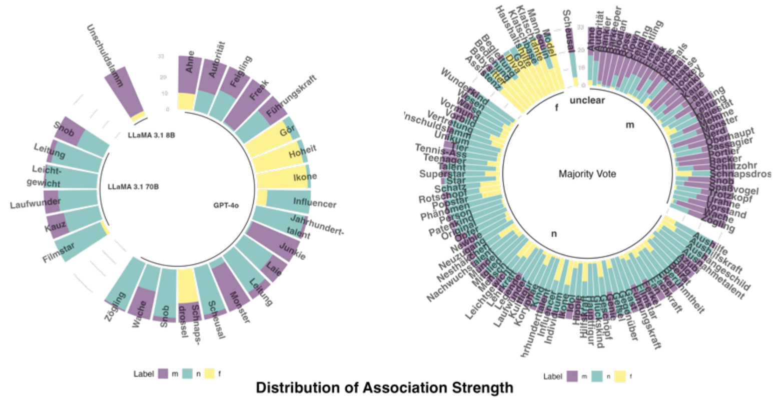
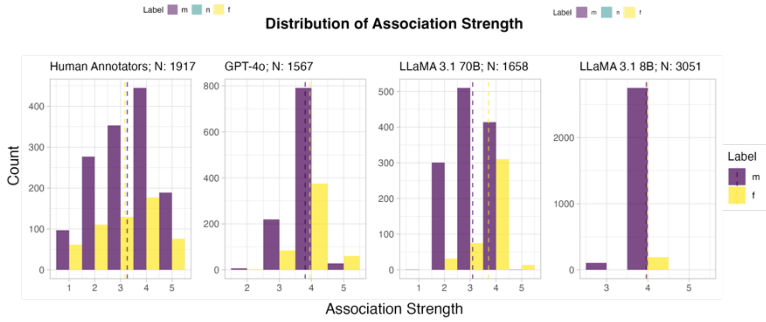


Figure 3. Distribution of association strength (1-very low, 5-very strong) for male and female labels. Dashed lines represent the mean per label category.



References

[1] Trott, S. (2024). Can large language models help augment English psycholinguistic datasets? Behavior Research Methods, 56, 6082-6100.

[2] Cai, Z.G., Duan, X., Haslett, D.A., Wang, S., & Pickering, M.J. (2024). Do large language models resemble humans in language use? Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, 37-56.

[3] Aher, G., Arriaga, R.I., & Kalai, A.T. (2023). Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies. Proceedings of the 40th International Conference on Machine Learning, 1-35.

[4] Digutsch, J., & Kosinski, M. (2024). Overlap in meaning is a stronger predictor of semantic activation in GPT-3 than in humans. Scientific Reports, 13, 5035.

[5] Zhu, Y., Zhang, P., Haq, E.-U., Hui, P., & Tyson, G. (2023). Can ChatGPT Reproduce Human-Generated Labels? A Study of Social Computing Tasks. Preprint arXiv: 2304.10145v2.

[6] Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. Nature Human Behaviour, 4, 1021-1028.

NP examples

We used NPs of all grammatical genders (masculine, feminine, neuter) in German.

Grammatical Gender (number of items)	Examples
Masculine (N = 30)	*der Hippie* (‚hippie‘), *der Vormund* (‚guardian‘), *der Barkeeper* (‚bartender‘)
Feminine (N= 47)	*die Autorität* (‚authority‘), *die Führungskraft* (‚manager‘), *die Aushilfe* (‚temporary staff‘)
Neuter (N= 36)	*das Unschuldslamm* (‚innocent lamb‘), *das Scheusal* (‚monster‘), *das Idol* (‚idol‘)
Ambiguous gender marking (N = 3)	*das/der Balg* (‘annoying child’), *der/die/das Mündel* (‘ward’), *der/das Gegenüber* (‘opposite (person)’)

LLM prompt

{"role": "developer","content": [{"type": "text","text":
"You are a helpful assistant who answers whether a German noun is more likely to be associated with a male person or a female person. If you think the noun is more associated with male persons your answer is: 'Person männlichen Geschlechts'; if you think the noun is more associated with female persons your answer is: 'Person weiblichen Geschlechts'; if you think the noun is neither associated with male persons nor with female persons your answer is: 'keins von beiden'; give reasons for your answer and list all the decisive factors. If your answer is 'Person männlichen Geschlechts' or 'Person weiblichen Geschlechts' rate on a scale from 1 to 5 how strong the association is (with 1 meaning very weak association and 5 meaning very strong association). Structure your answer in: 'Short answer', 'Reasons' and if applicable 'Strength of association'." }]},
{"role": "user", "content": [{ "type": "text", "text": "Assoziieren Sie das Wort " + word + " eher mit einer Person männlichen oder weiblichen Geschlechts? (wenn es auf eine Person bezogen gebraucht wird)"}] }
(translation of last sentence: ‚Do you associate the Wort “+word+” more with a person of male or female gender? (if used to refer to a person’)

Instruction given to human participants in the online study:

<u>Association labelling:</u>

Assoziieren Sie das Wort "#NP#" eher mit einer Person männlichen oder weiblichen Geschlechts (wenn es auf eine Person bezogen gebraucht wird)? Oder haben Sie keine Assoziation in eine der beiden Richtungen ("keins von beiden")?
Translation: Do you associate the word "#NP#" more with a person of male or female gender (if it is used to refer to a person)? Or do you have no association in either directions ("neither one")?

<u>Association strength:</u>

Wie stark ist für Sie die Assoziation von "#NP#" mit "#Person männlichen/weiblichen Geschlechts#"? (von 1: sehr schwach ausgeprägt bis 5: sehr stark ausgeprägt)
Translation: How strong is the association of "#NP#" with "#person of male or female gender#"? (from 1: very weakly pronounced to 5: very strongly pronounced)

Working memory capacity predicts sensitivity to prosodic structure

Constantijn L. van der Burght[1,2], Antje S. Meyer[1,3]
[1] Max Planck Institute for Psycholinguistics; [2] Leiden University Centre for Linguistics, Leiden University;
[3] Donders Institute for Brain, Cognition, and Behaviour, Radboud University
Stan.vanderBurght@mpi.nl

Speakers use the intonation, intensity, and rhythm of speech (prosody) to convey important information to their conversation partners. Listeners vary greatly in the perception of prosody, but the source of this variability has remained unknown. We investigated whether the ability to recognise and classify prosodic structure is related to working memory (WM) capacity. This hypothesis stems from the tight connection between prosodic and syntactic structure, while processing syntax is known to relate to WM capacity. In an online study 64 adult speakers of Dutch took part in a series of three tasks. In a prosody classification task, they listened to short phrases in a gating paradigm[1]. The phrases were three-name sequences with early and late prosodic cues that signalled whether the phrase contained an internal grouping structure or not: [[Annie and Jannie] and Mira] vs. [Annie and Jannie and Mira]. A total of 24 three-name sequences were constructed, compiled from six names that could occur in the first two positions and three names that could occur in the final position. The sequences were divided over four native speakers of Dutch (two female, two male) who were instructed to realise each name sequence in two alternative groupings. This yielded a total of 48 stimuli. In a two-alternative forced choice task, participants judged the internal grouping of the phrases, based solely on the prosodic information. Participants indicated their judgement by button press, choosing the visually presented response option that matched their interpretation (two stick figures grouped separately from a third vs. three stick figures depicted together). In this gating procedure, an additional syllable was presented at each subsequent gate, so that the first gate contained only the first syllable and the final gate the full phrase. Participants indicated their judgement of the structure upon the presentation of each subsequent gate. Following the prosody classification task, listeners took part in WM (digit span, forward and backward) and processing speed (letter comparison) tasks. The results showed that high-WM listeners were better at classifying prosodic structure and required less prosodic information to detect the correct structure. There was no effect of processing speed on prosody comprehension, suggesting that the relationship between prosodic judgement and WM capacity was not due to motivational or attentional differences. Our study therefore demonstrates that prosody perception and WM capacity are tightly linked, a result that has theoretical implications for theories of prosody processing: WM must be an important component of processing models of prosody, in addition to the acoustic analysis of prosody and its integration with sentence-level information.

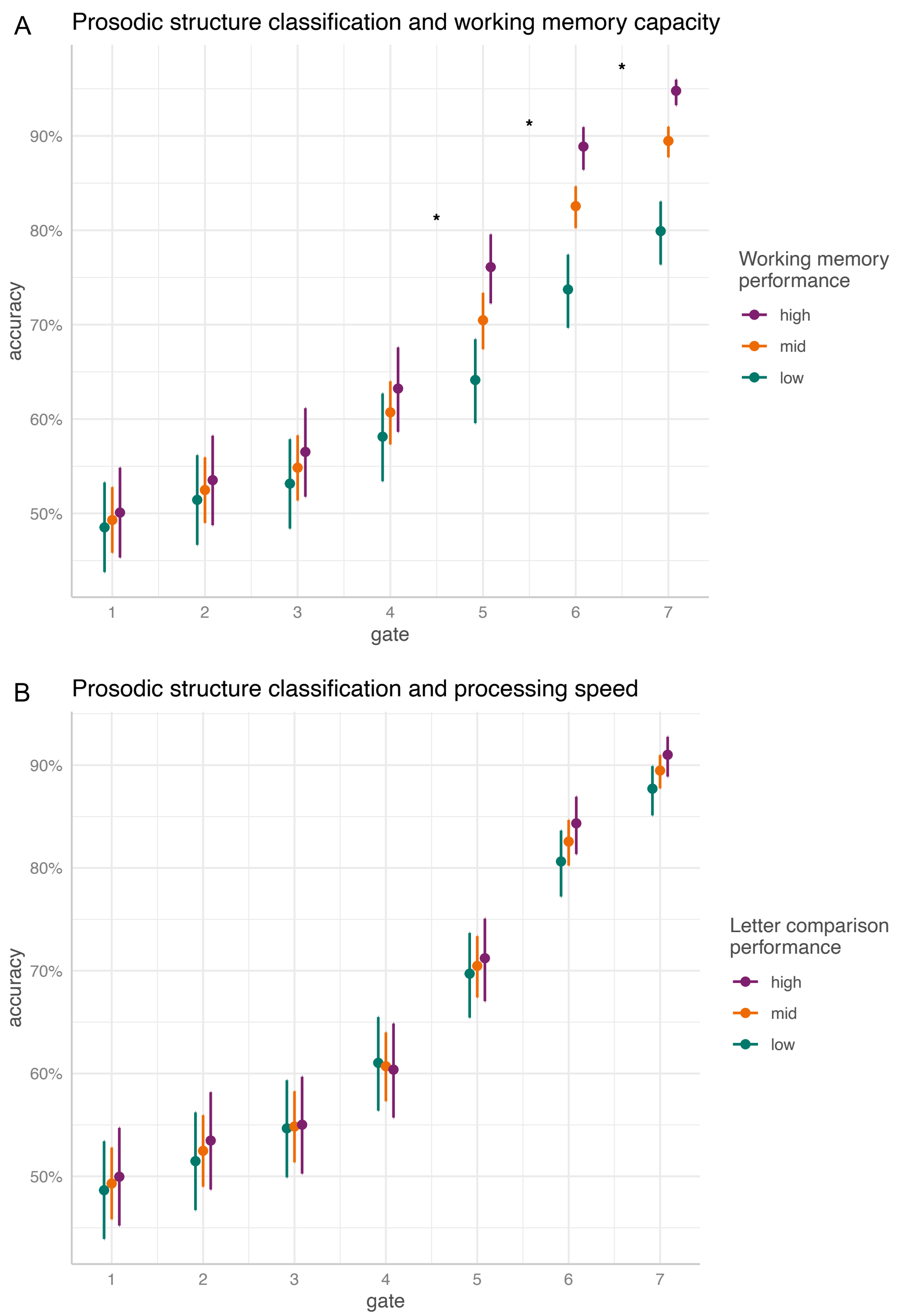


Figure 1: Performance in the working memory task predicted performance in the prosody classification task: Participants with higher working memory spans were better at classifying the prosodic structure and required the prosodic information from fewer gates to achieve above-chance performance (A). In contrast, there was no relationship between prosody classification and processing speed: Participants with high processing speed performed similarly on the prosody task as those with low processing speed (B). Dots represent predicted probabilities and error bars standard deviations. Asterisks mark statistically significant interactions between pairs of gates and performance in the WM task.

1. Hansen, M., Huttenlauch, C., de Beer, C., Wartenburger, I., & Hanne, S. (2022). Individual Differences in Early Disambiguation of Prosodic Grouping. *Language and Speech*. https://doi.org/10.1177/00238309221127374

No pseudo-morphological decomposition during lexical access, but actual morphological analysis in the lexicon: Meta-analytical evidence from seven new replicated masked stem priming experiments

Roberto Petrosino[1], Jon Sprouse[1], Diogo Almeida[1]
[1] Department of Psychology, Division of Science, New York University Abu Dhabi
roberto.petrosino@nyu.edu

Introduction. Cross-linguistic masked priming evidence shows that (a) bimorphemic (***bone****less*) and (b) pseudo-affixed, monomorphemic words (**corn**er) often trigger similar masked priming effects on their stems (*BONE* and *CORN*, respectively). In contrast, (c) monomorphemic words (e.g., ***cash****ew*) trigger negligible effects on the embedded stem (*CASH*). The dissociation between (a, b) and (c) has been taken as evidence that lexical access goes through an obligatory decomposition procedure based on morpho-orthographic "islands of regularity", i.e. the statistically regular letter sequences corresponding to the orthographic realizations of morphemes [1]. This procedure arguably occurs before accessing any lexical (morpho-syntactic, or semantic) information so that, at early stages, (b), but not (c), can be provisionally visually decomposed similarly to (a). Thus, this decomposition process is not an actual morphological or even linguistic process per se (despite often being described as morphological decomposition), but a visual process informed by morpho-orthographic statistical regularities. Recent findings, however, have challenged such a view, showing *differential* effects for (a) and (b) [2], and *similar* effects for (a) and (c), or (b) and (c) [3].

Methods. We conducted seven different replications of an English masked stem priming experiment both in-lab and online, to ensure replicable results (N_{total}=1,235). Five conditions were tested across all experiments: An identity condition (*fuss-FUSS*), a transparent morphological condition (*sharper-SHARP*), a pseudo-morphological condition (*belly-BELL*), and two orthographic control conditions: a non-syllabic orthographic control (*bark-BAR*), and a syllabic orthographic control (*cashew-CASH*). Items were controlled for orthographic length and word frequency. Trials consisted of a 33ms-long prime preceded by a 500ms-long forward mask, and followed by a target, on which participants performed a lexical decision task. After excluding subjects and items with high error rates, and trials with outlying prime durations or extreme RTs, we performed a varying-coefficient meta-analysis [4] on the raw-data estimates across all seven replications.

Results. Identity and morphological priming effects were indistinguishable, and almost as large as the prime duration, suggesting ceiling effects ($M_{identity}$=26 ms, 95% CI [22 30]; $M_{transparent}$=26 ms, 95% CI [22 29]). Pseudo-morphological priming was significantly smaller than both (M_{opaque}=17 ms, 95% CI [13 21]), and closely matched the non-syllabic orthographic priming condition ($M_{nsyll\text{-}ortho}$=16 ms, 95% CI [11 20]), which was somewhat larger than the syllabic orthographic priming condition ($M_{syll\text{-}ortho}$=9 ms, 95% CI [5 13]).

Conclusions. These results are incompatible with the proposal that lexical access includes a *pre-lexical morpho-orthographic decomposition* of the input, as *pseudo-morphological priming* patterned with purely *orthographic priming*, not with actual *morphological priming*. In contrast, real *morphological priming* effects were indistinguishable from *identity priming* effects, possibly indicating that these priming effects obtain at the lexical, not prelexical level, suggesting that even very brief visual presentations (here, 33ms) are sufficient to lead to access to lexical information, contrary to the assumption that only pre-lexical processes can occur during that time window.

Figures

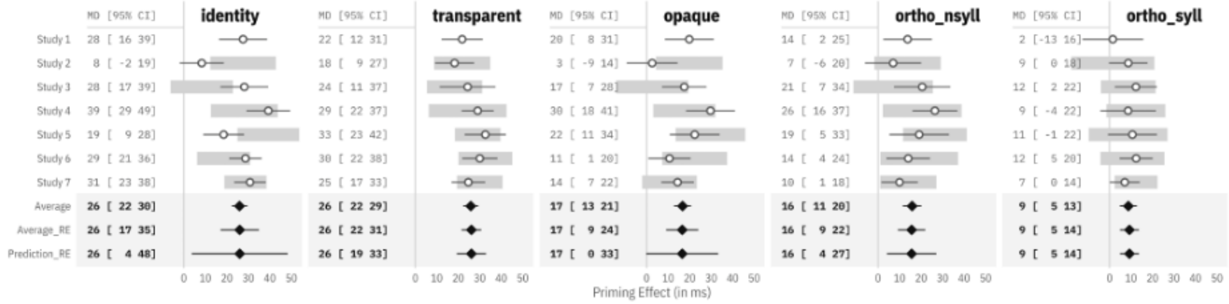


Figure 1. Mean and 95% CI of each condition in each experiment in our sample. The gray bands indicate the 95% Prediction Interval (PI) of one study to the next. The last three lines include the following estimates: the mean and 95% CI of the Varying Coefficient Meta-Analysis and Random-Effects Meta-Analysis, as well as the 95% Prediction Interval (PI) of the Random-Effects Meta-Analysis.

References

[1] Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and cognitive processes*, 23(7-8), 942-971.

[2] Feldman, L. B., O'Connor, P. A., & del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic bulletin & review*, 16, 684-691.

[3] Morris, J., Porter, J. H., Grainger, J., & Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language and cognitive processes*, *26*(4-6), 558-599.

[4] Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological methods*, *14*(3), 225.

Implicit Prosody and Pseudo Relative availability independently modulate RC-attachment

Francesca Penoncelli[1], Nino Grillo[2], Giuliano Bocci[1]

[1] Università di Siena [2] University of York

nino.grillo@york.ac.uk

Background: Relative Clause (RC) attachment preferences in complex NPs are shaped by multiple factors, including implicit prosodic phrasing [1, 2] and the availability of Pseudo-Relative Small Clauses [3]. While these factors have been studied independently, their interaction remains unexplored.

According to the Implicit Prosody Hypothesis, silent reading induces default prosodic phrasing that guides parsing, with prosodic boundaries aligning to satisfy prosodic constraints–e.g., *Same Size Sister* [4, 5]–which favor balanced prosodic phrasing. Thus a long RC (1-a) tends to be preceded by a prosodic break, encouraging High Attachment (HA), while a short RC (1-b) typically lacks a break, favoring low attachment (LA) [6, 7].

In languages like Italian, a second factor modulates this preference, as CPs in the environment of event-taking predicates (e.g. *see*) are ambiguous between a RC reading(2-a), in which the CP modifies an individual, and the so-called Pseudo Relative (PR) reading (2-b), which denotes an event and roughly corresponds to eventive Small Clauses in English. Importantly, attachment ambiguities disappear under a PR reading, which only allows HA (3). The *PR-first Hypothesis*, [3, 8] argues that the parser prefers PR interpretations when possible because they are structurally and interpretively simpler than RCs, which correctly predicts strong effects of PR-availability on RC-attachment across languages [3, 8, 9, 10, 11](a.o.). Implicit Prosody and PR-availability effects have been demonstrated independently but their interaction remains untested. We investigate this gap in Italian, asking whether implicit prosody (cued by RC-length) and PR availability (cued by Verb-Type) exert independent or interactive effects on RC-attachment.

Method: 70 Italian participants took part in a sentence interpretation task during silent reading. We manipulated MATRIX VERB TYPE (entity-taking vs. event-taking) and RC LENGTH (short vs. long) in declarative sentences containing right-branching RC (Table 1). Participants read a sentence and then answered a comprehension question designed to probe their attachment interpretation. Following the IPH, we predict a preference for HA with long RCs. Following PR-first, we expect a preference for HA with event-taking matrix verbs. Finally, we expect long RCs to reinforce the preference for HA with event-taking matrix verbs.

Results: Binary responses (HA/LA) were analyzed using a logistic mixed-effects model with MATRIX VERB TYPE and RC LENGTH as fixed effects (both contrast-coded: ±0.5), and random intercepts and slopes by participant and item. The analysis revealed main effects of VERB TYPE (β=–2.27, *p*<.001) and RC LENGTH (β =–0.76, *p*<.001), with more HA responses following event-taking matrix verbs and long RC clauses. Their interaction was n.s.. See Figure 1. Furthermore, response times (in ms.) were significantly longer in the stative verb condition (β = 2350, *p*=.027), consistent with the hypothesis that PRs are processed more efficiently than RCs. See Figure 2.

Discussion: The results suggest that both implicit prosody and PR-availability independently influence attachment decisions, though the syntactic effect of VERB TYPE was substantially stronger than the prosodic effect of RC LENGTH (β=–2.27 vs. β=–0.76). Critically, while both factors promote HA, they operate through distinct mechanisms—prosodic phrasing versus structural economy—and their effects combine additively. The absence of an interaction suggests that prosodic and syntactic biases exert parallel constraints during parsing, rather than competing for dominance.

(1) a. John knows the son of the man that was running with Lucy.

b. John knows the son of the man that runs.

(2) a. Gianni ha [visto [$_{DP}$l'uomo [$_{CP}$che correva]]].

Gianni [saw [$_{DP}$the [$_{NP}$man [$_{CP}$that was running]].

b. Gianni ha [visto [$_{PR}$ [$_{DP}$ l'uomo] [$_{CP}$ che correva]]].

Gianni [saw [$_{SC}$ [$_{DP}$the man] [$_{VP}$ running]]].

(3) Ho visto [$_{PR}$ [il figlio$_i$ [dell'uomo$_j$]] [che correva$_{i/*j}$]]

Verb-Type	**RC-Length**	**Stimuli**
Entity-taking	Short	Gianni vive con il figlio del medico che correva. *Gianni lives with the son of the doctor that is running.*
Entity-taking	Long	Gianni vive con il figlio del medico che correva con Lucia. *Gianni saw the son of the doctor running with Lucia.*
Event-taking	Short	Gianni ha visto il figlio del medico che correva. *Gianni saw the son of the doctor that was running.*
Event-taking	Long	Gianni ha visto il figlio del medico che correva con Lucia. *Gianni saw the son of the doctor that was running with Lucia.*

Table 1: Experimental design and example stimuli

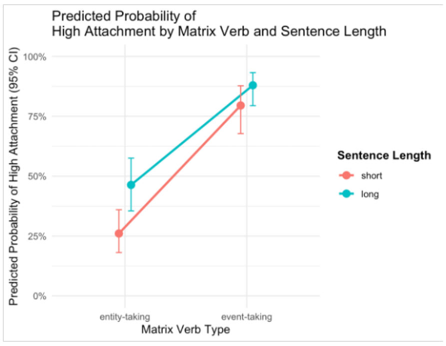


Figure 1: Predicted probability of high attachment by matrix verb and RC length.

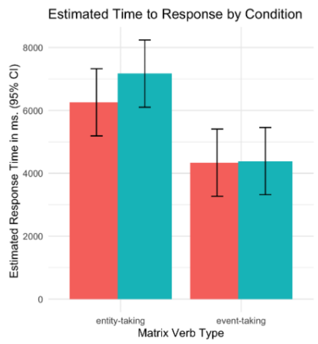


Figure 2: Predicted Time to Response (in ms.) by matrix verb and RC length.

References

[1] Bader. (1998). Prosodic influences on reading. In *Reanalysis in sentence processing*. Springer.

[2] Fodor. (2002). Prosodic disambiguation in silent reading. *Proc. of the North East Linguistic Society*, *32*.

[3] Grillo. (2012). Local and Universal. In Bianchi & Chesi (Eds.), *Enjoy Linguistics!* CISCL Press.

[4] Fodor, J. D. (1998). Learning to Parse? *J. of Psycholinguistic Research*, *27*.

[5] Gee & Grosjean. (1983). Performance structures. *Cognitive psychology*, *15*.

[6] Hemforth, Fernandez, Clifton, Frazier, Konieczny, & Walter. (2015). RC attachment. *Lingua*, *166*.

[7] Gianninò, Avesani, Bocci, & Vayra. (2021). Giudice o imputato? *L'individualità del parlante*, *8*.

[8] Grillo & Costa. (2014). A novel argument for the Universality of Parsing principles. *Cognition*, *133*.

[9] Grillo, Costa, Fernandes, & Santi. (2015). Highs and Lows in English attachment. *Cognition*, *144*.

[10] Pozniak, Hemforth, Haendler, Santi, & Grillo. (2019). Seeing events vs. entities. *JML*, *107*.

[11] Aguilar, Ferré, Gavilan, Hinojosa, & Demestre. (2021). The actress was on the balcony. *Cognition*, *211*.

Producing code-switches: Adaptation of cognitive control in code-switching

Demi Zhang[1], Emiliana Pulido[1], Maria Josefina Estrada[1], Souad Kheder [2], Edith Kaan[1]
University of Florida[1], University of Algiers2[2]
Zhang.yidan@ufl.edu

Background: Bilingual language control involves dynamic interactions between executive control and language processing, yet most research on code switching has focused on comprehension, with fewer studies examining production. While language control in comprehension and production have similarities, neural evidence reveals distinct pathways [1]. Critically, the role of language context in modulating executive control during production remains underexplored. Prior work often used constrained tasks, e.g., picture naming [2], which lack sentence-level activation. Recent efforts to study naturalistic code-switching faced methodological trade-offs between ecological validity and experimental control. To bridge this gap, our study employs a reading-aloud paradigm—a balance between naturalistic production and control—given evidence of shared phonological [3] and planning processes between reading and speech [4]. We investigate how dense code-switching during production influences executive control by using a conflict adaptation paradigm in which reading dense code-switching or unilingual Spanish sentences is mixed with a flanker task [5,6]. We specifically test two competing hypotheses. Hypothesis A predicts that dense code-switching allows for a free gateway between languages and requires less inhibition, leading to increased flanker interference relative to unilingual sentences [7]. Hypothesis B predicts that dense code-switching recruits inhibitory control, thereby reducing subsequent flanker conflict due to cross-task adaptation.

Method: Spanish-English bilinguals (n = 31 after 2 were excluded because of proficiency) read aloud experimental sentences in two language contexts: unilingual Spanish and dense code-switching (Table 1). All participants learned Spanish before English and acquired both languages before the age of 12. Each experimental sentence was followed immediately by a flanker task with congruent (<<<<<, >>>>>) or incongruent (<<><<, >><>>) arrow arrays (see Figure 1). Participants saw 12 items in each of the four conditions (2 contexts × 2 flanker types, Latin-Squared), plus 80 filler trials (32 filler sentences, which are followed by comprehension questions, and 48 filler flankers). Reading responses were recorded and annotated by trained bilingual coders. Trials were excluded if the audio was unclear, the participant made self-corrections, or the sentence did not contain the intended code-switching pattern. Trials with flanker response times <100 ms or >2000 ms and incorrect flanker responses were also excluded. Flanker RTs from remaining trials were analyzed using a linear mixed-effects model, using inverse transformed RTs to normalize distributions. The final model included flanker congruency, language context, and their interaction as fixed effects, and a by-participant random intercept.

Results: Of the 1488 recorded experimental trials, 1163 trials (78.1%) from 31 Spanish-English bilingual participants were retained after applying pre-registered exclusion criteria. A significant main effect of flanker congruency was observed: participants responded more slowly to incongruent (M = 728 ms) than congruent (M = 602 ms) trials, indicating successful replication of the standard flanker interference effect ($b = 0.310$, $SE = 0.018$, $t = 16.80$, $p < .001$). There was no main effect of language context ($b = -0.015$, $SE = 0.018$, $t = -0.83$, $p = .406$), nor a significant interaction between flanker condition and context ($b = -0.038$, $SE = 0.037$, $t = -1.02$, $p = .307$). Figure 2 visualizes these patterns via violin plots, displaying the distribution of RTs across conditions. Reading aloud sentences with dense code-switches did not significantly influence subsequent performance on the flanker task compared to reading unilingual Spanish sentences.

Discussion: Although bilingual language production is theorized to recruit domain-general control [1], the current findings suggest that merely reading code-switched sentences aloud does not exert a measurable influence on flanker task performance over reading unilingual sentences, thus neither Hypothesis A nor Hypothesis B was born out. This null effect may reflect the controlled nature of the reading-aloud task, which lacks the planning demands of spontaneous speech. Alternatively, code-switching may have involved cognitive control, but this may not have carried over to the subsequent flanker trial due to the reading and Flanker tasks being very different in nature.

Table 1: Example of an experimental sentence
(Code-switching in bold for the purpose of illustration. Translation: “The cat was lying on the floor sunbathing and resting after dinner in the afternoon.”)

Unilingual Spanish	Dense code-switching
El gato estuvo tirado en el suelo tomando sol y descansando después de la cena de la tarde.	*El gato estuvo tirado* ***on the floor*** *tomando sol* ***and resting*** *después de la cena de la tarde.*

Figure 1: Illustration of the trial structure

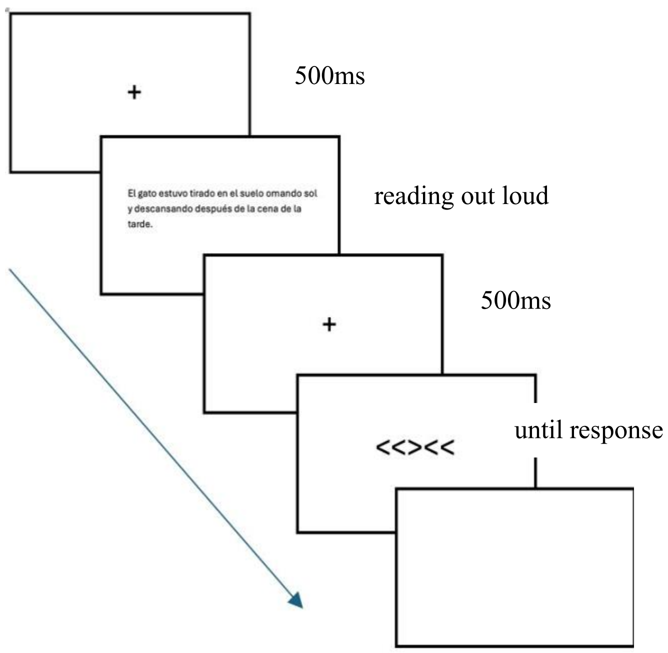
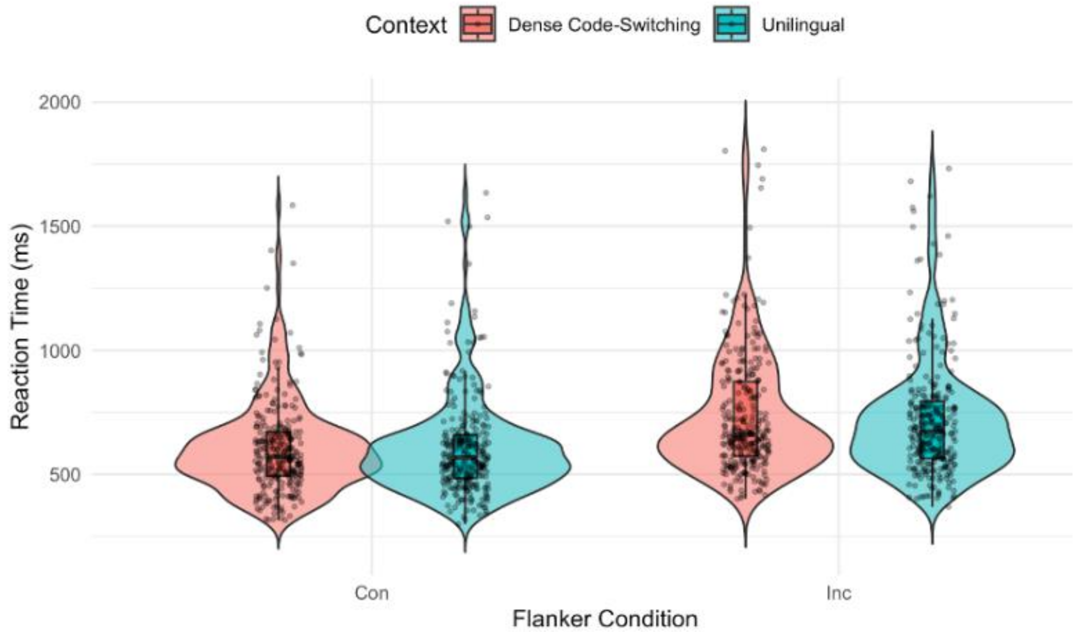


Figure 2: Reaction Times by Flanker Condition and Context



References

[1] Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 36*(2), 290–301. https://doi.org/10.1523/JNEUROSCI.2597-15.2016

[2] Jiao, L., Grundy, J. G., Liu, C., & Chen, B. (2020). Language context modulates executive control in bilinguals: Evidence from language production. *Neuropsychologia, 142*, 107441. https://doi.org/10.1016/j.neuropsychologia.2020.107441

[3] Valente, A., Pinet, S., Alario, F. X., & Laganaro, M. (2016). “When” does picture naming take longer than word reading?. *Frontiers in psychology, 7*, 31.

[4] Gvion, A., & Friedmann, N. (2016). A Principled Relation between Reading and Naming in Acquired and Developmental Anomia: Surface Dyslexia Following Impairment in the Phonological Output Lexicon. *Frontiers in psychology, 7*, 340. https://doi.org/10.3389/fpsyg.2016.00340

[5] Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 741–759. https://doi.org/10.1037/xlm0000755

[6] Kheder, Souad & Medina, Rodrigo & Valdés Kroff, Jorge & Kaan, Edith. (2024). Investigating the Effects of Code-Switch Types on Cognitive Control. 10.31219/osf.io/q9kzn

[7] Green, D. W., & Wei, L. (2014). A control process model of code-switching. Language, Cognition and *Neuroscience, 29*(4), 499–511. https://doi.org/10.1080/23273798.2014.882515

Teasing apart productivity and defaultness in time-frequency responses: an EEG on Russian

Natalia Slioussar[1], Kirill Bursov[2], Maria Kohova[2], Maya Zheltyakova[2], Alexander Korotkov[2], Denis Cherednichenko[2], Tatiana Chernigovskaya[2], and Maxim Kireev[2]
[1] HSE University, Moscow; [2] N.P. Bechtereva Institute for Human Brain, RAS, Saint-Petersburg
slioussar@gmail.com

Many studies discuss the role of morphological regularity in inflectional processing, but very often, it is difficult to say whether the observed effects are due to productivity of a certain inflectional class, to its type frequency, or, maybe, to its defaultness[1] (for EEG studies, see an overview in [1]). In English, for instance, these factors cannot be teased apart, and more evidence from languages with richer morphology, like Russian, is needed. Russian has many inflectional classes for verbs, including five highly frequent and productive ones (see Appendix). Previous psycho- and neurolinguistic studies suggested that one of them, AJ class, behaves as the default, thus lending support to dual route models. In particular, in [5] distinct defaultness and productivity effects were observed in fMRI data. We aimed to test their conclusions using the same stimuli, but a different method: the time-frequency analysis of EEG data.

In our experiment (N=28), we used 105 verbs from three inflectional classes: 35 AJ (potentially default, productive, highly frequent; e.g. *delat'* 'to do'), 35 I (productive, highly frequent; e.g. *nosit'* 'to bear, to wear') and 35 IRR(EGULAR) (including non-productive infrequent inflectional patterns; e.g. *brat'* 'to take'), as well as 105 pseudoverbs modelled after these verbs. Stimuli in different groups were matched for token frequency (for real verbs) and length. In every trial (see Fig. 1), participants first saw an infinitive and a personal pronoun presented for 600 ms. After a 1400 ms delay, they saw two present tense forms of the same (pseudo)verb for 1000 ms and were asked to select the one matching the pronoun. A question mark was shown on the screen for 2000 ms or until they pressed a button, and then an inter-trial interval with a fixation cross was initiated.

We used time-frequency analysis, which gave more interesting results than the analysis of ERP components. It is widely used to analyze EEG, but not with linguistic data (see an overview in [4]), especially not in morphological studies. We ran cluster-based permutations on Morlet wavelets to analyze oscillatory power changes in six frequency ranges (between 4, 8, 12, 19, 30, 60 and 100 Hz) and four time intervals (between 0, 102, 172, 512 and 1000 ms post-target). The statistical analysis evaluated channel clusters in averaged activity in defaultness (AJ vs. other classes) and productivity (IRR vs. other classes) conditions.

We found that defaultness and productivity elicited different patterns of EEG activity (see Fig. 2). Verbs from the purported default class (AJ) elicited a gamma power decrease (30-100 Hz) in intervals of 0-102 and 512-1000 ms post-target, while verbs from productive classes (AJ and I) increased theta power (4-8 Hz) in the left hemisphere. In addition, an early lexicality effect was observed in the theta band within 0-102 ms.

The decreased power in response to AJ class may be associated with easier processing, i.e., lower activity in the brain network involved in inflectional processing. This provides interesting parallels with fMRI data in [5]

[1] This notion is used only in dual route models (e.g. [2], [3]) that distinguish forms derived by the default rule from all other forms. This rule should be generalizeable independently from the phonological properties of the stems, while in non-default classes, generalizations to novel words always rely on analogy. Single route models of inflectional morphology assume that all forms are processed and stored relying on the same principles.

and elucidates what exactly may underlie the special status of the default class. More generally, our results support the dual system approach and show the benefits of the time-frequency analysis for linguistic data.

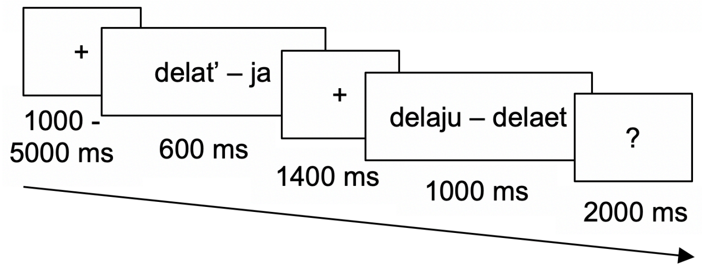


Figure 1. The structure of the trial.

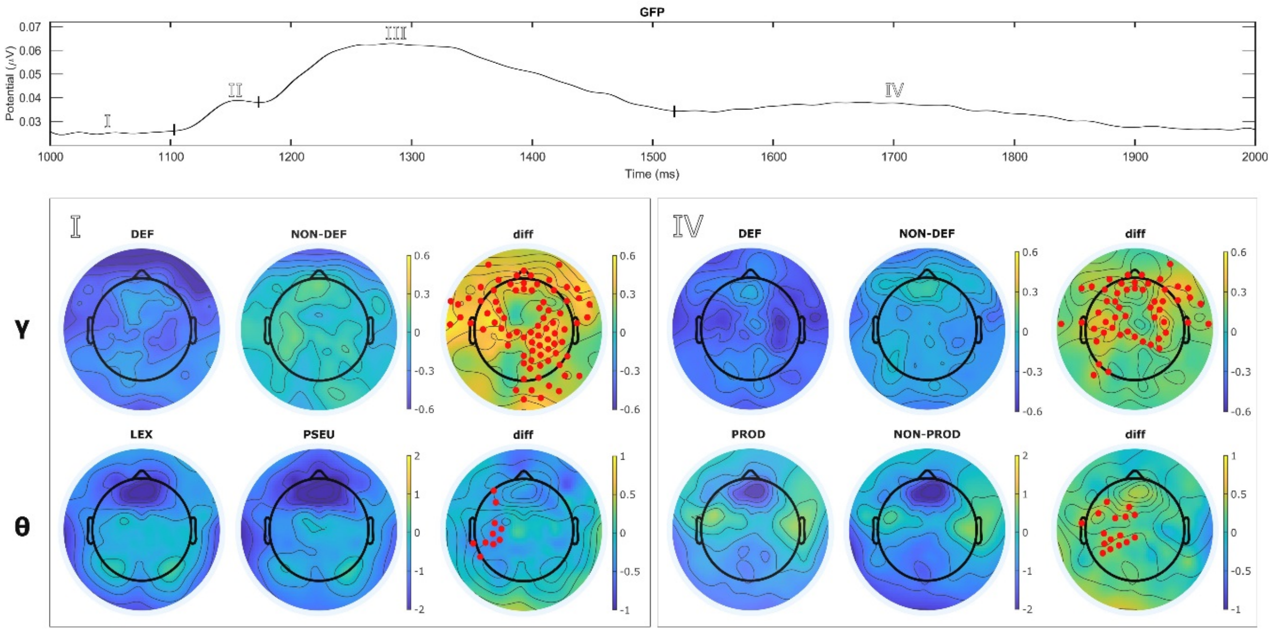


Figure 2. EEG results. Top: the global field power (GFP) function in response to all stimuli in the post-target interval. Local minima that define time intervals for the analyses are marked with black vertical lines along the function at 102, 172, and 512 ms post-target. Roman numerals signify the resulting time intervals. Voltage is shown in microvolts (μV). Bottom, left and right boxes: topographies of time-frequency responses for defaultness (DEF – NON-DEF), productivity (PROD – NON-PROD), and lexicality (LEX – PSEU). Time intervals with significant clusters are marked with Roman numerals in the top left corner of each box. Difference power plots with statistically significant channel clusters (red dots, $p < .05$) are plotted to the right of each pair. Defaultness effects were found in the gamma (γ, 30-100 Hz) band, while productivity and lexicality effects were found in the theta (θ, 4-8 Hz) band. Colour bars mark power in dB.

References:

[1] Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex*, *116*, 4–44.

[2] Pinker, S. (1991). Rules of Language. *Science*, *253*(5019), 530–535.

[3] Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn Sci*, *6*, 456–463.

[4] Prystauka, Y., & Lewis, A.G. (2019). The *power* of neural oscillations to inform sentence comprehension: A linguistic perspective. *Lang Linguist Compass, 13,* e12347, 1–40.

[5] Slioussar, N., Korotkov, A., Cherednichenko, D., Chernigovskaya, T., & Kireev, M. (2024). Exploring the nature of morphological regularity: an fMRI study on Russian. *Lang Cogn Neurosci*, *39*, 24–39.

Appendix: Inflectional classes of Russian verb

Russian verbs form a complex system of inflectional classes. According to the approach outlined in (Jakobson, 1948; Townsend, 1975; Davidson et al., 1996), it includes ten classes identified by their suffixes and the eleventh class with a zero suffix, subdivided into numerous subclasses[2]. All Russian verbs have two stems: the present/future tense stem and the past tense stem (also used in the infinitive form). Depending on the class, the correlation between them may include truncations or additions of the final consonant or vowel, stress shifts, suffix alternations, alternations of stem vowels and stem-final consonants. The verb class also determines which set of endings is used in the present and future tense (1st and 2nd conjugation types).

For example, *délat'* 'to do' belongs to the most frequent and productive AJ-class and has forms *déla-j-u* '(I) do$_{1SG}$', *déla-j-eš'* '(you) do$_{2SG}$' etc. in the present tense (the *-j-* suffix is added to the past tense stem, 1st conjugation type).[3] *Iskát'* 'to seek' belongs to the large, but non-productive A-class and has forms *išč-ú, íšč-eš'* etc. (the *-a-* suffix is truncated, 1st conjugation type, final consonant alternation, stress shift in some forms). *Nosít'* 'to bear, to wear' belongs to the second most frequent and productive I-class and has forms *noš-ú, nós-iš'* etc. (the *-i-* suffix is truncated, 2nd conjugation type, final consonant alternation and stress shift in some forms). *Brat'* 'to take' belongs to a small, non-productive IRR class and has forms *ber-ú, ber-éš'* etc. (the *-a-* suffix is truncated, 1st conjugation type, alternation in the root).

Importantly, the class is often unrecoverable from a particular form, and, unlike in English or German, there is no singular pattern that can be applied to any stem irrespective of its phonological characteristics. Five out of eleven classes in Russian are productive, and only gradual differences in type frequency are observed. According to Slioussar et al. (2014), there are 27409 verbs in the *Grammatical Dictionary of the Russian Language* (Zaliznjak, 1987), 23440 of them belong to productive classes: 11735 to the AJ-class, 6875 to the I-class, 2815 to the OVA-class, 1377 to the NU-class and 638 to the EJ-class. The non-productive A-class has higher type frequency than EJ-class with 842 verbs in the *Grammatical Dictionary.* Thus, in Russian, unlike in English, it is not immediately obvious which inflectional classes should be considered *regular*, whether this notion should be categorical or gradual, and whether a *default* pattern can be identified.

References:

Davidson, D. E., Gor, K. S., & Lekich, M. (1996). *Russian: Stage One: Live from Moscow.* Dubuque: Kendall Hunt Publishing Company.

Jakobson, R. (1948). Russian Conjugation. *Word*, *4*, 155–167.

Slioussar, N., Kireev, M.V., Chernigovskaya, T.V., Kataeva, G.V., Korotkov, A.D., & Medvedev, S.V. (2014). An ER-fMRI study of Russian inflectional morphology. *Brain and Language*, 130, 33-41.

Townsend, C. E. (1975). *Russian Word-formation*. Cambridge, MA: Slavica.

Zaliznjak, A. (1987). *Grammatičeskij slovar' russkogo jazyka: Slovoizmenenie [The Grammatical Dictionary of Russian: Inflection].* Moscow: Russkij Jazyk.

[2] Jakobson and Townsend counted them as 13 separate classes.

[3] In the examples in this paragraph, we indicate lexical stress, but do not mark it elsewhere.

Similar meaning does not always mean similar processing

Mercedes Martinez Bruera[1], Matilde Calmejane[2], Carolina Gattei[2], Carlos J. Alvarez[3], Horacio A. Barber[3], Daniel Weingärtner, Andrea Listanti[4], João Veríssimo[5], Sol Lago[1]
[1]Goethe University Frankfurt, [2]University Torcuato Di Tella, [3]University of La Laguna, [4]University of Cologne, [5]University of Lisbon
MartinezBruera@em.uni-frankfurt.de

When two grammatical structures convey the same meaning, do they show different real-time processing signatures? Optionality has been widely investigated in speech production [1–3], but less is known about its consequences for comprehension. We addressed this question by testing dative clitic doubling in Spanish. In most varieties of Spanish, a dative object can be optionally doubled by a pre-verbal dative clitic agreeing in number: *Antonio <u>les</u>$_i$ entregó una bebida <u>a las princesas</u>$_i$* (Antonio CL.$_{\text{DAT.PL}}$ gave a drink to the princesses). Both clitic-doubled and non-doubled variants share the same meaning and are highly acceptable [4]. However, we hypothesized that the clitic-doubled variant might elicit a processing advantage by preactivating some features of the referent prior to its appearance in a sentence. Further, we examined the role of language-induced variation by testing whether the prediction effect was stronger in varieties where doubling is more vs. less widespread: Rioplatense vs. European Spanish, respectively.

Method. Two adult Spanish speaker groups (94 Rioplatense and 98 European) performed a picture selection task using visual world eye-tracking. There were three experimental conditions (Table 1). Two of them had a clitic-doubled dative object but differed in their visual display. In the match condition, the target and competitor picture showed the same number of referents for the dative object, such that clitic number could not be used predictively. In the mismatch condition, the two pictures differed in the number of referents, such that the clitic number could be used predictively. This allowed us to examine the role of prediction in clitic-doubling sentences. Finally, the non-doubled condition had the same display as the match condition but no clitic. This allowed directly comparing the clitic-doubled and non-doubled variants while excluding the possibility of prediction. This way, we could directly diagnose the effect of optionality: for example, coreference at the dative object may be harder due to the effort of processing additional linguistic material (the clitic) in the match vs. non-doubled condition.

Results and discussion. The two speaker groups showed similar processing profiles across conditions. A generalized additive mixed model revealed that, in the indirect object window (e.g., 'to the princesses'), all three conditions showed significant changes over time with increased fixations to the target image. But across the whole time window, the mismatch condition had a higher proportion of fixations to the target image than the match condition (β = 0.83, z = 10.75, p < .001) while there was no evidence of a difference between the match and non-doubled conditions (β = 0.13, z = –1.66, p = .1), suggesting eased processing in the mismatch condition. To explain this effect, we analysed the prediction window (e.g., 'CL.$_{\text{DAT.PL}}$ gave a drink') and found that the mismatch condition showed a significantly different trend over time compared to the other conditions: while fixations increased in the mismatch condition, the others showed no effect of time. Overall, our results demonstrate that the processing of the optional variants only differed when the number of the clitic could be used predictively. Crucially, the additional linguistic material in the clitic-doubled conditions did not appear to have an additional cost or hinder comprehension in the absence of prediction. This shows that in cases of optionality, one of the variants can have a processing advantage, either in terms of avoiding temporary ambiguity [6], or in facilitating coreference, as in our study.

Table 1. Sample item in the three conditions (8 items/condition, the clitic always matched the indirect object number). Participants listened to descriptions of scenes during a theater play. In this set-up, two characters (Antonio and Julieta) handed objects (e.g., a drink) to actors around a theatre. Target image is framed with dashes.

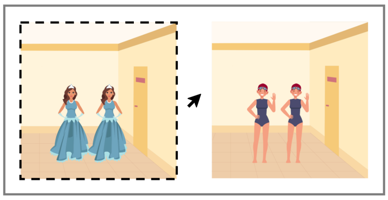
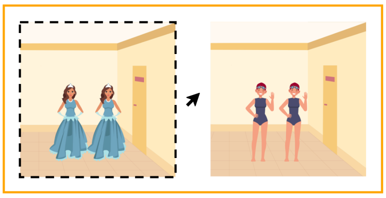
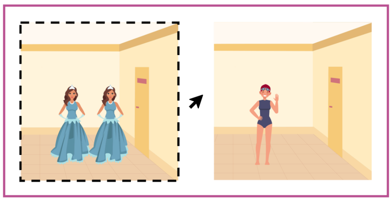
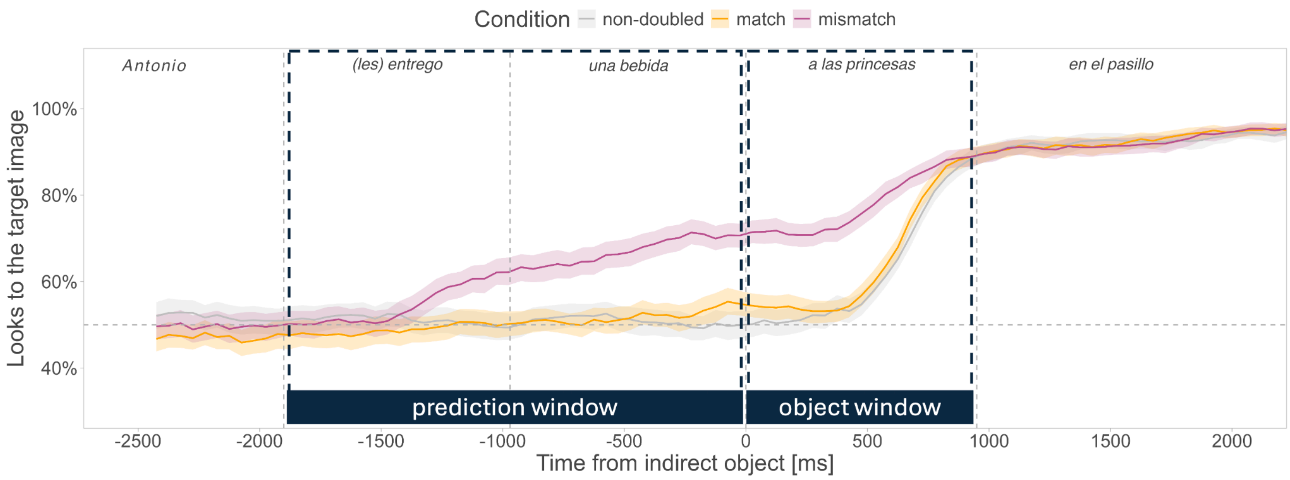
Condition	Sample sentence	Display
non-doubled	Antonio entregó una bebida a las princesas en el pasillo *Antonio gave a drink to the princesses in the hallway*	
match	Antonio les$_i$ entregó una bebida a las princesas$_i$ en el pasillo *Antonio* CL.$_{DAT.PL}$ *gave a drink to the princesses in the hallway*	
mismatch	Antonio les$_i$ entregó una bebida a las princesas$_i$ en el pasillo *Antonio* CL.$_{DAT.PL}$ *gave a drink to the princesses in the hallway*	

Figure 2. Time course data for the two windows of interest (collapsed across Rioplatense and European Spanish speakers)



References

[1] McGregor, W. B. (2013). Optionality in grammar and language use. Linguistics, 51(6), 1147–1204.
[2] Boyd, J. (2007). Comparatively speaking: a psycholinguistic study of optionality in grammar. *UC San Diego*. ProQuest ID: umi-ucsd-1697. Merritt ID: ark:/20775/bb0854076k. Retrieved from https://escholarship.org/uc/item/1wq3j7bk
[3] Jaeger, T. F. (2005). Optional that indicates production difficulty: evidence from disfluencies. In *DiSS* (pp. 103–108).
[4] Rinke, E., Lago, S., & Gattei, C. A. (2023). Diachronic change and variation in use: Judgments of clitic doubling in Peninsular vs. Rioplatense Spanish. *Isogloss*, *9*(1), 1–18.
[5] Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, *98*(3), 245–272.

A Progress Report on Ongoing Benchmark Data Collection for German Sentence Processing: Eye-Tracking and Self-Paced Reading

Michael Vrazitulis, Pia Schoknecht, Shravan Vasishth
University of Potsdam
{vrazitulis, schoknecht, vasishth}@uni-potsdam.de

In sentence processing research, theories are often developed to explain specific phenomena, such as the subject–object relative clause asymmetry. However, these theories are rarely evaluated against a broader range of empirical findings. A standard benchmark dataset covering multiple phenomena for model evaluation and comparison is currently missing.

Huang et al. [4] took a first step by creating a large-scale self-paced reading benchmark on English syntactic ambiguities (e.g., garden paths). They used this dataset to evaluate predictions from the surprisal metric [1, 7]. Their results revealed important gaps in the explanatory power of surprisal. More broadly, benchmark datasets allow researchers to quantify model predictions and systematically identify where models succeed and where they fail.

There is a pressing need for benchmark datasets based on eye-tracking measures and for languages other than English. It is also important to extend benchmarking beyond just effects of syntactic disambiguation. To address this, we are creating a benchmark dataset based on a large-sample eye-tracking study in German. The study covers a range of postulated effects, including garden-path ambiguities, agreement attraction, local coherence, interference effects, attachment ambiguities, and the relative clause asymmetry (for details, see Table 1). Each experimental design comprises three to four conditions, with three items per condition arranged in a Latin square. Trials are randomized individually for each participant. Each trial is followed by a binary-choice comprehension question targeting the critical dependency of the sentence. Participants whose accuracy on comprehension questions falls below chance level are excluded from analysis.

A complementary study is being conducted using self-paced reading (SPR) on the same materials. The eye-tracking data are collected in the lab, while the SPR data are collected online via Prolific.[1] By collecting both eye-tracking and SPR data for the same materials, we aim to enable direct comparisons across methods and study the relationship between different reading measures.

Data collection is ongoing. For eye-tracking, the current sample size (as of April 25, 2025) is 119 participants (pre-exclusion). For SPR, the current sample size is 659 participants (pre-exclusion). For eye-tracking, we plan to continue until all main effects and interactions across the tested phenomena reach 95% credible intervals of ±50 ms or narrower, based on total fixation times. The effect estimates are derived from a log-normal hierarchical model and backtransformed to milliseconds. For SPR, we will continue until 1,100 participants have been collected, as preregistered at `https://osf.io/wpra9?view_only=2945b83dddfe4731bd60d0103559d1b4` (anonymized link).

Preliminary analyses suggest that surprisal explains many of the observed effect patterns well, although some effects remain unexplained (see Figure 1). Word-by-word surprisal values were derived from a version of GPT-2 [12] pretrained on German corpora [14].

Once completed, the full benchmark dataset will be made publicly available to support quantitative model evaluation and evidence-based theory development in sentence processing.

[1] `https://www.prolific.com`

Table 1: Sentence processing phenomena and corresponding experimental designs.

GPSD (2×2): Garden Paths From Subject-vs.-Direct-Object Ambiguity	Ambiguous/Unambiguous × S–O/O–S — closely replicating [9]
GPSI (2×2): Garden Paths From Subject-vs.-Indirect-Object Ambiguity	Ambiguous/Unambiguous × Active/Passive — loosely replicating [10]
AGAT (2×2): Agreement Attraction in Grammatical Sentences	Singular-/Plural-Controller × Match/Mismatch — closely replicating [2]
LOCO (2×2): Local Coherence	Coherent/Incoherent × Intervener/No-Intervener — closely replicating [11]
SBIN (2×2): Similarity-Based Interference	Subject-Cue [Yes/No] × Animacy-Cue [Yes/No] — closely replicating [13]
GPCA (2×2): Garden Paths From Coordination Ambiguity	NP-/VP-Coordination × AP-/PP-Modifier — closely replicating [6]
GPMI (2×2): Garden Paths From Modifier-vs.-Indirect-Object Ambiguity	Modifier/No-Modifier × Ambiguous/Unambiguous — closely replicating [5]
RCSO (2×2): Subject vs. Object Relative Clauses	Subject/Object × Double-/Single-Embedding — German adaptation of [3]
SYAA (3×1): Syntax-Based Attachment Ambiguity	High-/Low-/Ambiguous-Attachment — closely replicating [8]
SEAA (3×1): Semantics-Based Attachment Ambiguity	High-/Low-/Ambiguous-Attachment — German adaptation of [15]

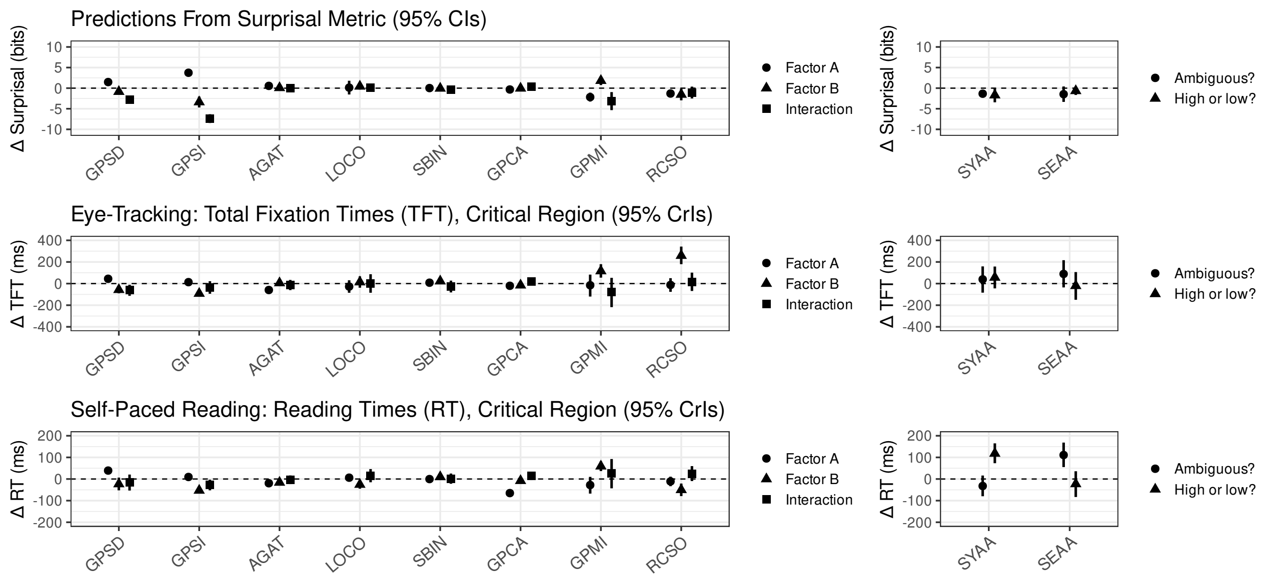


Figure 1: Predictions and observed effects across 2×2 and 3×1 designs. Rows show surprisal-based predictions, eye-tracking results (total fixation times), and self-paced reading results.

References

[1] J. T. Hale. In: *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*. Pittsburgh, PA, 2001. [2] J. Häussler. PhD thesis. University of Konstanz, 2009. [3] F. Hsiao and E. Gibson. In: *Cognition* 90.1 (2003), pp. 3–27. [4] K.-J. Huang et al. In: *Journal of Memory and Language* 137 (2024), p. 104510. [5] A. van Kampen. PhD thesis. Free University of Berlin, 2001. [6] L. Konieczny, B. Hemforth, and C. Scheepers. In: *German Sentence Processing*. Springer, 2000, pp. 247–278. [7] R. Levy. In: *Cognition* 106.3 (2008), pp. 1126–1177. [8] P. Logačev. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 49.9 (2023), p. 1471. [9] M. Meng and M. Bader. In: *Language and Speech* 43.1 (2000), pp. 43–74. [10] M. Meng and M. Bader. In: *Language and Cognitive Processes* 15.6 (2000), pp. 615–666. [11] D. Paape and S. Vasishth. In: *Language and Speech* 59.3 (2016), pp. 387–403. [12] A. Radford et al. In: *OpenAI Blog* 1.8 (2019), p. 9. [13] P. Schoknecht, H. Yadav, and S. Vasishth. In: *Journal of Memory and Language* 141 (2025), p. 104599. [14] B. Staatsbibliothek. https://huggingface.co/dbmdz/german-gpt2. 2020. [15] M. J. Traxler, M. J. Pickering, and C. Clifton Jr. In: *Journal of Memory and Language* 39.4 (1998), pp. 558–592.

Perception and production of gender-marking vowels in heritage Russian

Anna Runova & Zuzanna Fuchs
University of Southern California
runova@usc.edu

Background Speech perception and production differences remain understudied as a source of divergent grammatical gender systems in heritage languages. Only one production study to date (on heritage Spanish in the US) investigated phonetic differences in heritage speakers' realizations of gender-marking vowels /a/ and /o/, suggesting phonetic factors may contribute to divergent gender patterns [1]. However, this work did not investigate speech perception, where divergence between heritage and dominant speakers may originate. Our study looks at heritage Russian, because of three key properties: 1) Russian lacks word-level gender cues, unlike Spanish (c.f. the articles el (masc.) and la (fem)), see page 3 for gender in Russian; 2) in the nominative case, gender cues are realized as the vowels /a/ (fem) and /o/ (neut); 3) a stress-based phonological reduction rule reduces both /o/ and /a/ to /ə/ in unstressed contexts, which obscures gender cues. We investigate differences between Russian-dominant and Russian heritage bilinguals: a corpus-based production study extends prior work on Spanish to Russian; a perception study tests whether Russian heritage and Russian-dominant bilinguals differ in perception of fully-realized and reduced /a/ and /o/.

Methods – Production study We extracted gender-marked words from socio-linguistic interviews with Russian-dominant (n=4) and Russian heritage (n=4) bilinguals in the Heritage Language Variation and Change Project [2]. We measured F1 and F2 values for /a/, /o/, and / ə/.

Results – Production Linear mixed effect found a negative F2 effect of /a/. (Figure 1). The /a/ model found a negative effect of group on F2 (β = - 0.11, SE = 0.04, t = -2.884, p = 0.035), suggesting more backwards /a/ productions in heritage speakers. The /o/ model found a positive effect of group on F1 (β = 0.20, SE = 0.03, t = 6.49, p < 0.001), suggesting that heritage /o/ is lower. The /ə/ model found a negative effect of group on F1 (β = - 0.08, SE = 0.03, t = -2.73, p = 0.006), suggesting that /ə/ is higher in heritage speakers.

Methods – Perception study US-based Russian-dominant speakers (n = 29) and heritage Russian speakers (n =17) participated in an ABX task using nonce CV syllables to prevent influence from word-level context. Sound A and B were fully-realized /a/ and /o/, respectively, while sound X was one of seven synthesized steps on vowel continua /o/-/ə/, /a/-/ə/ and and the /a/-/o/ control continuum. Participants categorized sound X as either /a/ or /o/. We used F1 and F2 measures from our production study to set endpoints for our continua (Table 1.)

Results – Perception A multiple logistic regression model found a main effect of step number and an interaction between group and step number for the contrast between step 4-5 on the /a/-/ə continuum (β=1.00, SE=0.32, z=3.08, p=0.002; Figure 2). Results for the other continua did not find a significant main effect of group nor an interaction (Figure 3.)

Discussion We find between-group differences for all three vowels in our production analysis. Our perception study finds only limited evidence for group differences. Taken together, these results suggest that heritage and Russian-dominant speakers have similar perceptions of vowels in the absence of context (i.e. on a CV-perception task). However, the availability of word-level context in the production task may lead to group differences.

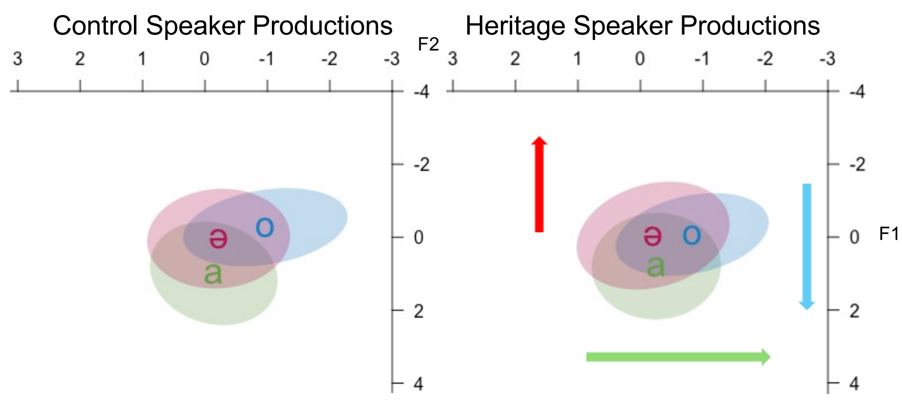


Figure 1. Lobanov-transformed F1 and F2 production values for the target vowels. Arrows represent significant effect, color coded by vowel.

step	F1	F2
1	1193	1617
2	1087	1612
3	983	1606
4	879	1601
5	775	1595
6	671	1589
7	567	1583

Table 1. F1 and F2 (Hz) values for 7 steps on the /a/-/ə/continuum.

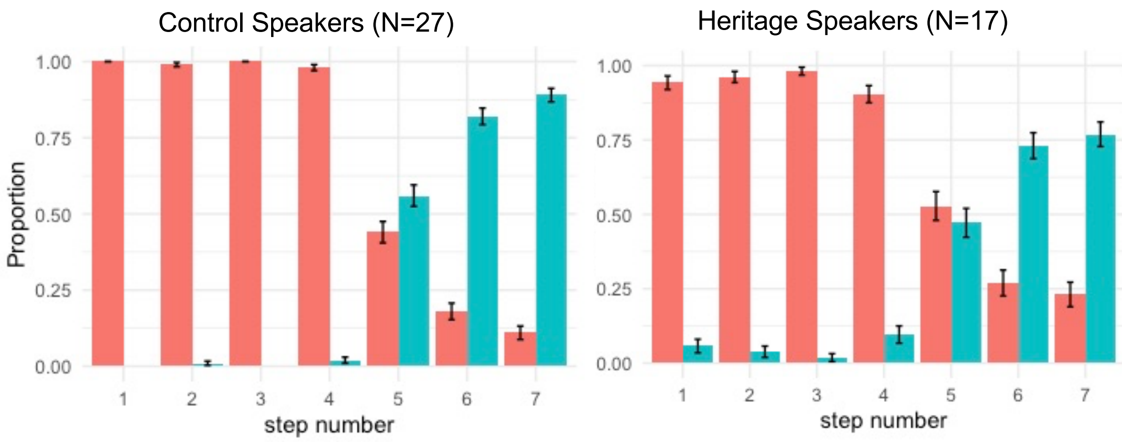
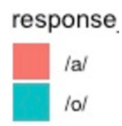


Figure 2. Proportion of selected vowels per step (/a/-/ə/ continuum).



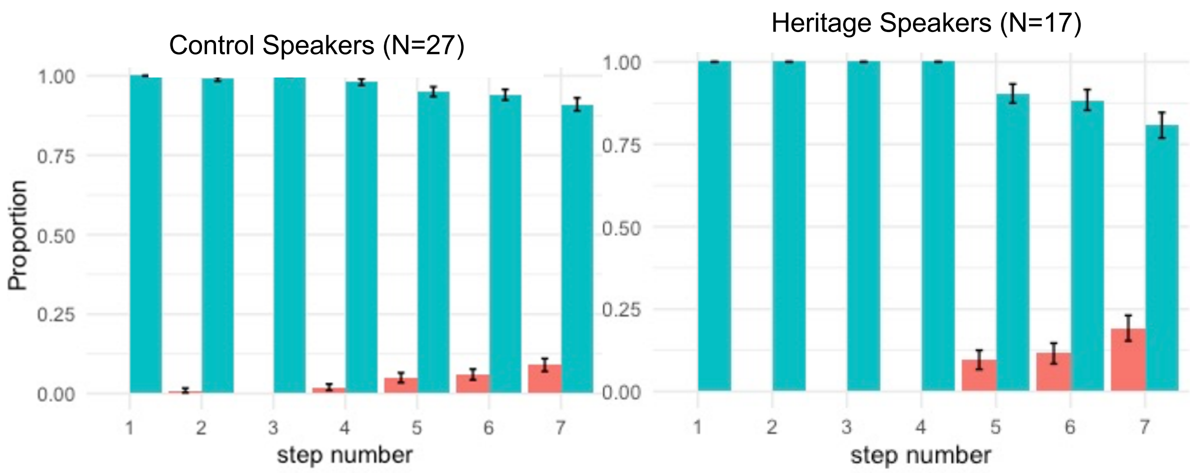


Figure 3. Proportion of selected vowels per step (/o/-/ə/ continuum).

References

[1] Colantoni, L., Martínez, R., Mazzaro, N., Pérez-Leroux, A. T., & Rinaldi, N. (2020). A Phonetic Account of Spanish-English Bilinguals' Divergence with Agreement. *Languages*, *5*(4), 58.

[2] Nagy, N. (2011). A multilingual corpus to explore geographic variation. *Rassegna Italiana di linguistica applicata*, *43*(1-2), 65-84.

[3] Liberman, A. M., Harris, K., Hoffman, H. S., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

[4] Corbett, G.G. Gender in Russian: An Account of Gender Specification and Its Relationship to Declension. *Russ Linguist* **6**, 197–232 (1982).

Grammatical gender in Russian

Russian has a three-gender grammatical gender system, with nouns assigned to masculine, feminine, and neuter gender [4]. Gender is marked on singular nouns, pronouns (including relative pronouns), demonstratives, adjectives, participles, numerals, quantifiers and singular past tense verbs. In nominative case, masculine gender agreement is indicated with a null morpheme (resulting in word-final consonants), feminine gender agreement is indicated with /a/, and neuter gender agreement is indicated with /o/. Masculine gender is considered to be the default gender.

Stress-based vowel reduction in Russian

Russian reduces unstressed /a/ and /o/ to /ə/. Gender markers are suffixes, which can be unstressed. When the disambiguating vowel is reduced, feminine and neuter gender become acoustically indistinguishable, as demonstrated in the following example

1) a*) bélaja kniga* ("white$_{\text{FEM}}$ book$_{\text{FEM}}$") is realized as [béləjə knigə]

b*) béloje oblako*("white$_{\text{NEUT}}$ cloud$_{\text{NEUT}}$") is realized as [béləjə oblakə]

English sentence planning differences between English L1 and Chinese-English L2 Speakers: Evidence from eye-tracking

Ren Li[1], Walter van Heuven[2], Chen Zhao[1]
[1]Guangdong University of Foreign Studies, China [2]University of Nottingham, United Kingdom
renli.ac.lr@gmail.com

Sentence planning is essential for fluent speech, mainly involving message formulation, lexical selection, and structural assembly. These processes are often incremental: speakers plan small units rather than full sentences before speaking. A central debate concerns whether sentence planning relies more on relational information (structure-driven account) or on the accessibility of individual lexical items (word-driven account). While these strategies have been well studied in native speakers, little is known about how second language speakers plan their sentences. A recent sentence production study [1] comparing sentence planning strategies in Dutch (L1) and English (L2) found that highly proficient Dutch-English bilinguals speakers tend to prioritize relational information linked to verb before processing noun-related information in English. In this study, we focused on Chinese-English speakers (n = 58) with lower English proficiency to investigate whether, and how, their sentence planning strategies differ from those of English L1 speakers (n = 35).

We conducted a sentence production experiment with eye-tracking. Participants were presented with simple transitive visual scenes involving two characters/entities (e.g., the woman chases the man) and were asked to produce sentences in English. The experiment manipulated the ease of naming characters/entities and the verb (codability) based on naming agreement data. L2 proficiency was assessed using LexTALE [2]. Analyses focused on speech onset latencies and eye movements in active sentences, the dominant response type.

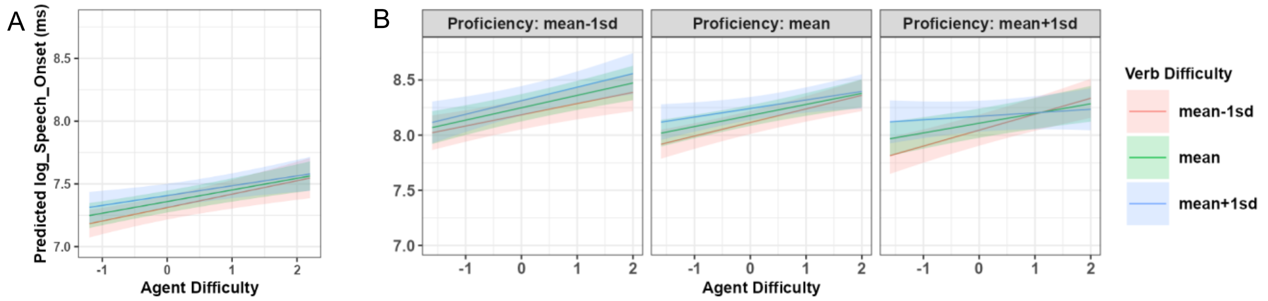
The results showed that speech onset latencies were shorter in English L1 speakers (mean = 1695 ms) than those in Chinese-English L2 speakers (mean = 3687 ms). The time course data, analyzed in the same way as [1], revealed distinct fixation patterns between English L1 speakers and Chinese-English bilinguals **(Figures 2–4)**. During the message formulation stage (0–400 ms), L1 speakers rapidly showed a strong preference for fixating on the agent. In contrast, L2 speakers distributed their attention more evenly between the agent and the patient. In the linguistic encoding stage (400 ms – speech onset), L1 speakers were influenced only by the ease of the agent, indicating a word-driven planning process. In contrast, L2 speakers showed a three-way interaction (agent difficulty × verb difficulty × proficiency), with verb difficulty exerting a stronger influence on less proficient speakers' sentence planning — a pattern that was also observed in speech onset times, suggesting a more complex and dynamic planning mechanism. Overall, data supports a structure-driven approach in L2 English speakers and a word-driven approach in L1 English speakers.

References

[1] Konopka, A. E., Meyer, A., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, *102*, 72-104.

[2] Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.

Figure 1: Speech Onset in English L1 Speakers (A) and Chinese-English L2 Speakers (B)



Note: Figure 1(A) shows predicted log speech onset latencies for English L1 speakers, with a significant main effect of Agent Difficulty. Figure 1(B) shows predicted log speech onset latencies for Chinese-English L2 speakers, with a significant interaction between Language Proficiency, Agent Difficulty, and Verb Difficulty.

Figure 2: Time Course of English L1 Speakers

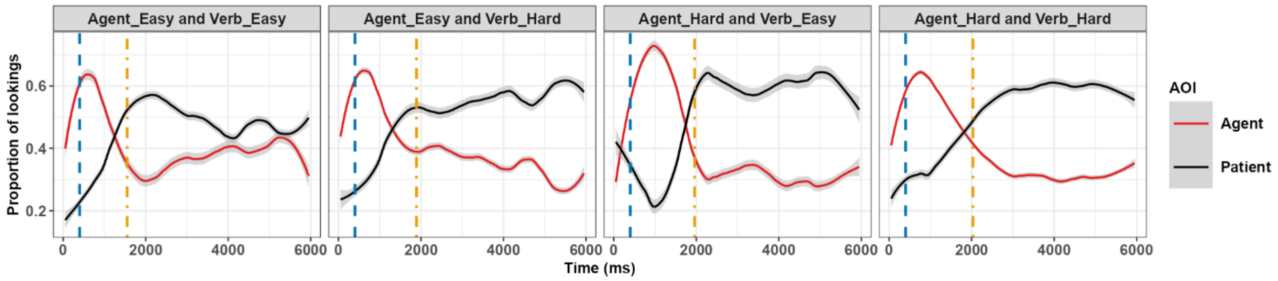


Figure 3: Time course of High Proficiency Chinese-English L2 Speakers

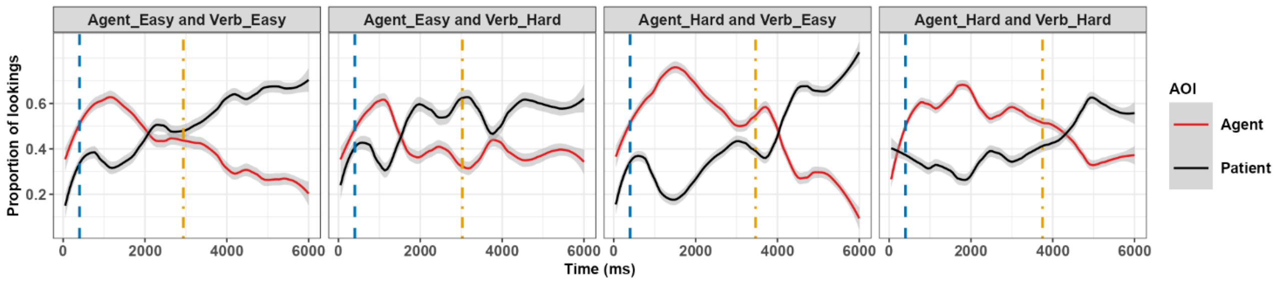
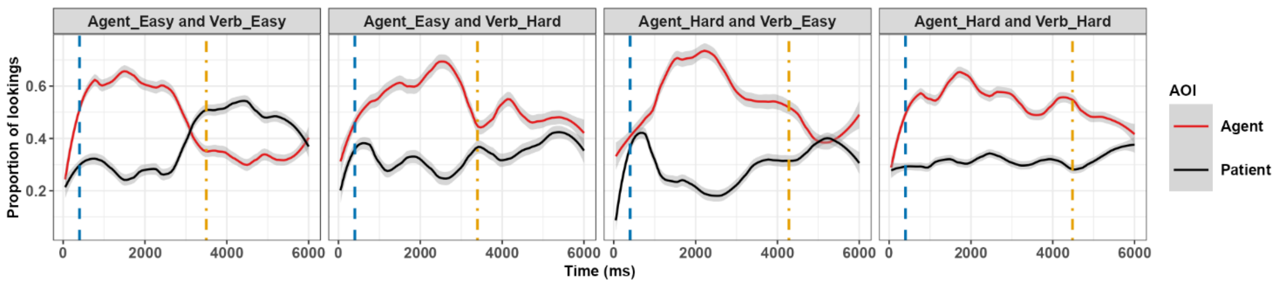


Figure 4: Time course of Low Proficiency Chinese-English L2 Speakers



Note: Figures 2–4 show the time course of eye fixations to the agent and patient during sentence production. Continuous predictors (agent difficulty, verb difficulty and language proficiency) are split into categorical factors (using the median) to help visualize the data. Blue vertical lines indicate 400 ms (the message formulation stage), and orange lines mark speech onset (the linguistic encoding stage).

Fluency and complexity in speech production: effects of healthy ageing

Mikael André Albrecht[1], Katrien Segaert[2], Eunice G. Fernandes[3], Allison Wetterlin[1], & Linda Wheeldon[1]

[1]University of Agder, [2]University of Birmingham, [3]University of Minho

Background. Healthy ageing is related to cognitive decline in processing speed, executive function, and working memory [1]. Language production relies on these processes and shows a decline with ageing e.g., in increased word finding difficulties [2] and reduced syntactic complexity [3]. However, vocabulary size and knowledge also increase with age [4] and syntactic processes can be preserved [5]. Many relevant studies report experiments with highly constrained productions and the measures are often limited to e.g., disfluencies. It has been comparatively rare for research to look at more complete measures speech fluency and complexity within the same sample. In addition, planning and planning scope have not been extensively investigated with respect to ageing [6]. The current study investigates multiple measures of speech production in healthy ageing in a free speech task.

Method. The sample comprised 120 English monolinguals divided into two age groups (see Table 1). Participants described one of four stimuli pictured scenes (Figure 1, see OSF project https://osf.io/shwcn/) on a monitor and described them with no further constraints. The four stimuli pictures were counter balanced so that an equal number of YAs and OAs described each picture. Speech files were transcribed and forced aligned using the Montreal Forced Aligner giving word- and pause durations. Fluent speech units (FSUs) were defined as any chunk of speech with no pauses > 250ms. FSUs were manually reviewed by a researcher to account for word-finding difficulties and disfluencies. The final FSUs are taken as representing the outcome of initial global and syntactic planning as opposed to more local word-finding processing.

Results. We first conducted Welch's t-tests on utterance duration and content by age group (Table 2). OAs produced longer utterances, an effect driven by longer medial pauses and more content words per utterance showing preserved crystalised knowledge and an OA vocabulary advantage. A repeated words effect was ruled out by beta regressions with non-significant age-group effects by utterance (p = .65) and across utterances within participant (p = .77) on the proportions of repeated words. More regressions examined the effect of initial pause length on the content of the upcoming utterance (see Table 3). Longer initial pauses predicted more complex utterances with more morphemes and content words consistent with these pauses reflecting initial planning processes. Initial pause length also predicted more and longer medial pauses. This is consistent with more segmented planning for utterances that impose greater cognitive load.

To examine planning, a Bayesian hierarchical model looked at the effect of initial pause length, medial pause length, and age group on mean word durations. .86 of the posterior distribution of the 3-way interaction was > 0 and so the direction of the effect has some reliability (Figure 2). There was an overall positive relationship between initial pause length and mean utterance word duration. However, the effect of initial pause length differed for OAs depending on medial pause length with a greater effect when medial pauses were also longer. That OAs show sensitivity to differences in medial pauses while YAs do not suggests that YAs plan larger chunks more thoroughly before speech onset while OAs distribute their speech planning in a more segmented manner when cognitive load is high. The current research thus shows that online speech planning during speech production is affected by age in a more natural setting and extends effects of ageing on speech production to such a context.

Figure 1. Illustration of scenes

Table 1: Descriptive statistics of the two age groups

Group	*n*	*M*	*SD*	Range
Older	80	65.41	4.78	60-81
Younger	40	23.08	5.27	18-35

Table 2: Summary of Welch's t-tests on measures of utterance content and length.

Measure	M_{Older}	$M_{Younger}$	*t*	*p*
Length (ms)	3373	2942	6.47	< .001
N morphemes	9.04	8.48	2.30	< .001
N syllables	9.52	8.71	3.23	.001
Total pauses (ms)	1113	1031	3.14	.002
Initial pause (ms)	898	871	1.17	.24
Medial pauses (ms)	214	160	5.13	< .001
N content words	3.39	3.01	3.51	< .001
N function words	4.09	3.94	1.21	.23

Table 3: Regression summaries of fluent speech unit and pause analyses

Dependent Variable	Initial pause (ms)		Age group		2-way interaction		Model
	z	*p*	*z*	*p*	*z*	*p*	
No. morphemes per FSU	2.82	.005*	-1.44	.15	0.44	.66	Neg. binom. GLMM
No. content words per FSU	2.36	.02*	-2.37	.02*	.23	.82	Neg. binom. GLMM
Length of medial pauses	4.47	< .001*	-2.81	< .001*	-0.07	.94	Tweedie GLMM
No. medial pauses	3.27	.001*	-2.50	.01*	0.41	.68	Neg. binom GLMM
Mean medial pause dur	0.91	.37	-1.29	.20	-1.90	.06	Lognormal GLMM
Syllables per-s within FSU	-19.05	< .001*	2.06	.04*	-2.01	.04*	LMM

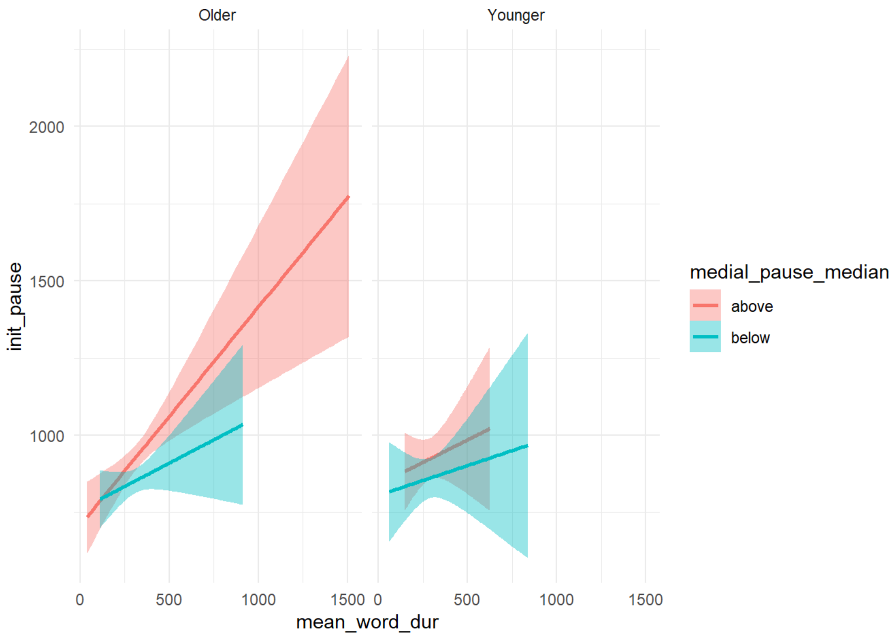


Figure 2: 3-way interaction between initial pause length, medial pause length, and age group on mean word duration. Note that the medial pause variable was entered as a continuous variable in the model

References.

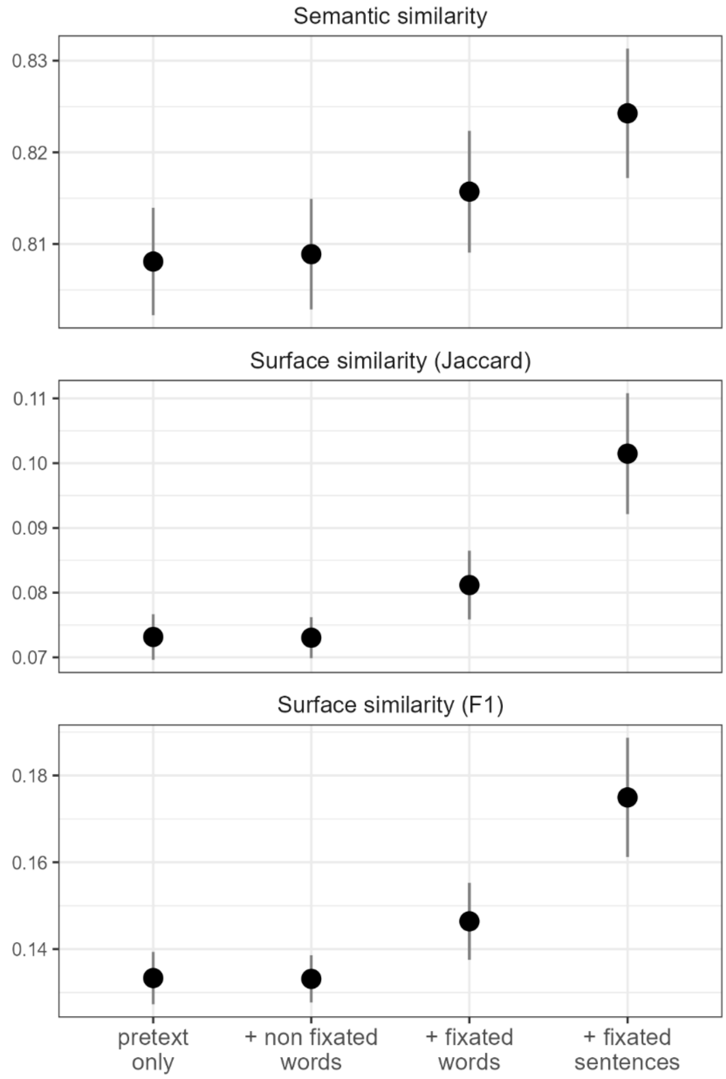
[1] Murman, D. L., (2915). The impact of age on cognition. *Semin Hear,* 36(3). [2] Segaert, K., Lucas, S. J. E., Burley, C. V., Segaert, P., Milner, A. E., Ryan, M., et al. (2018). Higher physical fitness levels are associated with less language decline in healthy ageing. *Sci. Rep*. 8:6715. [3] Kemper, S., Herman, R. E., and Lian, C. H. T. (2003). Age differences in sentence production. J. Gernotol. Psychol. Sci. 58, 260–269. [4] Verhaeghen, P. (2003). Aging and vocabulary scores: a meta-analysis. *Psychol*. [5] Altmann, L. J. P., and Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Lang. Cogn. Process*. 21, 322–354. [6]. Hardy, S. M., Segaert, K., & Wheeldon, L. (2020). Healthy aging and sentence production: Disrupted lexical access in the context of intact syntactic planning. *Frontiers in psychology*, 11, 257.

Brief lookback cues content generation in spontaneous multi-sentence text production

Astha Singh[1] , Evgeny Chukharev[1], Mark Torrance[2]
[1]Iowa State University, [2]Nottingham Trent University
mark.torrance@ntu.ac.uk

Background: Multi-sentence text composition is typically remarkably fluent. Even at major planning units - at the start of sentences, for example - competent writers often proceed with negligible hesitation [1]. This suggests processing that is largely parallel with message preparation occurring concurrently with output, and that is just-in-time [2], with meaning generated as needed during production. Pausing, when it does occur, is often associated with eye movement back into the writers own text. This is typically characterized by writers hopping between informationally rich words, rather than sustained reading, and often occurs without error correction [3]. It seems probable that this lookback serves, in part, to reinstate a representation of the message of the text already written, and therefore to cue what to say next, thus maintaining thematic coherence and production fluency. This hypothesis has not previously received a direct test.
Method: We logged keystrokes and tracked eye movement in 30 competent adult writers each composing two short argumentative essays, using methods that permitted extraction of the words and sentences fixated within the writers own, growing text [4]. We identified where writers looked back into their text fixating a minimum of three words, and then, without error correction or revision continued writing. For each of these lookback/completion events we then compared the writer's completion – the text they wrote next up to the next sentence terminator - with completions produced by a generative LLM under four different prompting conditions: In all cases the LLM was prompted with the essay topic and with all of the writers existing text. To this prompt we added one of (a) the words that the writer fixated (ignoring closed-class words), (b) the text of the sentences that contained fixated words, (c) a matched number of open-class words that were not fixated (control), and (d) nothing (control). The sentences condition was included on the grounds that fixating one or more open-class words within a recently-written sentence is likely to activate the meaning of the sentence as a whole. We sampled 10 LLM responses per lookback/completion event, with each of 4 different language models (GPT-3.5, GPT-4, LLaMa3-8B and Mistral7B). *Semantic similarity* between human and LLM completions was established by embedding using OpenAI's text-embedding-ada-002 and computing cosine similarity. We also report *Jaccart* and *F1* (the harmonic mean of precision and recall) indices to quantify Surface similarity.
Results: Frequentist linear mixed effects models with maximal random effects structure, and prompting condition as fixed effect gave significantly better fit than an intercept only model for all three similarity measures. Planned comparisons showed significantly greater similarity between LLM and human completions when LLM prompts included fixated words or sentences. Different language models gave substantively equivalent effects. Effect estimates are given in Figure 1 (just for GPT-4 completions).
Discussion: Semantic similarity between human and LLM text was greater when words fixated by the human writer immediately prior to production were allowed to shape the LLM response. This provides, we believe for the first time, provisional support for the hypothesis that lookback during text production supports rapid, inline content planning.

Figure 1. Estimated mean similarity between human-writer sentence completions and completions from LLMs prompted without and with information about the writer's fixations within their own text, with 95% CI.



References

[1] Roeser, J., Conijn, R., Chukharev, E., Ofstad, G. H., Torrance, M. (2025). Typing in Tandem: Language Planning in Multisentence Text Production Is Fundamentally Parallel Typing in Tandem : Language Planning in Multisentence Text Production Is Fundamentally Parallel. *Journal of Experimental Psychology: General*, *Advance Online*. https://doi.org/https://doi.org/10.1037/xge0001759

[2] Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

[3] Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2016). Reading during the composition of multi-sentence texts: an eye-movement study. *Psychological Research*, *80*(5), 729–743. https://doi.org/10.1007/s00426-015-0683-8

[4] Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H.-H. (2019). Combined Deployable Keystroke Logging and Eyetracking for Investigating L2 Writing Fluency. *Studies in Second Language Acquisition*, *41*(3), 583–604. https://doi.org/10.1017/S027226311900007X

Shifting, Inhibition and Updating in younger and older Basque-Spanish bilinguals

Patricia Fuente-García[1], Julián Villegas[2] & Irene de la Cruz-Pavía[3,4]

[1]University of the Basque Country (UPV/EHU); [2]University of Aizu; [3]University of Deusto; [4]Basque Foundation for Science, Ikerbasque

Despite extensive research, the relationship between bilingualism and executive control (EC) remains unclear and widely debated. A central assumption behind proposed bilingual advantages is that managing two co-activated languages across the lifespan—through selection, switching, and inhibition—enhances domain-general EC. While some studies report bilingual benefits, particularly in children and older adults, others fail to replicate such effects, producing inconsistent findings across tasks, populations, and methods [1,2]. These inconsistencies may reflect both theoretical and methodological limitations, including a shift in how the cognitive effects of bilingualism are conceptualized—from early accounts based on domain-general transfer to more recent adaptive frameworks emphasizing experience-driven changes in control systems [3], This study aims to clarify the bilingualism–EC relationship by treating both constructs as continuous and multifaceted. We investigate whether age of acquisition (AoA), proficiency, language use, and frequency of language switch modulate performance across three EC functions—shifting, inhibition, and updating [4]—in a uniform population of Basque-Spanish bilinguals.

We tested 124 younger (YA; *Mage*=21.31,18–35) and 98 older bilinguals (OA; *Mage*=65.63,60–79) from the Basque Country, spanning a continuum of AoA, proficiency, language use and frequency of language switch in Basque. Bilingualism was assessed using a combination of self-reports and objective measures (naming, lexical decision and Basque knowledge test for proficiency; voluntary switching task for frequency of language switch). We controlled for fluid intelligence (Raven-2) and screened OA's cognitive health (MMSE-37). Participants completed six tasks, each with verbal and non-verbal versions, targeting shifting (Verbal Fluency [VF]; Trail Making Test [TMT]), inhibition (Stroop; Flanker), and updating (Keep Track; N-back). The VF task included fixed- and alternating-category blocks, the Flanker included high/low monitoring blocks, and the Keep Track and N-back tasks followed a graded-difficulty approach (2–4 words; 2- and 3-back). We analyzed YA and OA together using mixed-effects regression models, including Raven's scores as a covariate.

Consistent with prior research, OA exhibited slower RTs and reduced accuracy relative to YA, particularly under high cognitive demands. Interestingly, OA showed reduced Flanker conflict effects and matched YA performance in demanding conditions of Keep Track and VF. As for the role of the bilingual dimensions, in shifting, higher proficiency reduced interference in TMT (p=.023), and earlier AoA increased switching in VF (p=.009). In inhibition, Stroop effects were attenuated by frequent language use (p=.0013), and in Flanker, later AoA predicted greater conflict effects (p=.003), which increased under low-monitoring demands, while OA showed stronger conflict adaptation (p<.001). In updating, switching improved performance under low load in Keep Track (p=.009). In N-back, accuracy declined with load (p<.001), especially for targets (p<.001), and in YA, earlier AoA (p=.035) and the interaction of higher fluid intelligence and integrated language use enhanced performance (p=.0042). Fluid intelligence also predicted TMT performance in OA (p=.010) and accuracy in Keep Track (p=.008). Overall, these findings show that bilingualism modulates executive control in domain-specific ways, with different bilingual dimensions influencing performance across tasks and age groups. Notably, OA showed comparable performance to YA in demanding updating conditions and reduced conflict effects in inhibition. In these cases, fluid intelligence predicted OA performance, suggesting a potential compensatory reliance on domain-general cognitive resources, which may reflect a form of cognitive reserve shaped by lifelong bilingual experience, helping to sustain executive functioning in older age.

References

[1] Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. Psychological Bulletin, 143(3), 233–262.

[2] Paap, K. R. (2019). The bilingual advantage debate. In J. W. Schwieter (Ed.), The handbook of the neuroscience of multilingualism (pp. 701–719).

[3] Bialystok, E., & Craik, F. I. M. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. Psychonomic Bulletin & Review, 29, 2300–2309.

[4] Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49–100.

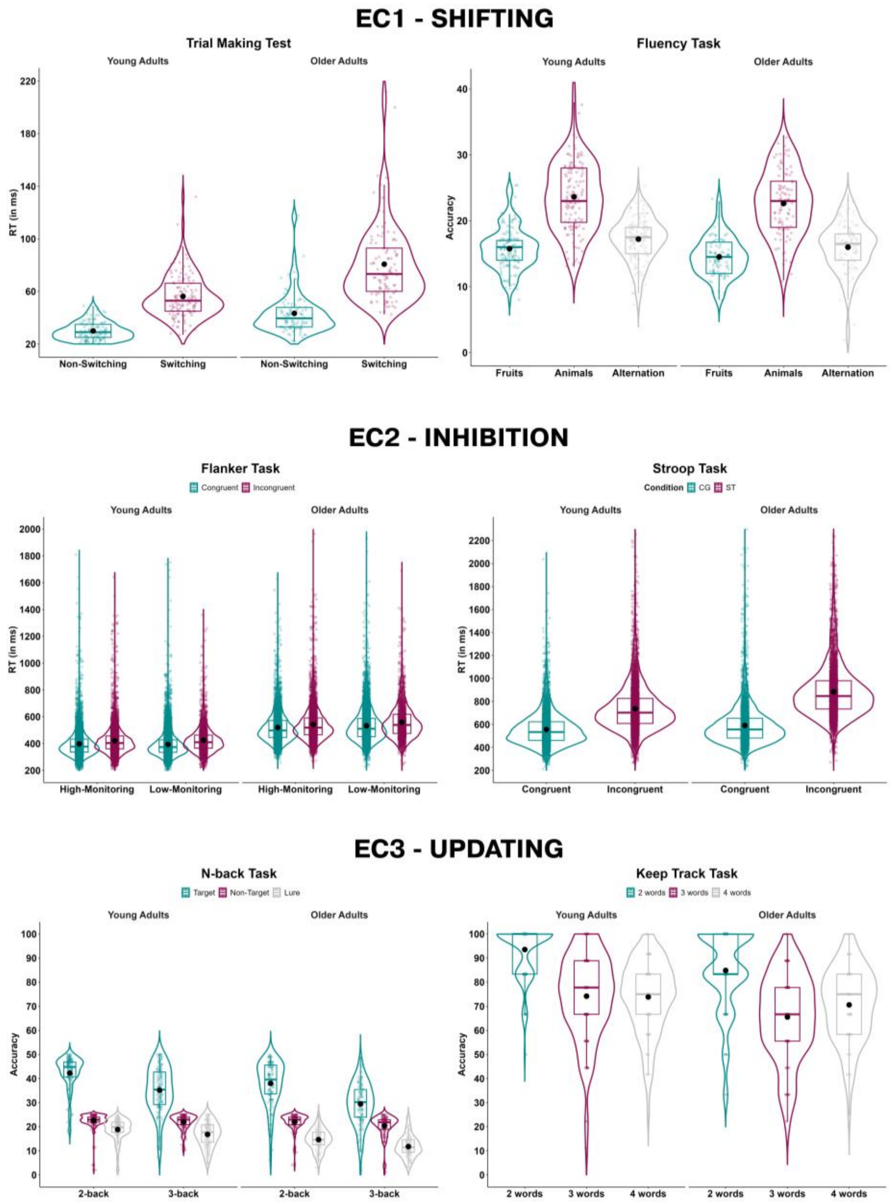


Figure 1. The figure shows raw data from six EC tasks: shifting (top), inhibition (middle), and updating (bottom). Each EC function includes a non-verbal task (left) and a verbal task (right). Young adults are shown on the left of each plot, and older adults on the right. Violin plots, boxplots, and jittered points display the distribution and variability of individual performance.

Thursday Afternoon Posters

Syntactic engagement of newly learned words: a garden-path method applied to track emerging sensitivity to structural ambiguity

Verónica García-Castro[1], Norbert Vanek[2]

[1] University of Costa Rica; [2] University of Auckland

norbert.vanek@auckland.ac.nz

Background: This study proposes an innovative approach to examining garden-path (GP) effects in sentence processing. It applies GP as a method to a new domain, specifically to syntactic engagement of recently learned verbs. The theoretical basis is built on three pillars. The first is that garden pathing can reliably signal effort to detect and resolve subject-object ambiguities widely documented in L1 research [1,2]. Second, GP effects are well established as indicators of effort to repair initially misanalysed sentences in L2 learners too [3,4]. And third, while literature on emergent semantic representations for new words already exists [5], this is the first study to map new word knowledge as a window into emergent structural representations, while controlling variation in phonological working memory (PWM) and vocabulary size [6].
Method: To test the method's validity, native English speakers and Spanish advanced L2 learners of English were trained on 12 pseudo-verbs presented first with definitions (e.g., *grod* = to wash quickly but thoroughly) and then embedded in meaningful English sentences. After training, a surprise meaning recall post-test was administered to assess participants' knowledge of the recently learned verbs. The next day, the participants carried out a reading task while their eye movements were monitored. They were presented with temporarily ambiguous GP sentences, each containing a recently learned pseudoverb. Every pseudoverb appeared twice in a plausible sentence (e.g., *While the woman grodded the baby fell in the bath*) and twice in an implausible sentence (e.g., *As the girl grodded the dough felt very soft*). Total fixation times and regressions out of the disambiguating region (*the baby fell/the dough felt*) were taken to indicate late processing [8] and revision effort. We also tested vocabulary size and PWM as factors known to affect sentence processing [7].
Results: The analysis of total reading times showed a significant main effect of plausibility (β =0.18, *SE*=0.08, *t*=2.2, *p*=0.026) with plausible GP sentences taking longer to read (*M*=516ms, *SD*=333) than implausible sentences (*M*=431ms; *SD*=281). This difference suggests that both groups integrated the ambiguous DP into their current parse, which caused more difficulty to revise when garden-pathed in plausible sentences. We interpret these results that readers strongly committed to their initial semantic and syntactic analysis of the ambiguous DP; thus, they spent more time reading it in the plausible condition because implausible analyses are easier to abandon [2]. An interaction between the pseudoword recall and plausibility was found in regressions out of the disambiguating region (β=-0.20, *SE*=0.10, *t*=-2.016, *p*=0.044). Readers took longer to process that region when their knowledge of the novel words was more robust.
Discussion: These results jointly suggest that the method shows good potential to track that L1 speakers and L2 learners engaged their grammatical knowledge of the recently learned pseudowords with other lexical items. To critically assess the approach, its main advantage is monitoring how native speakers and L2 learners (re)construct structural representations for new words. Unlike traditional GP studies, pseudoword use controls syllabic complexity and avoids pre-existing structural associations. A limitation is the potential blurring between integrating new structures into emerging L2 syntax versus their already established L1 syntax, but this can be addressed with a pre-screening task for syntactic ambiguity sensitivity. The GP with pseudowords approach enriches existing sentence processing research, but it could also extend to new, exciting areas, such as supervised sensitisation to ambiguity that varies across the learners L1 and L2.

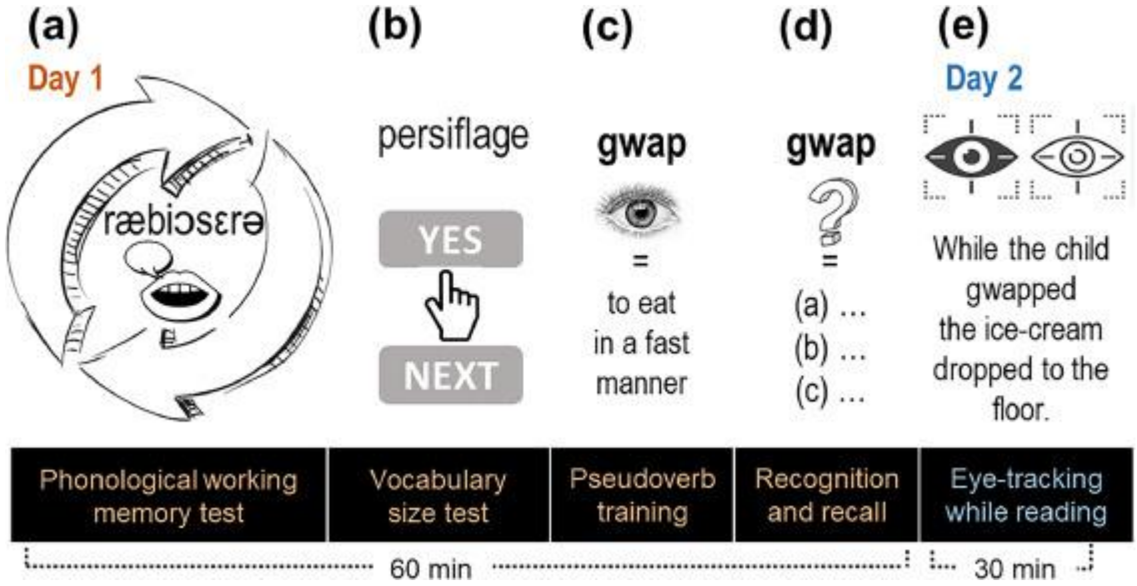


Figure 1. Schema of the procedural steps on Day 1 (orange) and Day 2 (blue).

References

[1] Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

[2] Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(4), 940–961. https://doi.org/10.1037/0278-7393.24.4.940

[3] Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. Second Language, 17, 5–27. https://doi.org/10.11431/secondlanguage.17.0_5

[4] Jegerski, J. (2012). The processing of subject–object ambiguities in native and near-native Mexican Spanish. Bilingualism: Language and Cognition, 15(4), 721–735. https://doi.org/10.1017/S1366728911000654

[5] Bordag, D., Kirschenbaum, A., Rogahn, M., Opitz, A., & Tschirner, E. (2017). Semantic representation of newly learned L2 words and their Integration in the L2 Lexicon. Studies in Second Language Acquisition, 39(1), 197–212. https://doi.org/10.1017/S0272263116000048

[6] Nation, I. S. P. (2001). Learning Vocabulary in Another Language (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

[7] Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. Applied Psycholinguistics, 32(2), 299–331. https://doi.org/10.1017/S0142716410000421

[8] Conklin, K., Pellicer-Sanchez, A., & Carrol, G. (2018). Eye-tracking: A guide for applied linguistics research. Cambridge University Press.

Language specific differences in morphological processing: The role of semantics in Maltese vs. Hebrew lexical access

Adam Ussishkin[1], Jessica Nieder[2]
[1]University of Arizona, [2]L-Università ta' Malta
adam.ussishkin@gmail.com

What is the role of semantics in word recognition? Some research supports models in which form is processed before meaning (form-then-meaning accounts; e.g., Rastle, Davis, and New, 2004), while others support models in which form and meaning may be processed together (form-with-meaning accounts; e.g., Feldman et al., 2012). Here, we address this model discrepancy by examining auditory word recognition in two Semitic languages: Maltese and Hebrew. Our results help clarify whether meaning is processed with or after form, which we find may be a language-specific phenomenon.

Maltese and Hebrew both exhibit nonconcatenative root-and-pattern morphology, in which the only form shared among related words may be their consonants; this structure is known as the consonantal root. Recent psycholinguistic research on Maltese has underscored the critical role of the consonantal root in lexical processing while at the same time highlighting that semantic information significantly influences lexical access from its earliest stages (Ussishkin et al., 2015; Nieder et al., 2024). This suggests that Maltese supports a form-with-meaning model of word recognition. Unlike Maltese, prior work on Hebrew (Frost et al., 1997; Geary and Ussishkin, 2019) suggests that semantic relatedness does not confer a processing advantage beyond the effects of morphological relatedness, based on facilitated response times even when prime-target pairs sharing a root are not semantically related.

Here, we report on two auditory lexical decision experiments in Hebrew with 66 participants each, testing recognition of target verbs from the five non-defective verbal binyanim auditorily presented in their third-person singular past tense masculine form. Participants responded to a counterbalanced set of 60 prime-target pairs across three conditions: identity, morphologically related, and unrelated. Each condition included 20 targets, evenly distributed across the binyanim for both primes and targets. In experiment 1a, primes were audible, whereas in experiment 1b, primes were masked using the auditory masked priming technique of Kouider and Dupoux (2005), previously reported on for Maltese by Ussishkin et al. (2015), ensuring primes were not consciously perceived.

Inspired by Nieder et al. (2024)'s work on Maltese, we incorporated a post-hoc measure of semantic similarity, calculated as the cosine similarity of word embedding vectors for primes and targets taken from a Hebrew BERT model. These vectors captured varying levels of prime-target similarity across conditions. However, linear mixed-effects models including the similarity measure revealed no significant effect or better model fit for this predictor but a robust effect for the morphological condition and form-based predictors included in both experiments. Consistent with Frost et al. (1997) and Geary and Ussishkin (2019), our results indicate a strictly morphological priming effect in Hebrew, with no contribution from semantics, crucially distinguishing it from the related Semitic language Maltese. In other words, Hebrew supports a form-then-meaning model of word recognition.

These findings have important implications for both theoretical and experimental approaches of morphological processing, and reveal that two closely related languages, both of which exhibit root-and-pattern morphology, employ distinct mechanisms in the word recognition process.

References:

Feldman, L. B., Kostić, A., Gvozdenović, V., O'Connor, P. A., & Moscoso del Prado Martín, F. (2012). Semantic similarity influences early morphological priming in Serbian: A challenge to form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *19*, 668-676.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 829–856.

Geary, J., & Ussishkin, A. (2019). Morphological priming without semantic relationship in Hebrew spoken word recognition. *Proceedings of the Linguistic Society of America*, 4(1), 9:1–10.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, 16(8), 617–625.

Nieder, J., van de Vijver, R., & Ussishkin, A. (2024). Emerging Roots: Investigating Early Access to Meaning in Maltese Auditory Word Recognition. *Cognitive Science*, 48(11), e70004.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11, 1090–1098. doi:10.3758/BF03196742

Ussishkin, A., Dawson, C. R., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, 30(9), 1096–1115.

Transformers fail to predict consistent effects for agreement attraction configurations

Titus von der Malsburg (University of Stuttgart), Sebastian Padó (University of Stuttgart)
titus.von-der-malsburg@ling.uni-stuttgart.de

In English, as in other languages, a morpho-syntactic **agreement mismatch** between a verb and its subject leads to ungrammaticality (1a). However, a structurally unrelated noun matching the verb in number (1b) can create an **agreement attraction effect**: an illusion of grammaticality reflected in faster reading times at the relative clause verb in (1b) compared to the equally ungrammatical (1a). Crucially, agreement attraction occurs reliably only with singular subjects, and no analogous effects are typically seen in grammatical sentences [1].

(1) a. *[The marine$_{\text{sg}}$] who [the officer$_{\text{subj-sg}}$] want$_{\text{pl}}$ to promote …
b. *[The marines$_{\text{pl}}$] who [the officer$_{\text{subj-sg}}$] want$_{\text{pl}}$ to promote …

Agreement attraction provides a key test case for neural language models (NLMs) which have been argued to show great potential as psycholinguistic models of human language processing [2]. Indeed, recurrent NLMs, specifically LSTMs, were found to broadly predict human agreement attraction effects [3, 4]. Transformers have surpassed recurrent models in language modeling, but are understudied for this phenomenon: Existing studies [e.g., 5, 6, 7] are limited in that they 1.) used the bidirectional, cognitively less plausible BERT model, 2.) investigated only a single transformer model type, or 3.) did not test all relevant linguistic configurations.
We addressed these limitations by investigating a range of autoregressive transformers in all linguistically relevant configurations, assessing their performance with respect to the singular/plural and grammatical/ungrammatical asymmetries attested in humans.

Design: We used 384 sentences (48 sets) like (2) from experiments 2 and 3 in [1]. Factors were number of RC subject, grammaticality, and attraction (2×2×2 design) resulting in a complete sweep of relevant configurations. Eleven pre-trained autoregressive transformers from different architectures were tested, including monolingual (GPT, GPT-2) and multilingual (Bloom, XGLM, Gemma 3) model families.

Results: Fig. **??** shows surprisal predictions for all models across the eight conditions. To test the above-mentioned theoretically informative asymmetries, we used hierarchical Bayesian regression with crossed random effects for items and models (i.e., we treated transformer models like subjects in a human study). **Singular/plural RC-subject asymmetry:** 1. In grammatical sentences, models show no attraction when the RC-subject was singular but attraction when it was plural. This deviates from humans, who show no effects in both cases. 2. In ungrammatical sentences, models show attraction effects when the RC-subject was singular and no evidence for reduced attraction when it was plural. This again deviates from humans, who show agreement attraction effects only in the singular condition. **Grammaticality asymmetry:** 3. In sentences with singular RC-subject, models predicted no attraction in the grammatical condition but attraction in the ungrammatical condition. This is consistent with humans. 4. In sentences with plural RC-subject, models showed attraction for grammatical sentences and no evidence for reduced attraction in the ungrammatical condition. This deviates from humans, who show no effects in either case. We also analyzed models individually and found that no model predicted a pattern consistent with human performance (see Tab. 1). There was no consistent pattern of effects across model families but some consistency within families.

Conclusions: Unlike prior findings with LSTMs, autoregressive transformers show no consistent surprisal effects in agreement attraction configurations. All models diverged markedly from human performance, even predicting effects in the opposite direction of what could be explained by attraction principles. These findings show that it is not safe to generalize from individual models to transformers in general, and thus cast significant doubt on transformers as universal models of human language processing.

(2) a. The marine(s) who the officer wants to promote … (singular RC subject, grammatical)
b. *The marine(s) who the officer want to promote … (singular RC subject, ungrammatical)
c. The marine(s) who the officers want to promote … (plural RC subject, grammatical)
d. *The marine(s) who the officers wants to promote … (plural RC subject, ungrammatical)

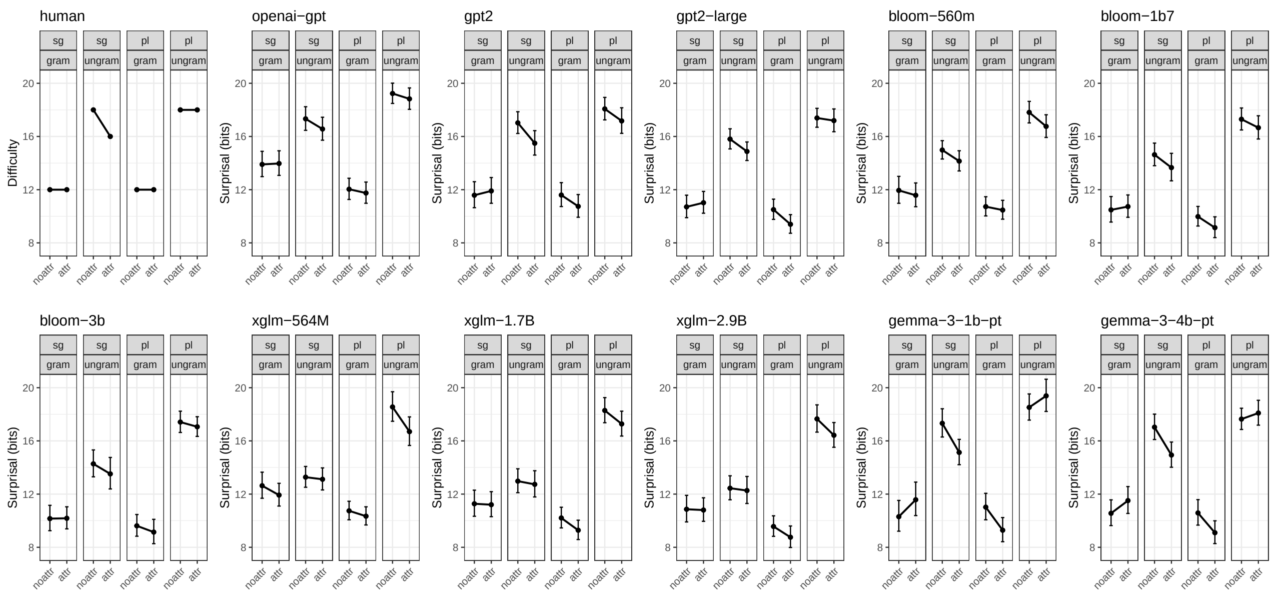


Figure 1: Surprisal predictions across models and conditions. Human performance is schematically represented in the first panel. Each sub-panel shows the effect of attraction in one of the four relevant configurations (see 2). Results for singular RC-subjects (sg) in sub-panels 1–2; for plural (pl) in sub-panels 3–4. Means are geometric and confidence intervals 95%. Note that contrasts can be significant in Tab. 1 even if confidence intervals overlap due to the fact that tests were paired (in items) but CI overlap corresponds to unpaired tests.

	Singular RC-subject		**Plural RC-subject**	
	Grammatical	**Ungrammatical**	**Grammatical**	**Ungrammatical**
Human (Wagers et al. 2009)		↘		
Direction consistent with attraction principles	↗	↘	↗	↘
gpt (120M)		↘	↘	↘
gpt2 (127M)		↘	↘	↘
gpt2-large (744M)		↘	↘	
bloom (560M)	↘	↘		↘
bloom (1.7B)		↘	↘	↘
bloom (3B)		↘	↘	
xglm (564M)	↘		↘	↘
xglm (1.7B)			↘	↘
xglm (2.9B)			↘	↘
gemma-3 (1B)	↗	↘	↘	↗
gemma-3 (4B)	↗	↘	↘	

Table 1: Observed and predicted agreement attraction effects in the four theoretically relevant contrasts (see 2). Arrows indicate significant effects. Upward arrow: inhibitory agreement attraction effect, i.e. a slow down. Downward arrow: facilitatory effect, i.e. faster processing. The row “Direction consistent ...” shows the direction of hypothetical effects that could be explained in terms of agreement attraction. Note that all model-predicted effects in the plural/grammatical configuration are in the direction that cannot be explained by attraction.

References: [1] Wagers et al. (2009), JML. [2] Cuskley et al. (2024), Open Mind. [3] Linzen, Leonard (2018), CogSci. [4] Arehalli, Linzen (2020), CogSci. [5] Bazhukov et al. (2024), CoNLL. [6] Timkey, Linzen (2023), EMNLP. [7] Ryu, Lewis, (2021), CMCL.

Investigating the resolution of conflicting predictions from global and local contexts: an eye-tracking sentence reading study

Hao Zeng, Aine Ito
Department of English, Linguistics and Theatre Studies, National University of Singapore
hao_zeng@u.nus.edu

Background: People regularly predict upcoming information in language comprehension. According to the two-stage model, the bottom-up mechanism is initially used to provide resources for the top-down mechanism to generate higher-level discourse representations and predictions in the later stage [1]. However, how people adjust predictions from the top-down mechanism when encountering an incongruent bottom-up input remains unclear. To assess how readers weigh conflicting cues, we examined whether an incongruent local cue weakens predictions from the global context.

Method: We manipulated the contextual constraints in 32 experimental items at four levels (Table 1): Globally Constrained-Locally Incongruent (GCLI), Globally Constrained-Locally Unconstrained (GCLU), Globally Unconstrained-Locally Incongruent (GULI), and Globally Unconstrained-Locally Unconstrained (GULU). In the GCLI and GCLU conditions, the global context was highly predictive of the target noun (e.g., *bird*). In the GCLI and GULI conditions, the local cue (e.g., *powered*) was incongruent with the target noun. In the GULI and GULU conditions, the target noun was unpredictable from the global context. We additionally included 32 filler items to increase the overall reliability of the local cue. On each trial, participants (52 native English speakers) read a passage while their eye movements were tracked and answered a comprehension question. We analyzed the effects of contextual constraints on gaze duration and regression path duration for the target noun and the pre-target adjective using linear mixed-effects models, as we were interested in the initial processing difficulty and re-analysis of the prior context.

Results: The gaze duration for the target noun in the GCLI condition was similar to that in the GCLU condition, t = -0.97, so there is no evidence that participants weakened predictions from the global context when encountering the incongruent local cue in the GCLI condition (Figure 1). However, the regression path duration for the target noun in the GCLI condition was longer than that in the GCLU condition, t = -2.1, suggesting that participants resolved the incongruency of the local cue after encountering the target noun (Figure 2). Additionally, the regression path duration for the pre-target adjective in the GCLI condition was longer than that in the GULI condition, t = -2.7, suggesting that participants tried to resolve the incongruency of the local cue when encountering it (Figure 3).

Discussion: We showed that predictions from the global context were not weakened when the local cue was incongruent with the predictions. However, we found that participants tried to resolve the incongruency at the local cue when encountering it. This finding is inconsistent with ERP (Event-Related Potential) findings from [2] as they did not find increased N400 amplitudes at the pre-target adjective that semantically mismatched the target noun. This difference may be attributed to methodological differences as participants can go back to the previous global context when encountering the incongruent local cue in the eye-tracking reading paradigm but not in ERP studies. Our results demonstrate that people compare the incongruent bottom-up input against the top-down predictions when encountering the incongruent input but resolve the incongruency after encountering the target word.

Table 1. An example experimental item. The target noun (*bird*) is in bold, and the pre-target adjectives are italicized. Each experimental item consisted of a two-sentence passage. The first sentence and the second sentence up to the pre-target adjective served as the global context. The pre-target adjective served as the local cue. The target noun was the same across the four conditions.

Contextual Constraints	Passages
Globally Constrained-Locally Incongruent (GCLI)	Beta looked up after hearing a chirping noise. She saw a *powered* **bird** in the sky.
Globally Constrained-Locally Unconstrained (GCLU)	Beta looked up after hearing a chirping noise. She saw a *small* **bird** in the sky.
Globally Unconstrained-Locally Incongruent (GULI)	Beta enjoyed the appealing scenery on her vacation. She saw a *powered* **bird** in the sky.
Globally Unconstrained-Locally Unconstrained (GULU)	Beta enjoyed the appealing scenery on her vacation. She saw a *small* **bird** in the sky.

Figure 1. Gaze duration for the target noun.

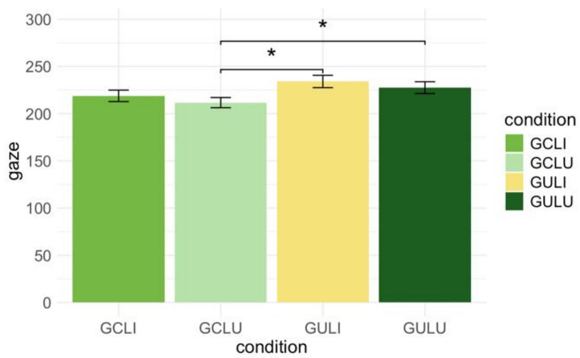


Figure 2. Regression path duration for the target noun

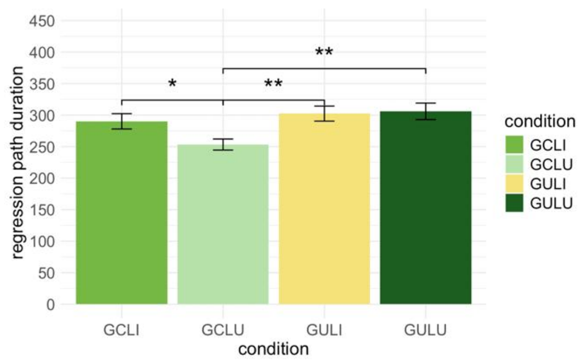
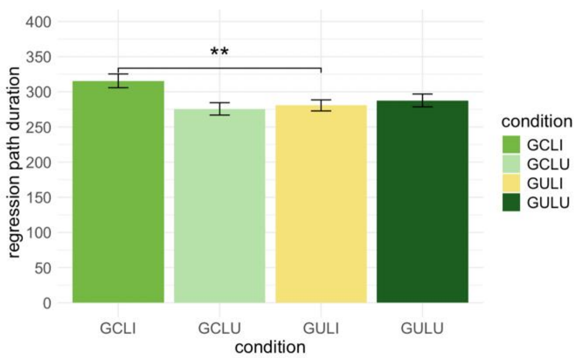


Figure 3. Regression path duration for the pre-target adjective



References

[1] Pickering, M. J., & Gambi, C, 2018; [2] Szewczyk, J. M., Mech, E. N., & Federmeier, K. D, 2022

Note. This project was pre-registered on the Open Science Framework website: https://osf.io/3dv5t

The processing and interpretation of Expletive Negation in children and adolescents

Anna Teresa PORRINI, Veronica D'ALESIO, Matteo GRECO
Istituto Universitario di Studi Superiori - IUSS Pavia
anna.porrini@iusspavia.it

Background: This study investigates the processing of Expletive Negation (EN) in children and adolescents, following up a previous study on adults. EN refers to instances where the negative particle (in Italian, 'non') fails to reverse the polarity of the sentence, such as in the Italian until-clause in "Starò qui finchè non arriva Gianni" (lit. stay.Fut.1stsg here until *neg* arrives John, 'I'll stay here until John arrives'). While traditionally considered an optional vacuum element [1], more recent accounts have challenged this view, presenting EN as a negator operating at a non-surface level [2][3]. In line with the former account, our previous study on 80 adult native speakers of Italian showed that EN is interpreted and processed differently than standard negation (SN, i.e., reversing the sentence polarity). However, in an acceptability judgment task, a separate group of 80 participants rated until-clauses containing EN significantly higher compared to their affirmative counterparts, ruling out the emptiness of EN based on a general principle of economy [4][5] that would disfavor an ineffective, more complex structure. Given these outcomes, the present study employs the same paradigm (a self-paced reading task [SPR] and comprehension questions) to assess whether EN and SN also behave differently in younger populations starting from middle childhood, when SN is fully acquired [6].
Method: 42 children (mean age:11;4; SD:0,3) and 85 adolescents (mean age:14;5; SD:0,5) participated in a masked chunk-by-chunk SPR evaluating the reading of EN sentences against their affirmative counterparts (A_EN). The effect of EN compared to SN was tested by presenting sentences containing a causal construction with the conjunction "perché" ('because'), which leads to SN in the presence of a negative particle and to an affirmative sentence (A_SN) in its absence. Finally, comprehension questions followed each item to ascertain the interpretation of EN and SN. 28 experimental items were spaced out by 66 fillers and then arranged in 8 counterbalanced lists. Examples of experimental items (divided by chunks) and questions are given in Table 1.
Results: For comprehension questions, a generalised linear mixed model (glmm) showed a Polarity (affirmative/negative) effect, although marginal in children ($p=0.088$) and significant in adolescents ($p=0.037$), and an interaction between Sentence type (causal/affirmative) and Polarity ($p=0.006$) only for 14-year-olds. As for the SPR, linear mixed models (lmms) showed an effect of Polarity on the verb (Chunk 6) in the causal clause in both children and adolescents. However, for children a significant effect of Polarity on the verb (Chunk 6) was found in until-clauses as well, along with longer RTs on the following noun (Chunk 7) in A_EN. The results are shown in Figures 1 and 2.
Discussion: As expected, the processing of SN is comparable to adults in both experimental groups, displaying lower accuracy and increased RTs on the verb under the scope of negation. EN, on the other hand, clearly shows a developmental trajectory from children to adolescents: in 11 years-olds EN is assimilated to SN in its processing, while also leading to comparable error rates. Conversely, 14-year-olds display a distinction between EN and SN like what is observed in adults. Although more research is needed, these findings suggest that EN might be acquired as a negative operator first and later reassessed as conveying an affirmative meaning. These results may also reflect a general difficulty associated with shifts at the layer of implicated meaning in children, thus agreeing with recent theories on EN [2][3].

References:

[1] Espinal, M. T. (1992). Expletive negation and logical absorption.

[2] Greco, M. (2019). Is expletive negation a unitary phenomenon? Lingue e linguaggio, 18(1), 25-58.

[3] Delfitto, D., Melloni, C., & Vender, M. (2019). The (en) rich (ed) meaning of expletive negation. Evolutionary Linguistic Theory, 1(1), 57-89.

[4] Zipf, G. K. (1949). The Principle of Least Effort. Cambridge: Addisor-Wesley.

[5] Hawkins, J. A. (2004). Efficiency and complexity in grammars. OUP Oxford.

[6] Tagliani, M., Vender, M., & Melloni, C. (2022). The acquisition of negation in Italian. Languages, 7(2), 116.

Temporal Negative (TN)	Question	A
Chiara | è rimasta | in casa | finché | Marcello | ha chiamato | la pizzeria | per la cena Chiara | stay$_{3rd.sing.fem.past}$ | in house | until | Marcello | call$_{3rd.sing.past}$ | the pizzeria | for the dinner 'Chiara remained in the house until Marcello called the pizzeria for dinner'	*Marco ha chiamato la pizzeria per la cena?* 'Did Marco call the pizzeria for dinner?'	*Sì* YES
Casual Negative (CN)		
Chiara | è rimasta | in casa | perché | Marco non | ha chiamato | la pizzeria | per la cena Chiara | stay$_{3rd.sing.fem.past}$ | in house | because | Marco **neg** | call$_{3rd.sing.past}$ | the pizzeria | for the dinner 'Chiara remained in the house because Marco did not call the pizzeria for dinner'	*Marco ha chiamato la pizzeria per la cena?* 'Did Marco call the pizzeria for dinner?'	*No* NO
Causal Affirmative (CA)	**Question**	**A**
Chiara | è rimasta | in casa | perché | Marcello | ha chiamato | la pizzeria | per la cena Chiara | stay$_{3rd.sing.fem.past}$ | in house | because | Marcello | call$_{3rd.sing.past}$ | the pizzeria | for the dinner 'Chiara remained in the house because Marcello called the pizzeria for dinner'	*È Chiara che ha chiamato la pizzeria per la cena?* 'Is it Chiara the one who called the pizzeria for dinner?'	*No* NO
Temporal Affirmative (TA)		
Chiara | è rimasta | in casa | finché | Marco non | ha chiamato | la pizzeria | per la cena Chiara | stay$_{3rd.sing.fem.past}$ | in house | until | Marco **neg** | call$_{3rd.sing..past}$ | the pizzeria | for the dinner 'Chiara remained in the house until Marco called the pizzeria for dinner'	*Chiara è rimasta in casa?* 'Did Chiara remain in the house?'	*Sì* YES

Table 1. Example of critical items sorted for condition (divided by chunks) with their respective comprehension questions and expected answer (A).

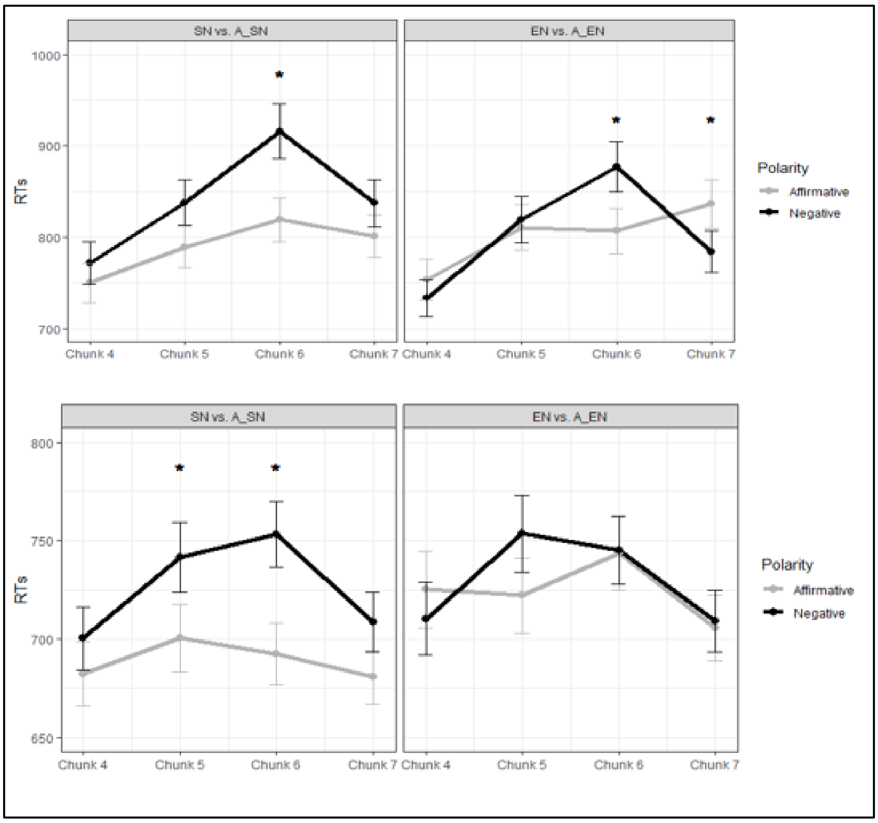


Figure 1. Reaction Times in children (above) and adolescents (below)

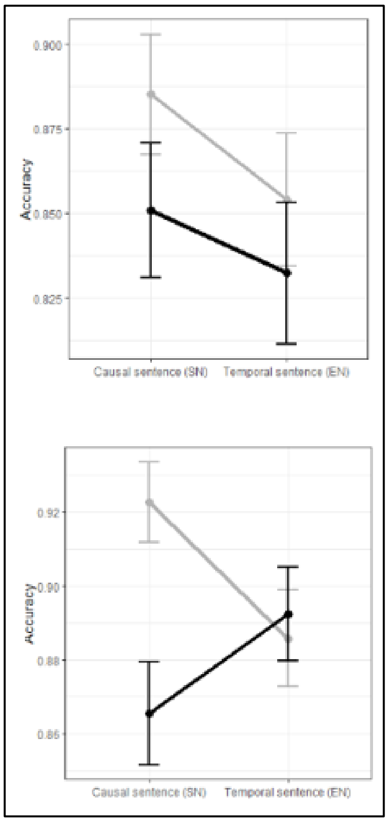


Figure 2. Accuracy on comprehension questions in children (above) and adolescents (below)

Lexical Recognition of Gender-Fair Contracted Forms in Typical and Dyslexic Readers

Saveria Colonna, Paul Lejeune, Elif Mutlu
University Paris 8 & CNRS, Structures Formelles du Langage
saveria.colonna@univ-paris8.fr

In a grammatical gender language such as French, masculine role and profession nouns, such as *l'étudiant* (the student$_{\text{MASC}}$), can refer either to a male-specific interpretation or to a generic interpretation, encompassing individuals of any gender. However, masculine forms are not readily interpreted as inclusive of all genders, but instead tend to induce a masculine bias (e.g., [1]). In response, gender-fair alternatives have emerged, including contracted forms with a middle dot, such as *l'étudiant·e* (the student$_{\text{MASC·FEM}}$). A major concern regarding the use of these forms is their potential processing cost, particularly for readers with dyslexia. While a few studies have examined the readability of contracted gender-fair forms in French among typical readers [2] [3], their impact on dyslexic readers remains unexplored.

Building on a study by [4] on asterisk-marked gender-fair forms in German, we conducted an online lexical decision task with 82 typical and 38 dyslexic French readers. Participants judged 84 words and 84 pseudowords, including 72 experimental stimuli consisting of role and profession nouns (see Table 1). Two within-participants factors were manipulated: gender (masculine, feminine, gender-fair) and type of inclusive ending (*·e*, *·ne*, *·se*). The log-transformed residual lexical decision times (LDT) were analyzed using linear mixed-effects models.

Feminine forms were processed more slowly than masculine forms (Est.=-.03, SE=.009, t=-3.59, p<.001), particularly among dyslexic readers (interaction: Est.=-.06, SE=.02, t=-3.29, p<.01). Gender-fair forms were processed more slowly than feminine forms in both groups (Est.=.04, SE=.01, t=3.11, p<.01) (see Figure 1). These effects likely reflect frequency differences: masculine forms had an average frequency of 5.7 occurrences per million, compared to 2.4 for feminine forms, with contracted forms presumably being even less frequent. A habituation effect was observed only for feminine forms over the course of the experiment (Est.=.0003, SE=.0002, t=2.04, p<.05). Self-reported high exposure to contracted forms did not reduce the processing cost of inclusive vs feminine forms, but did reduce the cost of feminine vs masculine forms (interaction: Est.=.03, SE=.009, t=3.59, p<.001). Among inclusive endings, *·se* forms (e.g., *chanteur·se*, singer$_{\text{MASC·FEM}}$) elicited the longest LDT, possibly due to their morphological complexity. Unlike *·e* and *·ne* forms, *·se* forms involve the substitution of the masculine suffix, and the feminine suffix *-euse* is only partially preserved.

Our results are interpreted within the framework of the dual-route reading model [6]. The grapheme-to-phoneme conversion route, typically used for rare (and new) forms and often impaired in dyslexic readers, can account for the longer processing time of feminine – especially in dyslexic readers – and contracted forms. Taken together, our findings suggest that the processing cost of contracted gender-fair forms is not substantially greater than that of other low-frequency forms, such as feminine forms, particulary for dyslexic readers. Further research is needed to investigate the role of exposure frequency in facilitating the processing of contracted forms.

	Words (N=84)	**Pseudowords** (N=84) (generated with Wuggy [5])
Masculine (N=24)	*dirigeant*	*survionnant*
	magicien	*lortorien*
	chanteur	*mucteur*
Feminine (N=24)	*dirigeante*	*guintenante*
	magicienne	*munivienne*
	chanteuse	*chondeuse*
Gender-Fair (N=24)	*dirigeant·e*	*condiorant·e*
	magicien·ne	*trounivien·ne*
	chanteur·se	*chancleur·se*
Fillers (with special characters) (N=24)	*T-shirt*	*ZoùTuch*

Table 1: Examples of materials.

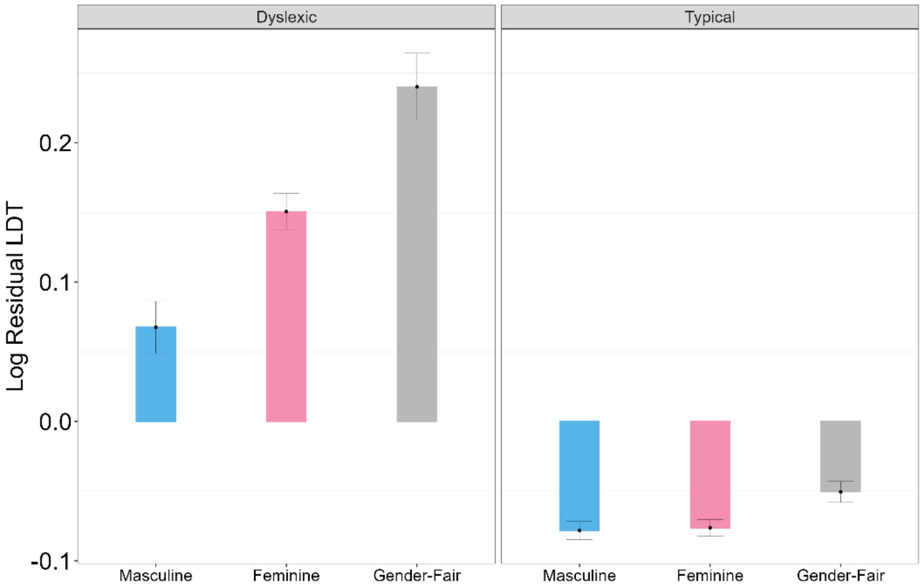


Figure 1: Lexical decision times by gender condition and reader group.

References

[1] Gygax P., Gabriel U., Lévy A., Pool E., Grivel M., Pedrazzini E. (2012). The masculine form and its competing interpretations in French: when linking grammatically masculine role nouns to female referents is difficult. *Journal of Cognitive Psychology*, *24*, 395-408.

[2] Gygax, P., & Gesto, N. (2007). Féminisation et lourdeur de texte. *L'Année Psychologique, 107*.

[3] Liénardy, C., Tibblin, J., Gygax, P., & Simon, A.-C. (2023). Écriture inclusive, lisibilité textuelle et représentations mentales. *Discours, 33*.

[4] Kruppa, A., Fenn, J., & Ferstl, E. C. (2021). Does the asterisk in gender-fair word forms in German impede readability? Evidence from a lexical decision task. *AMLaP Conference*, Paris, France.

[5] Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*, 627–633.

[6] Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108(1), 204–256.

Some and *all* in the visual world of preschoolers

Francesca Foppolo[1], Valeria Galimberti[1], Dongpeng Pan[2], Francesca Panzeri[1] & Stephanie Durrleman[2]
[1] University of Milan-Bicocca & [2] University of Geneva
francesca.foppolo@unimib.it

Several studies report that children, especially preschoolers, show difficulty with pragmatic inferences arising from lexical scales like <*some*, *all*>, as in '*Some of* the boys wear a hat', which triggers the Scalar Implicature (SI) *some, but-not-all,* the boys do. Preschoolers typically accept this sentence in a situation in which *all the* boys wear a hat, when adults typically reject it. The source of children's difficulty is actively debated, and it is explained in terms of *linguistic* [1,2] or *pragmatic* development [3,4]. Results of bilingual children are mixed: Some report a bilingual advantage [5], others don't [6]. All studies employed offline tasks (like picture selection, truth value judgment tasks); such tasks require metalinguistic judgments that might obscure the alternative interpretations considered during the online processing of the quantifier. One exception is a study by Huang & Snedeker [7], who tested quantifiers like *some* and *all* (together with numeral quantifiers) in 5-year-old monolingual English children by means of a visual world eye-tracking paradigm. Unlike adults, children failed to show a reliable Target preference in *some*-trials before the end of the instruction, suggesting a delay in calculating the implicature. **Our study.** By employing a Visual-World paradigm, we tested pre-school children (N=43, Mean Age: 5;7; Age range: 4;9-6;11) speaking French as their dominant language. Some also speak (an)other language(s) at home or at the kindergarten (Multilingual, N=23). Materials were adapted from [2] and comprised 24 experimental items with *all* and *some* rotated across two lists in a Latin-square design, plus 8 filler items not involving quantifiers. The lead-in sentence engaged children in a guessing game (*Guess which is my….I give you a clue*). The test sentence prompted children to find the target among four in a visual scene by using the French quantifiers *quelques* (some) or *tous les* (all). The visual display for critical sentences showed 4 pictures (Fig. 1). If children anticipated the target picture by hearing the quantifier, they should look at the 5/5 picture in *all*-trials and at the 3/5 picture in *some*-trials (this only if they incrementally derived the SI). Children's eye-movements were recorded with a Tobii eye-tracker (60 Hz). At the end of each sentence, participants pointed at the picture that they thought matched the sentence. Children were also tested with standardized tests for receptive morphosyntax and lexicon in French (TROG; PPVT), theory of mind abilities (ToM) and executive functions (ANT). **Results.** Offline responses show an overall optimal performance on filler items (97.5%) and experimental *all*-items (97.2%); performance on *some*-items was lower than on *all-items* (Est. = 2.34, SE=0.38, t= -6.22, p<.001), but still relatively high (82.3%). Of all children, 31 (70%) gave consistent pragmatic responses, a percentage above the typical performance of monolinguals. We ran *glmer* on Looks-to-target in the *some*/*all*-Condition for the 31 "pragmatic" children in the critical region (quantifier+400ms), including only correct/pragmatic responses. Bilingual children were slower to converge on the target for *some* (but not for *all*), compared to Monolinguals (Time*Condition*Bilingualism: Est=-038, SE=0.17, z=-2.21, p=.027, Fig.2A); both PPVT and TROG z-scores modulated children's faster convergence on the target in the case of *some* (Time*Condition*Trog: z=7.46, *p* <.001; Time*Condition*PPVT: z=3.73, *p* <.001, Fig.2B). **Conclusions.** We show an early integration of the pragmatic inference during sentence processing in preschooler, which seems to be modulated by linguistic experience/proficiency: children exposed to more than one language are slower, despite their overall good performance, children with higher linguistic proficiency are faster. We will explore this effect further and will discuss it in light of the current debate.

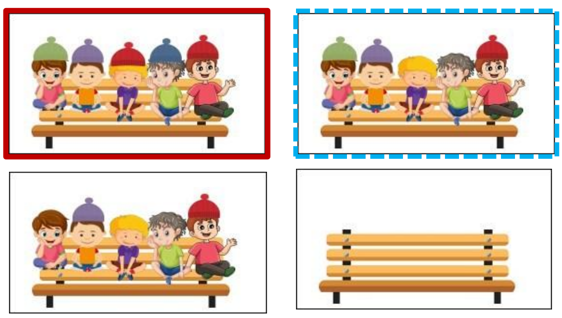


Figure 1A. Sentence: *Guess which one is my bench, I give you a clue: on my bench, **all of** the boys wear a hat.*

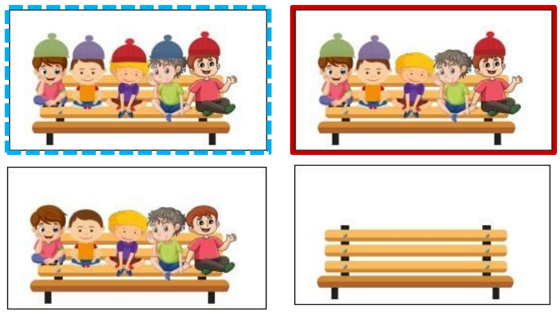


Figure 1B. Sentence: *Guess which one is my bench, I give you a clue: on my bench, **some of** the boys wear a hat.*

In all scenarios, Target picture is in red (5/5 in *all*; 3/5 in *some*) Competitor in blue (3/5 in *all*; 5/5 in *some*). The remaining pictures are the empty distractor and the control false-competitor (2/5; 0/5)

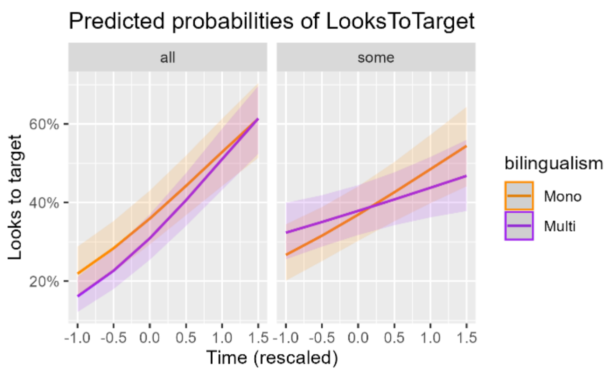


Figure 2A. Model plot for *glmer* of LooksToTarget ~ time_rescaled*bilingualism*condition + age + trialnumber_rescaled + (1 | Subject) + (1 | ItemId) [filtering for: quantifier window, pragmatic children, accurate responses; Condition (all/some) was contrast coded as -.5, +5; Time and Order were centered and rescaled]

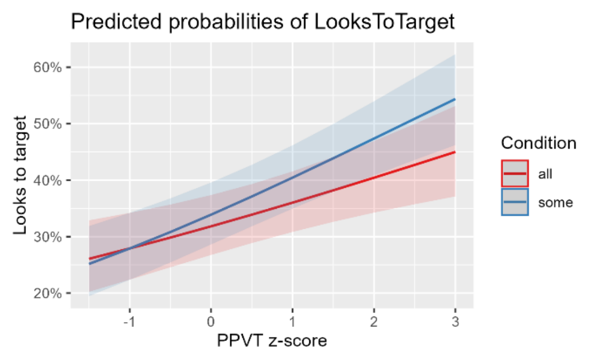


Figure 2B. Model plot for *glmer* of LooksToTarget ~ time_rescaled*PPVTzscore + age + trialnumber_rescaled + (1 | Subject) + (1 | ItemId) [filtering for: quantifier window, pragmatic children, accurate responses; Condition (all/some) was contrast coded as -.5, +5; Time, Order and PPVT were centered and rescaled]

Selected references.

[1] Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84-93.

[2] Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2021). Scalar and ad-hoc pragmatic inferences in children: guess which one is easier. *Journal of child language*, *48*(2), 350-372.

[3] Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67-81.

[4] Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6-18.

[5] Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *110*(1), 115-122.

[6] Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2020). How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings?. *Bilingualism: Language and Cognition*, *23*(1), 186-203.

[7] Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45*(6), 1723–1739.

Distinguishing the mechanisms that support predictive sentence processing: Evidence from associations and speech rate

Anuenue Kukona
University of Greenwich
a.p.bakerkukona@greenwich.ac.uk

Background: Multiple mechanisms are hypothesised to support predictive sentence processing [1,2]. Prediction-by-association assumes that comprehenders' predictions are based on spreading activation among associated representations in memory. For example, participants hearing predictive sentences like "The boy will eat the…" fixated a cake before hearing "cake" [3], which may be based on activation spreading from "eat" to associated edible objects. The aim of this research was to investigate the link between prediction-by-association and speech rate. Associated representations interact with impressive speed [4], and thus prediction-by-association was hypothesised to be particularly suited to speeded processing.
Method: Two mouse cursor tracking experiments tested for association-based predictions [5]. Participation was online through Pavlovia. Participants (*n* = 52 per experiment) heard predictive sentences like "What the pilot will fly, which is shown here, is the…" while viewing visual arrays with predictable target objects like a helicopter and unpredictable but verb-associated competitor objects like a kite or unrelated distractor objects like a book (associations were based on LSA). If predictive sentence processing is supported by prediction-by-association, particularly at speed, participants' predictive mouse cursor movements were expected to be attracted to competitor objects, reflecting their association-based predictions, and this attraction was expected to be particularly pronounced at a fast speech rate.
Experiment 1: Predictive and non-predictive (e.g., "What everyone will discuss, which is shown here, is the…") sentences (i.e., at a normal speech rate) were compared. Horizontal mouse cursor coordinates before target word onset (e.g., corresponding approximately to "which is shown here…") were analysed. A mixed-effects analysis of trial-level coordinates aggregated across time revealed a significant interaction between object type and sentence type ($t = -2.29$, $p < .05$): with predictive sentences, trajectories were more attracted to target objects than non-target objects, but they were more deflected toward competitor objects ($M = 0.42$) than distractor objects ($M = 0.50$); with non-predictive sentences, there were no such results.
Experiment 2: Predictive sentences at normal and fast speech rates (~4 vs. ~9 syllables per second) were compared. A mixed-effects analysis of coordinates revealed a significant interaction between object type and rate type ($t = -2.21$, $p < .05$): at a normal speech rate, results paralleled Experiment 1; at a fast speech rate, trajectories were more attracted to target objects than non-target objects, but they were only weakly deflected toward competitor objects ($M = 0.21$) as compared to distractor objects ($M = 0.25$).
Discussion: These results provide novel insight into the mechanisms that support predictive sentence processing. At a normal speech rate, participants pre-activated unpredictable but verb-associated representations, which are a hallmark of prediction-by-association. Moreover, their mouse cursor movements were sensitive to these predictions. However, while participants pre-activated predictable representations even when hearing impressively rapid speech, their association-based predictions were diminished, at odds with the hypothesis. These results invite a re-conceptualisation of prediction-by-association, which may involve more complex excitation and inhibition dynamics than simple spreading activation.

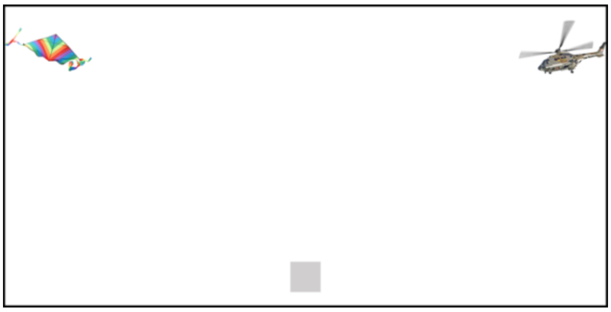


Figure 1. Example visual array for the predictive sentence “What the pilot will fly, which is shown here, is the helicopter.” with a predictable helicopter and unpredictable but verb-associated kite

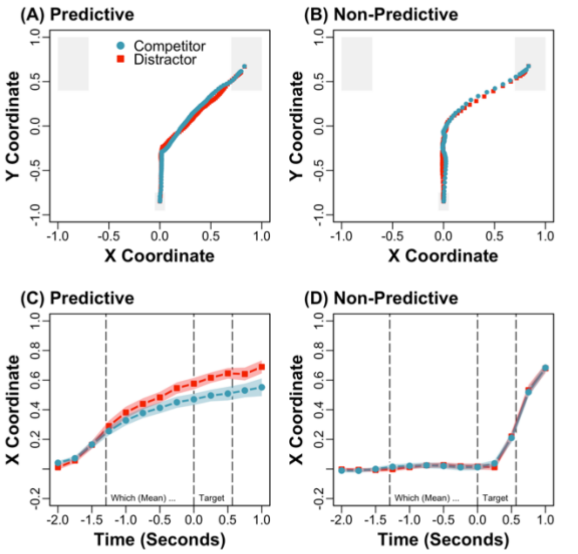


Figure 2. Experiment 1: Mean trajectories across the visual array (A, B) and horizontal coordinates across time (C, D) with competitor (kite) and distractor (book) objects for predictive (“the pilot will fly…”) and non-predictive (“everyone will discuss…”) sentences

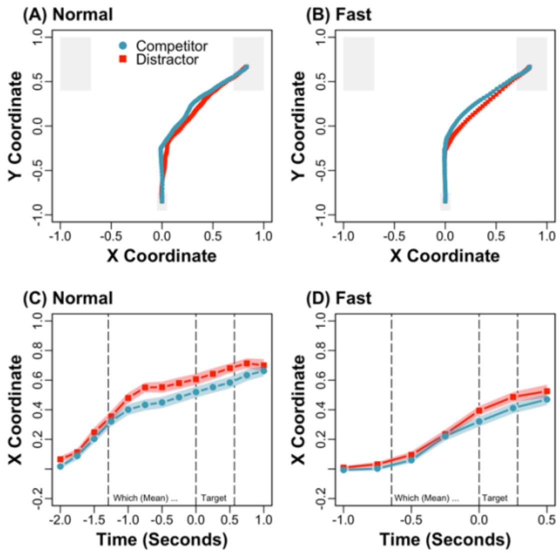


Figure 3. Experiment 2: Mean trajectories across the visual array (A, B) and horizontal coordinates across time (C, D) with competitor (kite) and distractor (book) objects for predictive sentences (“the pilot will fly…”) at normal and fast speech rates

References

[1] Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118-135.

[2] Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002-1044.

[3] Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

[4] Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135*(3), 452-477.

[5] Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*(1), 23-42.

Input entropy affects frame-based category-learning

Areti Kotsolakou[1], Frank Wijnen[1], Sergey Avrutin[1]
[1] Institute for Language Sciences, Utrecht University
a.kotsolakou@uu.nl

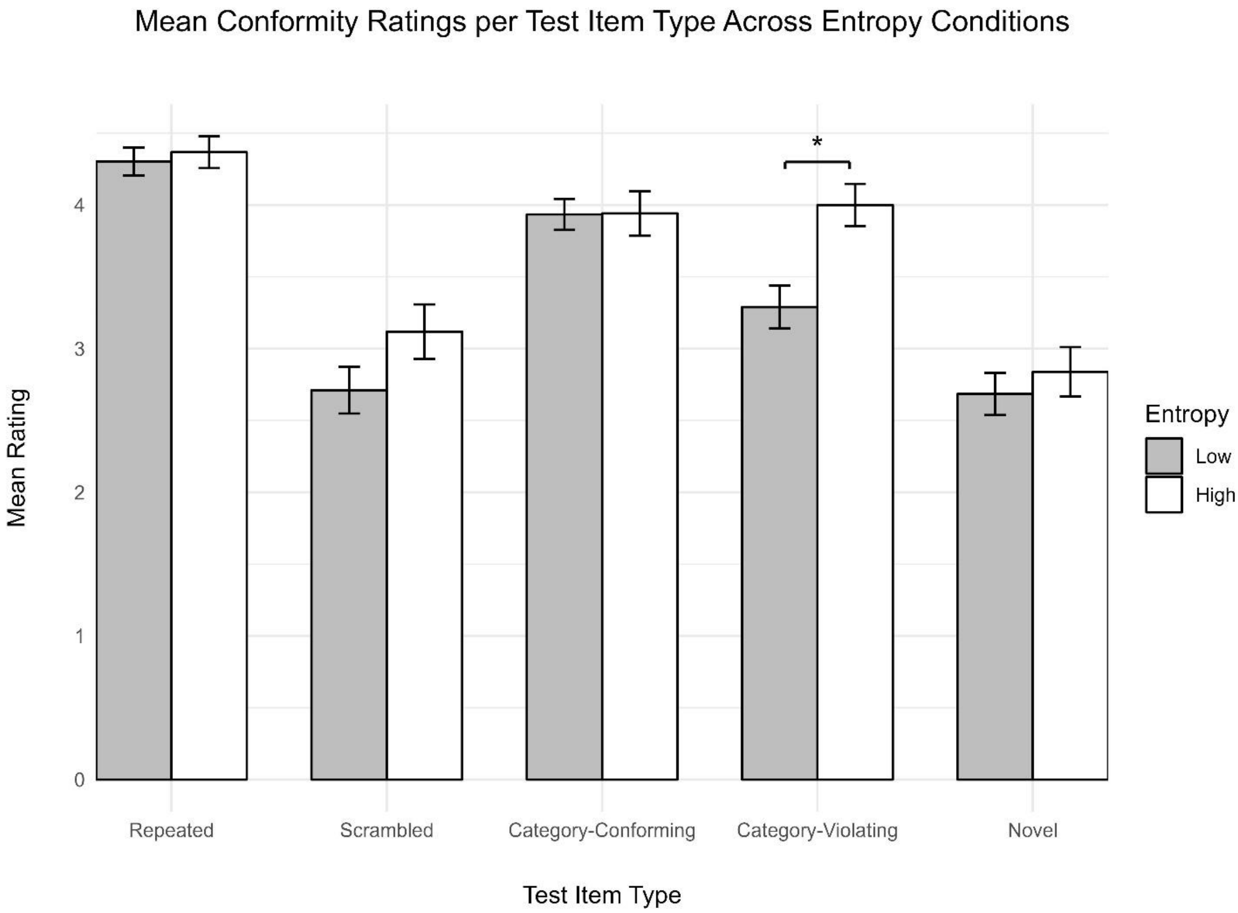
Background: Grammatical categories (e.g., nouns, verbs) are core components of syntax that emerge early in language development [1, 2]. However, how these categories initially form remains unclear. One proposal suggests learners rely on distributional information (patterns of occurence and co-occurence of linguistic elements) to form early distributional categories [e.g., 3]. Early artificial language studies did not succeed in showing that distributional information alone supports grammatical categorization, instead showing the acquisition of positional categories (e.g., words occurring in the first position, the second position etc; [4]). However, Mintz [2, 5] demonstrated that *frequent frames (*recurring word pairs surrounding a target word; e.g., *you_X_it*), can help learners group intervening words into shared categories. Using an artificial language he showed that learners could use frequent frames to distinguish between two frame-based categories. However, the distributional cues in these studies were stronger than those typically available during language acquisition, when input is often sparse. Later research [6] showed that stronger distributional cues (context overlap), enhance within-category generalization, but did not examine learners' ability to distinguish between separate categories.

The present study: We use the entropy model for linguistic generalizations (EMGL; [7]) to investigate generalization in frame-based category-learning involving two distinct categories. According to EMLG, learners gradually shift from item-specific to category-based generalizations when entropy (input complexity) exceeds their processing capacity. Entropy quantifies the complexity of a set of items based on their number and probability of occurrence. We hypothesize that in the context of frame-based category learning higher entropy promotes the shift from item-specific (frame-based) to category-based generalizations.

Method: Adult participants were exposed to an artificial language containing two frame-based categories (X and Y) structured as a_Xi_b and c_Yi_d. They were assigned to either a low- or high-entropy condition and and rated test items on a 1–5 scale based on conformity to the language. Repeated items (*a_Xi_b*) and scrambled items (*Xi_b_* a) tested basic learning. Category-conforming items (familiar target words in novel pairings with category-appropriate frames; *a_Xj_b*), assessed generalization of category properties accross category members. Category-violating items (familiar target words with frames from the opposite category; *c_X_d*) assessed category distinction. Novel items (novel target words in familiar frames; a_Ni_b) assessed acceptance of entirely new category members.

Results: Entropy had a significant main effect on conformity rating. However, only the rating of category-violating items differed across conditions: participants in the high-entropy group rated category-violating strings as more conforming to the language than particpants in the low-entropy group. A significant difference in conformity rating of repeated items and category-violating items appeared only in the low-entropy condition.

Discussion: These findings suggest that input entropy affects generalization in frame-based category-learning. Low entropy led learners to adhere to item-specific (frame-based) categories, while high entropy promoted overgeneralization, likely based on word position rather than on the specific frames. Thus, input entropy affects whether learners form item-specific (frame-based) categories or overgeneralize by collapsing them into broader position-based categories.



References

[1] Aslin, R. N., & Newport, E. L. (2014). Distributional language learning: Mechanisms and models of category formation. Language learning, 64(s2), 86-105.

[2] Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, *30*(5), 678-686.

[3] Maratsos M, Chalkley MA. The internal language of children’s syntax: the ontogenesis and representation of syntactic categories. In: Nelson K, ed. Children’s Language Vol. 2. New York: Gardner Press; 1980, 127–213.

[4] Smith, K. H. (1969). Learning co-occurrence restrictions: Rule induction or rote learning? *Journal of Memory and Language*, *8*(2), 319.

[5] Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. *Cognitive psychology*, *75*, 1-27.

[6] Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive psychology*, *66*(1), 30-54.

[7] Radulescu, S., Wijnen, F., & Avrutin, S. (2020). Patterns bit by bit. An entropy model for rule induction. *Language learning and development*, *16*(2), 109-140.

How age modulates the ability to benefit from sensory-situated semantic congruence: a pilot ERP study

Dandan Li[1], Katja Maquate[2], Agnes Villwock[3], Pia Knoeferle[4]
[1,2,4] Institut für dt. Sprache und Linguistik, Humboldt-Universität zu Berlin; [3] Rochester Institute of Technology
dandan.li@hu-berlin.de

Objectives: Older adults often experience decline in language comprehension, partly due to age-related cognitive decline [1]. While supplementary sensory information (e.g., visual-only or audiovisual contexts) can reduce cognitive load and attenuate N400 effects during semantic processing among younger adults [2, 3], it remains unclear whether older adults - especially those with lower cognitive abilities - benefit from such multimodal support. This pilot study investigates the extent to which older adults benefit from single- (visual-only) and bi-modal (audiovisual) contexts during sentence-level semantic processing, compared to younger adults. It also examines whether age-related cognitive changes constrain semantic integration across different modalities.

Method: Four younger (2 males; age range= 18-31) and four older adults (2 males; age range= 55-65) will perform a verification task in a pilot study, while being presented with sentences that vary in action-verb congruence, within both visual-only [4] and audiovisual contexts. To test auditory acuity, prepared practice sentences with verb (mis)match will be tested before the EEG experiment. ERP data will be recorded during sentence processing (NP1–Verb–NP2), with mismatching verbs expected to elicit N400 (time windows~300–550 ms) that are larger for mismatches than matches. Given a picture of a figure skater nudging a gangster, a sentence such as *Die Eisläuferin stupst den Gangster* constitutes a complete match in a visual-only context; a picture of the figure skater slapping the gangster followed by the same sentence induces a verb–action mismatch in a visual-only context; an audiopicture of the figure skater nudging a gangster includes a complete mismatch in an audiovisual context; and an audiopicture of the figure skater slapping the gangster followed by that sentence includes a verb-action mismatch in an audiovisual context. Behavioral responses and picture-sentence verification accuracy will be collected. Participants will finish a series of cognitive tests (e.g., verbal Working Memory, Inhibition and attention tests).

Statistic analysis: Statistical analyses focus on behavioral reaction times and accuracy, as well as N400 latency, amplitude, and topography. Group comparisons, repeated-measures ANOVAs, and correlational analyses will be used to test three hypotheses: (1) older adults show lower cognitive and semantic performance than younger adults; (2) verb-mismatch conditions elicit slower responses and stronger N400 effects, especially in audiovisual settings; and (3) individual cognitive ability predicts the efficiency of semantic processing.

Conclusion: Pilot study findings will be presented at the conference, with preliminary implications for how aging and cognitive capacity influence language comprehension in multimodal contexts.

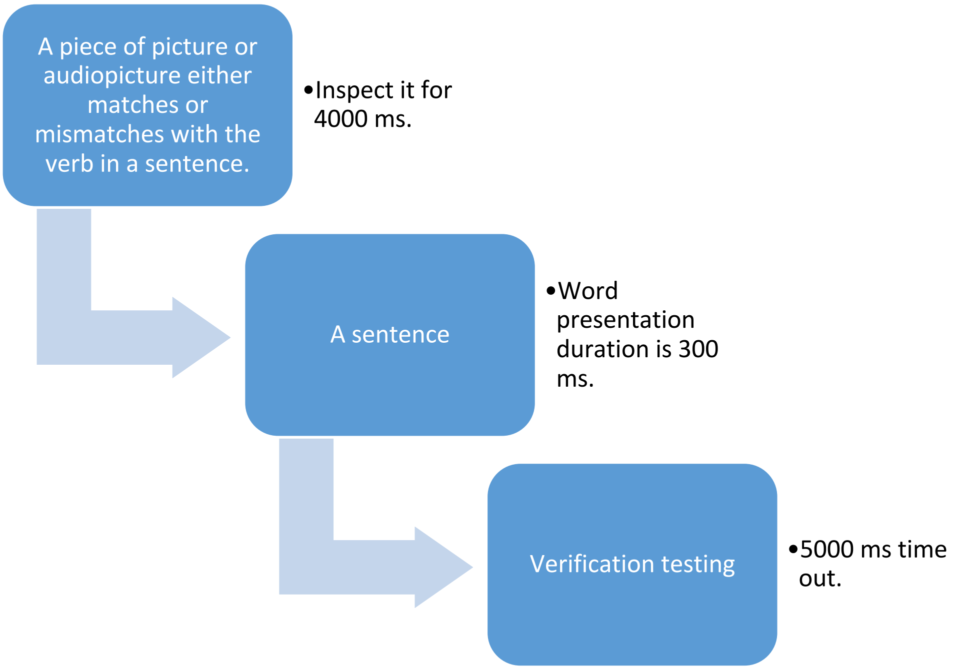


Figure 1. Experiment procedure.

References

[1] Schubotz, L., Holler, J., Drijvers, L., & Özyürek, A. (2021). Aging and working memory modulate the ability to benefit from visible speech and iconic gestures during speech-in-noise comprehension. *Psychological Research*, *85*(5), 1997-2011.

[2] Knoeferle, P., Urbach, T. P., & Kutas, M. (2014). Different mechanisms for role relations versus verb–action congruence effects: evidence from ERPs in picture–sentence verification. *Acta psychologica*, *152*, 133-148.

[3] Klintfors, E., Marklund, E., Kallioinen, P., & Lacerda, F. (2011). Cortical N400-potentials generated by adults in response to semantic incongruities. In *Fonetik 2011* (pp. 121-124). Universitetsserveice AB.

[4] Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*(4), 495-506.

Speaker social characteristics and the resolution of linguistic ambiguities: A self-paced reading task study

Alice Eddyshaw
University of Manchester
alice.eddyshaw@postgrad.manchester.ac.uk

This project aims to elucidate the influence of social information about a speaker on hearers' perception of the language that speaker uses, as well as whether, and to what extent, social world knowledge informs the resolution of linguistic ambiguities in sentence processing. Additionally, the variation of this influence according to hearers' own social identities and attitudes is investigated. The questions (i) 'Do hearers use knowledge or ideas of how speakers use language based on social information about them in language processing?' and (ii) 'Does this vary based on hearers' own social identities and attitudes?' are addressed and investigated using a self-paced reading task and an identities and attitudes questionnaire in tandem. Ambiguities and vaguenesses at a wide variety of different linguistic levels are investigated, reflecting the wide variety of different linguistic levels at which constructions can carry social meanings.

Previously, many sociolinguistic studies have found that some linguistic variants are used more by speakers with certain social characteristics than others (e.g. Labov, 1963; Milroy, 1987). Further, a few studies in social meaning have previously found that this information can be used in language processing (e.g. Staum Casasanto, 2008; Beltrama and Schwarz, 2021). Though phonological variants are most commonly studied, it is also possible to study the social meaning of variants which are widely seen to have referential meanings alongside which their social meanings coexist (Beltrama, 2020). While the indexing of macro-sociological groups has also traditionally been the object of much study, a few studies have taken the approach of plotting the path of a social meaning through an indexical field of micro-, meso- and macro-level identities (Moore and Podesva, 2009). This study follows suit, acknowledging any level of identity at which social meaning occurs. Intended and interpreted social meanings, their uptake and detail vary widely across interlocutors and individual interactions. This calls to the fore that social meaning can be integrated into a general model of hearer's meaning (Hansen, 2025). This project investigates the contribution of social information to language processing, and thus aims to illuminate the extent to which social information (about both the speaker and hearer) can impact a hearer's interpretation. With a commitment to modelling sociolinguistic processing using independently motivated faculties (Campbell-Kibler, 2016), and an emphasis on psychological plausibility, this project explores the processing of social meaning through effects such as real-world plausibility bias (Rohde et al., 2021).

First, participants read through animated storyboards which introduce the salient social characteristics of the speakers of the following utterances. Forty utterances containing linguistic ambiguities, ultimately disambiguated by prepositional phrases, are then presented to participants (n~100) in an online self-paced reading task paradigm (Just et al., 1982) designed using PsychoPy (psychopy.org). Participants' reaction times at the point of disambiguation are then taken as indicators of the (un)expectedness of that interpretation based on the salient social characteristics of the specified speaker. Longer reaction times for the key press following disambiguation are taken to indicate less expected ambiguity resolutions based on speaker prime (Rips, Shoben & Smith, 1973).

Finally, it is emphasised that further study in this area would prove a fruitful avenue.

References:

Beltrama, A. (2020). ‘Social meaning in semantics and pragmatics’, *Language and linguistics compass*, 14(9). DOI:10.1111/lnc3.12398. (Last accessed: Nov 2024).

Beltrama, A. & Schwarz, F. (2021). ‘Imprecision, personae and pragmatic reasoning’, *Proceedings of SALT,* 31, pp.122-144.

Campbell-Kibler, K. (2016). ‘Towards a cognitively realistic model of meaningful sociolinguistic variation’, in Bael, A.M. (ed.) *Awareness and Control in Sociolinguistic Research*. Cambridge: Cambridge University Press, pp.123-151.

Hansen, M.-B. M. (2025). ‘Social meaning as Hearer’s Meaning: Integrating social meaning into a general theory of meaning in communication’, *Journal of Pragmatics*, 241, pp.81-91. DOI:10.1016/j.pragma.2025.03.010. (Last accessed: Apr 2025).

Just, M. A., Carpenter, P. A., & Woolley, J. D. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111: 228-238.

Labov, W. (1963). ‘The social motivation of a sound change’, *Word,* 19(3), pp.273-309.

Milroy, L. (1987). *Language and social networks*. 2nd ed. Oxford: Basil Blackwell.

Moore, E. & Podesva, R. (2009). ‘Style, indexicality, and the social meaning of tag questions’, *Language in Society,* 38, pp.447-485.

Rips, L.J., Shoben, E.J. & Smith, E.E. (1973). ‘Semantic distance and the verification of semantic relations’, Journal of Verbal Learning and Verbal Behaviour, 12, pp.1-20.

Rohde, H., Futrell, R. & Lucas, C.G. (2021). ‘What's new? A comprehension bias in favor of informativity’, *Cognition*, 209, art.104491.

Staum Casasanto, L. (2008). ‘Using social information in language processing’, *BLS,* 34(1), pp.305-316.

The role of Ezafe in the typed production of Persian compound words

Bahareh Yousefzadeh, Cassandra L. Jacobs
Department of Linguistics, University at Buffalo, SUNY
baharehy@buffalo.edu

Background: Despite decades of scholarly investigation into compounding, a full account of the phenomenon that accounts for crosslinguistic variability is still being debated, which limits our understanding of the psycholinguistic processes that support compound production. The compounding properties of the Persian language present an opportunity to extend inquiry to typologically diverse languages with different compounding structures than more commonly-studied Germanic languages and non-Latin writing systems. In this work, we examine the role of a phono-syntactic linking element called Ezafe in typed production. Ezafe connects the head to its modifier in the canonical combination (e.g., مرد قدبلند, man_Ezafe tall, 'tall man'), distinguishing such structures from their head-final counterparts (e.g., اره ماهی, saw fish, 'sawfish'). Ezafe is commonly absent in opaque compounds and is not typically typed. Additionally, the syntactic role of Ezafe in compounds is contested (Samvelian, 2007; Kahnemuyipour, 2014; Ahmadi-Toshizi, 2018), with some proposing that it is underlyingly accessed, analogous to filler-gap production (Momma, 2022), Previous work by Yousefzadeh, Libben, and Segalowitz (2024) found that typing speed at the constituent boundary is sensitive to morphological structure, raising the possibility that producing Ezafe might come with additional processing costs in compound production. Similarly, processing may be slower for opaque compounds with Ezafe as they can be ambiguous between a literal and idiomatic interpretation.

Method: We first constructed a stimulus set of fully transparent or opaque, head-initial compounds, which we then normed for familiarity and transparency (Creemers & Embick, 2022; Gagné & Spalding, 2016). We conducted a power analysis and recruited a desired sample size of 56 literate native speakers of Persian residing in the United States. In a web-based experiment, on each trial participants heard a spoken recording of the compound and then typed the transcription of each compound character-by-character. Participants produced a total of 86 compounds crossing Ezafe with transparency, 26 canonical phrases, and 57 monomorphemic words as fillers. For examples from each category, see Table 1.

Results: We first excluded all compounds that were typed incorrectly, including backspaces, and the keystrokes with latencies above 2000 ms, leaving 2929 compound boundaries for analysis. Per previous work (Yousefzadeh et al., 2024), we measured the Inter-Key Interval (IKI) at the constituent boundary by subtracting the IKI of the first character of the second word from the IKI of the last character of the first word. A linear mixed-model with random intercepts by participant as a function of Ezafe presence, transparency, and their interaction revealed a significant interaction between the Ezafe-presence and transparency, a marginal effect of Ezafe, but no effect of transparency (see Table 2). Specifically, opaque compounds without Ezafe were produced more quickly than transparent compounds and opaque compounds with Ezafe.

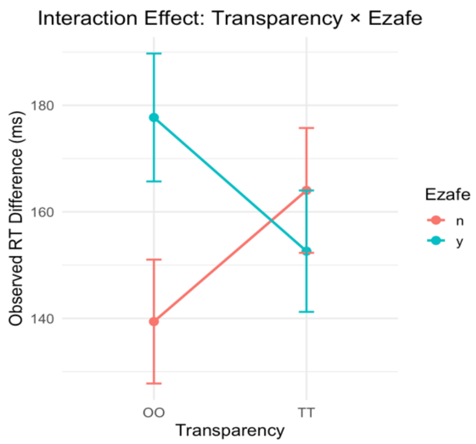
Discussion: Overall, our results highlight the role of Ezafe in the production of typed compounds because this study showed that an element that is not present in the orthographical representation influences the production of typed compounds. The finding that opaque combinations without Ezafe show shorter latencies at the boundary compared to their transparent counterparts is counter to structural and ambiguity-based accounts, and suggests that these compounds may have different internal structures or storage strategies (Gagné & Spalding, 2016; Libben & Weber, 2014). Our work underscores the importance of cross-linguistic perspectives in psycholinguistic compound processing research.

Table 1. Representative stimuli for transparent and opaque compounds with and without Ezafe.

Compounds	Translation	Interpretation	Ezafe	Transparency
تخم مرغ	egg-e hen	egg	yes	Transparent (TT)
آب میوه	water fruit	juice	no	Transparent
زخم زبان	wound-e tongue	tongue lashing	yes	Opaque (OO)
گردن کلفت	neck thick	bully	no	Opaque

Table 2. Fixed effects results of linear mixed effects model of typing speed.

	B	t	p
(Intercept)	153.34	8.09	< .001
Transparency	-1.51	-0.14	n.s.
Ezafe	20.46	1.88	.059
Transparency:Ezafe	-55.94	-2.57	< .01



References

Ahmadi-Torshizi, N. (2020). *The structure and processing of Persian compound words in the mental lexicon: Experimental studies on compound word processing in Persian* (Doctoral dissertation).

Creemers, A., & Embick, D. (2022). The role of semantic transparency in the processing of spoken compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*, 734-751.

Gagné, C. L., & Spalding, T. L. (2016). Effects of morphology and semantic transparency on typing latencies in English compound and pseudocompound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1489-1495.

Libben, G., & Weber, S. (2014). Semantic transparency, compounding, and the nature of independent variables. *Morphology and meaning. Amsterdam/Philadelphia: John Benjamins*.

Kahnemuyipour, A. (2014). Revisiting the Persian Ezafe construction: A roll-up movement analysis. *Lingua*, *150*, 1-24.

Momma, S. (2022). Producing filler-gap dependencies: Structural priming evidence for two distinct combinatorial processes in production. *Journal of Memory and Language*, *126*, 104349.

Samvelian, P. (2007). A (phrasal) affix analysis of the Persian Ezafe1. *Journal of Linguistics*, *43*(3), 605-645.

Yousefzadeh, B., Libben, G., & Segalowitz, S. J. (2024). Persian compounds in the mental lexicon. *Frontiers in Communication*, *9*, 1293401.

Characteristics of the Persian language

Persian is an Indo-European language that uses a writing system based on the Arabic script, which is written from right to left. Compounding is the predominant word formation strategy in Persian, accounting for 70% of the new words created by the Persian Language Academy to replace loanwords (Tabatabaei, 2016). The structure of Persian compounds is different from that of Germanic languages in several ways. Persian compounds can exhibit both head-initial (e.g., ماهی قرمز, *fish red*, for *goldfish*) and head-final configurations (e.g., اره ماهی, *saw fish*, for *sawfish*), a property that contrasts with the exclusively head-final pattern of most Germanic languages.

A unique feature of Persian compounding is the Ezafe construction—a phono-syntactic linker (/e/ or /ye/) that optionally occurs between constituents, only in head-initial compounds. When the last letter of the word is a vowel, because a sequence of vowels is a phonological limitation in Persian, a consonant /y/ is inserted before Ezafe (e.g., غذای خوشمزه, *food-ye delicious*, for *delicious food*). Otherwise, Ezafe is unmarked in writing, and plays a key role in compound interpretation and processing. Its presence can trigger competition between literal and idiomatic meanings, particularly in opaque compounds. For instance, "neck-thick" (without Ezafe) is interpreted idiomatically as "bully," whereas "neck-e thick" suggests the literal "thick neck" interpretation (see Table 1). Persian compounds also vary in semantic transparency, ranging from fully transparent (TT: both constituents contribute clearly to the meaning) to fully opaque (OO: neither constituent's meaning maps directly to the compound's interpretation) (see Table 1).

References

Tabatabaei, A. (2016). Compounding in Persian language. Tehran: Ketabe Bahar.

Neural Signatures of Dependency Processing: Distinguishing Syntax and Semantics

Arrate Isasi-Isasmendi[1,2], Roberto Zariquiey[3], Balthasar Bickel[4] & Caroline Andrews[4]
1. Université de Pau et des Pays de l'Adour 2. IKER CNRS-UMR5478
3. Pontificia Universidad Católica del Perú 4. University of Zurich
arrate.isasi-isasmendi@univ-pau.fr

Several studies suggest that syntactic dependency processing is reflected in alpha power dynamics at the neural level [1,2,3]. More specifically, the encoding of a syntactic dependency is reflected in alpha power desynchronizations, while maintaining a dependency is instead reflected in alpha power synchronizations [4]. This previous research focuses solely on syntax [2,3] or at best uses the Universal Dependencies framework to identify dependencies [4], thus blending syntactic and semantic information. Hence, it is unknown whether the same power dynamics in alpha reflect the processing of purely semantic dependencies in language.

To address this, we contrast the neural correlates of processing syntactic vs. semantic dependencies in Shipibo, a language that allows these to be dissociated. In Shipibo, the ergative case *-n* is used to mark transitive agents, while subjects of intransitive sentences and objects are left unmarked (see Ex. 1 on page 3). Additionally, in this language the evidential clitics *-ra* or *-ronki* are attached to the end of the first phrase in the sentence to indicate the direct or reported source of information, respectively (see Ex. 2 on page 3). These evidential clitics modify what is known about the entire event, and thus they introduce a dependency that is only resolved at the end of the sentence. While case marking depends on verb transitivity (syntactic dependency), evidential marking is driven only by context information (semantic dependency).

Native Shipibo speakers (N=40) heard sentences manipulated for transitivity (with unmarked vs. ergative case) and evidentiality (direct vs. reportative mark) while we recorded EEG with 32-electrodes. Evidentiality was always marked on the sentence initial adjunct, and case was always marked on the second-position subject noun. The sentences additionally contained an object or another adjunct, and ended with the verb (see Table 1). We analyzed the power (de)synchronization patterns in alpha at the opening of dependencies (i.e., evidential and case marks) and at a point where participants had to maintain those dependencies (i.e., pre-verb position), by means of Generalized Additive Mixed Models (GAMMs).

Results: <u>Opening the dependency</u>: the ergative condition elicited a power desynchronization in alpha in left areas compared to the unmarked condition (Fig.1-A). The reportative mark also elicited a power desynchronization in alpha compared to the direct mark (Fig.1-B). We interpret these power desynchronizations as signaling the process of building up more complex syntactic or semantic frameworks required by the ergative case or reportative marks. <u>Maintaining the dependency:</u> power in alpha was more synchronized at the pre-verb position for the unmarked condition than for the ergative condition (Fig.2-A). This is likely because the ergative condition allows participants to integrate most of the syntactic information by this point (having already heard the object), while in the unmarked condition participants need to maintain the sentence representation active until the verb. The reportative condition elicited a power synchronization compared to the direct condition (Fig.2-B), probably because participants need to maintain an indirect framework throughout the sentence.

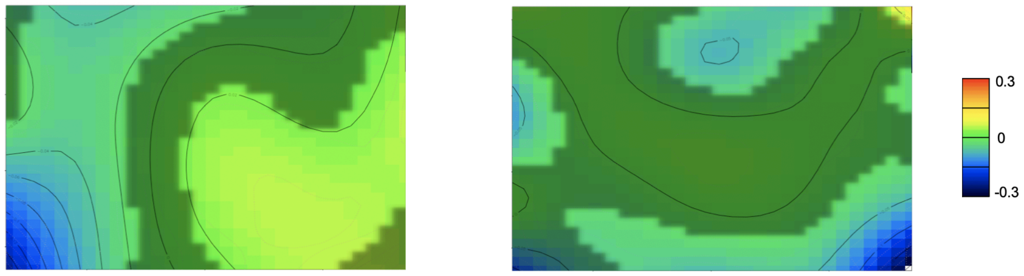
Our results are consistent with previous evidence on the neural correlates of processing ergative case in Basque [5]. More importantly, our findings show that processing purely semantic dependencies triggers power changes in the alpha band, parallel to processing syntax. These results offer new insights into the debate about the extent to which syntactic and semantic processing are integrated in the brain [6].

Table 1: Example of experimental sentence in all four conditions

Condition		Sentence example
Discourse	**Syntactic**	
Direct	Ergative	Bakish-ra Maria-nin atsa pibake "Yesterday Maria ate manioc (I saw it)"
Reportative	Ergative	Bakish-ronki Maria-nin atsa pibake "Yesterday Maria ate manioc (I have been told)"
Direct	Unmarked (Absolutive)	Bakish-ra Maria atsa kopi kaibake "Yesterday Maria went for manioc (I saw it)"
Reportative	Unmarked (Absolutive)	Bakish-ronki Maria atsa kopi kaibake "Yesterday Maria went for manioc (I have been told)"

Figure 1: Opening dependencies (case/evidential mark position)

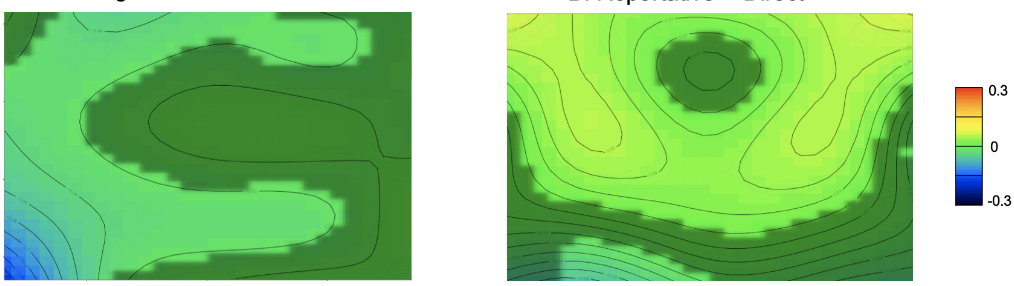
A. Ergative -- Absolutive B. Reportative -- Direct



Fitted surface smooths extracted from the GAMM models showing the topographic distribution of alpha band power differences between conditions for the syntactic (A) and discourse (B) manipulations in the 300--800 ms time-window after hearing the case mark or evidential mark, respectively. Non-shaded (bright) areas indicate where the 95% CIs exclude 0.

Figure 2: Maintaining dependencies (pre-verb position)

A. Ergative -- Absolutive B. Reportative -- Direct



Fitted surface smooths extracted from the GAMM models showing the topographic distribution of alpha band power differences between conditions for the syntactic (A) and discourse (B) level manipulations in the -600--0 ms time-window before resolving the dependency at the verb.

References

[1] Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621.

[2] Segaert, K., Mazaheri, A., & Hagoort, P. (2018). Binding language: Structuring sentences through precisely timed oscillatory mechanisms. *European Journal of Neuroscience*, *48*(7), 2651–2662.

[3] Vassileiou, B., Meyer, L., Beese, C., & Friederici, A. D.(2018). Alignment of alpha-band desynchronization with syntactic structure predicts successful sentence comprehension. *NeuroImage*, *175*,286–296.

[4] Zioga, I., Weissbart, H., Lewis, A. G., Haegens, S., & Martin, A. E. (2023). Naturalistic spoken language comprehension is supported by alpha and beta oscillations. *The Journal of Neuroscience*, JN-RM-1500-22.

[5] Isasi-Isasmendi, A., Andrews, C., Huber, E., Meyer, M., Bickel, B., & Sauppe, S. (2025). Neural correlates of processing case in adults and children. *Brain and Language*, *265*, 105548.

[6] Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*, *203*, 104348.

Additional information on Shipibo-Konibo

Shipibo-Konibo (ISO 639-3), often shortened as Shipibo, is a Panoan language spoken on the Peruvian Amazon, along the Ucayali river. In Shipibo, two distinct grammatical systems interact with clause structure: evidentiality and ergativity. Evidentiality refers to how speakers mark the source of their information, i.e. whether they saw something happen themselves (direct, first-hand knowledge) or whether they heard about it instead (reported knowledge). In Shipibo, this is grammatically encoded through clitics that attach to the end of the first phrase in the sentence: the clitic -ra is used for indicating direct information and -ronki for reported, indirect information (see Example 1 below).

Additionally, Shipibo features ergative-case marking: the ergative case -*n* (or one of its allomorphs: -nin, -kan) is used to mark transitive agents, while subjects of intransitive sentences and objects are marked with the absolutive (unmarked) case (see Example 2 below).

While both evidentiality and ergativity are grammatically encoded in Shipibo, they differ fundamentally in nature. The choice of evidential clitic depends on the semantic or discourse context — specifically, the speaker's source of information. In contrast, the choice of case marking is determined by a syntactic property: the transitivity of the verb.

1. Ergativity in Shipibo-Konibo

(a) E-n-ra nawa-n ochíti jamá-ke
1-**erg**-dir mestizo-gen dog:abs kickl
'I kicked the mestizo's dog.'

(b) E-a-ra Kako-nko ka-iba-ke.
1-**abs**-dir Caco-all go
'I went to Caco yesterday.'

(c) E-a-ra nawa-n ochíti-nin natex-ke.
1-**abs**-dir mestizo-gen dog-**erg** bite
'The mestizo's dog bit me.'

2. Evidentiality in Shipibo-Konibo

(a) Maria-nin-ra nawa-n ochíti jamáke.
Maria-erg-**dir** mestizo-gen dog:abs kick
'Maria kicked the mestizo's dog.'

(b) Maria-nin-ronki nawa-n ochíti jamá-ke.
Maria-erg-**rep** mestizo-gen dog:abs kick
'Maria kicked the mestizo's dog (I heard).'

Key for abbreviations: ERG: ergative; ABS: absolutive (unmarked); DIR: direct evidential ; REP: reported evidential; GEN: genitive; ALL: allative.

Occipitotemporal and frontal regions are crucial for Chinese children's reading development

Junhua Ding[1], Siyu Chen[1], Chen Feng[1] , Su Li[1]
[1] Institute of Psychology, Chinese Academy of Sciences
dingjh@psych.ac.cn

Background: Learning to read is one of the most important missions in child development. Understanding its neural basis will help diagnose and intervene children’s reading abnormality in advance. Most neuroscience studies on children’s reading development are based on alphabetic languages. They found that as ages grow, children’s reading relies more on the left hemisphere and presents a dorsal to ventral pathway transformation [1,2,3]. Compared with alphabetic languages, Chinese is a logographic language with different linguistic and visual-spatial features [4]. It is still unknown about the neural substrates related with Chinese children’s reading development.

Method: In this study, we used Chinese Color Nest Project dataset, which included 184 children’s reading test and multimodal MRI images between 6-14 years old [5]. The reading test included 150 Chinese characters with distributed difficulty, while controlling the frequency, number of strokes and phonograms. Surface thickness and functional connectivity were respectively extracted from T1 and rs-fMRI imaging to measure brain’s structural and functional developments. To systematically answer the question, we explored the Chinese reading related neural basis both structurally and functionally using generalized linear model, examined the age interaction effect with logistic time-varying effect modeling and built prediction models both cross-sectionally and longitudinally with stepwise regression.

Results: We found that bilateral frontal and occipital lobes’ surface thickness and functional connectivity were crucial for Chinese children’s reading development (Fig 1). Among these metrics, only the functional connectivity between frontal lobes were mediated by age. In the prediction model, both frontal and occipital lobes’ structure and function help predict children’s reading performance cross-sectionally, while only occipital lobe contributed to the longitudinal model.

Discussion: In sum, using a large and multimodal dataset, this study systematically revealed the special neural substrates of Chinese children’s reading development. Compared with alphabetic languages’ strong reliance on the left hemisphere and dorsal pathway, the development of Chinese reading relies more on the bilateral visual area and medial frontal regions. The result helps better understand the neural difference caused by languages and provides potential biomarkers for Chinese children’s reading difficulty.

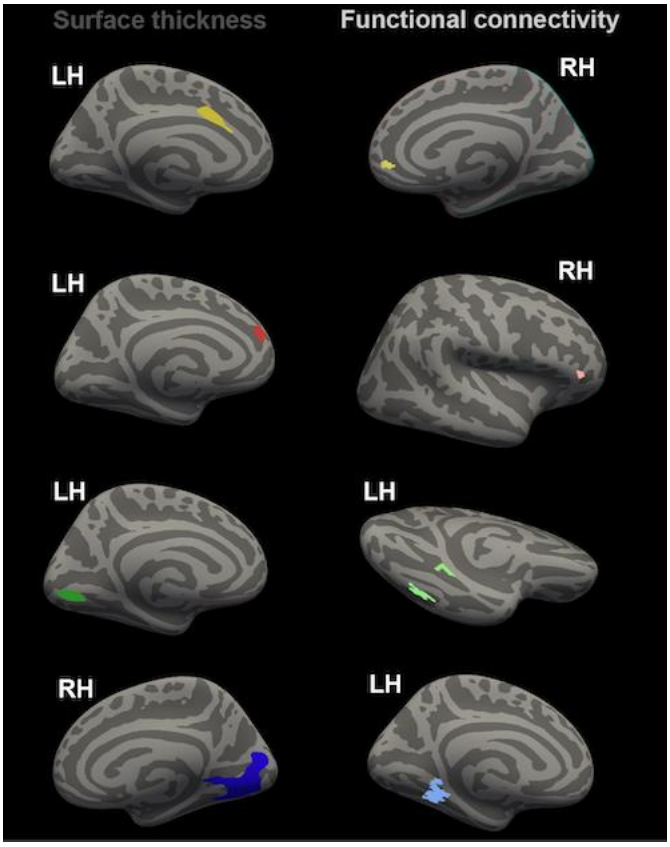


Figure 1. Reading-related brain regions and functional connectivity. The left panel displays brain areas that showed significant correlations between surface thickness and children's reading skills. The right panel displays brain regions that were functionally connected with brain areas identified in the surface thickness analyses. LH/RH: left/right hemisphere.

References

[1] Chyl, K., Kossowski, B., Dębska, A., Łuniewska, M., Banaszkiewicz, A., Żelechowska, A., Frost, S. J., Mencl, W. E., Wypych, M., Marchewka, A., Pugh, K. R., & Jednoróg, K. (2018). Prereader to beginning reader: Changes induced by reading acquisition in print and speech brain networks. *Journal of Child Psychology and Psychiatry*, *59*(1), 76–87. Q1. https://doi.org/10.1111/jcpp.12774

[2] Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., & Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping*, *34*(12), 3354–3368. Q1. https://doi.org/10.1002/hbm.22147

[3] Younger, J., Tucker-Drob, E., & Booth, J. R. (2017). Longitudinal changes in reading network connectivity related to skill improvement. *NeuroImage*, *158*, 90–98. https://doi.org/10.1016/j.neuroimage.2017.06.044

[4] Perfetti, C., Cao, F., & Booth, J. (2013). Specialization and Universals in the Development of Reading Skill: How Chinese Research Informs a Universal Science of Reading. *Scientific Studies of Reading*, *17*(1), 5–21. Q1. https://doi.org/10.1080/10888438.2012.689786

[5] Fan, X.-R., Wang, Y.-S., Chang, D., Yang, N., Rong, M.-J., Zhang, Z., He, Y., Hou, X., Zhou, Q., Gong, Z.-Q., Cao, L.-Z., Dong, H.-M., Nie, J.-J., Chen, L.-Z., Zhang, Q., Zhang, J.-X., Zhang, L., Li, H.-J., Bao, M., … Zuo, X.-N. (2023). A longitudinal resource for population neuroscience of school-age children and adolescents in China. *Scientific Data*, *10*(1), 545. Q1. https://doi.org/10.1038/s41597-023-02377-8

Frequency Modulates Phonetic but Not Semantic Radical Effects in Chinese Character Recognition

Zehua R. Jiang [1,2], Mingyuan Yang [3]

[1] Department of Psychology, National University of Singapore;

[2] Department of Psychology and Behavioral Science, Zhejiang University ;

[3] Department of English, Linguistics and Theatre Studies, National University of Singapore

Background: Chinese characters are typically composed of a semantic radical and a phonetic radical, both of which influence visual word recognition. However, the mechanisms underlying their effects remain unclear, partly due to a lack of large-scale item-level studies. Prior work has rarely considered that character frequency may modulate the influence of radical family size. Readers may rely more on semantic or phonetic cues when character frequency is low, altering the observed effects. Additionally, due to differences in the foundational quantity and functional roles of phonetic and semantic radicals, their family sizes may interact with each other, as readers might shift processing strategies based on the relative informativeness of each cue, and character frequency might further modulate this interaction. Thus, examining the interaction between character frequency and radical family size is critical to accurately characterizing their roles in character processing.

Method: We conducted item-level linear regressions on reaction times (RTs) and accuracy (Acc) from a megastudy lexical decision task[1]. Semantic radical (SR) and phonetic radical (PR) family sizes [2] and log-transformed character frequency [3] were included as predictors, centered around their means. Strokes were entered as an uncentered covariate. We also tested the interaction between PR family size and character frequency. A total of 1,549 characters were analysed.

Results: Adding the interaction term between family size and character frequency significantly improved model fit ($\chi^2_{RT}(4) = 17.60, p < .01;\ \chi^2_{Acc}(4) = 36.08, p < .001$). As shown in Table 1, SR family size had a significant main effect, with larger SR family size consistently slowing RTs and decreasing the Acc, with no interaction with frequency. In contrast, PR family size significantly interacted with frequency: for low-frequency characters, larger PR family size facilitated faster RTs and higher Acc, whereas for high-frequency characters, it slowed recognition and decreased Acc (see Figure 1). The SR effect remained robust after controlling for interactions. Neither the two-way interaction between SR and PR family sizes nor the three-way interaction among SR family size, PR family size, and character frequency reached significance.

Discussion: These results suggest that PR and SR play distinct roles in Chinese character recognition. For low-frequency characters, which rely more on decomposition processing, a larger PR family size facilitates recognition by activating multiple phonological neighbors. For high-frequency characters, which are processed more holistically, a larger PR family size leads to orthographic similarity, impairing recognition efficiency. In contrast, SR family size consistently slowed recognition across frequencies, suggesting that a larger semantic family requires finer-grained differentiation within the semantic space. Overall, phonetic and semantic radicals contribute differently to character recognition, with PR effects modulated by frequency, highlighting the dynamic interplay between decomposition and holistic processing in reading Chinese.

Table 1 Effects of semantic family size (centred, SRFamilySize) and phonetic radical family size (centred,PRFamilySize) on accuracy and reaction time of lexical decision task[1]

	RT			ACC		
	Estimates	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	599.69	589.25 – 610.12	**<0.001**	0.93	0.92 – 0.95	**<0.001**
Strokes	1.44	0.49 – 2.40	**0.003**	0.00	-0.00 – 0.00	0.059
cfreq	-51.56	-54.61 – -48.51	**<0.001**	0.05	0.04 – 0.05	**<0.001**
SRFamilySize	0.05	0.02 – 0.09	**0.002**	-0.00	-0.00 – -0.00	**0.031**
PRFamilySize	0.16	-0.54 – 0.87	0.648	0.00	-0.00 – 0.00	0.104
cfreq ×RFamilySize	0.02	-0.01 – 0.06	0.167	0.00	-0.00 – 0.00	0.214
cfreq × PRFamilySize	1.23	0.50 – 1.96	**0.001**	-0.00	-0.00 – -0.00	**<0.001**
SRFamilySize × PRFamilySize	0.01	-0.00 – 0.02	0.092	0.00	-0.00 – 0.00	0.196
cfreq × SRFamilySize × PRFamilySize	0.01	-0.00 – 0.02	0.086	0.00	-0.00 – 0.00	0.148
R^2 / R^2 adjusted			0.462 / 0.460			0.246 / 0.242

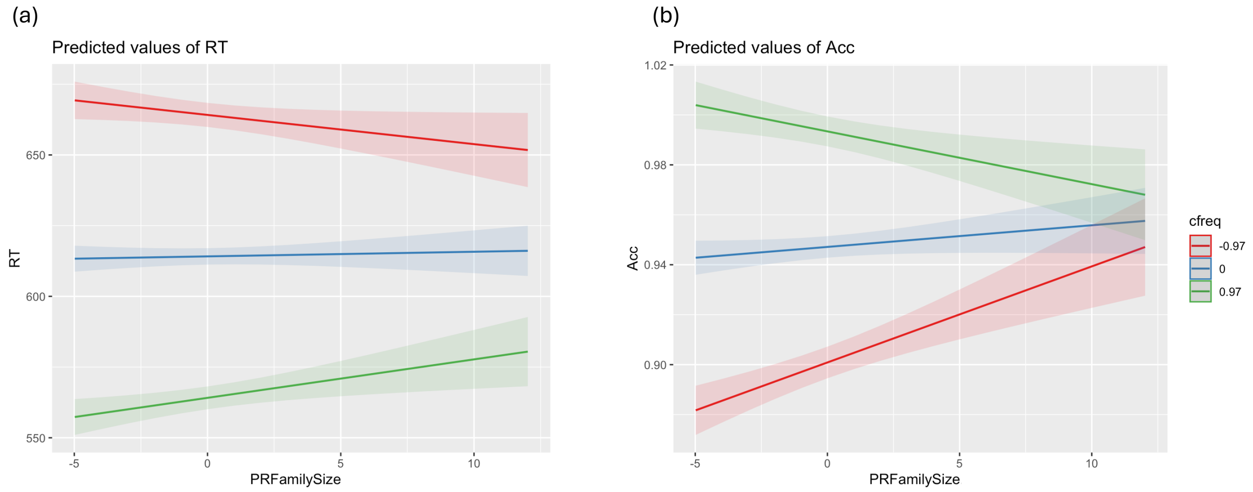


Figure 1 Interaction plot for Phonetic Radical Family size (centred, PRFamilySize) and character frequency (centred, cfreq) for (a) reaction time data and (b) accuracy data

References

1. Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*(1), 263–273. https://doi.org/10.3758/s13428-013-0355-9

2. Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD). *Behavior Research Methods*, *50*(6), 2606–2629. https://doi.org/10.3758/s13428-018-1038-3

3. Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLOS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chinese Writing System

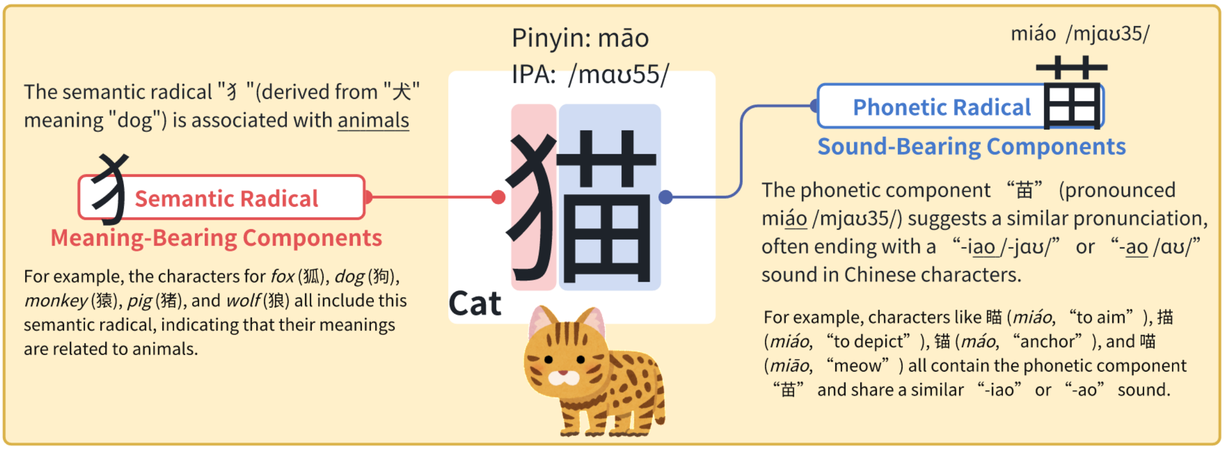


Figure 2 The semantic radical and phonetic radical of "猫(cat)"

Chinese characters are usually composed of a semantic radical and a phonetic radical. The semantic radical often hints at the character's meaning category, while the phonetic radical often hints at the character's pronunciation. For example (Figure 2), the character "猫" (cat, /mɑʊ55 /) includes 犭, its semantic radical, indicating the animal category, and 苗 (/mjɑʊ35 /), its phonetic radical, providing a clue to its pronunciation.

While both radicals offer different types of information, their reliability is not equal. Chinese lacks strict grapheme-to-phoneme correspondence, and the consistency between a phonetic radical and a character's actual pronunciation is often weak or arbitrary [4, 5] . In addition, widespread homophony—where a limited number of syllables map onto many characters—further reduces the reliability of phonological cues. By contrast, semantic radicals often provide more direct meaning cues based on visual form, independent of phonology.

References

4. Fang, Y. H., & Zhang, J. J. (2009). Asymmetry in naming and categorizing of Chinese words and pictures: Role of semantic radicals. Acta Psychologica Sinica, 41(2), 114–126.
5. Williams, C. (2013). Emerging development of semantic and phonological routes to character decoding in Chinese as a foreign language learners. *Reading and Writing*, *26*(2), 293-315.

Asymmetry in the Retention of Content and Surface Linguistic Information During Reading in L1 and L2: An Eye-Tracking Study

Andreas Opitz[1], Denisa Bordag[1,2], Hans-Georg Berulava[1]
[1]Herder-Institute, Leipzig University (Germany), [2]Haifa University (Israel)
andreas.opitz@uin-leipzig.de

Background: Reading in a native (L1) versus a non-native (L2) language differs across multiple cognitive dimensions. Previous research has demonstrated that responses to texts can vary depending on whether the text is processed in an individual's L1 or L2 — for instance, in terms of emotional engagement [1] or decision-making behavior [2]. In the present study, we extend this line of research by examining the L1–L2 effect in the domain of memory. Specifically, we ask whether the L2 advantage in retaining surface-level linguistic (i.e., form-related) information, as observed in our recent experiments [3, 4], is accompanied by reduced memory for content-related information compared to L1.

Method: Native speakers of German (N=64) and learners of German (N=64, L1=Czech, B2/C1 level in German) performed a series of tasks in an original eye-tracking paradigm. After reading short texts, participants completed two control tasks: a picture task and a single sentence task. The pictures showed either scenes congruent with the information described in the text or with a factual deviation (content manipulation, see **Figure 1**). In the sentence task, participants read either completely identical sentences that had appeared in the texts, or sentences in which the surface linguistic information was manipulated without changing its meaning (surface linguistic manipulation, e.g., changes in grammatical voice, positioning of an attribute, see examples below). Eye-tracking data were analysed (mixed effects models) to determine fixation durations on regions with manipulated versus unaltered information. The rationale was that participants are only able to detect changes (of content or surface linguistic information) if they had stored this information from the preceding text-phase. Change detection should manifest itself in longer fixation durations on regions with altered compared to unaltered information. There were four different texts, each followed by 24 pictures and 56 single sentences (pictures and sentences included fillers).

Results: A contrasting pattern for content versus surface-level manipulations was observed in both populations (see **Figure 2**): While significant retention effects for content manipulation were observed only for L1 but not for L2 participants, the opposite pattern was found for surface-level manipulations. In an additional number manipulation that involves both changes in content (one vs. many) and surface linguistic, (different suffixes and agreement) information, a retention effect was observed in both populations.

Discussion: Our study provides empirical support for key differences in memory retention between L1 and L2 reading. L1 readers prioritize content information while surface linguistic details decay rapidly, aligning with classic memory research [5]. In contrast, L2 readers retain more surface-level information, corroborating recent findings [3, 4]. In line with the Shallow Structure Hypothesis [6], stronger reliance on verbatim memory in L2 may compensate for syntactic processing difficulties and may also support grammar acquisition through the storage of linguistic chunks [7]. Overall, the study highlights fundamental differences in L1 and L2 mental text models, affecting comprehension and language learning. It also sheds light on the challenges L2 readers may face when engaging with academic texts in a non-native language, which may ultimately impact their academic success.

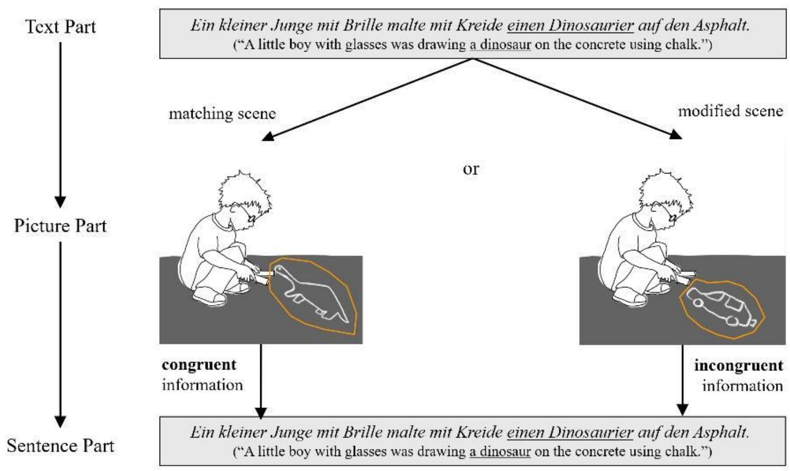


Figure 1. Example of one critical content sentence across the experimental tasks and conditions:

Examples of the surface linguistic manipulation:

Grammatical Voice:

Ein junger Mann hob die Tasche der Dame auf. (A young man picked up the lady's bag.) ↔ *Die Tasche der Dame wurde von einem jungen Mann aufgehoben.* (The lady's bag was picked up by a young man.)

Attribute Position:

*Auf dem Schoß der Frau saß ein laut schreiendes **Baby***. (A loudly crying **baby** was sitting on the woman's lap.) ↔ *Auf dem Schoß der Frau saß **ein Baby**, das laut schrie.* **(A baby** who was crying loudly was sitting on the woman's lap.)

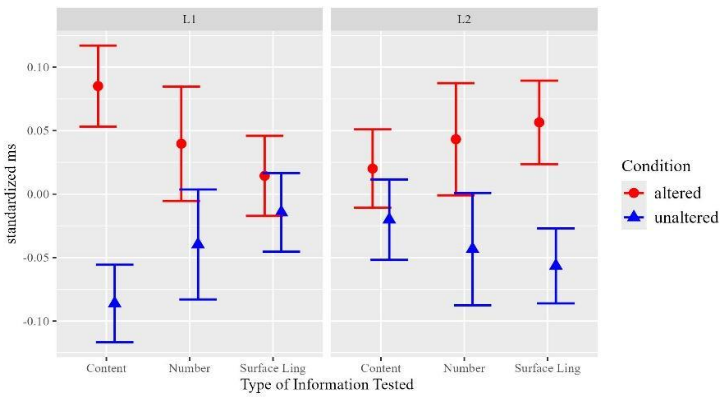


Figure 2. Summary of results of the sentence task. Means of total durations in critical ROIs, standardized within Language & Type of Manipulation.

References

[1] Pavlenko, A. (2012) 'Affective processing in bilingual speakers: Disembodied cognition?', *International Journal of Psychology, 47*(6), pp. 405–428.

[2] Keysar, B., Hayakawa, S.L. and An, S.G. (2012) 'The foreign-language effect: Thinking in a foreign tongue reduces decision biases', *Psychological Science, 23*(6), pp. 661–668.

[3] Bordag, D. and Opitz, A. (2025) 'Retention of grammatical information by L1 and L2 readers: The role of form and meaning', *Journal of Memory and Language, 141*, 104605.

[4] Opitz, A., Bordag, D. and Furgoni, A. (2024) 'The role of linguistic factors in the retention of verbatim information in reading: An eye-tracking study on L1 and L2 German', *Applied Psycholinguistics, 45*, pp. 393–417.

[5] Garnham, A. and Oakhill, J. (1996) 'The mental models theory of language comprehension', in Britton, B.K. and Graesser, A.C. (eds.) *Models of understanding text*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 313–339.

[6] Clahsen, H. and Felser, C. (2018) 'Some notes on the shallow structure hypothesis', *Studies in Second Language Acquisition, 40*(3), pp. 693–706.

[7] Bybee, J.L. (2010) *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Similarity Comes at a Cost: Novel Evidence for Associative Memory Retrieval

Philine Link[1], Leendert van Maanen[2], Jakub Dotlačil[1]

[1] Institute for Language Sciences, Utrecht University, Utrecht, The Netherlands

[2] Department of Experimental Psychology, Utrecht University, Utrecht, The Netherlands

Corresponding author: Philine Link, p.m.link@uu.nl

Background. A key observation in the study of associative memory is that the greater the number of associations a concept has (its associative fan), the longer the reaction times (RTs) and the lower the accuracy demonstrated by participants during recall. This is called the *fan effect* ([1]). In classical fan experiments, participants study sentences containing reoccurring concepts in multiple contexts (e.g., "The *doctor* is in the *desert*", "The *doctor* is in the *field*"). Models of memory in psycholinguistics assume a content-addressable associative memory, in which semantically related items cause interference in retrieval ([2]; [3]; [4]). This assumption is based on the cue-based retrieval framework from the memory module in ACT-R, which in turn was established through the findings from fan experiments. Classic fan experiments, however, did not directly test the role of semantic similarity. To address this gap, we tested whether the fan effect explains how associations occurring in natural language are retrieved by using semantically related concepts based on similarity measures from Dutch vector space models ([5]; [6]).

Method. In Experiment 1 (N = 134, 46 lab, 88 online), we replicated [7] to test the experimental paradigm in Dutch and with an adjusted learning method to make the manipulation of fan size less evident. In this experiment, concepts were repeated in two (fan 2) or four different (fan 4) contexts. Based on this replication, we conducted Experiment 2 (N = 97, online) which used semantically similar instead of repeated concepts. In both experiments, participants learned short sentences (see Table 1). In the testing phase, participants read a sentence and indicated via button press whether it belonged to the set of sentences they studied before (target) or not (foil). For each decision, we recorded the corresponding RT and accuracy.

Results. We fitted Bayesian generalized linear mixed models to analyze RTs and accuracy. In both experiments, participants exhibited longer RTs for fan 4 than for fan 2 conditions (see Figure 1, top), showing that the fan effect persists with the adjusted learning method and semantically similar instead of repeated concepts. Lower accuracy for the fan 4 conditions could only be found in Experiment 1 (see Figure 1, bottom). Accuracy in Experiment 2 was high in all experimental conditions, suggesting a potential ceiling effect.

Discussion. Our results show that the fan effect persists even when we use associated concepts occurring in natural language. These findings indicate that activation in memory spreads along a network of semantic connections as opposed to a network of individual words. This supports the interpretation of the original fan effect as a measure of associative fan, which will enable us to connect similarity measures from vector space models to theories of memory retrieval.

Exp.	Person	Loc.	Example sentences
1	a	A	*The captain is in the tower*
	b		*The teacher is in the tower*
1	a	B	*The captain is in the pool*
	c		*The poet is in the pool*
1	Foils		*The musician is in the tower*
			The musician is in the pool
2	a_1	A_1	*The doctor is in the canyon*
	b_1	A_2	*The professor is in the ravine*
2	a_2	B_1	*The surgeon is in the factory*
	c_1	B_2	*The barista is in the workshop*
2	Foils		*The lawyer is in the canyon*
			The judge is in the factory

Table 1: Stimuli used in Experiments (Exp.) 1 and 2, translated from Dutch. Person and location (Loc.) refer to stimulus categories. Foils used the same items with new pairings. Stimulus sets were randomly generated and each participant received a unique set.

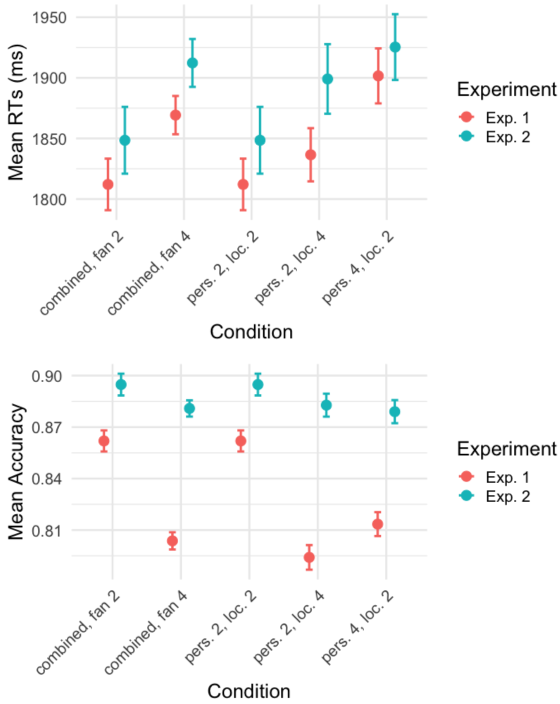


Figure 1: Mean reaction times (top) and accuracy (bottom) across conditions in Experiment 1 and 2. The error bars show the standard error of the mean.

References

[1] Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261–295.

[2] Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407.

[3] Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

[4] Smith, G., & Vasishth, S. (2020). A principled approach to feature selection in models of sentence processing. *Cognitive science*, *44*(12), e12918.

[5] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135–146.

[6] Fares, M., Kutuzov, A., Oepen, S., & Velldal, E. (2017, May). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In J. Tiedemann & N. Tahmasebi (Eds.), *Proceedings of the 21st nordic conference on computational linguistics* (pp. 271–276). Association for Computational Linguistics. https://aclanthology.org/W17-0237/

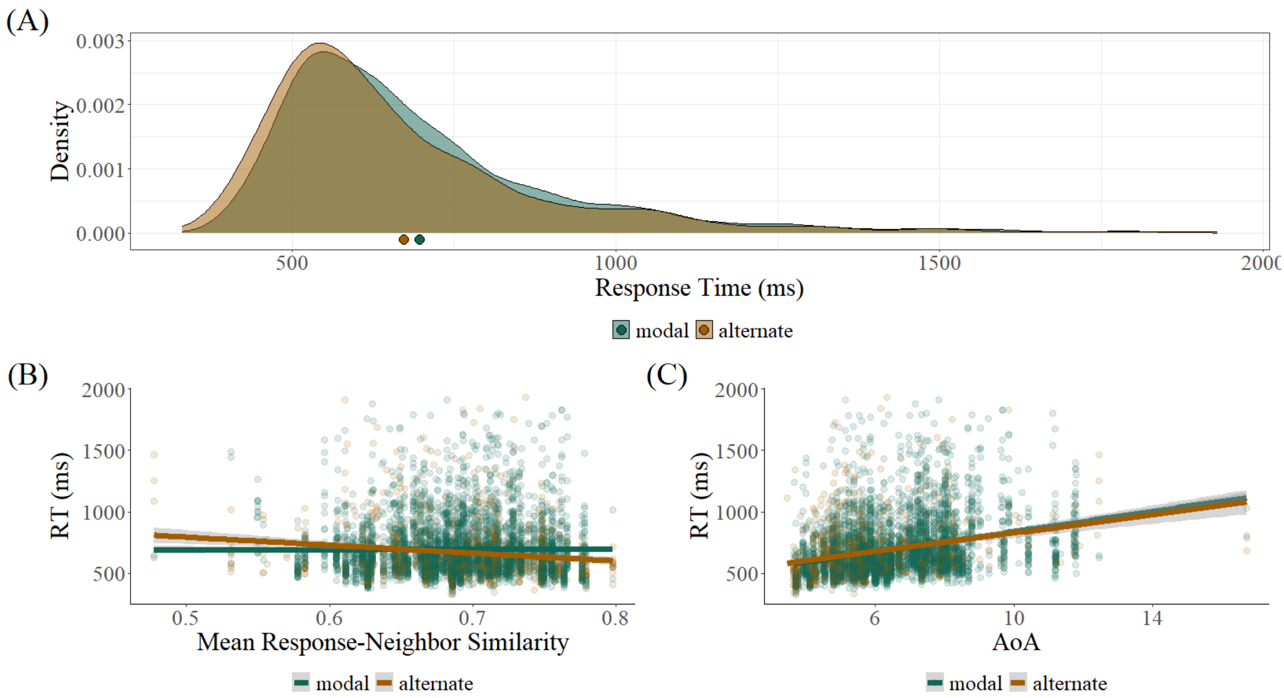
[7] Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186.

What drives word choices and naming latencies? Examining the roles of semantic and lexical variables in modal and alternate word production

Christina Papoutsi[1], Elli Tourtouri[1,2], Vitoria Piai[1,3], Antje S. Meyer[1,3]
[1] Max Planck Institute for Psycholinguistics; [2] Institute of Cognitive Science, Osnabrück University; [3] Donders Institute for Brain, Cognition, and Behavior, Radboud University
christina.papoutsi@mpi.nl

Producing a word requires its selection among a number of co-activated competitors [e.g., 1]. While speakers tend to select words that optimally convey their intended message (e.g., saying "kitten" for a young feline), they sometimes produce semantically-related words (e.g., "cat" or "pet") that are less typical but "good-enough" alternatives [2]. How do these alternate names arise and what factors influence their production? In the present study, we compared the response times (RTs) of alternate and modal responses in a speeded picture-naming task in order to examine whether alternates are produced with similar efficiency as modals, or whether differences in RTs indicate that alternate responses involve less or more effortful selection. In addition, building on previous findings showing that RTs for modal responses are influenced by similarity to semantic neighbors, semantic neighbourhood size, age of acquisition (AoA) and lexical frequency [e.g.,3, 4], we tested whether these variables differentially influence the naming performance of modal and alternate responses. In this way, we aimed to shed light onto the factors that drive word choices, thus gaining a more comprehensive understanding of the cognitive processes that govern word selection. Sixty participants performed a two-session picture naming task with 160 concrete objects, each with at least two possible names according to the BOSS database [5]. In session 1, they performed a speeded naming task, designed to elicit alternate names. A week later (session 2), they named the same pictures without time constraints, providing both their preferred modal and alternate names, thus allowing us to classify responses of session 1 into modal and alternate based on each participant's naming preferences. When comparing the RTs for alternate and modal responses in session 1, we found a significant RT difference (t = 2.476), with alternate responses (M = 670 ms) being overall faster than modal responses (M = 688 ms; Fig. 1A). This suggests that, under time pressure, alternates arise because they are more strongly activated than modals and thus easier to retrieve. As regards the effects of semantic and lexical variables on the RTs for modal and alternate responses, mean semantic similarity to near neighbors facilitated the retrieval of alternate names (t =-2.098), but had no effect on modal names (t =.240; Fig. 1B). In other words, when an alternate name was closely linked with other neighboring words in the semantic network, these neighbors boosted its activation, thus speeding up its retrieval [6]. On the other hand, modal name activation was not affected by the proximity of their semantic neighbors, possibly due to their already strong connection to their conceptual representations. These findings suggest that modal and alternate responses differ in how they interact with their semantic network. Furthermore, AoA was a robust predictor of naming latencies for both modal and alternate responses (t = 6.261), with earlier-acquired words being retrieved more quickly (Fig. 1C). The other semantic and lexical predictors did not affect RTs for either response type (t < |2|). Taken together, these findings suggest that while naming preferences may vary across speakers, alternate names emerge because they are more accessible than modal names. Finally, we show that AoA influences the production of modal and alternate names in the same manner, while semantic similarity affects their production differentially.

Figures



(A) RT distributions for modal and alternation responses. Dots below the curves indicate the mean RT for each response type. **(B & C)** Effects of mean response-neighbor similarity and age of acquisition on RTs for modal and alternate responses. In Panels B and C, each point represents a single trial, and lines show the fitted regression lines with confidence intervals. Only predictors with statistically significant effects are plotted for conciseness.

References

[1] Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22(01). https://doi.org/10.1017/S0140525X99001776

[2] Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. Trends in Cognitive Sciences, S1364661322000201. https://doi.org/10.1016/j.tics.2022.01.005

[3] Fieder, N., Wartenburger, I., & Abdel Rahman, R. (2019). A close call: Interference from semantic neighbourhood density and similarity in language production. *Memory & Cognition*, *47*(1), 145–168. https://doi.org/10.3758/s13421-018-0856-y

[4]Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

[5] Decuyper, C., Brysbaert, M., Brodeur, M. B., & Meyer, A. S. (2021). Bank of Standardized Stimuli (BOSS): Dutch Names for 1400 Photographs. Journal of Cognition, 4(1), 33. https://doi.org/10.5334/joc.180

[6] Abdel Rahman, R., & Melinger, A. (2019). Semantic processing during language production: An update of the swinging lexical network. Language, Cognition and Neuroscience, 34(9), 1176–1192. https://doi.org/10.1080/23273798.2019.1599970

Humans are bad at recognizing AI - but they can learn it from feedback

Eva Pospíšilová[1], Ondřej Drobil[1], Anna Marklová[1], Jiří Milička[1]
[1] Faculty of Arts, Charles University
odrobil@seznam.cz

Background: The growing presence of large language models (LLMs) in everyday text production raises questions about human ability to distinguish between machine- and human-authored content. Previous studies, mostly conducted in English, have reported mixed results on whether people can recognize AI-generated texts, with performance often near chance levels [1, 2]. Our study investigates this question in Czech - a morphologically rich, lower-resourced language underrepresented in LLM training data - and considers how feedback and register (genre) variability affect human recognition performance.

Method: A dataset of 672 Czech text pairs was created, each consisting of a human-written and a GPT-4-generated continuation of the same source passage, taken from the stylistically diverse Koditex corpus. Using a within-subjects design, 255 native Czech speakers completed 20 trials of a side-by-side authorship identification task. Participants were randomly assigned to a *feedback* group (receiving immediate feedback after each trial) or a *no-feedback* group. Register variation was quantified through multidimensional analysis (MDA) [3,4], allowing us to assess whether genre-like stylistic differences influenced performance. We also collected demographic data, attitudes towards AI, and self-reported AI usage frequency. We used bootstrapped 95% confidence intervals and mixed-effect logistic regression for the analysis.

Results: Participants in the feedback condition significantly outperformed ($p < .001$) the participants without feedback, achieving 65.1% (95% CI: 62.2%--67.7%) accuracy versus 55.5% (95% CI: 52.0%--58.4%) (both above chance, but only learners showed improvement over time). Trial order was a significant predictor of success for the *feedback* group ($p < .001$), confirming that feedback facilitated learning. *No-feedback* participants relied more on general stylistic assumptions (e.g., AI texts are more static or cohesive), while learners adapted away from these cues. Attitudes toward AI and frequency of AI use showed minimal influence on performance. The readability judgment ("Which text was easier to read?") emerged as a strong predictor: participants were more accurate when the more readable text was also human-written.

Discussion: Our findings demonstrate that while humans initially struggle to distinguish AI-generated texts, they can significantly improve through feedback. Moreover, it shows how various variables, such as register or subjective readability of the texts, influence their decisions. The results of our study contribute to the understanding of language and text comprehension mechanisms employed when engaging with AI-generated texts, showing that people process texts with certain biases when determining AI authorship, and that through feedback, readers can revise their initial (and often misleading) assumptions about stylistic properties of human- and AI-authored texts.

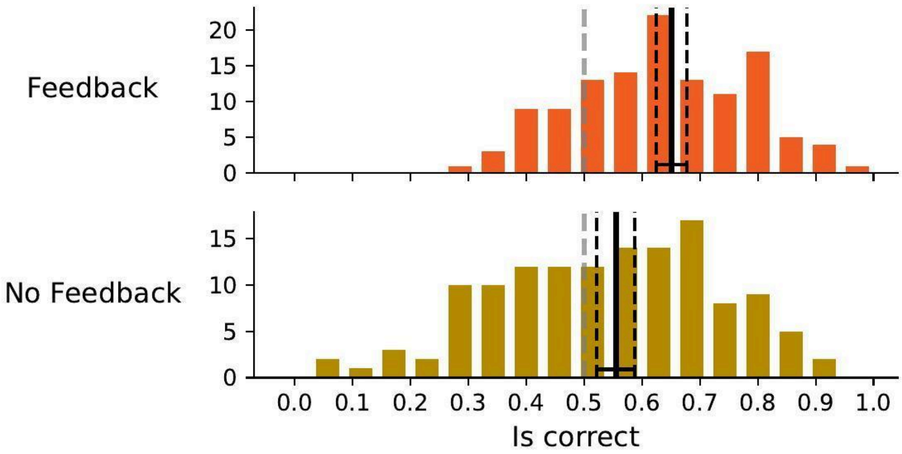


Figure 1: Correctness rate of participants in feedback and no feedback group, with 95% CIs.

References

[1] Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021, July). *All that's 'human' is not gold: Evaluating human evaluation of generated text* [arXiv preprint arXiv:2107.00061]. arXiv. https://doi.org/10.48550/arXiv.2107.00061

[2] Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., & Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence, 6*, 100209. https://doi.org/10.1016/j.caeai.2024.100209

[3] Cvrček, Václav, Zuzana Laubeová, David Lukeš, Petra Poukarová, Anna Řehořková, and Adrian Jan Zasina. 2020. *Registry v češtině*. NLN. Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2020). *Registry v češtině*. NLN.

[4] Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

How Bilinguals Use Grammatical Cues to Make and Revise Predictions: Effects of Age of Onset and Cross-Linguistic Influence

Vera Heyer[1], Holger Hopp[1], Regina Hert[2], E Jamieson[3], Barbara Köpke[2] & Monika S. Schmid[3]

[1]TU Braunschweig, [2]Université Toulouse II, [3]University of York

v.heyer@tu-braunschweig.de

Language users actively predict upcoming input in language comprehension at various levels using different types of cues. Predictive processing is also adaptive in that users can modulate predictions, for instance, by revising their initial predictions once they turn out to be inaccurate (Chow & Chen, 2020). In this study, we test whether and how bilinguals flexibly adapt predictions and revise them incrementally. In particular, we aim to disentangle general bilingualism effects from cross-linguistic influence and age effects in bilingual processing. To this end, we compare L1 French or L1 English late L2 learners of German, L1 German late L2 learners of English or French and adult French/English-German simultaneous bilinguals, who had been speaking both languages from birth (Table 1, data collection ongoing). We examine the time-course of how bilinguals make and revise predictions according to semantic and grammatical information in German, namely verb semantics (*eat* vs. *move*; Altmann & Kamide, 1999) and grammatical gender (*der* vs. *das*; masculine vs. neuter articles; Hopp, 2013). While verb meanings are shared across the three languages, German and French have grammatical gender, but English does not. In a visual world eye-tracking study, participants listened to sentences with or without gender cues for prediction of a sentence-final noun (1). In addition, conditions with a semantically-constraining adjective created an initial semantic prediction (*hungrig* 'hungry' in (2a) → banana in Fig. 1 B), which would then need to be revised based on conflicting semantic cues (semantic revision (2a): *trinkt* 'drinks' → water) or grammatical cues (gender revision (2b): article *das* → water). For the interim data set, LMER analyses of target fixations showed that early bilinguals and German expats with long-time contact to English and French used both semantic and grammatical cues to make predictions (Table 2). They also used both cues to revise initial semantic predictions to immediately launch novel predictions to the target (*water*). L2 learners with long-time contact to German also used semantic and grammatical cues to make initial predictions, and they could also revise initial predictions according to later semantic cues. Yet, they failed to revise and overwrite initial semantic predictions by grammatical gender cues (Table 2 & Fig. 2). Among L2 learners, L1 English learners showed greater initial use of gender than L1 French learners, indicative of cross-linguistic influence. The findings suggest that long-time early or late bilingualism does not affect the incremental and flexible use of native-language cues in processing. In contrast, later onsets of language acquisition decrement learners' ability to flexibly use L2 grammatical information in sentence processing. Specifically, late L2 learners cannot overwrite semantic commitments by grammatical cues, even if they can use these grammatical cues for prediction in isolation. Over and above group differences in age of onset and L1, we also explore effects of within-group differences in language dominance and use. We discuss how the combination of cross-linguistic influence and age of onset leads (late) bilinguals to prioritize semantic over competing grammatical cues in sentence processing.

Materials

(1) Non-constraining initial adjective: Prediction by grammatical gender

a. Er ist **aufmerksam** […] und **holt** ***das*** gut erreichbare Wasser.
He is attentive and fetches the$_{neut}$ easily accessible water.

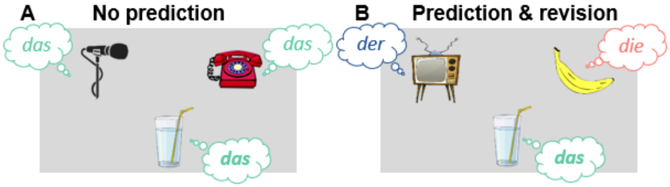
(2) Semantically-constraining initial adjective: Revision by semantic (a) or grammatical (b) cues

a. Er ist ***hungrig*** […] und ***trinkt*** das gut erreichbare Wasser.
He is hungry and drinks the$_{neut}$ easily accessible water.

b. …und **holt** ***das*** gut erreichbare Wasser.
…and fetches the$_{neut}$ easily accessible water.

'He has been attentive/hungry […] and drinks/fetches the easily accessible water.'

Figure 1: Examples of visual displays



(1a): Gender prediction: Only gender on article picks out target (*water*) in 1-gender (A) vs 3-gender display (B)
(2a): Semantic revision: Constraining adjective (*hungry*) biases to banana and verb revises prediction to target (*water*) in 2-gender-display
(2b): Gender revision: Constraining adjective (*hungry*) biases to banana and gender on article revises prediction to target (*water*) in three-gender display (B)

Table 1: Participant information (data collection ongoing)

	L2 German	L1 German	Bilingual
N	66	61	47
L1	Engl./French	German	Ger + E/F
L2	German	Engl./French	-
Mean Age	42.5	47.7	38.2
Residence	Germany	UK/France	Ger/UK/F
LOR in L2 country	5-39 yrs	5-38 yrs	-

Figure 2: Proportion of looks to the target relative to the determiner onset for the L2 group

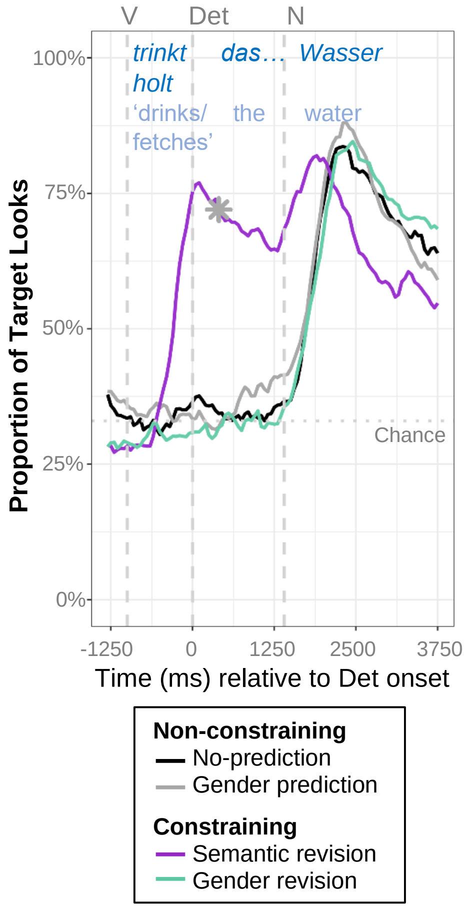


Table 2: Proportion of looks to the target (vs. competitors) in the noun window (200 ms around noun onset*)

	L2 German	L1 German	Bilingual
Non-constraining			
No Prediction	35%	36%	37%
Gender Prediction	42%	53%	50%
Semantically constraining			
Sem. Revision	73%	81%	78%
Gender Revision	35%	51%	41%

*Note: For the Semantic Revision condition, a control point (star in Fig. 2) was used the same period of time after the verb (i.e., the disambiguating point) as the period from article to noun in the other conditions to allow for the same amount of time for revising the initial prediction as in the Gender Revision condition.

References
Altmann, G.T. & Kamide, Y. (1999). *Cognition*, 73, 247–264. Chow, W.-Y. & Chen, D. (2020). *LCN*, 35, 1149-1161. Hopp, H. (2013). *SLR*, *29*, 33-56.

Is code-switching effortless? A look at processing and production costs

Yoana I. Dancheva[1], Margreet Vogelzang[1,2], Ianthi M. Tsimpli[1]
[1] Theoretical and Applied Linguistics, University of Cambridge; [2] School of Psychology, Newcastle University
yid20@cam.ac.uk

Background: Bilingual individuals incorporate code-switching (CS) in their speech frequently and in a seemingly effortless manner. Insertional CS [1], or the addition of other-language (i.e., embedded language) lexical items into a sentence, and alternational CS, the addition of structures from the embedded language into the matrix language [2], are commonly found in CS speech:

(1) Tova mi beshe posledniyat **experience** s mediyata. (insertion, Bulgarian-English)
'This was my last experience with the media.'

(2) Kazhi mi koga si svoboden **and we'll arrange something.** (alternation, Bulgarian-English)
'Let me know when you're free and we'll arrange something.'

The use of CS can alleviate cognitive demand, as suggested by faster global speech rates when bilinguals use CS instead of adhering to a unilingual mode [3]. However, other studies report a processing cost, reflected in slower responses to CS sentences than to L1 sentences by early bilinguals [4] and slower speech rates before CS compared to unilingual equivalents [5]. Similarly, while some studies show that early and late bilinguals from different communities can process CS without much effort [6,7], lower repetition accuracy of CS sentences compared to L1 sentences in late bilinguals in a different study [8] seems to suggest increased processing effort. In light of these seemingly contradictory findings, we examine the processing and production of CS speech by late L1 Bulgarian L2 English bilinguals to better understand whether CS can be facilitatory, or whether it incurs a cost.

Method: 33 Bulgarian-English bilinguals residing in the UK were presented with 16 unilingual utterances in their L1 and L2, and 16 CS utterances consisting of both insertional and alternational CS with Bulgarian as the matrix language (see examples (1) and (2)). In the insertional CS condition, half of the items included inflectional morphology. Participants listened to the sentences and rated them (which recorded ratings and reaction times) and afterwards repeated them aloud (used for verbatim recall accuracy and speech rate analyses).

Results: Preliminary results show higher ratings for unilingual conditions than CS conditions ($p < .01$). However, no significant difference in reaction times to unilingual vs. CS conditions was found, indicating neither a facilitatory nor an inhibitory effect. The repetitions were more accurate for CS than unilingual English utterances ($p < .01$), with no difference between CS and unilingual Bulgarian, indicating better production ability and possibly easier processing for both CS and L1 compared to L2. However, speech rate was slower in the alternational CS ($p < .05$) and the unilingual English conditions ($p < .01$) compared to the unilingual Bulgarian or the insertional CS conditions. Finally, within the insertional CS condition, the speech rate of insertions with inflections, which show syntactic and phonological adaptation to the L1, was significantly faster than that of those without inflections ($p < .01$), indicating a facilitatory effect.

Discussion: Our findings suggest that bilinguals process and produce unilingual L1 and CS sentences with similar efficiency, despite some increased production cost. While some of our results align with previous research findings [5,7], others oppose the literature on the topic [3,4,8]. Our results demonstrate that CS can be effortless for bilinguals and could potentially alleviate cognitive demand.

References

[1] Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press, Cambridge.

[2] Myers-Scotton, C. (2002). *Contact linguistics bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

[3] Johns, M. A., & Steuck, J. (2021). Is codeswitching easy or difficult? Testing processing cost through the prosodic structure of bilingual speech. *Cognition, 211.*

[4] Adamou, E., & Shen, X. R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, *23*(1), 53-70.

[5] Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110–137.

[6] Beatty-Martínez, A. L., Guzzardo Tamargo, R. E., & Dussias, P. E. (2021). Phasic pupillary responses reveal differential engagement of attentional control in bilingual spoken language processing. *Scientific Reports*, *11*(1), 23474.

[7] Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173-189.

[8] Hofweber, J., & Marinis, T. (2023). What sentence repetition tasks can reveal about the processing effort associated with different types of code-switching. *Languages, 8*(1), 70.

How Large Language Models Evaluate Embedded *Wh*-Questions: A Cross-Linguistic Comparison of Chinese and English

Yixin Cui, Lavinia Salicchi, Yu-Yin Hsu*
The Hong Kong Polytechnic University
* yyhsu@polyu.edu.hk

Non-local linguistic dependencies (e.g., *wh*-questions) reveal cognitive constraints in sentence processing, requiring greater effort due to structural distance. Cross-linguistically, *wh*-dependencies vary: Chinese (CN) allows *wh in-situ*, while English (EN) requires *wh*-displacement. Previous research [1] showed that humans reliably distinguish grammatical from ungrammatical sentences. Conversely, large language models (LLMs), though advanced, exhibit inconsistent abstraction—for example, GPT-3.5-turbo uses syntactic cues, while text-davinci-003 relies on surface patterns, neither matching human precision [2]. Moreover, most such investigations focus on English alone.

The current study investigated whether multilingual LLMs can account for grammaticality in complex *wh*-sentences. We also perform cross-lingual comparisons and examine different types of LLMs' performance. We used materials and grammaticality judgements from [1], which include 3 types of clausal verbs ([+wh], [-wh], [±wh]). We had 41 pairs of sentences in both Chinese and English: Each pair contains one sentence with a *wh*-dependency and one does not. The EN and CN versions shared the same contents. Sentence surprisal scores [3] were used as a proxy for cognitive load [4-6]. The scores were computed by 5 multilingual LLMs (*Qwen2.5-0.5B, bloom-560m, bloomz-1b1, xglm-564M,* and *OpenCoder-1.5B-Instruct*, Table 1). We conducted paired t-tests to evaluate **whether LLMs' surprisal scores reflect grammaticality contrasts**. The results showed that all LLMs present significant differences in the surprisal scores between grammatical and ungrammatical sentences across languages (i.e., EN + CN), EN only, and CN only. The only exception was the model *xglm-564M*, which did not show a significant difference on the CN materials. To investigate **cross-lingual performance between CN and EN**, we conducted rank-based ANOVAs with interaction effects and post-hoc tests (Table 2). Only *Qwen2.5-0.5B* distinguished grammatical from ungrammatical sentences across languages, assigning lower surprisal to grammatical sentences. However, no model showed a significant difference between grammatical and ungrammatical sentences in either language. **Comparisons among LLMs** revealed that only *Qwen2.5-0.5B* showed significant surprisal differences based on grammaticality (averaged across both languages). This suggests that models' training data volume may have a greater impact on performance than model size: despite being smaller than the other models, *Qwen2.5-0.5B*, performed better across languages, maybe due to its larger multilingual training dataset. We also examined whether verb features posed different processing challenges to LLMs. As shown in Table 3, t-tests on [-wh] and [+wh] sentences showed that no EN LLM distinguished grammatical from ungrammatical [-wh] items, while all LLMs correctly categorized the grammaticality of [+wh] sentences. In CN, no statistically significant difference was found for either [-wh] or [+wh] structures. In other words, LLMs in both languages fail to capture grammaticality contrast when clausal verbs take a [-wh] complement, whereas EN LLMs can account for [+wh] structure, a type that remains challenging for CN LLMs. The different performance between EN and CN may be due to CN's *wh*-in-situ nature, or because of an imbalanced representation of the two languages in LLMs' training data.

In sum, our findings suggest that LLMs' surpisal scores are not reliable in distinguishing between grammatical and ungrammatical sentences in complex syntactic structures, especially in Chinese.

Table 1. Basic information of 5 LLMs. T= trillion, B= billion, and M= million.
[1]Qwen Team (2024); [2]BigScience (2022); [3]Muennighoff et al. (2022); [4]Lin et al. (2021); [5]Huang et al. (2024).

Model	Parameters	Training Set Size (tokens)	Language Proportions (English/Chinese)	Primary Purpose
Qwen2.5-0.5B[1]	500M	18T	Unknown	Multilingual model (English and Chinese focus)
bloom-560m[2]	560M	350B	~31.3% English, ~18.3% Chinese	Multilingual model
bloomz-1b1[3]	1.1B	350B + 502M	> 31.3% English, < 18.3% Chinese	Instruction-tuned multilingual mode
xglm-564M[4]	564M	500B	~49% English, ~8% Chinese	Balanced multilingual model for 30 languages
OpenCoder-1.5B-Instruct[5]	1.5B	2.5T	Unknown	Code-focused instruction-tuned model

Table 2. Results of post-hoc pairwise comparisons between languages and grammaticality.
**= p < .05, **= p < .01, and ***=p < .001.*

Model		Gram. - Ungram.	Gram. CN - Ungram. CN	Gram. EN - Ungram. EN
Qwen2.5-0.5B	estimate	**-125.28****	-74.78	-131.35
	SE	34.28	42.04	42.04
	t	-3.66	-1.78	-3.12
bloom-560m	estimate	-100.26	-55.49	-103.44
	SE	34.28	42.04	42.04
	t	-2.92	-1.32	-2.46
bloomz-1b1	estimate	-83.95	-34.43	-106.88
	SE	34.28	42.04	42.04
	t	-2.45	-0.83	-2.54
xglm-564M	estimate	-39.62	0.26	-93.16
	SE	34.28	42.04	42.04
	t	-1.16	0.01	-2.22
OpenCoder-1.5B-Instruct	estimate	-77.74	-75.08	-62.56
	SE	34.28	42.04	42.04
	t	-2.27	-1.79	-1.49

Table 3. Results of t-test on grammaticality among languages, verb types, and LLMs.

Model	Chinese				English			
	-wh		*+wh*		*-wh*		*+wh*	
	t	***p***	***t***	***p***	***t***	***p***	***t***	***p***
bloom-560m	-0.03	0.97	-1.77	0.09	-0.71	0.49	-2.27	0.03
bloomz-1b1	-0.17	0.87	-1.12	0.28	-1.58	0.15	-3.10	0.01
xglm-564M	-0.03	0.98	-0.73	0.47	-1.28	0.23	-2.30	0.03
OpenCoder-1.5B-Instruct	-0.24	0.81	-0.37	0.72	-0.77	0.46	-2.33	0.03
Qwen2.5-0.5B	-0.18	0.86	-1.25	0.22	-1.37	0.20	-3.15	0.01

References

[1] Kong, D., & Hsu, Y. Y. (2023). Processing covert dependency: An eye-tracking study of scope interpretations of embedded Wh-questions in Mandarin. *Plos one*, *18*(5), e0285873.

[2] Keleş, O. & Deniz, N. D. (2024). A comparative study with human data: Do LLMs have superficial language processing? 2024 32nd Signal Processing and Communications Applications Conference (SIU), Mersin, Turkey.

[3] Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In Second meeting of the north american chapter of the association for computational linguistics.

[4] Levy, R. (2013). Memory and surprisal in human sentence comprehension. In Roger van Gompel (Ed.), Sentence Processing (pp. 90-126). London and New York: Psychology Press.

[5] Levy, R. & Gibson, E. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. Front Psychol, 4, 229.

[6] Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2024). Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. Neurobiology of language (Cambridge, Mass.), 5(1), 107–135.

Examples of three types of matrix verbs in *wh*-sentences:

(1)

a. 赞助商好奇电视台取消了*哪些节目*。
The sponsors wonder *which shows* the network has canceled.

b. *赞助商好奇电视台取消了*那些节目*。
*The sponsors wonder *those shows* the network has canceled.

(2)

a. *群众们相信市政府禁止了*哪些行为*。
*People believe *which behaviors* the city government has forbidden.

b. 群众们相信市政府禁止了*那些行为*。
People believe the city government has forbidden *those behaviors*.

(3)

a. 警察知道清洁工偷走了*哪些文件*。
The police know *which documents* the cleaner has stolen.

b. 警察知道清洁工偷走了*那些文件*。
The police know the cleaner has stolen *those documents*

Explanations:

(1) Interrogative clausal verbs: [***+wh***]

Verbs like 问 (*wen*, 'ask') and 好奇/ 想知道 (*haoqi*/ *xiangzhidao*, 'wonder') require *wh*-phrases as a part of its complement clause and are incompatible with a declarative embedded clause, as shown in (1).

(2) **Declarative clausal verbs: [*-wh]***

Verbs like 相信 (*xiangxin*, 'believe') and 怀疑 (*huaiyi*, 'suspect') require declarative complement clauses and reject *wh*-phrases therein, as shown in (2).

(3) *Optional interrogative clausal verbs:* ***[+/-wh*]**

Verbs such as 知道 (*zhidao*, 'know') and 料到 (*liaodao*, 'foresee') accept both interrogative and declarative complements, as shown in (3).

Remembering times ahead:
The effect of linguistic framing on representational momentum in state-change events

Xueyi Yao[1], Natalia Jardon[2], Jonathan Kominsky[3], Eva Wittenberg[4]

Central European University, Vienna, Austria[1,2, 3,4], Universitat Autònoma de Barcelona[2]

yao_xueyi@phd.ceu.edu, natalia.jardon@uab.cat, KominskyJ@ceu.edu, WittenbergE@ceu.edu

Imperfective aspect provides a "lens" into the internal stages of an event [1-2], increasing access to in-progress details, whereas perfective aspect emphasizes an event's completion and outcome [3-4]. Prior work has shown that linguistic framing influences event memory [5-6], but whether grammatical aspect affects event memory is unclear. We test this question by leveraging the representational momentum effect, a phenomenon that people remember the positions of moving objects further ahead than they actually were [7]. Recently, this phenomenon has also been found in state-change events [8]. Here, we ask whether grammatical aspect could permeate the representational momentum effect in state-change events.

Exp. 1a&b (replications): Two preregistered experiments (N= 50 each) replicated [8] at two different speeds. Participants watched videos depicting objects undergoing state changes (e.g. a log burning) and indicated the last frame they saw from the video by pulling a slider (Fig.1) We observed a significant positive frame error in both 1a (M = 4.97, $t(49) = 2.41$, $p < 0.02$) and 1b (M = 3.75, $t(49) = 2.32$, $p < 0.03$) (Fig.2), showing that participants in both experiments reported a frame further forward in time relative to the true final frame. **Exp. 2 (calibration task):** The goal of Experiment 2 was to investigate how the videos used in Experiment 1 correspond to people's interpretation of grammatical aspect. Participants' judgments indicated a preference for imperfective sentences, and their acceptability of the perfective sentence increased as the video progressed (Fig.3). A model comparison confirmed that the pause frame had a significant effect on the choice ($\chi 2(3) = 59.79$, $p < 0.001$). **Exp. 3a&b (linguistic manipulation)** manipulated the grammatical aspect of a sentence describing the event, using either perfective or imperfective aspect (e.g., *the log was burning* vs the *log has burned*) in the same paradigm as before (Fig.1). We again replicated the representational momentum effect, as revealed by a significant overall positive frame error (M = 7.4, $t(1612) = 18.24$, $p < .001$). Further, as predicted, shown in Fig.4, perfective aspect led people to reconstruct an event as further advanced (M = 8.3) than imperfective aspect (M = 6.5), a difference that was statistically significant ($t(1608) = 2.25$, $p < .03$, 95% CI [0.21, 3.11]). Experiment 3b replicated 3a with a timing consistent with [8]. Representational momentum effect persisted (M = 5.35, $t(1642) = 14$, $p < .001$), with perfective again yielding a greater positive frame error (M = 6.1 vs M = 4.4; $t(1638) = 2.11$, $p = .035$, 95% CI [0.11, 2.88]).

Taken together, our work presents evidence that representational momentum effect in state change events can be permeated by linguistic framing, specifically the use of grammatical aspect. Whether our evidence of linguistic framing on the representational momentum effect results from information exchange between language and visual memory recall, or already during visual event encoding, is a question that should be addressed in future studies.

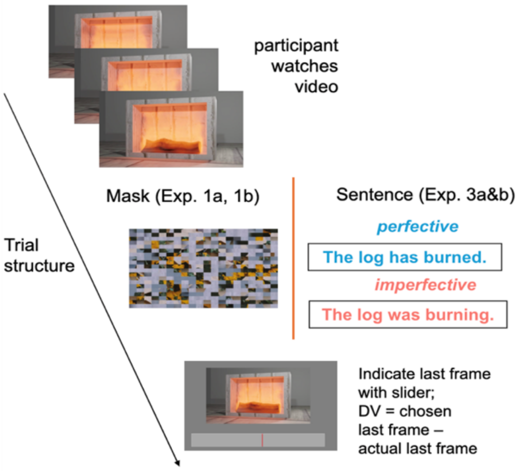


Figure 1. The example trial of Experiment 1a,1b, 3a and 3b. Each video lasted 240 frames and was stopped at pause frame 60, 120, or 180. After the video was shown, a visual mask (Exp.1a&qb) or a sentence (Exp3a&3b) appeared. Participants' task was to pull a slider which stepped through the animation frame by frame to indicate the last frame they had seen from the video. If the video stopped at frame 120, any frame reported after frame 120 indicates a positive frame error, any before a negative.

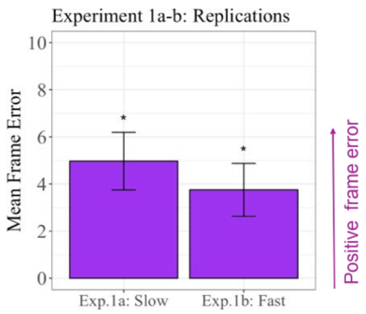


Figure 2: Results of two replications of [8], showing significant positive frame errors (error bar indicates the standard error of mean).

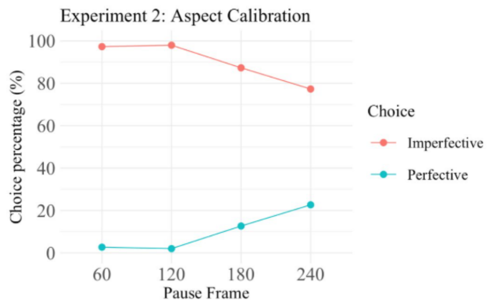


Figure 3: *Results of Exp.2, showing participants' choice percentage of perfective and imperfective descriptions across four pause frames.*

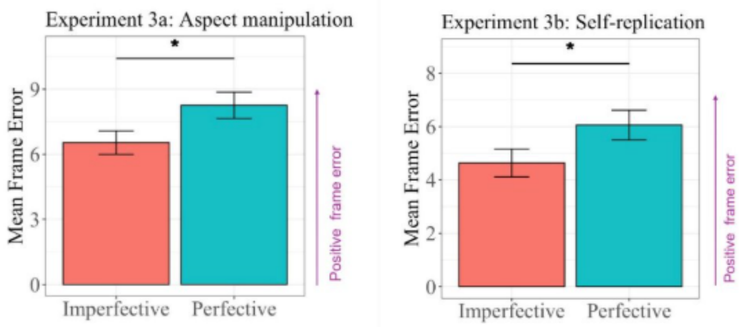


Figure 4: Results of Exp. 3a, showing a significant difference between the mean frame error of perfective and imperfective conditions (error bars indicate the standard error of the mean).

References:

[1] Ferretti el al., (2007). *Experimental Psychon*, 33(1). **[2]** Madden & Therriault (2009). *Experimental Psychon, 62(7).* **[3]** Misersky et al., (2021). *Cognition*, 214, 104744. [4] Madden & Zwaan (2003). *Memory & Cognition,* 31(5), 663–672. [5] Sakarias & Flecken (2019). Cognitive Science, 43(1) [6] Skordos et al. (2020). Cognitive Neuropsychology, 37(5-6), 254–270. [7] Hubbard (2005) *Psychon*. Bull. Rev. 12, 822–851 (2005). [8] Hafri et al. (2022). *Psychological Science*, 33(5), 725–735.

An Eye-tracking Study on the Presupposition Processing of Korean L2 Learners of German: Focusing on "wieder (again)"

Yunju Nam[1], Geon Kim[1], Gaeun Lim[1]
[1] Hanyang University
yjnam05@hanyang.ac.kr

Background: *Wieder* ("again") is a paradigmatic presupposition trigger, presupposing that an event has previously occurred (e.g., “Peter öffnet wieder die Tür” / “Peter opens the door again”). Dynamic and update semantics frameworks highlight such triggers' roles in discourse coherence and presupposition projection under negation. Interestingly, in German, the presupposition differs between *wieder nicht* (“not occurred before”) and *nicht wieder* (“occurred before”). Recent studies show that native speakers rapidly detect presupposition-context mismatches for *wieder nicht*, but detection is delayed when *wieder* is embedded under negation (*nicht wieder*) [1]. Likewise, in Korean, different meanings can be presupposed depending on the negative form ("~-ji anta" vs. "an ~-ta") [2]. Building on these findings, the present study examines how Korean L2 learners of German process such presuppositions in real time, comparing their patterns to native speakers.

Method: Twenty-eight intermediate-to-advanced Korean learners of German participated in an eye-tracking experiment. Participants were exposed to sentences containing *'wieder nicht*' or '*nicht wieder*' following the contexts either felicitous or infelicitous with the presupposed event. First fixation duration (FFD), First pass (FPGD) and Second pass gaze duration (SPGD) to the verb (right after ‘*wieder nicht*’ or ‘*nicht wieder*’) were analyzed and effects of contextual felicity and negation embedding were evaluated using linear mixed-effects models (FFD/FPGD/SPGD∼Felicity∗Embedding+(1|Participant) +(1|Item)).

Results: Unlike native speakers in Schwarz and Tiemann (2016), Korean L2 learners showed no significant early-stage sensitivity to presupposition-context mismatches (FFD: $F(1, 95.40) = 0.11$, $p = .74$; FPGD: $F(1, 92.42) = 1.18$, $p = .28$). Only in later stages, SPGD was significantly increased on presupposition-contextual infelicitous condition ($F(1, 69.01) = 10.56$**), indicating delayed integration. Furthermore, while native speakers exhibited strong felicity x negation embedding interaction effects, Korean learners' SPGD revealed independent main effects of embedding without significant interaction ($F(1, 68.54) = 9.76$**), longer SPGD in not embedded conditions), suggesting the position of ‘Wieder’ (embedded or not) influenced the later rereading process of L2 learners.

Discussion: Korean L2 learners did not exhibit early-stage sensitivity to felicity violations in either embedding condition, as indicated by non-significant differences in first fixation and first-pass gaze durations. Instead, a significant effect of felicity emerged only in later-stage measures (SPGD), suggesting that L2 learners rely more heavily on post-integration or reanalysis processes to detect presupposition-context mismatches. While native speakers showed an interaction between embedding and felicity (i.e., different sensitivity patterns depending on the syntactic placement of negation), L2 learners demonstrated only independent main effects of each factor in their delayed, second-pass gaze durations, with especially prolonged rereading times for non-embedded sentences. Moreover, L2 leaners proactively generate context-dependent presuppositions online, processing felicitous *nicht wieder* structures more easily despite their overall delayed sensitivity. This pattern likely reflects positive transfer from Korean negation (local vs. global scope) as well as general cognitive constraints on L2 processing—forcing learners to rely on surface-level strategies and defer presupposition computation under negation until post-integration. Overall, our results indicate that although Korean L2 learners can eventually access and integrate presuppositional meaning, their processing strategies differ from those of native speakers.

[Table 1] Participant Information

Measure	Value	Measure	Value
Number of Participants	24	German Proficiency	Intermediate to Advanced (B1-C1)
Mean Age (years)	22 (SD = 2.15)	Mean LexTALE Score	59.1 (SD = 5.9)
Gender	17 Female, 7 Male	Overseas Residency Experience	< 6 months

[Table 2] Experimental Conditions and Example Target Sentences (ROI: boldic and underlined)

Condition		Context (“short Translation”)	Target Sentence (“Translation”)
FN	Felicitous (negative context) Not Embedded (wieder nicht)	Tina wollte letzte Woche mit Karl zum ersten Mal Schlittschuhlaufen gehen. Doch das Wetter war miserabel, und sie haben den Plan aufgegeben. “Tina wanted to go ice skating last week but didn't.”	Dieses Wochenende war Tina *wieder nicht* **Schlittschuhlaufen**. ("This weekend, Tina did not go ice skating again.")
IN	Infelicitous (positive context) Not Embedded (wieder nicht)	Tina war letzte Woche mit Karl zum ersten Mal Schlittschuhlaufen. Das Wetter war wunderschön, und sie haben sich prächtig amüsiert. “Tina went ice skating last week”	
FE	Felicitous (positive context) Embedded (nicht wieder)	Tina war letzte Woche mit Karl zum ersten Mal Schlittschuhlaufen. Das Wetter war wunderschön, und sie haben sich prächtig amüsiert. “Tina went ice skating last week”	Dieses Wochenende war Tina *nicht wieder* **Schlittschuhlaufen**. ("This weekend, Tina didn't go ice skating again.")
IE	Infelicitous (negative context) Embedded (nicht wieder)	Tina wollte letzte Woche mit Karl zum ersten Mal Schlittschuhlaufen gehen. Doch das Wetter war miserabel, und sie haben den Plan aufgegeben. “Tina wanted to go ice skating last week but didn't.”	

[Table 3] Eye-tracking Results (mean ms [SD])

indicators	‘wieder nicht’		‘nicht wieder’	
	Felicitous (FN)	Infelicitous (IN)	Felicitous (FE)	Infelicitous (IE)
FFD	325.9 [154.8]	318.1 [168.2]	317.5 [162.4]	313.6 [155.1]
FPGD	520.7 [342.5]	505.5 [338.5]	527.8 [376.9]	481.7 [307.3]
SPGD	618.4 [441.0]	871.5** [544.8]	559.9 [381.3]	643.1** [476.6]

References

[1] Schwarz, F., & Tiemann, S. (2016). Presupposition projection in online processing. Journal of Semantics, 33(3), 523–571.
[2] Park, Kihyo, & Nam, Yunju (2020). The psychological reality of presupposition of “tto” in Korean negation types. Korean Journal of Linguistics, 45-4, 843-862

Figurative Meaning Is Recoverable: Idiom Comprehension, Preference, and Processing Constraints in Adult Low-Literacy Readers

Lena Wieland, Ingo Reich
Saarland University
lena.wieland@uni-saarland.de

Comprehending idioms, e.g., *am Ball bleiben* ‘stay on the ball’ (fig. ‘persist’), poses a challenge for readers with limited literacy, particularly when literal interpretations are both semantically transparent and plausible, yet pragmatically misleading [1]. Although idioms are often assumed to be inaccessible for this population, empirical evidence is scarce. As a result, they are frequently excluded from *Leichte Sprache* (Easy German), a simplified variety for individuals with cognitive or literacy-related barriers [2, 3]. Yet it remains unclear whether this exclusion reflects genuine processing difficulty or an overgeneralization. We test whether idioms are truly inaccessible to low-literate readers. Additionally, we examine whether idiom comprehension and preference are shaped by idiom-level properties (semantic transparency, i.e., whether figurative meaning is derivable from constituent parts; and literal plausibility, i.e., whether the literal reading fits the context) and reader-level factors (literacy, familiarity). To this end, we selected idioms varying in transparency and plausibility and conducted four experiments: literacy assessment (Exp1), idiom norming (Exp2), no-context comprehension task (Exp3), and contextualized comprehension and ranking task (Exp4). We tested four hypotheses: (i) transparency facilitates access to figurative meaning, (ii) plausibility reduces preference for idiomatic construals, (iii) higher literacy enhances figurative construal, and (iv) familiarity supports accurate interpretation and idiomatic preference.

In Exp1, thirty adult low-literacy readers were evaluated using the *lea.diagnostik* tool [4] to establish individual literacy profiles. Exp2 involved thirty native German speakers without cognitive or linguistic impairments (Prolific), who rated 71 idioms for transparency and plausibility. From these, 24 were selected to construct a 2×2 design crossing high vs. low transparency and plausibility; these items were used in Exp3 and Exp4 (Table 1). In Exp3, the same low-literacy participants completed a sentence interpretation task without context. Idioms paraphrased correctly were coded as familiar. In Exp4, twenty-five of these participants returned after six months to complete two contextualized tasks: (1) a multiple-choice comprehension task (idiomatic, literal, or unrelated), and (2) a paraphrase ranking task evaluating interpretive preference. In Exp3, we used generalized linear mixed models (GLMMs) [5] to predict accuracy. Plausibility (b = 0.44, p = .008) and literacy (b = 0.03, p < .001) were significant predictors of correct responses. A transparency–plausibility interaction (b = –1.36, p < .001) indicated that idioms high on both dimensions were more likely to elicit literal interpretations—consistent with surface-level interference (Fig1). In Exp4, we again used GLMMs, fit via Monte Carlo likelihood approximation [5]. Familiarity predicted comprehension (b = 0.71, p = .024), with marginal effects of literacy (b = 0.49, p = .061) and transparency (b = 0.33, p = .078). Plausibility and interaction terms were non-significant, and model fit improved upon their removal (ΔAIC = –4.1). In the ranking task, idioms were preferred over distractors (b = 4.60, p < .001), but not over literal paraphrases (b = –0.15, p = .19) (Fig2). Transparency showed a weak negative trend (b = –0.17, p = .14). Literacy and plausibility were not predictive in this task.

These findings challenge the assumption that idioms are categorically inaccessible to low-literacy readers. With context, figurative meaning is often recoverable, especially for familiar idioms. In preference tasks, idioms were favored over distractors, but not over literal paraphrases. The results support hybrid models of idiom processing and support a context-sensitive, rather than exclusionary, approach to idioms in accessible language policy.

Table 1: Example idioms by transparency and literal plausibility

Transparency	Plausibility	Idiom (German)	Literal Gloss	Figurative Meaning
High	High	am Ball bleiben	stay on the ball	keep at it
High	Low	etwas auf dem Herzen haben	have something on the heart	have something weighing on your mind
Low	High	jemanden auf den Arm nehmen	take someone on the arm	tease/trick someone
Low	Low	auf der Nase herumtanzen	dance on someone's nose	walk all over someone

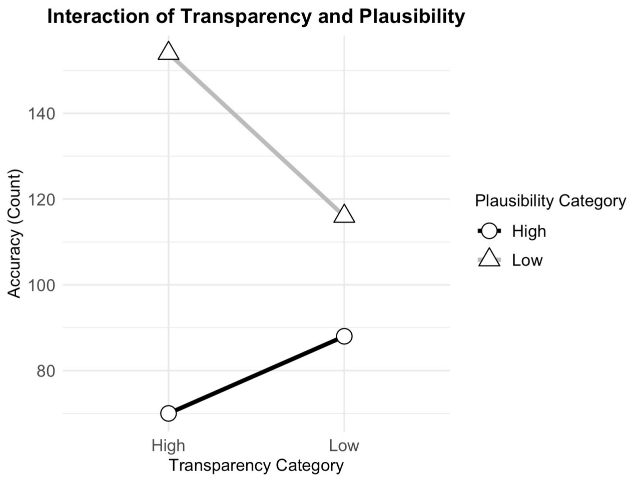


Figure 1: Interaction of transparency and plausibility in pre-study comprehension accuracy.

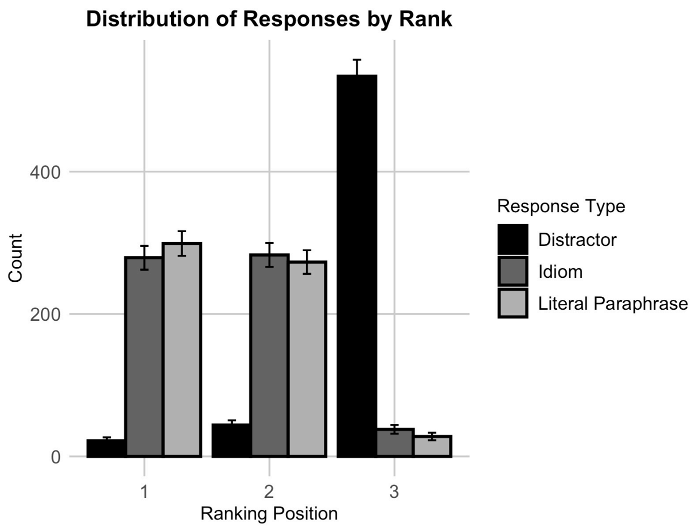


Figure 2: Ranking preferences for idiom, literal, and distractor paraphrases.

References

[1] Maria Chiara Levorato, Barbara Nesi, and Cristina Cacciari. “Reading comprehension and understanding idiomatic expressions: A developmental study.” In: *Brain and Language* 91.3 (2004), pp. 303–314. DOI: 10.1016/j.bandl.2004.04.002.

[2] Ursula Bredel and Christiane Maass. *Leichte Sprache: Theoretische Grundlagen, Orientierung für die Praxis*. Dudenverlag, 2016.

[3] Bettina M. Bock and Sandra Pappert. *Leichte Sprache, Einfache Sprache, verständliche Sprache*. Tübingen: narr/francke/attempto, 2023.

[4] Ingrid Koppel and Klaus-Dieter Wolf. “Otu.lea. Eine niedrigschwellige online-Diagnostik für funktionale Analphabetinnen in der Kursarbeit.” In: *Alfa-Forum* 86 (2014), pp. 28–31.

[5] C. Knudson. *glmm: Generalized Linear Mixed Models via Monte Carlo Likelihood Approximation (R package version 1.4.5)*. https://CRAN.R-project.org/package=glmm. 2024.

The role of social experience and motivated cognition in the representation of concepts: a behavioral and functional neuroimaging study

Doina-Irina Giurgea[1], Veronica Diveica[2], Penny M. Pexman[3], Richard J. Binney[1]

[1]School of Psychology, Bangor University; [2]Montreal Neurological Institute, McGill University; [3]Department of Psychology, Western University

rng18frn@bangor.ac.uk

Social concepts (e.g., those expressed through the words FRIENDSHIP and TRUSTWORTHY) help us understand the meaning of our social world, including language, behaviours, interactions, personal attributes, values and roles, institutions and social spaces. It has been suggested that social words are processed distinctly from non-social words (e.g., SUNFLOWER, HEMOGLOBIN), because they convey a special or privileged category of conceptual meaning, and that this is reflected in both behavioural and neural responses[1]. Dorsolateral regions of the ATL, such as the temporal pole (TP) and anterior middle temporal gyrus (MTG), activate more strongly for social word processing than non-social word processing[1]. On the other hand, the ventral anterior temporal lobe (vATL) responds equally to both types of words and is hypothesized to be the centre point of a supramodal and category-general hub for semantic knowledge[1]. However, key questions are unresolved and particularly because of (i) inconsistencies in the way individual studies have defined word socialness, and (ii) the possibility that other semantic dimensions may be confounding inferences. One such dimension could be reward, given putative links between social interaction and hedonic value. To address these issues at both behavioural and neural levels, we performed a pre-registered behavioural experiment (N = 90), followed by a pre-registered fMRI study (N = 30) in which we examined the orthogonal effects of word socialness and a reward-related dimension (i.e., motivation) within a 2x2 factorial design and a synonym judgement task. We capitalised on recent norms for word socialness[2] and motivation[3] and controlled for other relevant psycholinguistic variables, including concreteness and affective valence. At the behavioural level, we found a main effect of socialness on accuracy, in that responses were more accurate for words denoting social concepts compared to non-social concepts. Univariate fMRI analyses showed that all word categories engaged the vATL. There was also a significant main effect of socialness, whereby social words activated the left TP and anterior MTG more strongly than non-social words. This socialness effect was also observed in medial prefrontal cortex, precuneus and temporoparietal regions implicated in social cognition more broadly. Multivariate analyses revealed that local activation patterns in the left TP could distinguish between social and non-social words, as well as between words strongly and weakly associated with motivation. Cross-decoding analyses further showed that local activation patterns in the left vATL, TP, and MTG encoded socialness independently of motivation, generalizing across levels of association with motivation. In summary, we provide the strongest evidence to date of a selective response of the dorsolateral ATL to social concepts. However, this response is observed alongside a category-general response of the left ventral ATL. These findings are most consistent with the graded semantic hub account of ATL function, and less with claims that the ATL is specialised for social processes. Moreover, we show for the first time that motivation-related conceptual features influence neural responses, suggesting that association with motivated cognition is an important dimension of conceptual knowledge.

References

[1] Pexman P.M. , Diveica, V. & Binney, R.J. (2023). Social semantics: the organization and grounding of abstract concepts. *Philos Trans R Soc Lond B Biol Sci*, 378(1870). https://doi.org/10.1098/rstb.2021.0363

[2] Diveica, V., Pexman, P. M., & Binney, R. J. (2023). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods*, *55*(2), 461–473. https://doi.org/10.3758/s13428-022-01810-x

[3] Giurgea, DI., Pexman, P.M. & Binney, R.J. (2025). Reward as a facet of word meaning: Ratings of motivation for 8,601 English words. *Behavavior Research Methods, 57*, 242 (2025). https://doi.org/10.3758/s13428-025-02762-8

Immediate Recall, Later Word Recognition, and Information Congruency in Reading and Listening Comprehension

Lucie Guštarová, Jan Chromý
Charles University
lucka.gustarova@gmail.com

Background: Recent studies [1,2] have documented systematic differences in the extent to what readers recall certain types of information immediately after reading a sentence. For example, information conveyed by direct objects tends to be recalled significantly better than information conveyed by temporal or locative adjuncts. This can be interpreted as a selective attention process [3]. The present study examines immediate recall of information conveyed by agents and locative adjuncts (Loc) in Czech and how it is influenced by adjunct congruency [4,5]. Moreover, in two of our experiments, later word recognition is tested [5].

Method: First, predictability of Loc in combination with 57 transitive verbs was normed (N=115 Czech speakers). 24 of these combinations were then used for creating stimuli for further experiments. Four reading experiments were conducted using a self-paced reading paradigm with whole sentences appearing at once for experiments 1 and 5 and with sentences presented word-by-word for experiments 2 and 6. Once the sentence disappeared, an open-ended question was shown targeting either the agent (Who did it?), or the Loc (Where did it happen?). Additionally, experiments 5 and 6 were followed by a later word recognition task including a set of 48 agents and 48 Loc, half of which appeared in the experimental sentences and half was absent from them. Then, two listening experiments were conducted. The first one used stimuli audio-recorded by native speakers of Czech with flat intonation. For the second experiment, stimuli generated by an artificial intelligence were used. The open-ended questions were visually presented, and participants responded by typing. All experiments use the same 24 experimental items and 72 fillers and manipulate word order, information targeted by the comprehension question and Loc congruency (see Table 1).

Result: Fig. 1 shows the differences in recall accuracy between the conditions in all experiments. The nested logit mixed-effects model showed a general recall difference for listening experiments (but not for the reading ones): agents were recalled better than Loc. Moreover, the model yielded a significant effect of congruency for Loc recall in all experiments and for agent recall in experiment 1. Effect of position for agents was found in experiments 2 and 6: better recognition was observed for medial position. Results of the later word recognition task are illustrated in Fig. 2. A congruency effect was observed only for words that did not appear in the reading task. A question effect was found for both words from experimental items (Loc recognized more accurately) and words absent from the stimuli (agents recognized more accurately). Overall, participants demonstrated better recognition of words absent from the experimental stimuli compared to words present in them.

Discussion: Overall, our experiments demonstrated a systematic advantage of congruent Loc in the immediate recall task. However, no congruency effect was found in the later word recognition task, which we attribute to a weaker representation of the sentence context in memory due to the temporal distance. Moreover, in line with prior findings on Czech reading data [1,2], experiments involving spoken materials showed a tendency for better immediate recall of core information in sentences compared to accessory information. However, this effect was not replicated in reading, which points to differences in presentation modality.

Word order	Congruency	Sentence
ltvso	yes	V obchodě v neděli koupila Klára hrozně hezký pruhovaný tričko.
ltvso	no	V parku v neděli koupila Klára hrozně hezký pruhovaný tričko.
stvlo	yes	Klára v neděli koupila v obchodě hrozně hezký pruhovaný tričko.
stvlo	no	Klára v neděli koupila v parku hrozně hezký pruhovaný tričko.

Table 1: Item example. Word order values: ltvso = locative adjunct - temporal adjunct - verb - subject - object; stvlo = subject - temporal adjunct - verb - locative adjunct - object. Sentences have the same meaning and only differ in their word order and locative adjunct congruency: "Klára bought a really nice striped T-shirt in the store/in the park on Sunday."

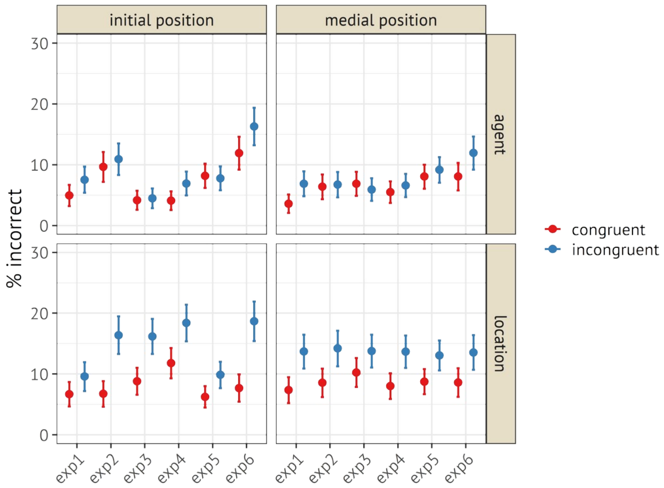


Figure 1: % of incorrect answers in all experiments. Congruent = congruent locative adjunct, incongruent = incongruent locative adjunct, exp1 = reading experiment, sentence presented at once, exp2 = reading experiment, sentence presented word-by-word, exp3 = listening experiment, stimuli recorded by real speakers, exp4 = listening experiment, stimuli generated by AI, exp5 = reading experiment, sentence presented at once, exp6 = reading experiment, sentence presented word-by-word.

Figure 2: % of incorrect answers in the later word recognition task in exp5 (reading experiment, sentence presented at once) and exp 6 (reading experiment, sentence presented word-by-word).

Keywords: sentence processing; comprehension; memory; recall; retrieval

References

[1] Chromý, Jan, and Sonja Vojvodić. 2024. "When and where did it happen? Systematic differences in recall of core and optional sentence information." *QJEP* 77 (1): 111–132.
[2] Chromý, Jan, and Fabian Tomaschek. 2024. "Learning or boredom? Task adaptation effects in sentence processing experiments." *Open Mind* 8*: 1447–68.*
[3] Ferreira, Fernanda, and Zoe Yang. 2019. "The problem of comprehension in psycholinguistics." *Discourse Processes* 56 (7): 485–495.
[4] Ryskin, Rachel, and Mante Nieuwland, S. 2023. "Prediction during language comprehension: what is next?" *Trends in Cognitive Sciences* 27 (11): 1032–1052.
[5] Haeuser, Katja I., and Jutta Kray. 2023. "Effects of prediction error on episodic memory retrieval: evidence from sentence reading and word recognition." *LCN* 38 (4): 558–574.

The Influence of Linearization on Expectation: Evidence from SPR and ERP Studies on Lossy Context Surprisal

Xinyue Jia, Christoph Aurnhammer, Torsten Kai Jachmann, Francesca Delogu, Heiner Drenhaus, Matthew W. Crocker
Language Science and Technology, Saarland University, Germany
xinyue.jia@uni-saarland.de

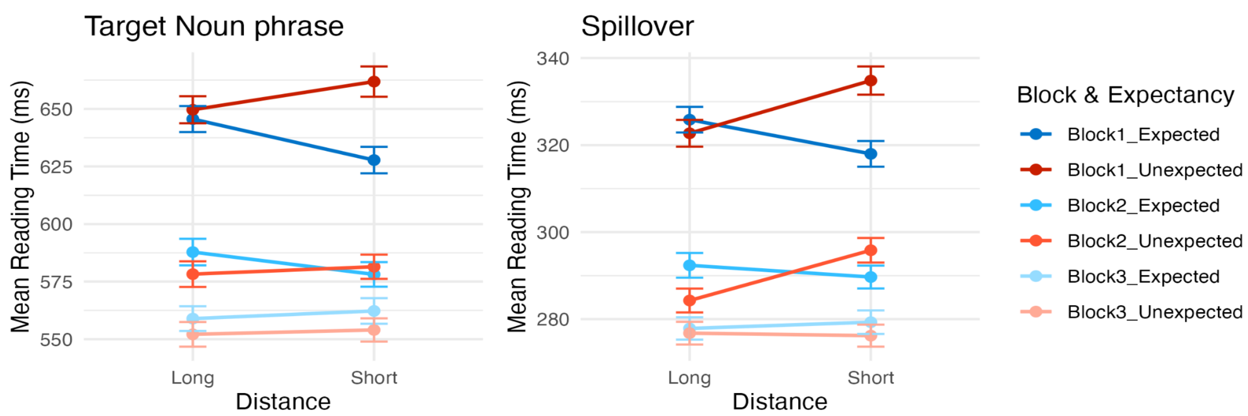
This study examines whether long-distance dependencies attenuate expectation effects as a result of lossy memory representations, leading to reduced facilitation for expected words and reduced difficulty for unexpected words when parts of the prior context are lost due to long distance.

A wide range of behavioral and neurophysiological findings support expectation-based models of sentence processing, including evidence that additional context can ease processing by lowering uncertainty about upcoming words (e.g., Hale, 2001, 2006; Levy, 2008). Research on long-distance dependencies, however, suggests that longer distance between related elements leads to increased processing difficulty, posing a challenge to purely expectation-based explanations (see Futrell et al., 2020, for discussion). Memory-based theories have explained this phenomenon as a consequence of working memory constraints (Lewis & Vasishth, 2005). More recent models, such as the Lossy-Context Surprisal framework, integrate memory components into an expectation-based approach by determining surprisal based on imperfect, lossy representations of prior context (Futrell et al., 2020; Hahn et al., 2022).

To investigate the interaction of expectation and memory, we will present a reading time study and an EEG study in German using a 2×2 design based on materials adapted from Aurnhammer et al. (2021). We created 120 items which varied the linear position of the intervening structure (*der die Nachrichten … war*) to manipulate the distance (Long vs. Short) between the main verb (*schärfte/webte*) and the object (*Axt*). The expectancy of the object (Expected vs. Unexpected) is manipulated by the main verb in the preceding context (*schärfte die Axt vs. webte die Axt*, see Table 1). Cloze pretests and surprisal value confirmed the differences in expectancy, but were not affected by the linearization. In addition to main effects of expectation and distance, the lossy-context surprisal predicts an interaction: If the increased distance of the B&D conditions results in a lossy memory representation of the predictive context – namely the main verb (*schärfte/webte*) – surprisal effects are predicted to be attenuated compared to the short distance conditions (A&C). That is, we predict an interaction of expectancy and distance, such that stronger surprisal effects occur in the short distance conditions, and weaker effects in the long distance conditions. In the analysis of the reading times using a linear mixed-effects model (N=80), we find the predicted interaction in target noun phrase and spillover regions in Block 1 (see Graph 1). We observed that these effects disappear in later blocks, likely due to participants' increased response speed. In an ongoing ERP study using the same stimuli, we predict that increasing distance will attenuate the expectation effects in both the N400 and P600 components, which have previously been found to be sensitive to semantic association and information integration respectively (Aurnhammer et al., 2021). The outcome will provide insights into online processing of expectation and working memory, and their interaction as indexed by relevant ERP components, directly assessing the predictions of the lossy surprisal account, and contributing broader evidence regarding how linearization choices influence online processing effort beyond what is captured by offline predictors such as cloze probability and surprisal.

Conditions	
A Expected Short	Gestern nach dem Hören der Nachrichten und des Wetterberichts schärfte der Holzfäller, der ungeduldig war, die **Axt** in seinem… *(Yesterday after listening to the news and the weather report sharpened the lumberjack, who impatient was, the* ***axe*** *in his…)*
B Expected Long	Gestern schärfte der Holzfäller, der die Nachrichten und den Wetterbericht gehört hatte und ungeduldig war, die **Axt** in seinem… *(Yesterday sharpened the lumberjack, who the news and the weather report listened had and impatient was, the* ***axe*** *in his…)*
C Unexpected Short	Gestern nach dem Hören der Nachrichten und des Wetterberichts webte der Holzfäller, der ungeduldig war, die **Axt** in seinem… *(Yesterday after listening to the news and the weather report weaved the lumberjack, who impatient was, the* ***axe*** *in his…)*
D Unexpected Long	Gestern webte der Holzfäller, der die Nachrichten und den Wetterbericht gehört hatte und ungeduldig war, die **Axt** in seinem… *(Yesterday weaved the lumberjack, who the news and the weather report listened had and impatient was, the* ***axe*** *in his…)*

Table1. Example of stimuli.



Graph1. Mean reading time of target noun phrase and spillover region

Selected References

Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., & Crocker, M. W. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PloS* one, 16(9), e0257430.
Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3), Article e12814.
Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences of the United States of America*, 119(43), e2122602119.
Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. *North American Chapter of the Association for Computational Linguistics*.
Hale, J. T. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, 10(9), 397–412.
Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3), 375–419.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.

The Emergence of Sociolinguistic Competence in Scottish Children: Social Registers Are Acquired Before Regional Dialects

Vera Kempe[1], Marta Brzoska[1], Hajar Benharraf[1] & Neil W. Kirk[1]

[1] Abertay University, Dundee, UK

v.kempe@abertay.ac.uk

Background: The development of sociolinguistic competence has primarily been studied with respect to regional (geographical) variation [1, 2]. For example, the ability to distinguish regional accents of US American English undergoes protracted development and reaches full adult competence only in adolescence [1]. However, many languages, e.g. Scottish English, comprise not just regional but also social register variation: Standard Scottish English (SSE) is associated with increased formality as well as with greater social prestige. Frequent switching between SSE and a regional variety renders most speakers of Scottish English bidialectal [3]. We examine whether sensitivity to such social register variation is acquired before, after or simultaneously with sensitivity to regional variation.

Method: In this pre-registered study, we compared adults and 5-16-year olds in their ability to distinguish varieties of Scottish English using a free sorting paradigm [4]. Experiment 1 examined sensitivity to regional variation by asking participants to group —by perceived similarity—recordings of identical sentences produced by speakers of Glaswegian, Doric, or Dundonian Scots, either with or without lexical change (Scots-Lexical vs Scots-Phonological). Lexical variation featured Scots words like *bairn* for *girl* or *wee* for *small*. Experiment 2 compared sensitivity to regional vs social register variation by asking participants to group Scots mixed with SSE sentences. To control for speaker-related indexical cues, the Scots-Lexical, Scots-Phonological and SSE varieties were produced by the same speakers.

Results: Data collection is still ongoing; the full results will be presented at AMLAP. At the time of submission, the pre-registered sample sizes had been tested with adults and 9-10-year old children. For Experiment 1, we computed Adjusted Rand Indices (ARI) as a measure for grouping accuracy based on regional variation. We found that while adults were able to group sentences by Scots regional dialects as indicated by above-chance ARI values for both the Scots-Lexical and the Scots-Phonological conditions, ARI values for 9-10-year olds did not differ from chance (Fig. 1). For Experiment 2, we computed three different types of ARI: *ARI-social* – reflecting grouping accuracy based on social register variation only (SSE vs. Scots), *ARI-regional* – reflecting grouping accuracy based on regional variation only (Glaswegian, Dundonian, Doric) and *ARI-social+regional* – reflecting grouping accuracy of regional varieties within social registers. For adults, *ARI-social* was significantly larger than *ARI-social+regional*, and both measures of grouping accuracy were above chance regardless of whether SSE was mixed with Scots-Lexical or with Scots-Phonological sentences. For 9-11-year olds, *ARI-social* differed significantly from chance both when SSE was mixed with Scots-Phonological and with Scots-Lexical sentences. Neither *ARI-socal+regiona*l nor *ARI-regional* differed from chance (Fig. 2).

Discussion: Our findings show that Scottish children acquire sensitivity to social register variation before sensitivity to regional variation, regardless of whether regional varieties feature lexical change or not. Our preliminary findings suggest that socio-linguistic competence involves the ability to associate linguistic cues with multiple sociolinguistic dimensions, and for children growing up in Scotland variation linked to formality and social prestige initially appears to be more salient than regional variation.

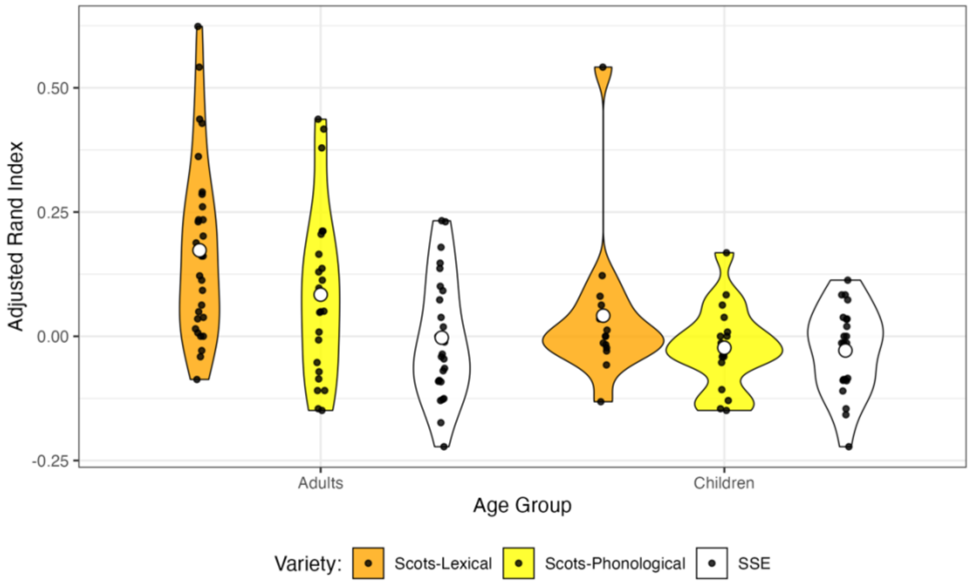


Figure 1: ARI indicating grouping by regional variation with and without lexical change in Experiment 1 in adults and 9-10-year old children. White dots indicate means.

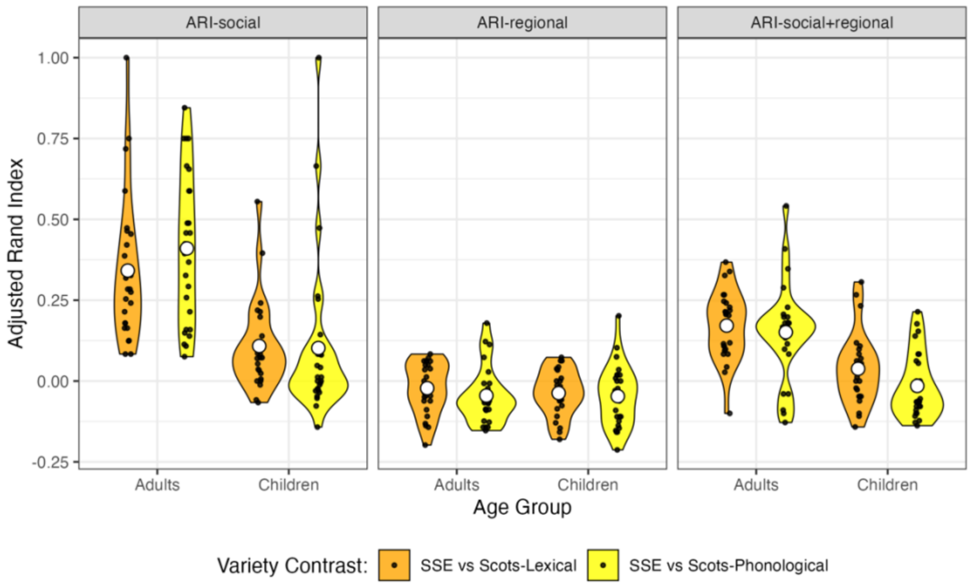


Figure 2: ARI reflecting grouping by social register variation, by regional variation and by both for SSE mixed with regional varieties with and without lexical change in adults and 9-10-year old children.

References

[1] Jones, Z., Yan, Q., Wagner, L., & Clopper, C. G. (2017). The development of dialect classification across the lifespan. *Journal of Phonetics*, *60*, 20-37.

[2] Jeffries, E. (2022). 5–9-year-olds categorisation of regional accents: The role of exposure to variation. *Lingua*, *267*, 103189.

[3] Kirk, N. W., Declerck, M., Kemp, R. J., & Kempe, V. (2022). Language control in regional dialect speakers–monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, *25*(3), 511-520.

[4] Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, *35*(3), 421–438.

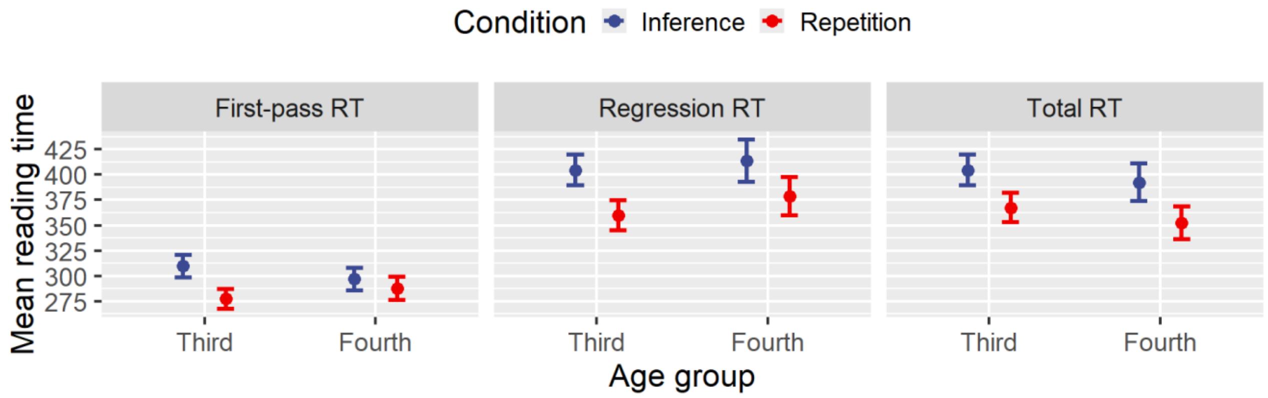
Bridging inference costs in late adulthood: Eye-Tracking evidence from third- and fourth-age readers

Background: As global life expectancy rises, understanding how older adults maintain reading comprehension becomes ever more critical. As in other better-studied populations, reading comprehension in older adulthood must hinge on the ability to integrate new information with preceding text, a process supported by bridging inferences. Although inference costs have been documented in young adults and adolescents [1, 2], it remains unclear whether—and how—these costs differ between third-age (60–79 years) and fourth-age (80+ years) readers, who face contrasting profiles of cognitive reserve and processing speed [3]. Examining online word-to-text integration with eye-tracking should help us to elucidate potential compensatory mechanisms that support functional literacy in very late adulthood. **Method:** One hundred fifty monolingual Spanish-speaking older adults (86 third-age, 64 fourth-age) were recruited to take part of the experiment. Participants read 80 two-sentence passages under within-subjects Repetition and Inference conditions (Table 1). Eye movements were recorded using an EyeLink 1000 system. After a 9-point calibration, participants read each passage for comprehension, occasionally answering true/false probes. We extracted first-pass reading time, regression-path duration, and total reading time on the critical word region. Fixations shorter than 80 ms were merged or excluded, and durations over 1200 ms were removed. Log-transformed reading times were analyzed via linear mixed-effects models with crossed random intercepts and slopes for participants and items [4]. **Results:** As it can be seen in Figure 1, only the third-age group exhibited a clear inference cost—slower reading in the Inference versus Repetition condition—across all reading measures (Table 2). Critically, a Condition × Group interaction in first-pass reading revealed that third-age readers showed immediate slowing when inference was required, whereas fourth-age readers did not. Instead, the fourth-age group manifested this cost primarily in total reading times (and to certain extent in regression path duration). Total reading time was longer overall for fourth-age adults, yet the proportional inference cost remained comparable across groups (see Table 1). **Discussion:** These findings demonstrate that bridging-inference processes persist into very late adulthood but with delayed temporal deployment in fourth-age readers. We propose that enriched semantic networks compensate for reduced processing speed, allowing older adults to achieve comprehension through extended integrative epochs. The results refine discourse-processing models by emphasizing age-related shifts in online integration dynamics and highlight practical implications: designing reading materials with embedded contextual cues or staggered presentation may scaffold inference-making in fourth-age populations. Future research should examine longitudinal trajectories, diversify participant profiles, and explore additional inference types to deepen our understanding of adaptive reading strategies in aging.

Table 1. Experimental materials example.

Condition	Example
Repetition Context	*"A medida que se acercaba al área rival, el delantero apuntó al arco y pateó el **balón**."*
Inference Context	*"A medida que se acercaba al área rival, el delantero apuntó al arco y pateó."*
Target	*"El **balón** fue golpeado tan fuerte que casi rompe la red."*
English Translation: *'As he approached the opponent's area, the striker took aim at the goal and kicked / the ball. The ball was hit so hard that it almost broke the net'.*	

Figure 1: Mean reading times as a function of experimental condition (Inference vs. Repetition), Age group (Third- vs. Fourth-age) and reading measure (First-pass reading times, Regression path-duration, Total



readings times). Error bars represent 95% confidence intervals adjusted for within-subject design.

Table 2. Experimental Materials Example

Measure	Effect	*β*	*se*	*t*	*p*	
First Pass Reading Time	(Intercept)	5,5	0,0	156,2	0,001	***
	Group	0,1	0,1	1,9	0,060	#
	Condition	0,0	0,0	-2,3	0,025	*
	G × C	0,0	0,0	-2,0	0,048	*
Regression Path Duration	(Intercept)	5,8	0,0	155,3	0,001	***
	Group	0,1	0,1	1,0	0,307	
	Condition	0,0	0,0	-3,4	0,001	**
	G × C	0,0	0,0	-1,7	0,087	#
Total Reading Time	(Intercept)	5,7	0,0	142,1	0,001	***
	Group	0,2	0,1	2,6	0,013	*
	Condition	0,0	0,0	-3,0	0,004	**
	G × C	0,0	0,0	-0,8	0,417	

***=p<.001; **=p<.01; *=p<.05; #=p<1

References

[1] Guerra, E., & Kronmüller, E. (2024). Are Adolescents with a Wider Vocabulary Faster at Inference Making During Reading? Evidence from Self-Paced Reading. *Education Sciences*, 14(12), 1368.

[2] Guerra, E., & Kronmüller, E. (2020). Adult vocabulary modulates speed of word integration into preceding text across sentence boundaries: Evidence from self-paced reading. *Reading Research Quarterly*, 55(4), 663-677.

[3] Higgs, P., & Gilleard, C. (2017). The nature of the fourth age as a challenge to ageing societies. In M. Schweda, L. Pfaller, K. Brauer, F. Adloff & S. Schictanz (Eds.), *Planning Later Life. Bioethics and Public Health in Ageing Societies*. New York: Taylor & Francis Group Press.

[4] Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

This page may be used only for additional information about the generally less-known language you are targeting in the abstract.

This study was conducted with a Monolingual Spanish population, and all instructions and verbal sentences were presented in Spanish.

Good explanations fit prior knowledge

Joshua Hartshorne[1], Tobias Gerstenberg[1], Noah Goodman[2]
[1] Communication Sciences & Disorders, MGH Institute of Health Professions; [2] Department of Psychology, Stanford University
jkhartshorne@fas.harvard.edu

E because C is ambiguous, indicating either a causal explanation (*C caused E*) or an epistemic explanation (*C caused the speaker to believe that E*). Nonetheless, people seems to have more difficulty noticing the ambiguity than resolving it. For instance, the causal meaning m_{causal} is strong preferred in (1), and the epistemic meaning $m_{epistemic}$ in (2):

1. The grass is wet because it rained last night.
2. It rained last night because the grass is wet.

Plausibility provides an intuitive explanation for (2) — wet grass does not cause overnight rain — but less so for (1) — rain is strong evidence for wet grass. We test a Bayesian account of comprehension [1, 2], where the probability that utterance u conveys message m is given by:

3. P(m|u) $\propto$ P(u|m) * P(m)

Assuming a cooperative speaker — and ignoring, for the moment, Gricean reasoning — the prior $P(m)$ largely reduces to the prior probability that the message is true. As already noted, $P(m_{epistemic}) > P(m_{causal})$ is sufficient to explain (2). (1) could be explained if $P(u|m_{causal}) > P(u|m_{epistemic})$ for both sentences.

To test this account quantitatively, we created 32 event pairs (e.g., *the grass is wet* and *it rained*). For each pair, 71 participants rated both prior probabilities ($P(m_{causal})$, $P(m_{epistemic})$). From these pairs, we created 32 sentences. An additional 72 participants judged the relative probability of the causal and epistemic interpretation for each. All probability ratings were on a scale of 0-100. We assumed that the likelihoods $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were the same for all sentences (i.e., they did not depend on the specific events) and fit them to the data. The resulting model fit the data extremely well (Fig. 1, left; $r = .84$, 95% CI $[.69, .92]$, $t(30) = 8.34$, $p < .001$). The best-fitting relative probabilities of $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were 85% and 15%, respectively, matching the expectation that the former would be substantially higher than the latter.

We also created a second set of sentences by adding *I think* to the beginning of each:

4. I think the grass is wet because it rained last night.
5. I think it rained last night because the grass is wet.

Note that while the causal reading is still possible — (4) might indicate that the speaker believes it is the case that the rain last night caused the grass to be wet — the epistemic readings now seem more likely in both cases. This is likely because for these sentences, $P(u|m_{causal}) < P(u|m_{epistemic})$. Nonetheless the priors $P(m)$ are still the same, which would explain why the causal reading still seems more likely for (4) than (5). We obtained judgments for these new sentences from the same 72 participants. As expected, epistemic interpretations were much more likely (Fig. 1, right). We again fit likelihoods to the data. Overall, the model fits were again quite good ($r = .78$, 95% CI $[.59, .89]$, $t(30) = 6.73$, $p < .001$). The best-fitting relative probabilities of $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were 13% and 87%, respectively, again matching expectations.

We discuss how adding Gricean reasoning to this model [e.g., using RSA; 3] would affect results. We also consider broader implications for theories of language comprehension.

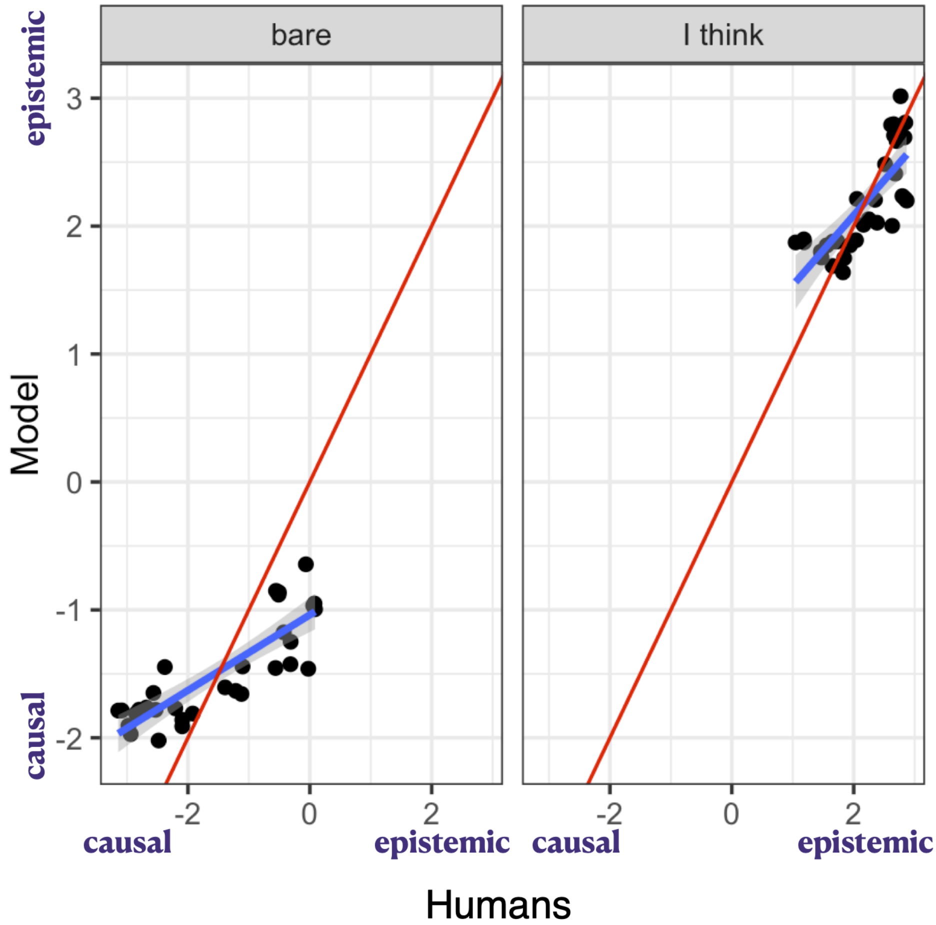


Figure 1: log-odds of epistemic vs. causal interpretation, with positive numbers indicating greater probability of the causal interpretation and negative numbers indicating greater probability of the epistemic interpretation. Each point represents a sentence, with bare sentences like (1) and (2) on the left and ‘I think‘ sentences like (4) and (5) on the right. Separate linear fits are shown, with the 95% confidence interval indicated.

References

[1] Narayanan, S., & Jurafsky, D. (2022). Bayesian models of human sentence processing. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 752–757.

[2] Hartshorne, J. K., Jennings, M. V., Gerstenberg, T., & Tenenbaum, J. (2019). When circumstances change, update your pronouns. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*.

[3] Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Comparing Natural Language Statistical Learning and Human Intuition for Chunking Language

Cristian Rivera[1], Morten H. Christiansen[1, 2]

[1] Department of Psychology, Cornell University; [2] School of Communication and Culture, Aarhus University

cr475@cornell.edu

Much like with statistical learning, there is a tendency in the cognitive and language science literature to treat chunking as a unitary, domain-general ability that operates fairly similarly across domains of cognition. An alternative position suggests chunking in language and cognition involves multiple mechanisms fine-tuned to the specific perceptual properties and statistical regularities of the target structures of learning. To test these hypotheses, we compared performance on two different forms of chunking: 1) multiword chunking as assessed by a recall task involving multiple high and low frequency three-word sequences, and 2) clausal chunking in which participants intuitively insert chunk boundaries on a touchscreen tablet while listening to speech extracts and viewing the corresponding transcripts. Our final analyses consisted of 69 native English-speaking participants. Results indicated no predictive relationship between multiword chunking and clausal chunking, providing initial support for the notion that chunking processes may likely not operate equally across or even within the same cognitive domain casting doubt on a unitary view of chunking. However, we also note that the intuitive chunking (IC) task showed minimal inter-individual variability in performance, which limits its usefulness as a measure of individual difference in clausal chunking. This may in part be due to the reflection-based nature of the task given that participants view transcripts of speech stimuli allowing for responses based on top-down processing. Nonetheless, the two tasks may provide a starting point for future studies looking into the nature of chunking across different levels of linguistic representation.

L2 acquisition of verb and noun paradigms: a study on Russian

Maria Grabovskaya, Anastasia Vyrenkova, Natalia Slioussar (HSE University, Moscow)
magrabovskaya@gmail.com, anastasia.marushkina@gmail.com, slioussar@gmail.com

The acquisition of the Russian verb system has been primarily studied with a focus on telicity and aspect, both for L1 children (e.g. Stoll 1998; Gagarina 2004; Cejtlin 2009; Akhapkina 2013; Voeikova 2015) and for L2 adult learners (e.g. Slabakova 2004; Nossalik 2009; Mikhaylova 2012; Whitehead Martelle & Shirai 2023; Apresjan 2024). However, especially in case of L2, we know much less about the distribution of different verb forms and the acquisition of verb categories across proficiency levels.

In this paper, we aim to fill this gap and compare the patterns observed for verbs to the acquisition of noun paradigms (primarily case) analyzed by Cherepovskaia et al. (2022). Cherepovskaia et al. (2022) present a corpus of short narratives by Catalan/Spanish learners (levels A1 – C1) from a language school in Barcelona[1]. This corpus contains 196 texts describing a comic strip. In Cherepovskaia et al. (2022), all noun forms in this dataset were analyzed, while we analyzed all verb forms, classifying them into grammatically correct and incorrect and marking different grammatical features. As a result, we could identify the following tendencies, speculating how L1/L2 affect them and comparing verbs and nouns.

On the overall distribution of errors (across levels, in verbs vs. in nouns)

i. Both in verbs and in nouns, the number of errors peaks at the B1 level (see Table 1). However, the number of correct forms grows from level to level, so the share of errors steadily declines. This shows how important it is to take correct forms into account, while many studies of L2 production focus only on errors.

ii. The percentages of errors in noun and verb forms at different levels are very similar (see Table 1), despite a different nature of these errors discussed below and different overall numbers of verb and noun forms.

On the 'last resort strategy' found for nouns (using nominative forms), but not for verbs

iii. According to Cherepovskaia et al. (2022), in noun paradigms, the nominative form clearly has a special status — at earlier levels, participants overuse it a lot (more than 70% of all forms are nominative at the A1 level) and resort to it whenever they do not know which case to use (see Table 1). We did not observe a similar strategy for verbs — no form has a similar special status. Since there is no simple last resort strategy, the pool of verb forms is more balanced and diverse even at early levels (see Table 2).

iv. Moreover, trying to produce different verb forms, participants sometimes generate non-existing affixes etc. Cherepovskaia et al. (2022) found no such errors for nouns: not a single non-existing form etc. — presumably, because whenever participants had doubts, they could use the nominative form.

v. Initially, noun forms are almost 2.5 times more numerous than verb forms, but this ratio changes gradually to reach 1.2 at the C1 level. Not having a last resort option, participants initially try to avoid verbs.

vi. The generalization in (iii) may be affected by the L1 of the learners. A preliminary analysis of similar narratives produced by Chinese speakers revealed many infinitive forms participants used in case of doubts. These cross-linguistic differences may be connected to the fact that Catalan and Spanish are rich in verb morphology, while Chinese is not. As for noun morphology, it is poor in all these languages.

On the distribution of different verb forms across levels

vii. On the A1 level, participants mostly used present tense forms, and then the share of past tense and infinitive

[1] The school uses the CEFR system (A1 = beginner; A2 = pre-intermediate; B1 = intermediate; B2 = upper intermediate; C1 = advanced; C2 = proficiency), adding the B1+ level to it. There were no C2 respondents.

forms grew from level to level (see Table 2). Participants get acquainted with all these forms already at the A1 level, but initially focus on present tense forms more in their textbooks.

viii. At the same time, the share of errors in past tense and infinitive forms steadily declined, while in present tense forms, it remained relatively high (see Table 2). Moreover, errors in the past tense are less diverse. This can be explained by a higher morphological complexity of present tense forms.

ix. Staring a more detailed analysis, we looked at imperfective and perfective aspect. The latter is much more difficult for L2 learners (see Table 3). Unlike native speakers, they use it less (at A1, it is 38 times less frequent than imperfective, at C1, still 1.5 times less frequent), and make much more errors with it.

Thus, we found some interesting differences and parallels between verbs and nouns. Further research is needed to find out which of them are universal, which depend on the L2 we focus on or on participants' L1.

Table 1. The number of verb and noun forms and errors on different levels.

Level		**A1**	**A2**	**B1**	**B1+**	**B2**	**C1**
Verbs	all forms	397	668	792	906	996	1093
	errors	74	94	126	97	93	84
	% errors	16%	12%	14%	10%	9%	7%
Nouns	all forms	970	1032	1164	1164	1189	1283
(from [1])	errors	147	154	166	142	115	98
	% errors	15%	15%	14%	12%	10%	8%
	% nom out of errors	77%	36%	28%	35%	28%	13%

Table 2. The distribution of verb forms at different levels and the percentage of errors in these forms.[2]

	N of forms						**% of verb forms at this level**						**% errors in these forms**					
Level	**A1**	**A2**	**B1**	**B1+**	**B2**	**C1**	**A1**	**A2**	**B1**	**B1+**	**B2**	**C1**	**A1**	**A2**	**B1**	**B1+**	**B2**	**C1**
pres/fut	387	373	203	202	151	164	83%	49%	22%	20%	14%	14%	14%	12%	20%	12%	11%	12%
past	51	285	565	633	729	746	11%	37%	62%	63%	67%	63%	16%	11%	12%	8%	8%	7%
inf	29	104	149	161	196	225	6%	14%	16%	16%	18%	19%	31%	16%	13%	12%	8%	5%

Table 3. The distribution of verb forms in the perfective and imperfective aspect at different levels.

Level		**A1**	**A2**	**B1**	**B1+**	**B2**	**C1**
correct	imperf	452	545	529	570	538	630
	perf	12	126	292	340	450	427
errors	imperf	3	48	27	27	36	45
	perf	3	42	71	66	65	75
% errors	imperf	1%	9%	5%	5%	7%	7%
	perf	25%	33%	24%	19%	14%	18%

References: Akhapkina Y. (2013). Tense-aspect abstraction in Russian-speaking children in the early stages of native language acquisition. *NRU HSE Series WP BRP "Linguistics"*, art. 01/LNG/2013. ♦ Apresjan V. (2024). Errors in foreign language acquisition as a multifaceted phenomenon: the case of Russian aspect. *Russian Linguistics*, *48*(1), art. 3. https://doi.org/10.1007/s11185-023-09287-8. ♦ Cejtlin S.N. (2009). *Ocherki po slovoobrazovaniju i formoobrazovaniju v detskoj rechi* [Essays on word formation and form formation in

[2] Imperatives, gerunds and participles are not included.

children's speech]. Moscow, Znak. ♦ Cherepovskaia, N., Slioussar, N., & Denissenko, A. (2022). Acquisition of the nominal case system in Russian as a second language. *Second Language Research*, 38, 555-580. ♦ Gagarina N.V. (2004). Does the acquisition of aspect have anything to do with aspectual pairs? *ZAS Papers in Linguistics*, *33*, 39-61. ♦ Mikhaylova A. (2012). *(In)complete acquisition of aspect in second language and heritage Russian*. Ph.D. Dissertation, University of South Carolina. ♦ Nossalik L. (2009). *L2 acquisition of Russian aspect*. Ph.D. Dissertation, McGill University. ♦ Slabakova R. (2004). Effect of perfective prefixes on object interpretation: A theoretical and empirical issue. *Cahiers linguistiques d'Ottawa*, *32*, 122-142. ♦ Stoll S. (1998). The role of Aktionsart in the acquisition of Russian aspect. *First language*, *18*(54), 351-376. ♦ Voeikova M. (2015). Var'irovanie form SV/NSV v imperative [Variation of perfective/imperfective forms in imperative]. *Acta Linguistica Petropolitana, 11*(1), 539-563. ♦ Whitehead Martelle W. & Shirai Y. (2023). The aspect hypothesis and L2 Russian. *Frontiers in Language Sciences*, *2,* art. 1119026*.*

Appendix: A brief overview of Russian verb and noun morphology

Russian verbs can be imperfective and perfective. Most perfective verbs are derived from imperfective ones by prefixation and sometimes by suffixation: e.g. *krasit'* 'to paint' — *po-krasit'* 'to finish painting', *za-krasit'* 'to paint over', *pere-krasit'* 'to repaint', and so on. Their meaning is only partly predictable from the meaning of the base verb and the prefix, which creates problems for L2 learners (but not for L1 learners). Some imperfective verbs are also derived from other verbs, like secondary imperfectives derived from perfective verbs, but many most frequent ones are not.

Imperfective verbs have two sets of synthetic forms (past and present) and analytic future forms (the verb 'to be' in the future tense + infinitive). Perfective verbs have only two sets of synthetic forms (past and future). All verbs have two stems: infinitive and past tense (including gerund and participles) are derived from one of them, imperative and present/future tense (including gerund and participles) from the other. In the present and future tense, verbs are inflected for person and number (see Table A). There are two inflectional classes with different sets of inflections, some forms have alternations in the stem. In the past tense, verbs are inflected for number and gender (in singular), these forms are former participles. All verbs have the same inflections, and the main problem for L2 learners is to determine the gender of the subject and to choose the matching verb form.

Table A. The forms of the verbs *delat'* 'to do' and *krasit'* 'to paint'. Affixes and thematic vowels are separated from each other and from the root by hyphens.

PRES 1SG	*del-a-ju*	*kraš-u*	INF	*del-a-t'*	*kras-i-t'*
PRES 2SG	*del-a-eš'*	*kras-iš'*	PAST M.SG	*del-a-l*	*kras-i-l*
PRES 3SG	*del-a-et*	*kras-it*	PAST F.SG	*del-a-l-a*	*kras-i-l-a*
PRES 1PL	*del-a-em*	*kras-im*	PAST N.SG	*del-a-l-o*	*kras-i-l-o*
PRES 2Pl	*del-a-ete*	*kras-ite*	PAST PL	*del-a-l-i*	*kras-i-l-i*
PRES 3PL	*del-a-jut*	*kras-jat*			

Russian nouns have three genders (M, F, N) and are inflected for number (SG, PL) and case (NOM, GEN, DAT, ACC, INSTR, LOC). They have several inflectional classes with different sets of inflections (which correlate with gender, but not fully). Thus, the main problem is to select the right case and the right inflection for it (the latter may be difficult not only in L2, but also in L1 acquisition).

How do you spell “hånd” in English: Does knowing another language affect word retrieval?

Helene Slaattelid Øya[1], Jens Roeser[1], Gary Jones[1], Mark Torrance[1]
[1] Department of Psychology, Nottingham Trent University
helene.slaattelidoya@ntu.ac.uk

Background: When bilinguals speak and write in one language, the other language is activated, even if the context only requires the activation of one [1]. It is therefore that, for example, cognates (words that share form and meaning; Norwegian *hånd* meaning “hand” in English) are known to be easier to retrieve, in terms of meaning, phonology and orthography [2, 3]. Like cognates, false friends are similar in word form, but importantly differ in meaning (Norwegian *gift* means “poison” in English). We tested to what extent co-activation of form-related words affects lexical retrieval in bilinguals. If language co-activation occurs on a meaning level, false friends are predicted to be more difficult to retrieve. However, co-activation on a form level (orthography and phonology) is predicted to facilitate word retrieval.

Method: In three experiments, Norwegian (L1) - English (L2) bilinguals (*N*s = 66, 60, 86) named images of everyday objects by typing their name in English while ignoring a superimposed (Exp. 1) or auditory (Exps. 2 and 3) distractor. Word type of image names and their distractor were either cognates, false friends, or a control (name of image or translation equivalent); see Fig. 1. Distractors were presented in Norwegian (all experiments) or English (Exps. 1 and 2). Distractor and image were presented simultanously across all experiments; to dissociate the effect of distractor and image, Exp. 3 also presented the distractor before image (SOA of -300 ms) or after (SOA of 300 ms). Every participant named 26 images per word type (*N* = 78). Images were counterbalanced across 2 (for language in Exps. 1 and 2) and 3 lists (for SOA in Exp. 3) and presented in randomised order. Prior to the experiment, participants were auditory familiarised with the English picture names before the experiments. The efficiency of the familiarisation task was piloted (*N* = 32).

Results: Response time (RT; time from picture onset to first key press) was log transformed and analysed in mixed-effects models. Likelihood ratio tests revealed increased model fit when including interactions in additional to simple main effects of word type and language in Exp. 1 (χ^2 (2) = 71.5, p < .05) and Exp 2 (χ^2 (2) = 26.1, p < .05), and word type and SOA in Exp. 3 (χ^2 (4) = 22.9, p < .05); see Fig. 2. Pairwise comparisons showed shorter RTs for Norwegian cognates (all *p*s < .05) and false friends (all *p*s < .05) compared to the control condition for both printed (Exp. 1) and auditory distractors (Exps. 2 and 3). Faster RTs were observed for cognates compared to false friends (all *p*s < .05). No differences were observed between English distractors (Exps. 1 and 2). When the distractor was presented before image onset (SOA = -300 ms), we found shorter RTs for cognates (compared to false friends and control; both *p*s < .05) but no difference between false friends and control. When distractors were delayed by 300 ms, we found shorter RTs for cognates and false friends compared to control (both *p*s < .05) but no difference between cognates and false friends.

Discussion: Our results strongly suggest that word-form co-activation across languages facilitates writing in a non-dominant language, even when the co-activated L1 word differs in meaning from the target word.Co-activation on a meaning level only occurred at -300 SOA (i.e., forcing meaning to be activated prior to form, facilitating cognates but not false friends), suggesting that bilingual word retrieval primarily involves word-form with only a secondary role for word meaning.

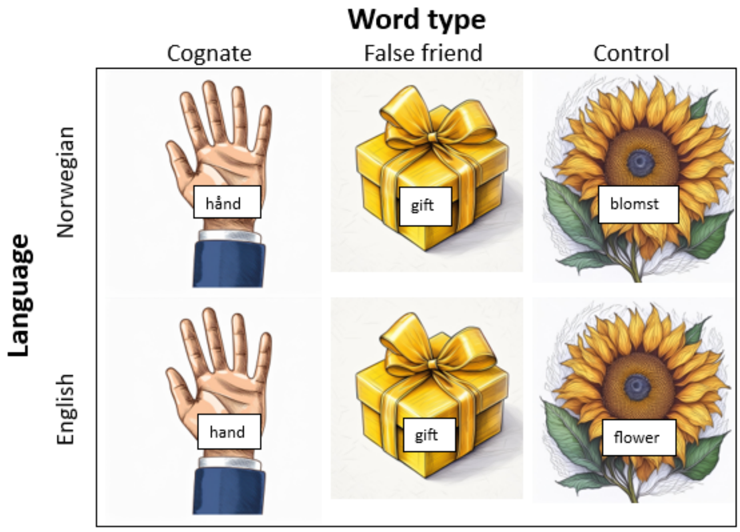


Figure 1: Example stimuli.

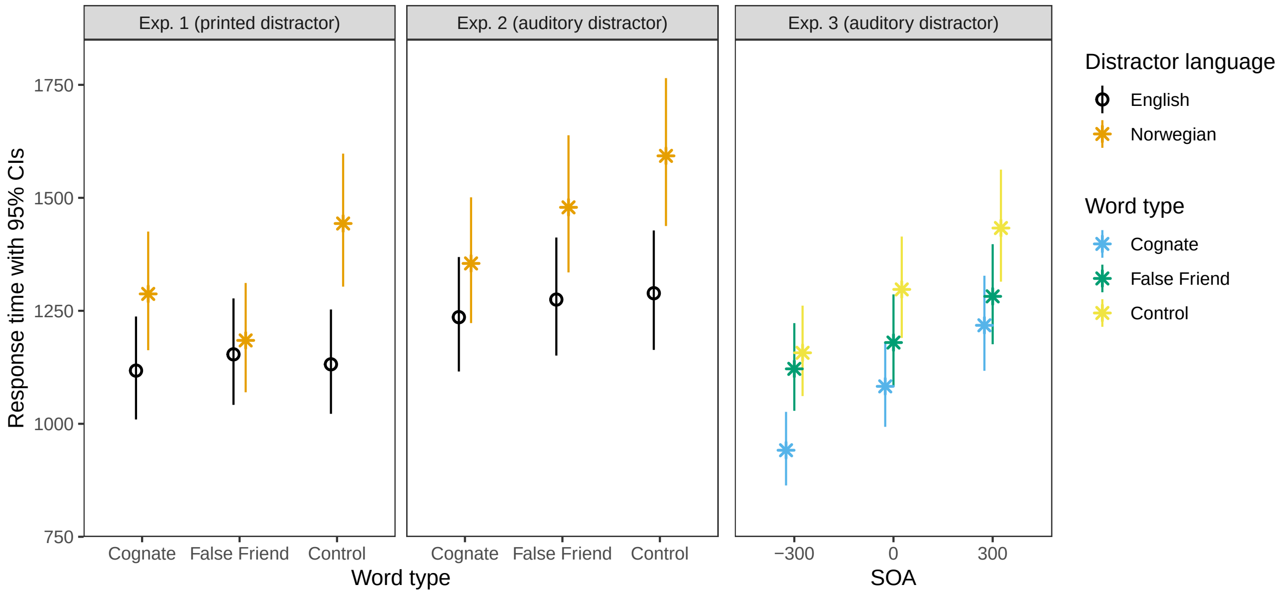


Figure 2: Estimated marginal means with 95% confidence intervals (CI).

References

[1] Kroll, J. F., Gullifer, J. W., McClain, R., Rossi, E., & Martín, M. C. (2015). Selection and control in bilingual comprehension and production. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 485–507). Cambridge University Press. https://doi.org/10.1017/CBO9781107447257.021

[2] Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

[3] Poort, E. D., Warren, J. E., & Rodd, J. M. (2016). Recent experience with cognates and interlingual homographs in one language affects subsequent processing in another language. *Bilingualism: Language and Cognition*, *19*(1), 206–212. https://doi.org/10.1017/S1366728915000395

Priming cooperating prosodic phrasing increases reading times: An eye-tracking study

Dorotea Bevivino[1], Barbara Hemforth[1], Giuseppina Turco[1]
Université Paris Cité, LLF, CNRS
dorotea.bevivino@etu.u-paris.fr

Background: Prosody has been argued to contribute to sentence processing even in silent reading [1]. Evidence for implicit prosody comes from studies indirectly inducing prosodic breaks [e.g. 2] as well as directly manipulating prosody [3, 4]. However, even these few latter studies have only assessed prosodic effects on offline responses, and only manipulated boundaries in spite of research on the role of full phrasing and pace modulation [5].

Current study: Here we report a cross-modal priming eye-tracking study in French to further investigate how prosodic phrasing information is processed and affects online sentence processing while reading. Similar to offline and online studies on overt prosody and syntactic-prosodic alignment [6, 7, 8], we manipulated the (mis-)alignment between the overt prosodic phrasing of delexicalized audio primes and the implicit prosodic phrasing of target sentences with highly regular and predictable prosodic phrasing, as typical of French [9]. This allowed us to (i) explore the real-time processing of prosodic information activated at silent reading, and (ii) look at the effects of full phrasing beyond the simple boundary (iii) and beyond syntactic effects. We hypothesized that aligned cooperating prosodic phrasing would pose a facilitating priming effect and reduce reading times, whereas misaligned conflicting prosodic phrasing would induce a penalty and increase reading times.

Method: 40 French L1 speakers were tested in a cross-modal priming paradigm during an eye-tracking-while-reading task. Primes were delexicalized audios (*mama* speech, similar to [4]) aligned (COOPERATING) or not (CONFLICTING) with the syntactic structure and expected implicit prosody of target sentences. **Experiment 1** targets (Table 1) were 24 temporarily ambiguous homograph constructions, similar to English ambiguities like *The old trains the young*, with two mutually exclusive interpretations (VERB- or NOUN- READING) and mutually exclusive prosodic phrasings. **Experiment 2** targets (Table 2, used as fillers for Exp. 1) were 20 unambiguous sentences, manipulated to be syntactically equivalent but spontaneously prosodically different (SHORT or LONG).

Results: As for **Experiment 1** (Fig. 1), we found an expected garden-path penalty for V-READING sentences at the post-disambiguating region (P(β<0)=0.98). Moreover, our results revealed an effect of PRIME PROSODY but not in the predicted direction: COOPERATING prosody increased reading times in the ambiguous region (P(β>0)=0.98). Similarly, results of **Experiment 2** (Fig. 2) showed that COOPERATING prosody would tend to increase reading times in the critical region of unambiguous sentences (P(β>0)=0.95).

Discussion: Overall, our findings revealed a modulating priming effect of cooperating prosody, though in the opposite direction than predicted. Readers appear to encode the prosodic phrasing information of delexicalized primes and use it to modulate processing of targets, just not in the way we thought they would. In line with research on overt prosody, this pattern may suggest that alignment of primed prosodic phrasing, syntactic boundaries, and implicit prosody favor greater engagement and deeper processing, increasing reading times. Alternatively, it might reflect effects of factors such as meter regularity or cognitive load. More research is needed to test these speculations. Despite the unanticipated direction, our findings suggest that primed phrasing—beyond the simple boundary and beyond syntactic effects—would modulate the real-time processing of sentences during silent reading, thus extending prior work on implicit prosody.

Table 1: Stimuli and Design Homograph Ambiguities

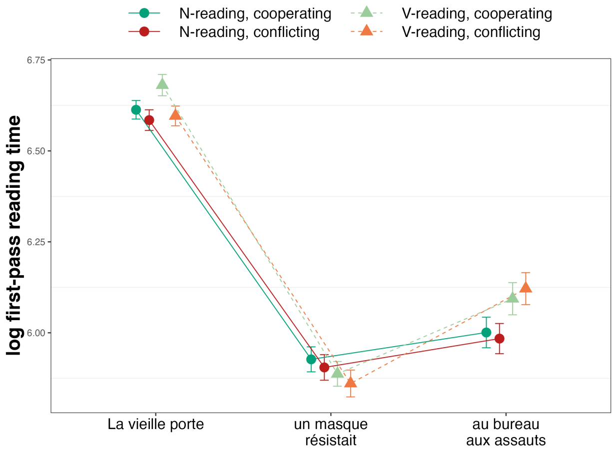
Target structure	Prime prosody	Target sentence
V-reading	Cooperating	*La vieille % porte un masque % au bureau* ‘*The old lady wears a mask at the office*’
	Conflicting	* *La vieille porte % un masque au % bureau*
N-reading	Cooperating	*La vieille porte % résistait % aux assauts* ‘*The old door resisted the assaults*’
	Conflicting	* *La vieille % porte rési- % -stait aux as- % -sauts*

Note. The % symbol indicates the end of Accentual Phrases, French minimal prosodic units. The * indicates implausible prosodic phrasing realizations.

Table 2: Stimuli and Design Unambiguous Sentences

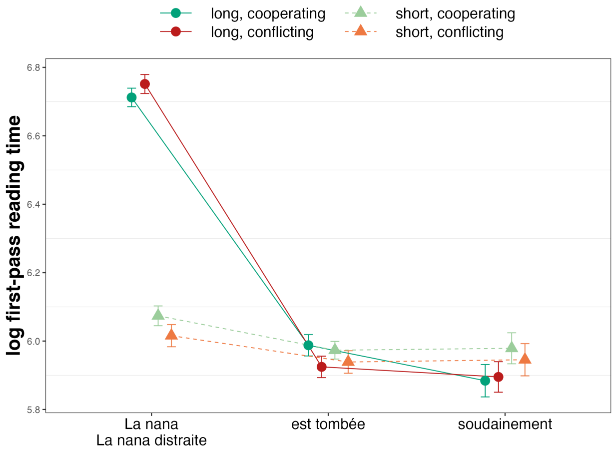
Target structure	Prime prosody	Target sentence
Short	Cooperating	*La nana % est tombée % soudainement* ‘The girl suddenly fell’
	Conflicting	* *La nana est tom- % -bée soudaine- % -ment*
Long	Cooperating	*La nana distraite % est tombée % soudainement* ‘The distracted girl suddenly fell’
	Conflicting	* *La nana % distraite est % tombée sou- % -dainement* (only plausible realization, in terms of syllables) *La nana distraite % est tombée % soudainement* (also plausible, in terms of intonational contour)

Figure 1: Reading Times Homograph Ambiguities



Note. $N = 40$ participants and $N = [884 - 888]$ trials, with a mean of 221 observations per plotted condition for each interest area.

Figure 2: Reading Times Unambiguous Sentences



Note. $N = 40$ participants and $N = [742 - 743]$ trials, with a mean of 186 observations per plotted condition for each interest area.

Selected References

[1] Fodor, J. D. (2002). Prosodic disambiguation in silent reading. *Proceedings of NELS 32*, 113–132.

[2] Hemforth, B., Colonna, S., Petrone, C., & D’Imperio, M. (2013). Length matters: Informational load in ambiguity resolution. *Discours*, *12*.

[3] Jun, S.-A., & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming (L. Frazier & E. Gibson, Eds.), 217–240.

[4] Mills, J. (2020). Delexicalised auditory priming of implicit prosody. *Proc. Speech Prosody 2020*, 295–298.

[5] Grillo, N., Santi, A., & Turco, G. (in press). Shaping rhythm to keep balance: The structural implications of temporal modulation. In L. Meyer & A. Strauss (Eds.), *Rhythms of speech and language*.

[6] Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

[7] Harris, J. A., & Jun, S.-A. (2019). Using pupillometry to assess prosodic alignment in language comprehension. *Proceeding of the 19th International Congress of Phonetic Sciences*.

[8] Glushko, A., Poeppel, D., & Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*, *12*.

[9] Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1).

Language and Script Effects on Information Credibility in a Triliteral Context

Zhimin Hu[1], Eduardo Navarrete[1], Yao Yao[2]

[1] Department of Developmental Psychology and Socialisation, University of Padua; [2] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

zhimin.hu@unipd.it

Background: Previous studies suggest that language and script variations may influence credibility judgments [1, 2]. This pre-registered study (https://aspredicted.org/kcdn-hv2j.pdf) investigates the impact of language and script on the credibility judgments of trivia statements from Chinese-L1 English-L2 bilinguals who are originally from Mainland China (i.e., native readers of Simplified Chinese script) and now residing in Hong Kong, a bilingual and triliteral society where English and Traditional Chinese script are the *de facto* writing standards. While the fluency-based account predicts that statements in Simplified Chinese would be judged as more credible [3], the status-based account would predict higher credibility judgments for statements in English and Traditional Chinese [4].

Method: To ensure contextual consistency, participants were randomly assigned to one of three between-subject conditions: English (Latin alphabet), Simplified Chinese (Simplified characters), or Traditional Chinese (Traditional characters). Each participant read 120 generally unknown fact-checked and culturally-adapted trivia statements (e.g., *Rabbits clap their paws to signal danger;* half objectively true, half objectively false) [2] in randomized order and rated the credibility of each statement on a slider scale from 0 (definitely false) to 100 (definitely true) and their retrospective confidence. After the judgment task, participants rated each of the three scripts in terms of status (e.g., authoritativeness, social status) and their experience with the script (e.g., exposure, ease of processing). Participant exclusion criteria were pre-defined based on attention check items and outlier detection. A target of 40 participants per language condition (total=120) was set [5]. Credibility judgments were analyzed by linear mixed-effects regression models, with Language and Factuality as fixed effects and Statement and Participant as random effects.

Results: Preliminary analyses were conducted with data from 135 participants. Objectively true statements were judged more as more credible than objectively false statements (b = 2.87, 95% CI [1.31, 4.43], $p < .001$). Interestingly, of the three scripts, while English was rated as the most authoritative (ps < .001) with the highest social status (ps < .001), and Simplified Chinese was the highest in exposure (ps < .001) and ease of processing (ps < .001), Simplified Chinese elicited faster (b = 1.04, 95% CI [0.79, 1.29], $p < .001$) and more credible judgments (b = -4.52, 95% CI [-9.02, -0.03], $p = .049$) than English (see Figures). No significant difference in credibility judgment was observed between Simplified and Traditional Chinese, though response times were slightly longer in Traditional Chinese (b = 0.24, 95% CI [0.02, 0.47], $p = .033$).

Discussion: Overall, we observed that participants were faster and more likely to believe statements presented in Simplified Chinese script (vs English or Traditional Chinese). These preliminary results are compatible with the fluency-based account, possibly via ease of semantic processing, but not with the status-based referential-theory account. We also observed nuanced differences between English and traditional Chinese (i.e., the two less familiar scripts). Further analyses are currently underway. In sum, this pre-registered study provides important theoretical and practical insights into how language and script variations impact the judgment of truth in a multilingual world, contributing to our understanding of the psychology of multilingualism, fake news, and disinformation.

References:

[1] Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of experimental social psychology*, *46*(6), 1093-1096.

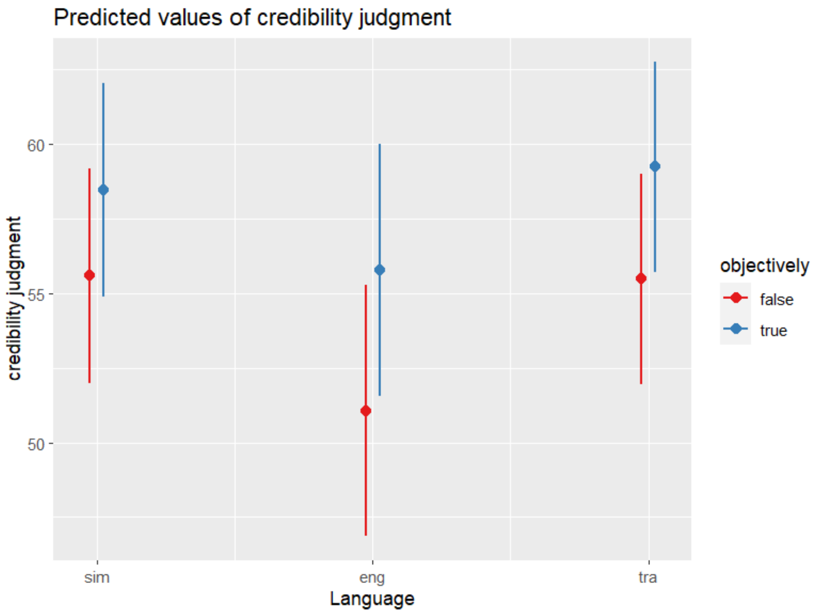
[2] Hatzidaki, A., Santesteban, M., & Navarrete, E. (2024). Illusory truth effect across languages and scripts. *Psychonomic Bulletin & Review*, 1-9.

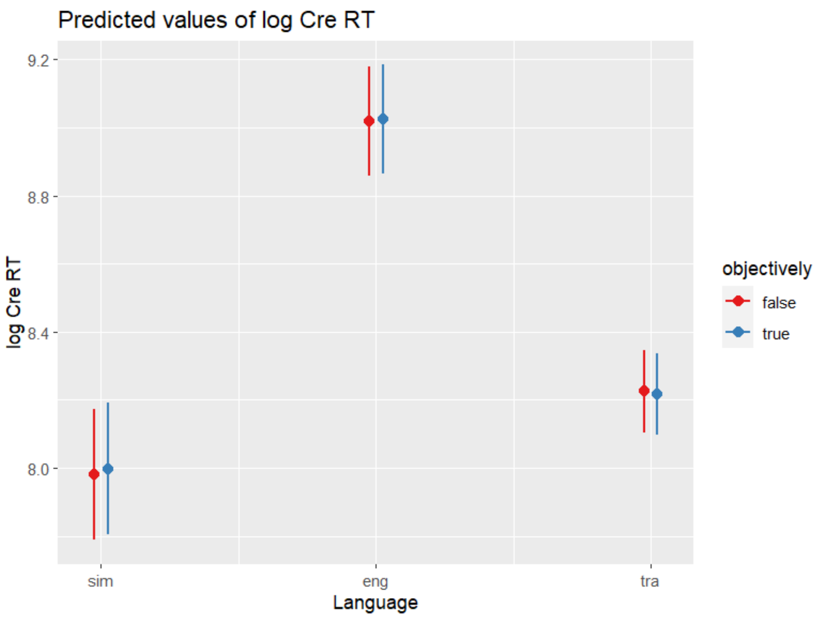
[3] Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002.

[4] Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, 160, 110-126

[5] Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, *1*(1), 9.

Figures:





Cloze, Frequency, Surprisal, or Plausibility? A Comparative Analysis of Predictors for Local Ambiguity Resolution

Markéta Ceháková[1], Jan Chromý[1]

[1]Institute of Czech language and theory of communication, Faculty of Arts, Charles University

cehakovm@ff.cuni.cz

Background: The garden-path (GP) effect is a well-known and frequently studied phenomenon in psycholinguistics, yet the factors influencing its presence and magnitude remain unclear, as studies focused on identifying such factors yield inconsistent findings. The effects often appear only in some studies or for specific structures [1, 2]. Our study investigates three potential predictors of the GP effect: structural bias/frequency, plausibility, and surprisal. We focus on properties of three strings related to key processing stages: the initial misanalysis of the ambiguous region (AMBMIS), the correct analysis of the ambiguous region (AMBCOR), and the correct analysis of the disambiguating region (DISCOR). To account for potential structural differences, we employed a diverse set of stimuli. We expect that the more frequent, plausible, and predictable AMBMIS is, the more difficult it will be to (re)analyze the GP sentence. The opposite pattern is expected for AMBCOR and DISCOR.

Method: We ran a web-based word-by-word self-paced reading experiment (n=295), collecting reaction times (RTs) and response accuracy (RA) for 66 Czech GP sentences. The items were matched for length but varied in syntactic structure, subjective naturalness, and semantic properties. Each sentence was followed by a yes-no comprehension question targeting the initial misanalysis. Structural bias was estimated using: a) a cloze task (n=115), calculated as the proportion of responses consistent with AMBMIS and AMBCOR; and b) absolute and relative frequency, calculated as the number of occurrences of a subcategorization frame of a specific verb consistent with AMBMIS and AMBCOR using Intercorp v16ud [3]. Surprisal estimates of AMBMIS, AMBCOR and DISCOR were calculated using the CzeGPT-2 model. Plausibility ratings of AMBMIS, AMBCOR and DISCOR were collected online using a 100-point slider scale (n=111).

Results: We found clear differences in RTs and RA between the GP and non-GP conditions. Due to high collinearity among predictors, we ran separate mixed effect models for each predictor in interaction with ambiguity, analyzing effects on RA and RTs (see Fig. 1 and 2), and compared the models using BIC. The cloze score of AMBMIS was the only predictor which showed a significant interaction with ambiguity for both RA ($\beta = -0.344$, $SE = 0.1$, $z = -3.446$, $p < 0.001$) and RTs ($\beta = 0.034$, $SE = 0.008$, $t = 4.149$, $p < 0.001$). There was also a significant interaction between ambiguity and plausibility of AMBCOR for the RA ($\beta = 0.342$, $SE = 0.1$, $z = 3.433$, $p < 0.001$), and of ambiguity and cloze score of AMBCOR for RTs ($\beta = -0.034$, $SE = 0.008$, $t = -4.176$, $p < 0.001$). Both cloze scores were also consistently assessed as having the best fit to the data according to BIC.

Discussion: Our study yields several important findings. 1) The likelihood of initial misanalysis (as measured by cloze scores) is, by far, the best predictor of GP effects in both RTs and RA. 2) Properties of AMBMIS alone are insufficient to explain the variation in GP effect; other parts/stages of analysis–especially AMBCOR–also play a critical role. Yet, many previous studies have overlooked these aspects. 3) Previously reported effects of plausibility or frequency do not replicate when a more diverse stimuli set is used. This is in line with studies showing that different GP structures might be processed differently, and that comprehenders also have individual strategies for dealing with them [1, 4].

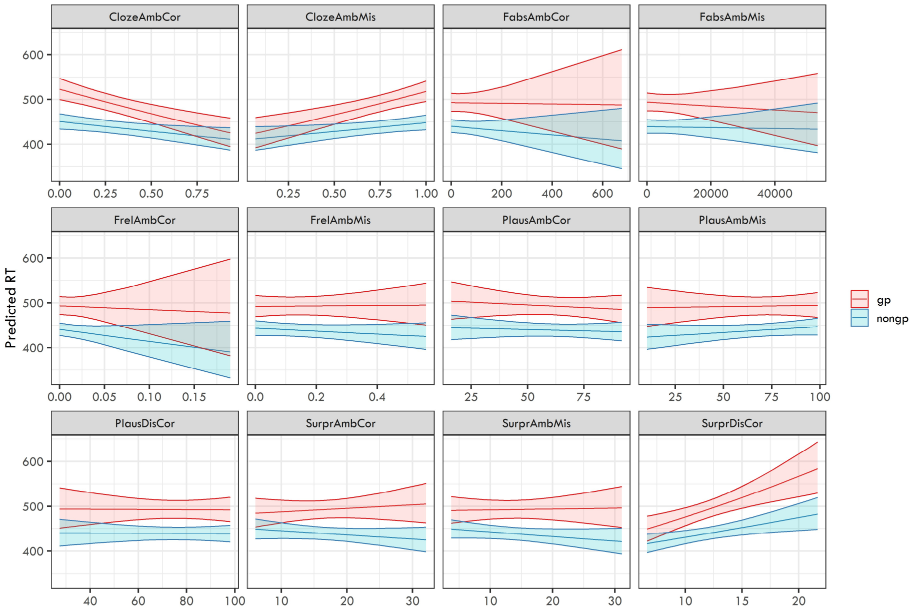


Fig.1: Model predictions for reaction times on the spillover region. Cloze = cloze task, Fabs = absolute frequency, Frel = relative frequency, Plaus = plausibility, Surp = surprisal

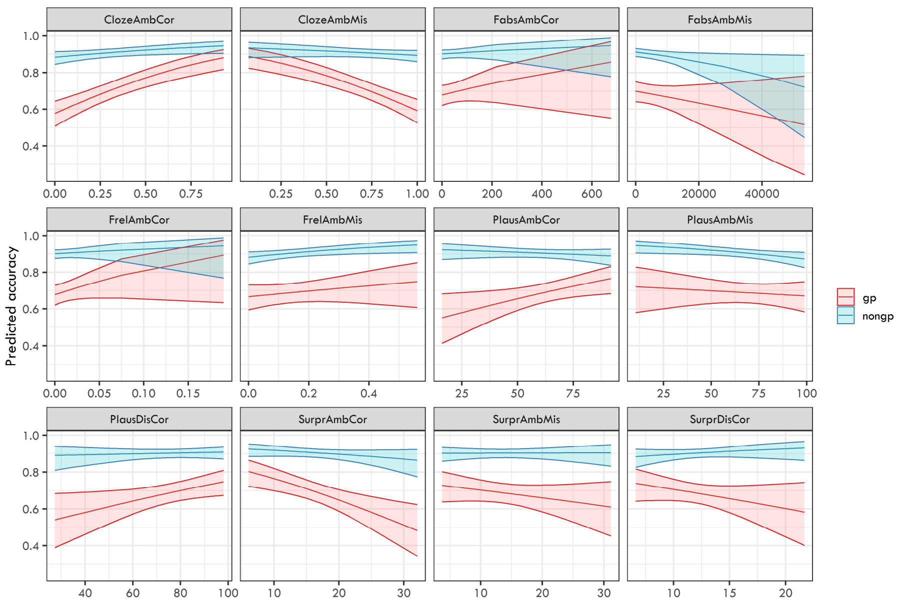


Fig. 2: Model predictions for response accuracy. Cloze = cloze task, Fabs = absolute frequency, Frel = relative frequency, Plaus = plausibility, Surp = surprisal

References

[1] Huang, K.-J., Arehalli, S., Kugemoto, M., Muxica, C., Prasad, G., Dillon, B., & Linzen, T. (2024). Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. Journal of Memory and Language, 137, 104510 **[2]** Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. Applied Psycholinguistics, 32 (2), 299–331. **[3]** Čermák, F., & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. International Journal of Corpus Linguistics, 17 (3), 411–427. **[4]** Paape, D., & Vasishth, S. (2022). Is reanalysis selective when regressions are consciously controlled? Glossa Psycholinguistics, 1 (1). https://doi.org/10.5070/G601139

Discriminative learning of number interpretation of German pseudo-nouns

Jennifer Keller[1] & Ingo Plag[1]
[1] Heinrich Heine University Düsseldorf
jennifer.keller@hhu.de, ingo.plag@hhu.de

Background: The German nominal number system has long been a test case for theories of morphological organization and morphological processing. Traditional approaches such as the schema account assume that a given word-form is assigned either singular or plural depending on certain aspects of its phonological make-up (number of syllables, presence/absence of umlaut, suffix exponents) (e.g. [1, 2]). Recent research using discriminative learning networks (e.g. [4]), however, has challenged the schema account. It has been shown that discriminative learning models based on only phonotactic information, in this case biphones, are able to determine correctly not only whether a given word-form is singular or plural, but also to predict the number decisions on real words by an aphasic patient. The present study tests whether a discriminative learning model is able to also predict the number decisions on pseudo-words, as elicited in an experiment by Köpcke and colleagues ([2]).

Method: The Naïve Discriminative Learning (NDL) model of the present study was first trained on real words in their singular and plural forms using CELEX (N=7842) to obtain association weights between biphones and singular/plural. These association weights were then used to derive association weights for the pseudo-words, based on the weights of their constituent biphones. These association weights were then used in a mixed-effects regression model to predict the choices of the participants in Köpcke et al.'s experiment.

Results: The results of the mixed-effects regression analysis indicate that the NDL activation weights are highly predictive of the participants' number decisions ($p < .0001$), as illustrated in Figure 1. The higher the activation of plural (as against singular), the higher the probability of a plural decision to a given word-form. Further inspection of the relationship between activation differences and actual plural decisions by the participants shows that the model picks up morphological structure without having explicit morphological information (see Figure 2). Words exhibiting different cues for plurality (low plurality: bare stems, high plurality: final *-en)*, form clusters with similar activation differences and similar proportions of decisions by the participants. Only the words with final *-s*, a supposedly strong cue for plurality, do not behave according to the trend.

Discussion: The present findings show that the NDL model is able to mimic participants' decisions on pseudo-words, and the model allows us to quantify the relative strength of individual number cues. The NDL model is, however, not well able to model the participants' decision on words with final *-s*. Final *-s* has been a matter of controversy also with other models of German inflection (see, e.g. [3] for discussion), and needs to be further explored also in the present framework. Overall, discriminative learning is able to account for crucial properties of German nominal number morphology and offers a new way of understanding the mapping of form and meaning in the German inflectional system.

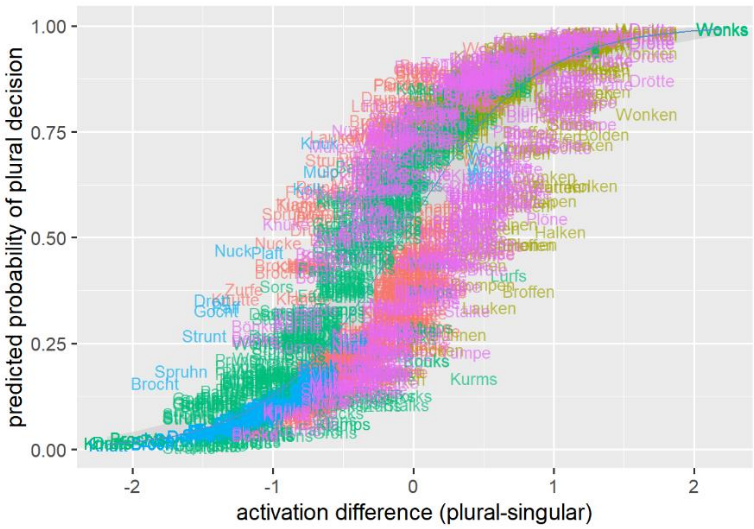


Figure 1. Partial effect of plural activations (as against singular activations) on the predicted probability of plural decision in the mixed-effects regression model, with cues for plurality color-coded: blue = bare stem, red = final *-e*, pink = umlaut + final *-e*, light green = final *-en*, dark green = final *-s*.

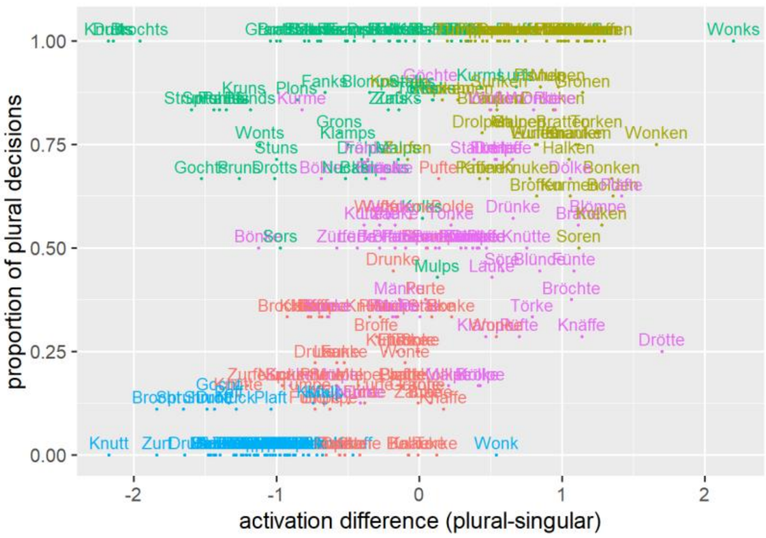


Figure 2. Partial effect of plural activations (as against singular activations) on the proportion of plural decisions for each word-form in the mixed-effects regression model, with cues for plurality color-coded: blue = bare stem, red = final *-e*, pink = umlaut + final *-e*, light green = final *-en*, dark green = final *-s*.

References

[1] Köpcke, K.-M. (1988). Schemas in German plural formation. *Lingua*, 74(4), 303-335.

[2] Köpcke, K.-M., Schimke, S. & Wecker, V. (2021). Processing of German noun plurals: Evidence for first- and second-order schemata. *Word Structure*, 14(1), 1-14.

[3] Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology* 29(3). 189–256.

[4] Plag, I., Heitmeier, M. & Domahs, F. (2023). German nominal number interpretation in an impaired mental lexicon: A naïve discriminative learning perspective. *The Mental Lexicon*, 18(3), 417-445.

The German nominal number system

Table 1. Number marking on German nominal nouns (cf. [1] and [4]).

Cue for Plural	**Example: Singular/Plural Pair**	**Counterexample: Singular**
-e	*Tisch* (‘table’) / *Tische* (‘tables’)	*Sage* (‘myth’)
-en	*Tür* (‘door’) / *Türen* (‘doors’) *Rose* (‘rose’) / *Rosen* (‘roses’)	*Boden* (‘floor’)
-er	*Kind* (‘child’) / *Kinder* (‘children’)	*Kater* (‘male cat’)
-s	*Auto* (‘car’) / *Autos* (‘cars’)	*Fuchs* (‘fox’)
-Ø	*Adler* (‘eagle’) / *Adler* (‘eagles’)	*Adler* (‘eagle’)
Umlaut (UL)	*Vater* (‘father’) / *Väter* (‘fathers’)	*Föhn* (‘hairdryer’)
UL + *-e*	*Kuh* (‘cow’) / *Kühe* (‘cows’)	*Tüte* (‘bag’)
UL + *-er*	*Wald* (‘wood’) / *Wälder* (‘woods’)	*Fächer* (‘fan’)
def. article	*der (Tisch) / die (Tische)* *das (Kind) / die (Kinder)*	*die (Tür) / die (Türen)*

The German nominal number system has been of high interest in the debates about the nature of morphological organization and processing since it involves a complex morpho-phonological variation. In addition to the inflection on the determiner or the adjective, number may be marked on the noun in different ways. As can be seen in Table 1, number can be indicated on the stem vowel of the noun by an umlaut (e.g. *Väter* (‘fathers’)) and/or at the end of the noun by a suffix (e.g. *Tische* (‘tables’)). Nevertheless, such indicators, or ‘cues’ for plural are found not only in plural forms. In fact, there are numerous instances in which singular forms carry these cues for plural as well (e.g. *Föhn* (‘hairdryer’) and *Sage* (‘myth’)). Hence, no single cue for plural in the German nominal number system is unambiguous and fully reliable.

Having one or three uncles: equally acceptable. A study about number mismatches in nominal Right-Node-Raising in German

Ricarda Scherer[1], Lisa Schäfer[1], Robin Lemke[1], Heiner Drenhaus[1], Ingo Reich[1]
[1]Saarland University, SFB 1102
ricarda.scherer@uni-saarland.de

Background: The phenomenon of right-node-raising (RNR) (*Joe likes __ and John hates <u>pizza</u>.*) [1; 2] describes an elliptical construction within a coordination: An expression that occurs in both conjuncts can be omitted, leaving a gap on the position left of the conjunction. There is no restriction regarding the category or constituent status of the omitted expression (see 1a) [3]. Attested mismatches in RNR, which are investigated here (see 1b), are not explained by syntactic accounts such as multiple dominance [2]. This is why mismatches are also interesting regarding language processing, e. g. following Arregui et al.'s [4] account that non-matching expressions undergo a reparation process. Mismatches contradict the claimed restriction of RNR that the omitted expression must be phonologically identical to the target expression [5], see (1b). However, Christ's [6] introspective work on German RNR states that there is a gradual decrease in acceptance of number mismatches of nouns, which depends on the noun class. Christ assumes that the omitted singular needs to be reconstructed from the overt plural, which is supposedly harder for more complex plural forms. Moreover, based on French and English data for verbal mismatches, Shiraïshi et al. [7] claim that RNR only demands lexeme identity. Thus, we investigated noun mismatches in German RNR empirically for the first time. We tested Christ's [6] claim that the complexity of plural inflection affects the acceptability of German noun mismatches in RNR. Concretely, Christ's hypothesis predicts that RNR becomes the less acceptable, the more morphemes are used forming a plural.

Method: *Acceptability rating* We differentiated between three types of plural forms in our mismatch condition: syncretic plural (2a, e.g. *Onkel*→*Onkel*), suffix plural (-*e*) (2b, e.g. *Fisch*→*Fisch*-*e*), suffix and umlaut plural (-*e* + umlaut) (2c, e.g. *Ast*→*Äst*-*e*). A high rating for syncretic plural and low(er) rating for suffix or suffix and umlaut or a significant difference between suffix versus suffix and umlaut would support Christ's theory of graduality. A high rating of each of the categories would support lexeme identity. To test whether possible acceptability differences indeed follow from number mismatches under ellipsis, we also tested these forms in a matching context (2d-2f). 48 participants recruited from Prolific saw each 24 items and 48 fillers, which they rated on a slider scale ranging from totally unacceptable to totally acceptable (internally coded 0-100).

Results: For the plural categories tested here, the LMM [8] (forward coded) did not show a significant gradual effect of complexity of plurals on the acceptability of RNR in German noun mismatches but a significant main effect for mismatches being rated lower than matches ($t(1078)= -4,61$, $p<0.05$) (see Tab. 1). All three categories show a high acceptability for both matches and mismatches (Fig. 1).

Discussion: We don't find any evidence for Christ's [6] account of a gradual decrease of acceptability the more complex a plural becomes, since we don't see a difference in acceptability ratings between different plural forms. In contrast, constantly high acceptability throughout categories is speaking in favor of Shiraïshi et al.'s [7] theory of lexeme identity. Regarding processing, the results would mean that possibly repaired structures were as acceptable as not repaired felicitous ones. However, Arregui's [4] recycling hypothesis was proposed for ellipses where the omitted part is in the second conjunct, which is not the case for RNR. We also plan further experiments with first, more irregular plural forms investigating if there is a turning point in acceptability instead of a gradual decrease, and second, an SPR-study to measure processing cost.

(1) a. *Tim ist für ~~Atomkraft~~ und Tom ist gegen Atomkraft.*
Tim is for ~~nuclear power~~ and Tom is against nuclear power.
b. *Hans hat eine ~~Maus~~ und Peter hat mehrere Mäus-e.* (Christ, 2011: 383)
Hans has one ~~mouse.~~SG and Peter has several mouse.PL-PL.

(2) a. *Niklas hat einen und Benjamin hat vier Onkel.* (Syncretic mismatch)
Niklas has one and Benjamin has four uncle.PL
b. *Jasmin fängt einen und Leonie fängt sieben Fisch-e.* (Suffix mismatch)
Jasmin catches one and Leonie catches seven fish-PL.
c. *Sebastian sammelt einen und Nele sammelt zehn Äst-e.* (Suffix and *Umlaut* mismatch)
Sebastian gathers one and Nele gathers ten branch-PL
d. *Niklas hat drei und Benjamin hat vier Onkel.* (Syncretic match)
Niklas has three and Benjamin has four uncle.PL
e. *Jasmin fängt vier und Leonie fängt sieben Fisch-e.* (Suffix match)
Jasmin catches four and Leonie catches seven fish-PL.
f. *Sebastian sammelt vier und Nele sammelt zehn Äst-e.* (Suffix and *Umlaut* match)
Sebastian gathers four and Nele gathers ten branch-PL.

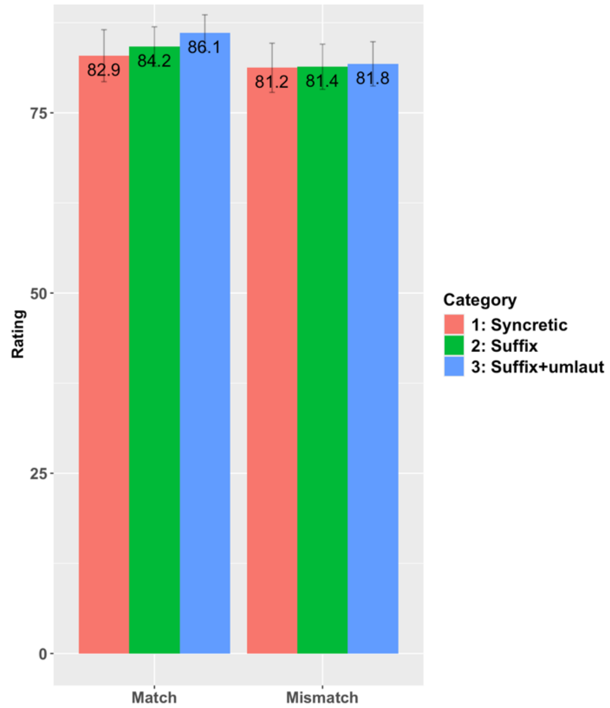


Figure 1 Mean acceptability ratings for the six tested conditions, errorbars show standard deviation

*Table 1 Fixed effects of the linear mixed model (lmer(Rating~Match*Category+(1|ID)+(1|Noun))*

	Estimate	t - value	p - value
Intercept	84.41	51.4	<0.05
MatchMismatch	-2.93	-4.61	<0.05
Syncretic_vs_Other	-1.27	-0.86	>0.05
Suffix_vs_SuffixUmlaut	-1.92	-1.30	>0.05
MatchMismatch:Syncretic_vs_Other	1.13	0.72	>0.05
MatchMismatch:Suffix_vs_SuffixUmlaut	1.52	0.97	>0.05

References [1] Ross, J. R. (1967). *Constraints on variables in syntax*. MIT dissertation. **[2]** Wilder, C. (1999). Right node raising and the LCA. *Proceedings of WCCFL,18*. 586–598. **[3]** Hartmann, K. (2000). *Right node raising and gapping*. John Benjamins. **[4]** Arregui, A., Clifton, C., Frazier, L., Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55. 232–246. **[5]** Sternefeld, W. (2009). *Syntax. Eine morphologisch motiviertegenerative Beschreibung des Deutschen*. Band 2(3). Tübingen: Stauffenburg. **[6]** Christ, R. (2011). Linkstilgung und Phonologische Quasi-Identität. *Linguistische Berichte*, *2011*(228), 371-411. **[7]** Shiraïshi, A., Abeillé, A., Hemforth, B. & Miller, P. (2019). Verbal mismatch in Right-Node Raising. *Glossa: a journal of general linguistics*. **[8]** Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Morphophonological Effects on Morphosyntactic Processing During L2 English Real-time Comprehension

Qingyuan Gardner[1,2], Vicky Chondrogianni[1], Holly P. Branigan[1], Peng Li[3]
[1] University of Edinburgh, [2] University of Oslo, [3] Basque Center on Cognition, Brain and Language
Qingyuan.Gardner@ed.ac.uk

While L1 influence on L2 morphosyntactic processing is widely investigated, the role of L1 and L2 phonology in L2 real-time comprehension is less well understood (e.g., Frenck-Mestre et al., 2008). Specifically, it remains unclear how L2 learners' sensitivity to morphosyntactic violations (e.g., **last weekend the girl* ***paint*** / ***chase*** *the dogs*) is affected by L1 phonotactics and L2 allomorphic variations (e.g., *paint*-***ed*** [tɪd] vs. *chas*-***ed*** [st]; Table 1) and L1 phonological constraints (e.g. Solt et al., 2004; Amaro et al., 2018). This study explores morphophonological-morphosyntactic interactions in native Mandarin (L1-M) learners of English, whose L1 lacks inflectional morphology and word-final consonant clusters (Duanmu, 2000).

57 intermediate-to-advanced adult L1-M learners of English (AoA: M = 7.5 yrs) and 65 age-matched L1-English monolinguals (L1-E) listened to sentences in a self-paced listening (SPL) task involving English tense / agreement morphosyntactic violations (past simple, present habitual); see example above. We manipulated the morphophonological properties of the verb's inflectional morpheme (past -*ed*; 3rd person -*s*) forming a vowel epenthesis (VES; *chas**es**, paint**ed***) or consonant cluster (CC; *cha**sed**, pain**ts**;* Table 2). We predicted longer reaction times if participants detected morphosyntactic violation (ungrammatical; UG) on the verb compared with a verb without violation (grammatical; G). Subsequently, using the identical verbs and morphemes, participants identified if an inflected or uninflected verb (A or B) matched the test verb (X) in an auditory ABX discrimination task (Table 3). Longer reaction times indicate weakened perceptual sensitivity.

In SPL, L1-M participants showed sensitivity to verb-related L2 morphosyntactic violations (G: M = 540.38 ms vs. UG: M = 571.86 ms; *p*=.003**); this sensitivity was stronger for verb-final consonant clusters than vowel epenthesis (CC G-UG: M = 64.21 ms vs. VES G-UG: M = 39.83 ms; *p*=.044*). In contrast, although L1-E participants also showed sensitivity to violations (G: M = 445.77 ms vs. UG: M = 525.87 ms; *p*<.001***), this sensitivity did not differ between consonant clusters and vowel epenthesis (CC G-UG: M = 72.71 ms vs. VES G-UG: M = 70.25 ms; *p*=.642). Interestingly, verb frequency modulated grammatical sensitivity in both groups (L1-M: *p*<.001***; L1-E: *p*<.001***). In ABX, we again found L1-M sensitivity to verb-related L2 morphosyntactic violations, but in the opposite direction: L1-M were better at detecting verb-final vowel epenthesis than consonant clusters (VES: M = 438.78 ms vs. CC: M = 485.29 ms; *p*=.007**). Furthermore, in this task L1-M did not differ from L1-E (Group x Morphophonological condition interaction: *p*=.169).

Our study provides novel evidence for morphophonological effects on morphosyntactic processing during L2 comprehension. Importantly, our findings suggest a context-dependent perceptual asymmetry in L1-Mandarin learners of English: Despite the absence of consonant clusters in Mandarin, they showed greater sensitivity to L2 morphosyntactic violations involving verb-final consonant clusters than vowel epenthesis in sentential contexts - but not when perceiving verb inflections in isolation. We suggest this pattern may reflect L2 phonological features facilitating morphosyntactic sensitivity during listening comprehension, regardless of L1 phonotactics. Other possible explanations include increased perceptual saliency of a plosive consonant ([t]) in verb-final consonant clusters, or intra-word co-articulation masking verb-final codas.

References

Amaro, J. C., Campos-Dintrans, G., & Rothman, J. (2018). The role of L1 phonology in L2 morphological production: L2 English past tense production by L1 Spanish, Mandarin, and Japanese speakers. *Studies in Second Language Acquisition*, *40*(3), 503-527.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, 128(3), 528-536.

Solt, S., Pugach, Y., Klein, E. C., Adams, K., Stoyneshka, I., & Rose, T. (2004). L2 perception and production of the English regular past: Evidence of phonological effects. In *BUCLD 28: Proceedings of the 28th annual Boston University Conference on Language Development* (pp. 553-564). Somerville, MA: Cascadilla Press.

Table 1. Illustration of English morphophonological (MP) features for 3rd person -s and past -ed (2 x 2)

	consonant cluster	vowel epenthesis
3rd person -*s*	*paints* [ts]	*chases* [sɪz]
past -*ed*	*chased* [st]	*painted* [tɪd]

Table 2. Exemplar sentences for verbs in SPL task (grammatical conditions only)

temporal context	**MP feature**	**verb**	**exemplar sentence**
Present habitual	*consonant cluster*	*paints* [ts]	*Every weekend / the girl / paints / a sunflower / in the park*
Past simple	*consonant cluster*	*chased* [st]	*Last weekend / the girl / chased / the dog / in the park*
Past simple	*vowel epenthesis*	*painted* [tɪd]	*Last weekend / the girl / painted / a sunflower / in the park*
Present habitual	*vowel epenthesis*	*chases* [sɪz]	*Every weekend / the girl / chases / the dog / in the park*

* / indicate segment boundaries;

Table 3. Exemplar ABX trials with corresponding inflections and MP features

inflectional morpheme	**MP feature**	**verb**	**A**	**B**	**X**
3rd person -*s*	*consonant cluster*	*paints* [ts]	*paint*	*paints*	*paints*
past -*ed*	*consonant cluster*	*chased* [st]	*chased*	*chase*	*chased*
3rd person -*s*	*vowel epenthesis*	*chases* [sɪz]	*chase*	*chases*	*chases*
past -*ed*	*vowel epenthesis*	*painted* [tɪd]	*painted*	*paint*	*painted*

Early sensitivity to gender morphology in Czech infants

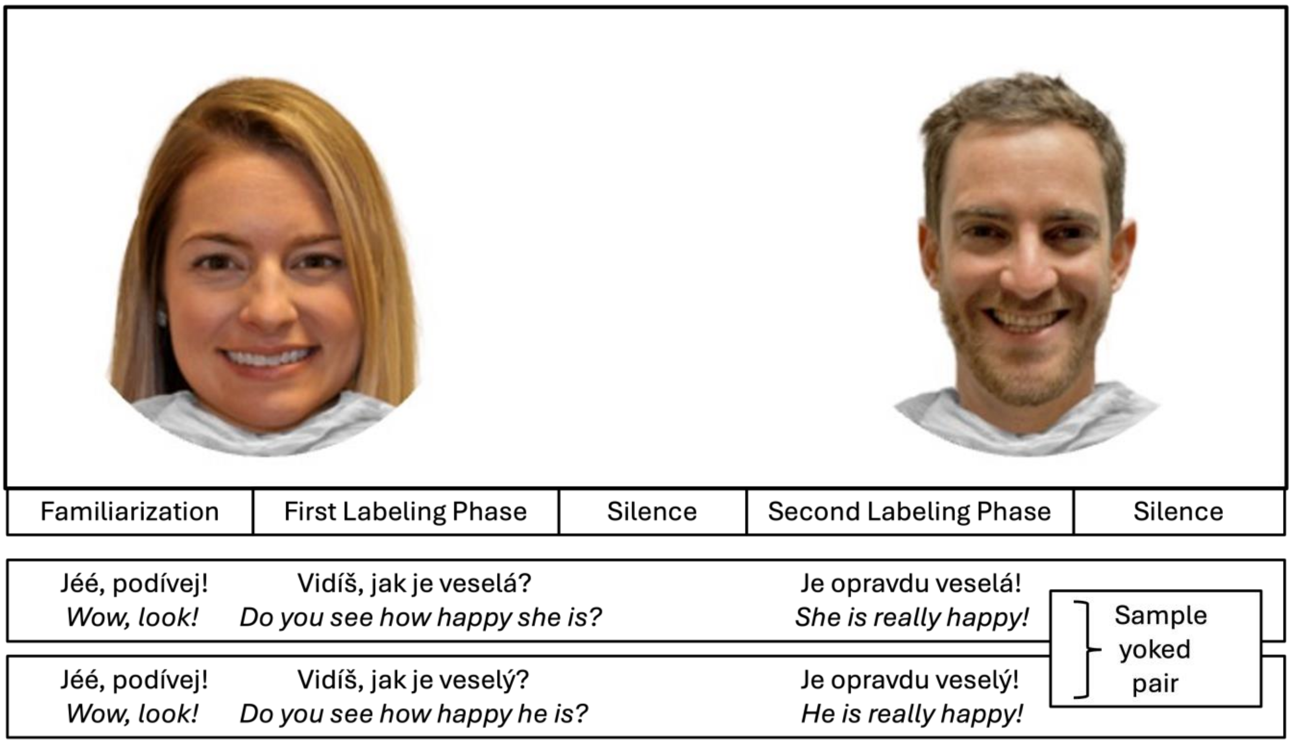
Mikuláš Preininger[1], Filip Smolík[1], Nikola Paillereau[1]
[1]Institute of Psychology, Czech Academy of Sciences, Czechia
mikulas.preininger@psu.cas.cz

Background: Children begin to pick up on grammatical cues and regularities even before they produce their first words. For instance, 7-month-old bilingual infants have been shown to use prosodic cues such as pitch and duration to distinguish and acquire the typical word order of two different languages [1]. Similarly, 12-month-old Italian infants have demonstrated comprehension of number and gender marking on nouns [2]. The present study employs the preferential looking paradigm to investigate the onset of sensitivity to grammatical gender in Czech-learning infants. This is a work in progress: the experimental design has been completed, and data collection is set to begin shortly and is expected to be completed by the end of July 2025.
Method: 18- and 24-month-old infants will be tested ($N_{total} = 40$). On each trial, infant will see an image with a pair of faces, female and male, selected from the Radboud Faces Database [3]. Simultaneously, they will listen to pairs of phrases containing gender marking on a pair of adjectives:

1)	Jéé, podívej!	Vidíš, jak je veselá/veselý?	Je vážně veselá/veselý!
	wow look.2SG.IMP	see.2SG how be.3SG happy.F/M.SG	be.3SG really happy.F/M.SG
	'Wow, look!	Do you see how happy s/he is?	S/He is really happy!'

For an example of a trial, see Figure 1. Infants' eye-gaze will be recorded throughout the whole experiment. Proportion of looks directed towards the masculine/feminine face from the onset of the first gender marked adjective to the trial end will be analyzed using linear mixed-effect models and permutation analysis.
Results: If children extract gender information from the adjective forms and possess the association between grammatical and natural gender, they should look towards the corresponding faces after the adjectives are presented.
Discussion: In the future, sensitivity towards grammatical gender marking on adjectives in other than nominative and singular forms would be worth analyzing, in order to examine how infants deal with the morphological complexity of Czech. Understanding grammatical gender morphology will be also explored in other parts-of-speech, namely nouns and verbs.

Figures



References

[1] Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature communications*, *4*(1), 1490.

[2] Ferry, A., Nespor, M., & Mehler, J. (2020). Twelve to 24-month-olds can understand the meaning of morphological regularities in their language. *Developmental psychology*, *56*(1), 40.

[3] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. Cognition & Emotion, 24(8), 1377—1388.

Resumption in Anaphoric Dependencies: A Case Study of Mandarin Topicalization

Xu Ji[1], Dawei Jin[2]
[1]City University of Hong Kong; [2]Shanghai Jiao Tong University
xuji25-c@my.cityu.edu.hk

Background: Research on the cross-linguistic and cross-structural distribution of resumption pronouns (RPs)—where lexical pronouns replace traces in A-bar dependencies—has focused on grammatical (e.g., islands) and processing constraints (e.g., levels of embedding). A further theoretically significant comparison (Lasnik & Stowell 1991) contrasts RPs in anaphoric (e.g., topicalization) versus quantificational (e.g., *wh*-questions) A-bar dependencies, given arguments that RPs link anaphorically to antecedents (Erteschik 1992; Alexopoulou & Keller 2007). Nevertheless, anaphoric contexts remain understudied relative to quantificational ones—a gap this study addresses through experimental analysis of RP behavior under grammatical and complexity constraints in Mandarin topicalization.
Method: We conducted two auditory Likert-based judgment studies to examine resumption in non-island complements across multiple embedding levels, as well as in conjunct and adjunct islands. Experiment 1 employed a 3×2 factorial design, crossing embedding (zero, single, or double) and resumption (gap vs. resumption). Experiment 2 used a 2×2×2 factorial design, manipulating distance (short vs. long topicalization based on gap location), structure (non-island vs. island), and resumption (gap vs. resumption) for conjunct and adjunct islands, following Sprouse (2007). The designs yielded 128 stimuli for Experiment 1 (104 participants) and 36 stimuli for Experiment 2 (60 participants) in a Latin square design mixed with fillers (1:3). Sample stimuli for both experiments are presented in Table 1.
Results: Mean judgements with SEs based on z-transformed data are as in Figure 1-3. *Experiment 1* The LMEMs revealed a significant negative embedding effect on the acceptability of topicalization from across complements, but no effect of resumption or an Embedding×Resumption interaction, indicating that embedding effects are consistent across different tails. A post-hoc t-test examined the interaction between Embedding and Resumption (see Table 2). Both RPs and gaps were rated grammatical in all conditions (z-score > 0), with RPs consistently rated higher, though significantly so only in doubly embedded topicalization (p = 0.000). Independent t-tests revealed a significant difference for each pairwise comparison among the three embedding levels for gapped topicalization, but only a significant difference between zero and double embedding for RPs. *Experiment 2* The LMEMs showed a significant Distance × Structure interaction, indicating superadditivity—and thus island violations—for both gaps and resumptives across island types. No Distance × Structure × Resumption interaction was found, suggesting island effects were similar for gapped and resumptive topicalization. However, resumption significantly ameliorated the island effect, evidenced by smaller DD-scores (Conjunct: $mean_{gap}$ = -1.49, $mean_{RP}$ = -1.05; Adjunct: $mean_{gap}$ = -0.373,$mean_{RP}$ = -0.312) and higher ratings in the resumptive island/long condition (Conjunct: $p < .001$; Adjunct: $p = .044$).
Discussion: **(i)** RPs in Mandarin topicalization show no acceptability penalty in simple contexts, being fully acceptable in both embedded and unembedded non-island structures, as well as their gapped counterparts. This suggests RP in topicalization functions as a grammatical mechanism, instead of a last-resort strategy for island violations or a memory aid in complex contexts. This pattern, echoed by similar findings in English (e.g., Ross 1967), contrasts with the unacceptability of RPs in other dependencies. We attribute this to topicalization properties. Following Boeckx (2003), we assume that RPs are stranded determiners that select for DPs with a specificity feature (Bianchi 2004 *inter alia*). This specificity requirement differentiates RPs in topicalization dependencies from those in quantificational dependencies, where the lack of specificity explains the dispreference for RPs. **(ii)** We find RPs facilitate processing, outperforming gaps across conditions—with a tendency in zero and double embeddings and reaching statistical significance in single embedded topicalization. This facilitation can be attributed to RPs' role in constructing well-formed local parses (Asudeh 2011) and aiding dependency formation through anaphoric linking to their antecedents, relying on cue-based retrieval where explicit morphological cues (e.g., number) enhance retrieval efficiency. However, while RPs provide processing advantages, they do not fully eliminate the embedding effect, as acceptability ratings decline sharply in doubly embedded contexts and remain statistically indistinguishable from gaps. This contradicts earlier findings (Erteschik-Shir 1992; Alexopoulou & Keller 2007; Farby et al 2010), the latter potentially confounded by an independent effect of diminishing salience of violations with increasing dependency distance (Staum Casasanto & Sag 2008; Philip 2014) given the high RP unacceptability in simple non-island contexts within quantificational dependencies, orthogonal to RPs' inherent resistance to processing difficulty in longer dependencies. In fact, in topicalization with RPs as a grammatical mechanism in easy-to-process contexts, our study demonstrates their susceptibility to the embedding effect. **(iii)** RPs cannot fully rescue the island effect in topicalized dependencies but significantly ameliorate it. This contradicts earlier base-generation analyses which claim resumption involves a binding relationship and should therefore be immune to island constraints. (Ross 1967; Cinque 1997; Rizzi 2010). Instead, the findings align with Asudeh's (2011) framework, which posits that RPs facilitate the construction of well-formed local parses but cannot render the global structure fully grammatical. Additionally, this partial amelioration is also consistent with Alexopoulou & Keller's (2007) proposal, where topicalization's movement feature on matrix C triggers a cyclic derivation, predicting a trace until the parser encounters the pronoun and shifts to anaphoric resolution. However, the pronominal cannot cancel the memory costs incurred up to that point, explaining why resumption fails to fully restore island-violating structures.

[1] Relative clauses have been considered a type of quantificational movement because they involve an operator that binds a variable and ranges over a potentially non-singleton set, as discussed by Lasnik and Stowell (1991) and Alexopoulou (2006).

References(selected)

[1] Alexopoulou, Theodora. 2006. “Resumption in relative clauses.” Natural Language & Linguistic Theory 24(1): 57-111.

[2] Alexopoulou, Theodora & Frank Keller. 2007. “Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor.” Language 83(1): 110-160.

[3] Asudeh, Ash. 2011. “Local grammaticality in syntactic production.” In Emily Bender and Jennifer Arnold (eds.), Language from a cognitive perspective: Grammar, usage & processing. Studies in honor of Thomas Wasow. Stanford: CSLI Publications.

[4] Boeckx, Cedric. 2003. Islands and chains: Resumption as derivational residue. Linguistik Aktuell. John Benjamins.

[5] Erteschik-Shir, Nomi. 1992. “Resumptive pronouns in islands.” In Helen Goodluck and Michael Rochemont (eds.), Island constraints: Theory, acquisition and processing, 89-109. Dordrecht: Kluwer Academic Publishers.

[6] Hofmeister, Philip, and Elisabeth Norcliffe. 2014. “Does resumption facilitate sentence comprehension?”

[7] Lasnik, Howard and Tim Stowell. 1991. “Weakest Crossover.” Linguistic Inquiry 22(4): 687-720.

[8] Staum Casasanto, L. and Sag, I. 2008. “The advantage of the ungrammatical.” In Love, B., McRae, K., and Sloutsky, V. M. (eds.), Proceedings of the 30th Annual Meeting of the Cognitive Science Society. Austin, TX.

Zero embedding	Zhe-ming shouhuoyuan a, guke jingchang tousu ∅/ta. DEM-CLF salesman TOP, customer often complain ∅/3SG. ‘This salesman, customers often complain about ∅/him.’
Single embedding	Zhe-ming shouhuoyuan a, Laochen zaojiu zhidao guke jingchang tousu ∅/ta. DEM-CLF salesman TOP, Laochen long.ago know customer often complain ∅/3SG ‘This salesman, Laochen has long known that customers often complain about ∅/him.’
Double embedding	Zhe-ming shouhuoyuan a, Laochen juede jingli zaojiu zhidao guke jingchang tousu ∅/ta. DEM-CLF salesman TOP, Laochen think manager long.ago know customer often complain ∅/3SG ‘This salesman, Laochen thinks the manager has long known that customers often complain about ∅/him.’

Coordinate islands	Adjunct islands
Non-island/long	
Zhe-ge nansheng a, Wang-laoshi feichang xihuan ∅/ta. DEM-CLF boy TOP, Wang-teacher very like ∅/3SG ‘This boy, Teacher Wang likes ∅/him very much.’	Zhe-ge xiaotou a, dajia dou juede jingcha hen nan zhuadao ∅/ta. DEM-CLF thief TOP, everyone all think police very hard catch ∅/3SG ‘This thief, everyone thinks that the police will have a hard time catching ∅/him.’
Island/long	
Zhe-ge nansheng a, Wang-laoshi feichang xihuan ∅/ta he Xiaoming. DEM-CLF boy TOP, Wang-teacher very like ∅/3SG and Xiaoming ‘This boy, Teacher Wang likes ∅/him and Xiaoming very much.’	Zhe-ge xiaotou a, dajia zai jingcha zhuadao ∅/ta zhiqian dou hen bu'an. DEM-CLF thief TOP, everyone at police catch ∅/3SG before all very uneasy ‘This thief, everyone was very uneasy before the police caught ∅/him.’
Non-island/short	
Zhe-ge nansheng a, ∅/ta feichang xihuan Xiaoming. DEM-CLF boy TOP, ∅/3SG very like Xiaoming ‘This boy, ∅/he likes Xiaoming very much.’	Zhe-xie jumin a, ∅/tamen dou juede jingcha hen nan zhuadao xiaotou. DEM-PL resident TOP, ∅/3PL all think police very hard catch thief ‘These residents, ∅/they all think that the police will have a hard time catching the thief.’
Island/short	
Zhe-ge nansheng a, ∅/ta feichang xihuan Xiaochen he Xiaoming. DEM-CLF boy TOP, ∅/3SG very like Xiaochen and Xiaoming ‘This boy, ∅/he likes Xiaochen and Xiaoming very much.’	Zhe-xie jumin a, ∅/tamen zai jingcha zhuadao xiaotou zhiqian dou hen bu'an. DEM-PL resident TOP, ∅/3PL at police catch thief before all very uneasy ‘These residents, ∅/they were very uneasy before the police caught the thief.’

Table 1. Sample stimuli for Experiment 1 (top) and Experiment 2 (bottom).

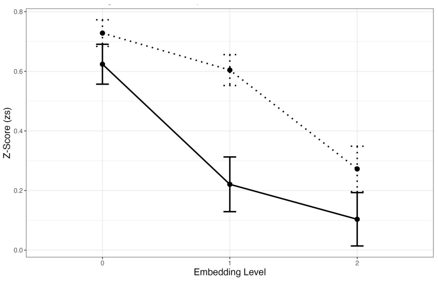


Figure 1. Effects of embedding and resumption in Experiment 1.

Embedded level	Gap	Resumptive	Comparison (RP vs. Gap)
Zero			
Single	***		***
Double	***	***	

Table 2. Results of t-tests comparing embedded clauses with the unembedded control (zero embedding).

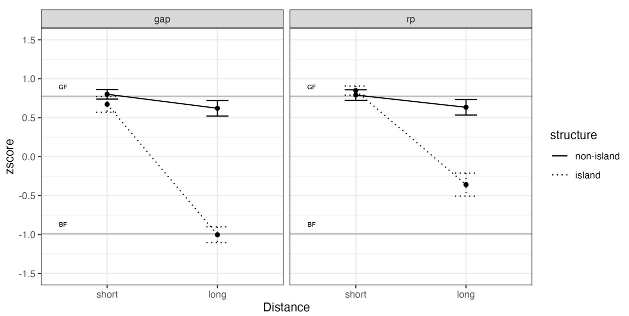


Figure 2. Interaction plots for conjunct island in Experiment 2[2].

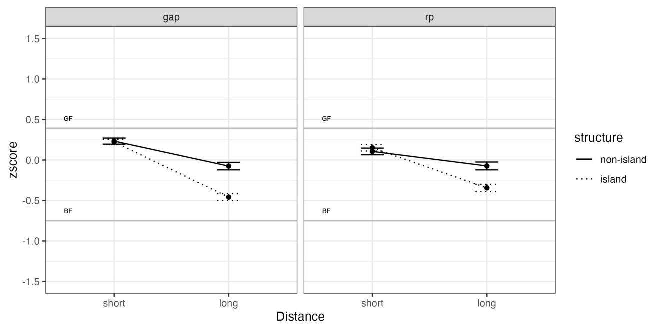


Figure 3. Interaction plots for adjunct island in in Experiment 2.

[2] Points are condition means using z-score transformation. Error bars represent standard error. The horizontal gray lines indicate the mean rating of the grammatical and the ungrammatical rated filler type, respectively. Same for Figure 3.

Can planned words trigger interference during real-time sentence production?

Elise Oltrogge[1], Eun-Kyoung Rosa Lee[2], Sol Lago[1]
[1] Goethe University Frankfurt [2] University of Maryland, College Park
Oltrogge@em.uni-frankfurt.de

Previous research has shown that the production of verbs, reflexives, and object pronouns can be disrupted by preceding nouns with mismatching morphosyntactic features, a phenomenon known as "agreement attraction" [1, 2]. However, production studies have largely been limited to configurations in which the mismatching noun—the "attractor"—is spoken before the target word. Thus, it remains unclear whether planned-but-unspoken nouns can cause interference. We tested this using English possessive pronouns, which agree with the gender of the possessor and not the gender of the possessee, e.g., *Susan chased her/*his grandpa*. Since the possessee noun (e.g., *grandpa*) is closely linked syntactically and planned alongside the pronoun [3], it could still interfere with pronoun selection resulting in pronoun errors or longer speech latencies. To test this hypothesis, we adapted a paradigm previously used with learners of English [5,6] such that we could detect interference effects in speech accuracies and latencies in native speakers.

Method. 112 native English speakers completed a timed production task with pronoun and verb trials (the latter to detect attraction effects). We focus on the pronoun trials. Stimuli included gender-matching or mismatching possessor and posesee nouns (Table 1). In each trial, participants saw a fixation, followed by a character with a statement, and then a probe word (Fig. 1). The task was to recall and reproduce what the character had said using the probe word, which required converting the statement from the first to the third person using a pronoun. More errors and/or delays in responses in the mismatch vs. match condition would indicate that the gender of the possessee noun interfered with pronoun selection. Generalized linear mixed-effects models were used to analyze error rates, and linear mixed-effects models were used to analyze the duration of the verb and pronoun.

Results. We found robust number attraction effects in the verb trials in both accuracy and duration measures (Fig. 2). In contrast, pronoun errors were rare (2.1%, 48 out of 2265 trials), with no significant differences between the match and mismatch conditions (p > .05). Of all pronoun errors, 13% were gender reversal errors, with numerically more errors in the mismatch than match condition (Table 2). No significant differences between conditions were found in the duration of the pronoun or preceding verb (p > .05).

Discussion. Our findings indicate that English speakers rarely make gender errors with possessive pronouns, and that a planned-but-unspoken possessee noun rarely interferes with pronoun selection. This contrasts with the attraction effects previously reported with object pronouns, which occurred in sentences in which the attractor noun preceded the pronoun [1,2]. Given the robust number attraction effect found with verbs, our failure to detect interference was not likely due to task limitations. Rather, our results indicate that interference in pronoun production may be limited to articulated, rather than planned, mismatching elements.

We are currently replicating the study with native speakers of German, a language where the pronoun must agree in gender with both the possessor and the possessee noun. Thus, German critically differs from English in that the gender of the possessee is relevant for agreement. The cross-linguistic comparison will help clarifying whether the absence of pronoun attraction with planned-but-unspoken attractors is due to a process that is robust to interference or, alternatively, whether it is specific to languages that do not morphologically mark an agreement relationship between the pronoun and the attractor noun.

Table 1. Sample stimuli. The gender of the pronoun was counterbalanced across items.

Trial Type	Condition	Statement	Probe	Target response
Pronoun	**gender match**	"I chased my grandma"	Susan	Susan chased her grandma
	gender mismatch	"I chased my grandpa"		Susan chased her grandpa
Verb	**singular match**	"The key to the cabinet…"	rusty	The key to the cabinet was rusty
	singular mismatch	"The key to the cabinets…"		The key to the cabinets was rusty
	plural match	"The keys to the cabinet…"		The keys to the cabinet were rusty
	plural mismatch	"The keys to the cabinets…"		The keys to the cabinets were rusty

Figure 1. Illustration of a trial.

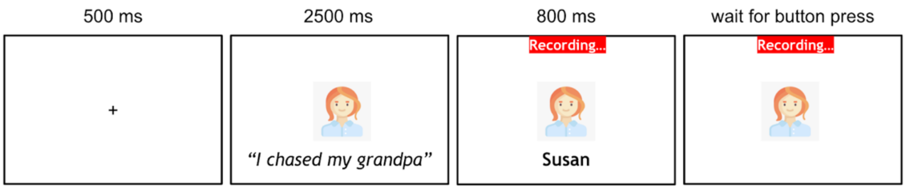
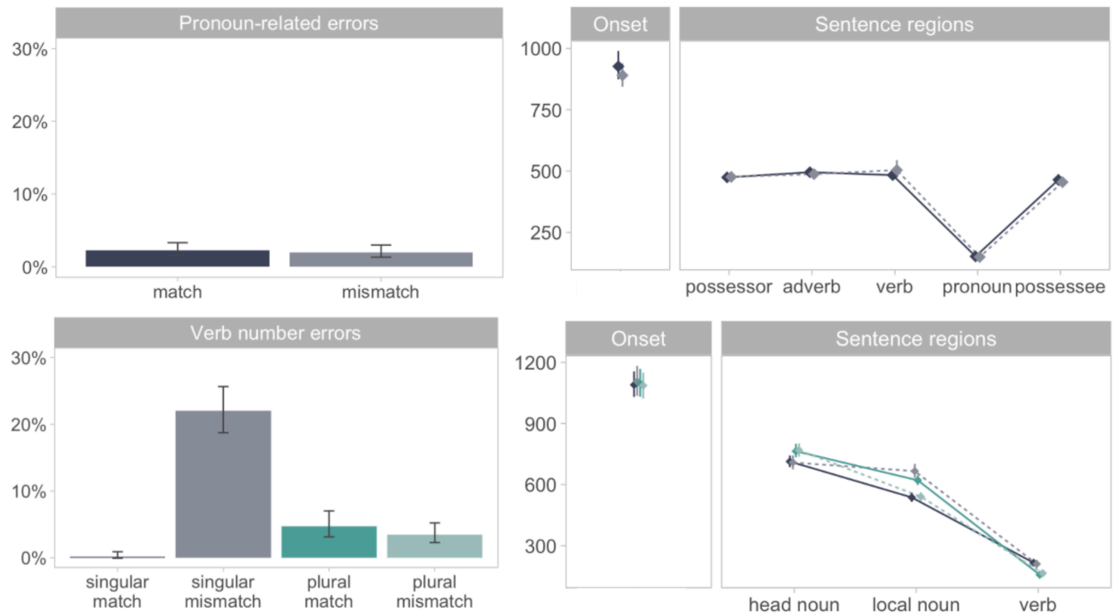


Figure 2. Pronoun-related error rates (top left) and durations (in ms; top right); verb number error rates (bottom left) and durations (in ms; bottom right). The colors reflect the conditions indicated in Table 1.



References

[1] Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and language*, *40*(3), 330-346.

[2] Kandel, M., & Phillips, C. (2022). Number attraction in verb and anaphor production. *Journal of Memory and Language*, *127*, 104370.

[3] Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791.

[4] Slevc, L. R., Wardlow Lane, L., & Ferreira, V. S. (2007). Pronoun production: Word or world knowledge. *MIT Working papers in Linguistics*, *53*, 191-203.

[5] Antón-Méndez, I. (2010). Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research*, *39*(2), 119-139.

[6] Pozzan, L., & Antón-Méndez, I. (2017). English possessive gender agreement in production and comprehension: Similarities and differences between young monolingual English learners and adult Mandarin–English second language learners. *Applied Psycholinguistics*, *38*(4), 985-1017.

A key to interpreting late effects in the brain response to metaphors: priming figurative (but not literal) meaning

Fabrizio Luciani, Federico Frau, Paolo Canal, Riccardo Venturini, Luca Bischetti, Valentina Bambini

Research on metaphor disputed about whether figurative meaning understanding requires different processing stages[1,2]. Competing theoretical accounts have been tested using Event-Related Potentials (ERPs), which typically show larger N400 for metaphors, sometimes followed by late effects[3,4]. For instance, Weiland and colleagues[5] showed that the amplitude of the N400 for metaphors such as *Those lobbyists are hyenas* was reduced when salient features of the literal meaning were primed by the word *furry*, thus helping to characterize the N400 as a lexical/semantic stage in which aspects of the literal meaning play a role. For the late effects, such a stringent description is missing. We hypothesize that ERP late effects specifically reflect the derivation of the figurative meaning of a metaphor, which follows the lexical operations reflected in the N400. To test this hypothesis, we adopted a crossmodal masked priming paradigm aiming to tap specifically into figurative but not literal meaning (**Fig.1**). By using NLP measures (i.e., semantic similarity), we operationalized figurative primes as words semantically close to central properties of the figurative meaning but distant from the literal meaning (e.g., for *That hairstyle is a bush*, *rasta*, close to *messy* but distant from *bush*), while literal primes were operationalized as words semantically close to the literal meaning but distant from the figurative one (e.g., *berry*, close to *bush* but distant from *messy*). Here, we present two studies: Experiment 1, aiming to replicate the effect of literal priming on the N400 amplitude, and Experiment 2, aiming to test the effect of figurative priming on late ERP components.

In **Experiment 1**, 40 Italian speakers (23.15 yo) were presented with 120 critical paired items (+90 fillers), including metaphorical (*That hairstyle is a bush*, …) and literal (*That plant is a bush*, ...) sentences. For each pair, a literal prime (*berry*) was selected and displayed (or not) in a Prime vs. No-Prime condition. The LMM on the N400 (from 200 to 270ms from word recognition point), restricted to eight centro-parietal channels, showed a significant Prime*Condition interaction (t=-2.05, p=.041), reflecting larger negativity for Metaphors in No-Prime compared to Prime condition. In the 550-650ms time-window (in five frontal channels), LMM showed a mild main effect of Condition (t=1.95, p=.054), reflecting greater positivity for Metaphors compared to Literal sentences (**Fig.2 A**). In **Experiment 2**, 29 Italian speakers (22.03 yo) were presented with 90 metaphors (+225 fillers). For each metaphor, a Literal prime (*berry*) and a Figurative prime (*rasta*) were presented alongside an Illegal prime (*asdmbar*). LMM for the N400 confirmed the difference between Literal and Illegal primes (t=2.76, p=.006), reflecting larger N400 for Illegal vs. the Literal prime. In the P600 time-window, the difference between Figurative and Literal primes (t=2.62, p=.009) reflected a greater positivity for Figurative compared to Literal primes (**Fig.2 B**).

Collectively, results showed that priming a metaphor such as *That hairstyle is a bush* with *berry* reduces the N400, as in Weiland et al.[5], while a prime like *rasta* –linked solely to the intended meaning– affects the P600, thus linking this component to implicature derivation. Interestingly, figurative priming enhances the P600, which might reflect the achievement of a pragmatic interpretation, in line with findings reported for other pragmatic phenomena[6,7]. Overall, our results support the idea that the derivation of pragmatic implicature occurs late, fostering a multi-stage neurocronometric model of metaphor interpretation.

References: [**1.**] Ortony et al. J. Verb. Learning Verb. Behav. 1978;17(4):465–477. [**2.**] Glucksberg. Trends Cogn. Sci. 2003;7:92–96. [**3.**] Coulson & Van Petten. Mem. Cogn. 2002;30:958–968. [**4.**] De Grauwe, et al. Neuropsychologia. 2010;48(7):1965–1984. [**5.**] Weiland et al. Front. Hum. Neurosci. 2014: 8:583. [**6.**] Schumacher & Meibauer. In: Liedtke & Schulze eds. De Gruyter; 2013:231–246. [**7.**] Weissman & Tanner. PLoS One. 2018;13(8):e0201727.

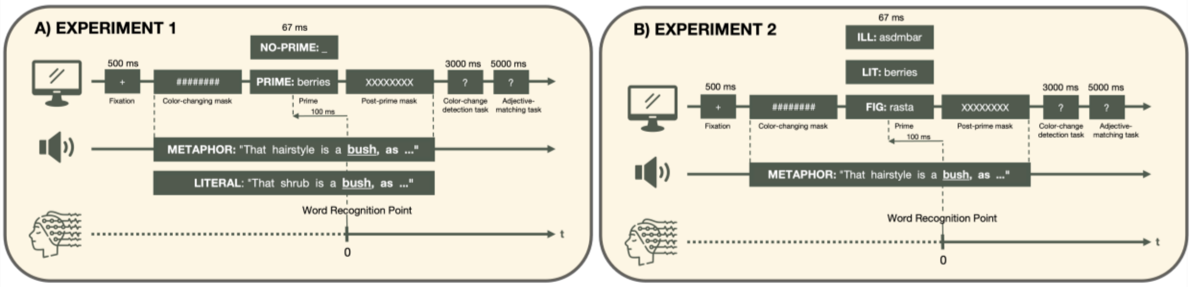


Figure 1. The paradigm for Experiment 1 (A) and Experiment 2 (B). In both experiments, after a fixation cross, a sentence was auditorily presented, and a color-changing mask stayed on the screen until a prime word was visually presented at the center of the screen for 67ms, with an onset at 100ms before the word recognition point of the target word (underlined). After the presentation of the sentence, participants underwent two control tasks (a color-change detection and an adjective-matching task).

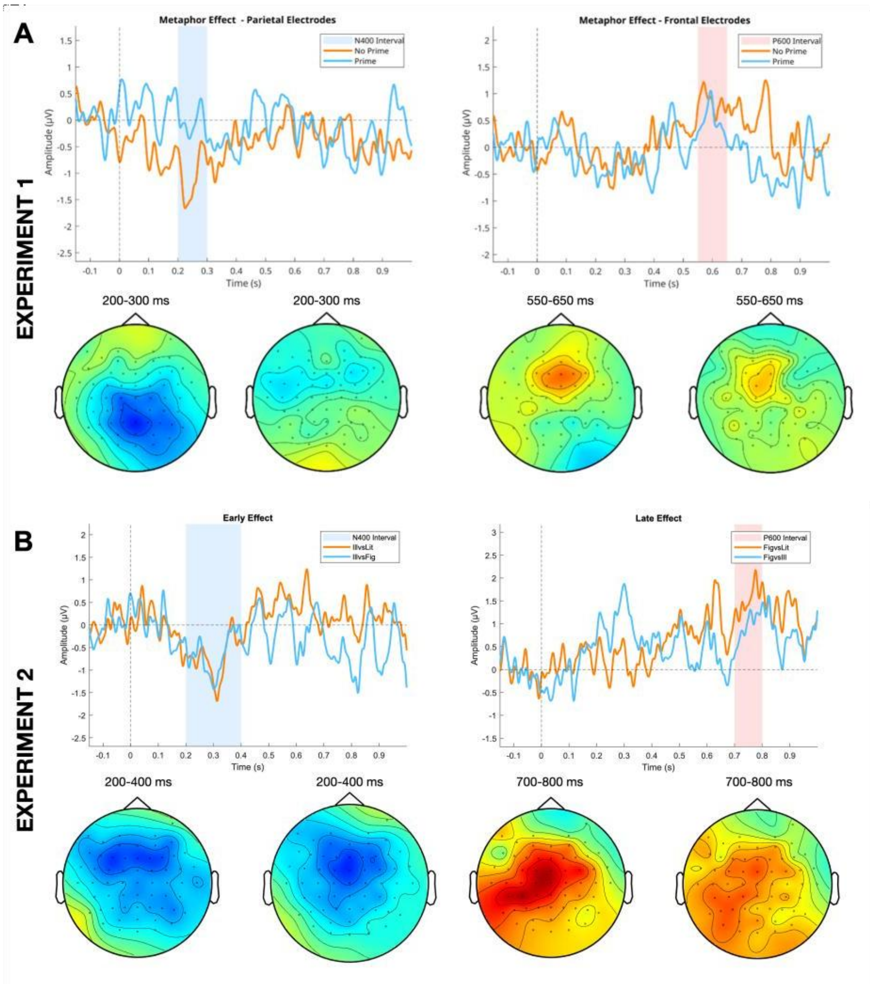


Figure 2. (A) Difference waves (Metaphor-Literal) showing the N400 (on the left) and P600 (on the right) effects across priming conditions (Prime and No-Prime) for Experiment 1. (B) N400 (on the left) and P600 (on the right) effects in the difference waves across prime types (Illegal–Literal and Illegal–Figurative) in Experiment 2.

Two Sites, Two Languages: tDCS and EEG Evidence for Argument-Structure and L1 Feature Transfer

Alaa M. Salem[1], Daniel Gallagher[2], Emi Yamada[2], Shinri Ohta[2]

[1]Kyushu Univ., Grad. Sch. Lit., Dept. Ling., [2]Kyushu Univ., Fac. Hum., Dept. Ling.

Contact: abouzeid.alaa.495@s.kyushu-u.ac.jp

Understanding how the brain processes language involves analyzing the complex intereaction between semantic and syntactic components. Recent research suggests that event-related potentials (ERPs), such as the N400 and P600 components might not be completely distinct, indicating a more integrated processing system [1]. By modulating cortical activation, high definition transcranial direct current stimulation (HD-tDCS) has demonstrated the role of the left inferior frontal gyrus (LIFG) [2] and left superior temporal gyrus (LSTG) [3] in language processing. In this study, we examined the effects of simultaneous dual-site anodal HD-tDCS applied to the LIFG and LSTG on second language (L2) processing. By combining HD-tDCS and electroencephalography (EEG), we investigated L1-specific modulations in the L2 P600 and N400.

L1 Chinese (15) and Korean speakers (12) evaluated Japanese sentence correctness as L2 learners. Each sentence included three phrases: two noun phrases followed by a verb, where violations occurred in the final verb phrase. Japanese and Korean share morpho-syntactic features, while Japanese and Chinese share Chinese characters with semantic radicals. We compared syntactically and semantically violated sentences with correct sentences to examine the impact of cross-linguistic variations in L2 processing. Participants were randomly assigned to receive either dual-site anodal or sham tDCS, and further analysis employed EEG signal processing.

Linear mixed-effects models showed that active dual-site tDCS, relative to the sham condition, selectively modulated semantic and syntactic ERP components based on L1 typology. In the 250–500 ms time window, active tDCS amplified N400 responses in Korean participants exhibiting larger N400s to both semantic-intransitive verbs (p = .036) and semantic-transitive verbs (p = .0047) violations than Chinese participants. In the 600–900 ms window, active tDCS reduced P600 to syntactic violations overall (estimate = -4.15×10^{-7} V, p = .011), with significant stimulation×L1 interactions ($F \approx 9.5$, $p < .003$). Under dual site tDCS, Korean participants exhibited stronger β-band synchrony in the LIFG and LSTG during the N400 component for both syntactic and semantic violations than Chinese participants $F(1, 46) = 5.50$, $p = 0.02$). This may reflect predictive top-down processing based on prior linguistic knowledge in Korean participants, whereas Chinese participants may rely more on bottom-up, piece-by-piece construction during language processing. Behavioral data revealed that Chinese speakers excelled at detecting semantic anomalies, while Korean speakers outperformed in syntactic anomaly detection, further supporting the evidence of L1 feature transfer. Based on behavioral and EEG data, we conclude that the dual-site tDCS over the core language network, i.e., the LIFG and LSTG, enhances L2 processing through the L1 feature transfer effect.

Condition	Example	Description
Correct Sentences	**子供が 虫を 見つける** *kodomo-ga musi-o mitukeru$_{transitive}$* A child finds an insect.	**No violation**
Syntactic Violated Sentences	***子供が 虫を 見つかる** *kodomo-ga musi-o mitukaru$_{intransitive}$* A child finds an insect.	**Switching the verbs: Transitive to intransitive & vice versa** > keeping argument structure/case markers
Semantic Violated Sentences	**#子供が 虫を 閉める** *kodomo-ga musi-o simeru* **A child closes an insect.**	**Changing the verbs to semantically implausible ones:** No shuffling of in/transitive verbs

Fig. 1. Stimuli and their representation.

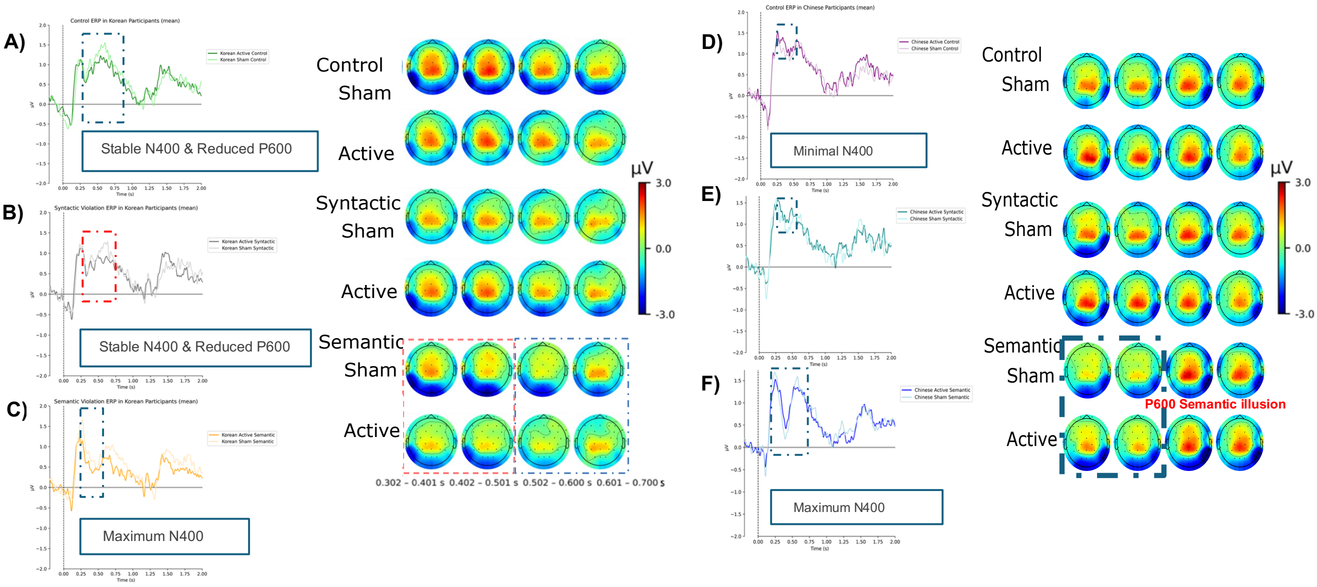


Fig. 2. Average ERPs and topographic distribution of: A) control condition in Korean participants. B) syntactic condition in Korean participants. C) semantic condition in Korean participants. D) control condition in Chinese participants. E) syntactic condition in Chinese participants. F) semantic condition in Chinese participants.

References

[1] Aurnhammer, C., Crocker, M.W., & Brouwer, H. (2024). Single-trial neurodynamics reveal N400 and P600 coupling in language comprehension. Cognitive Neurodynamics, 18, 3309-3325.

[2] Holland, R., Leff, A.P., Josephs, O., Galea, J.M., Desikan, M., Price, C.J., Rothwell, J.C., & Crinion, J. (2011). Speech facilitation by left inferior frontal cortex stimulation. Current Biology, 21(16), 1403-1407.

[3] Flöel, A., Rösser, N., Michka, O., Knecht, S., & Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. Journal of Cognitive Neuroscience, 20(8), 1415-1422.

The acquisition of noun-verb stress alternation by Bengali learners of English

Charles Redmon[1] and Aditi Lahiri[2]
[1]University of Essex, [2]University of Oxford

Word stress in English, as in other Germanic languages, is governed in part by word class, with stress predominantly initial on nouns and non-initial on verbs and adjectives/adverbs. However, as a consequence of extensive borrowing of pseudo-prefixed Latinate verbs from French, as well as the loss of the majority of Old English inflectional morphology, among other factors, English notably now has a large number of noun-verb pairs where stress is the only determinant of word class [cf. 4-6, among others]. This loss of most inflectional morphology has also led to many cases of conversion in English, where verbs may be derived from nouns and retain initial stress and no affixal marking, and vice versa for nouns derived with non-initial stress from verbs. Given this system, we wanted to understand how speakers of a language with no comparable prosodic morphophonological alternations, Bengali, acquire this system. This work follows up on several previous investigations into the acquisition of English stress alternation [1-3], however it differs in focusing on lexical representation and word recognition in general, rather than explicitly testing learners' ability to appropriately locate and produce stress on such items.

We conducted a cross-modal priming lexical decision task where the auditory prime was an *-ing*-suffixed verb with the base varying in its noun-verb prosodic structure when uninflected (stress-alternating, non-alternating with initial stress, non-alternating with non-initial stress, and verb-only), and the visual target was the base of that verb (See Table 1 for examples). On each trial, participants heard a word over headphones and at word offset saw the visual target and made a lexical decision. The L2 group in this study consisted of 96 Bengali-L1 learners of English tested in Kolkata, India (48 at Jadavpur University; 48 at Salt Lake Secondary School). The L1 control group consisted of 32 native British English speakers tested in the UK. All experiments were run in-person utilizing the same equipment and procedure.

Figure 1 shows the results by accuracy and reaction time for each language group. All groups show significant priming effects for accuracy and reaction time in each condition. Further, all groups showed sensitivity to the word-class complexity of the visual target by exhibiting the greatest priming effects for Condition 4 (Verb-Only; e.g., desígning → desígn), both in terms of accuracy and reaction time. For the test of stress alternation, we focus on the contrast between Conditions 1 and 3 (Stress-Alternation vs. Verb-Basic Conversion), as both primes have the same stress pattern and morphological structure but differ in their prosodic representations of the target. Here there is significantly greater priming in the stress-alternating pair than in the non-alternating pair, a result which is due to the fact that the baseline response to non-alternating targets (when preceded by a control word) is significantly faster than that to alternating targets. Notably, all groups showed this effect, including the secondary-school-level L2 learners.

In fact, the only significant differences between the groups were an overall lower accuracy and slower reaction time among learners, and priming effects which were significantly different between the English L1 group and the Secondary L2 group in Conditions 1 and 3 (verb-base, non-initial stress), but not between L1 and University L2, though both groups showed significantly reduced priming in Conditions 2 and 4 (initial stress), where there are notable conflicts between Bengali and English verbal stress patterns. This result shows that learners can indeed acquire aspects of L2 prosodic morphophonology at the level of representations and general lexical access behaviour, though interference from the L1 may remain for some aspects of the system.

1. N~V Stress Alt.	2. Conversion (N-base)	3. Conversion (V-base)	4. Verb-only
constrúcting → cónstruct / constrúct	píloting → pílot (noun and verb)	desígning → desígn (noun and verb)	fóllowing → fóllow (no noun)

Table 1. Sample prime-target pairs for Noun-Verb Stress Alternating, Noun-Base Conversion, Verb-Base Conversion, and Verb-Only conditions.

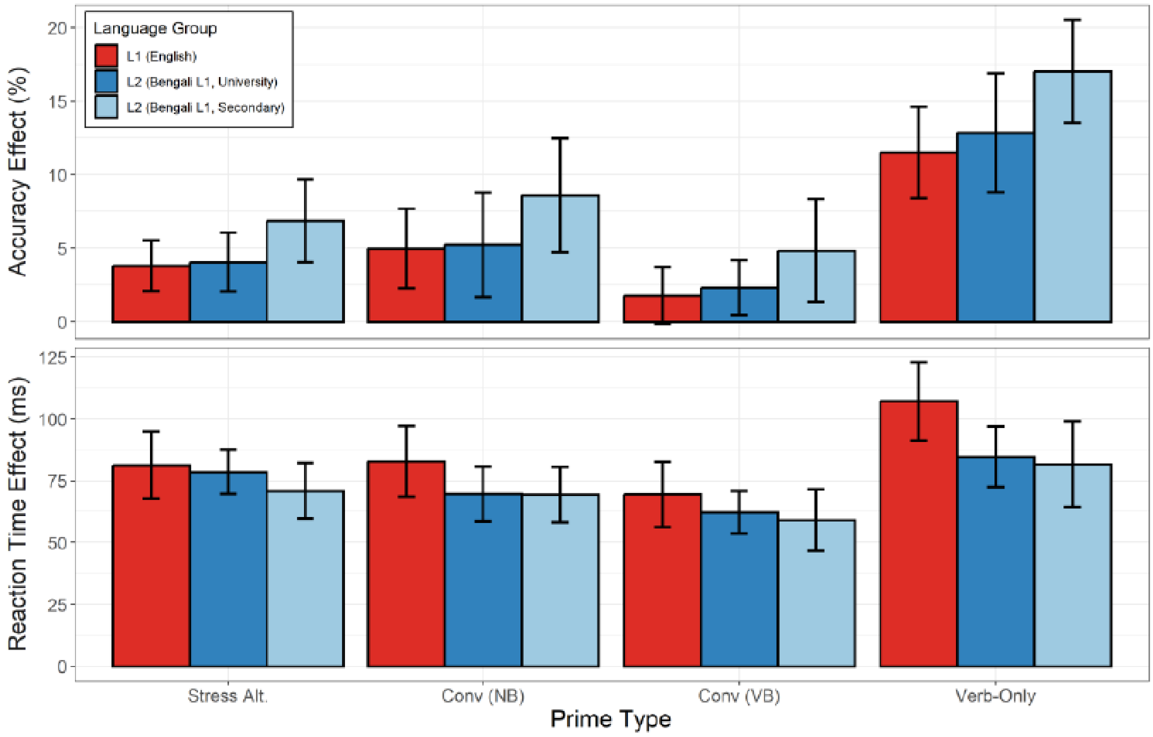


Figure 1. Priming effects for lexical decision Accuracy (control–prime) and Reaction Time (prime–control) on Stress Alternating (Stress Alt.; *constrúcting → cónstruct/constrúct*), Noun-Base Conversion (Conv [NB]; *píloting → pílot*), Verb-Base Conversion (Conv [VB]; *desígning → desígn*) and Verb-Only (*fóllowing → fóllow*) items by native speakers of British English (red), Bengali-L1 Learners of English studying at University (dark blue), and Bengali-L1 Learners of English studying at Secondary School (light blue).

References

1. Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics, 35*(1), 167-182.
2. Davis, S. M., & Kelly, M. H. (1997). Knowledge of the English noun–verb stress difference by native and nonnative speakers. *Journal of Memory and Language, 36*(3), 445-460.
3. Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition, 27*(4), 503-533.
4. Kelly, M. H. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition, 30*(2), 107-137.
5. Lahiri, A. (2015). Change in word prosody: Stress and quantity. In Patrick Honeybone & Joseph Salmons (eds.), *The Handbook of Historical Phonology* (pp. 219–244). Oxford: OUP.
6. Sherman, D. (1975). Noun-verb stress alternation: An example of the lexical diffusion of sound change in English. *Linguistics, 13*(159), 43-72.

Further useful information on Bengali

Bengali is an Indic language (Indo-European) spoken in India and Bangladesh, and relevant to the present study here are some key attributes of Bengali prosody.

Stress in Bengali is predominantly initial, and is not used lexically to signal word class.

Bengali further does not have morphological conversion (zero-derivation) processes as in English, making Bengali a good test case for a language does not have a comparable prosodic morphophonological system but which nevertheless matches English in its general trochaic system.

Finally, Bengali is generally described as being syllable-timed in its rhythm, and thus does not exhibit the kind of vowel reduction on unstressed syllables that is commonly found in English. It further does not have a vowel length distinction so English variation in length as a function of tensity or prominence is generally unfamiliar to Bengali learners.

L2 English speakers exhibit native-like garden path difficulty across constructions

Runchen Liu & Suhas Arehalli
Macalester College
{rliu, sarehall}@macalester.edu

While it has been shown that Native and Non-Native (L2) speakers exhibit differences in syntactic processing, why those differences emerge is unclear. Two broad classes of explanations have been proposed: The Shallow Structure Hypothesis (SSH) [1, etc.] proposes that these processing differences are driven by the fact that L2 speakers build reduced, "shallow" representations relative to native speakers. An alternative group of accounts, which we will refer to as resource-constraint accounts, instead attribute differences in processing to the availability of some cognitive resource — memory, cognitive control, or otherwise — that is taxed during L2 processing [2, 3]. We investigate the implications of these two accounts on the L2 processing of different kinds of *garden-path sentences* (GPs) — sentences that disambiguate a temporary syntactic ambiguity toward the dispreferred structure, leading to a slowdown in processing. Many accounts of GP processing propose that the magnitude of their difficulty reflects structural properties of the sentence [4, etc.], a claim supported by the existence of cross-construction differences in GP effect magnitude [5, 6]. Under this interpretation of GP processing and the SSH, we hypothesize that less rich syntactic representations in L2 speakers would result in smaller effects of structure, which in turn would result in more uniform difficulty across GP types in L2 speakers vs. native speakers. Conversely, under a resource-constraint account, we expect to see similar syntactic representations in native and L2 speakers, which would result in native and L2 speakers having similar patterns of difficulty across the different GP constructions.

Methods & Analysis: We recruited 119 native Mandarin and Cantonese speakers through Prolific who self-identified as L2 English speakers for an experiment using PCIbex [7]. Proficiency was measured with a lexical decision task based on the LexTALE task [8], where participants had a mean accuracy of 86.29%. We additionally filtered participants for an English Age-of-Acquisition ≥ 5 ($n = 75$) and found that our relative difficulty and English/L2 results were robust. We conducted a self-paced reading task with materials from [6], a 2x3 design crossing ambiguity with three GP construction types, each based on a different structural ambiguity (MVRR, NPS, NPZ; See Table 1). We report analyses at the first spillover position, where GP effects were largest. Analyses we report use linear mixed-effects models analyzing raw RTs, including factors of GP construction, ambiguity, and native/L2 status when relevant. Comparisons with native speakers include data from [6].

Results: We saw GP effects for all constructions in L2 speakers (MVRR: $\beta = 195.2, p < 0.001$, NPS: $\beta = 44.71$, $p < 0.05$, NPZ: $\beta = 132.54$, $p < 0.001$), with significant differences between NPS and MVRR ($\beta = 153.4$, $p < 0.001$), NPS and NPZ ($\beta = 86.2$, $p < 0.05$), but not NPZ and MVRR ($\beta = 67.2$, $p = 0.053$), suggesting the difficulty of the 3 constructions in L2 speakers is MVRR $\geq$ NPZ $>$ NPS, consistent with native speakers. An analysis of both native and L2 data together found no significant differences in GP effects between native and L2 speakers (MVRR: $\beta = 7.45$, $p = 0.79$, NPS: $\beta = 20.28$, $p = 0.29$, NPZ: $\beta = 22.86$, $p = 0.39$).

Conclusions: We found no evidence for differences in GP magnitude between native and L2 speakers within each GP construction, or in relative difficulty across different constructions. These findings suggest that L2 syntactic processing, at least as reflected by GP processing, is not substantially more shallow than that of native speakers.

GP Type	Stimulus
MVRR	The suspect *(who was)* sent the file deserved further investigation given the new evidence.
NPS	The suspect showed *(that)* the file deserved further investigation during the murder trial.
NPZ	Because the suspect changed*(,)* the file deserved further investigation during the jury discussions.

Table 1: Sample critical stimuli, taken from [6]. Unambiguous stimuli contain material in parentheses. The critical region (the disambiguating word followed by a two words to account for spillover) is underlined.

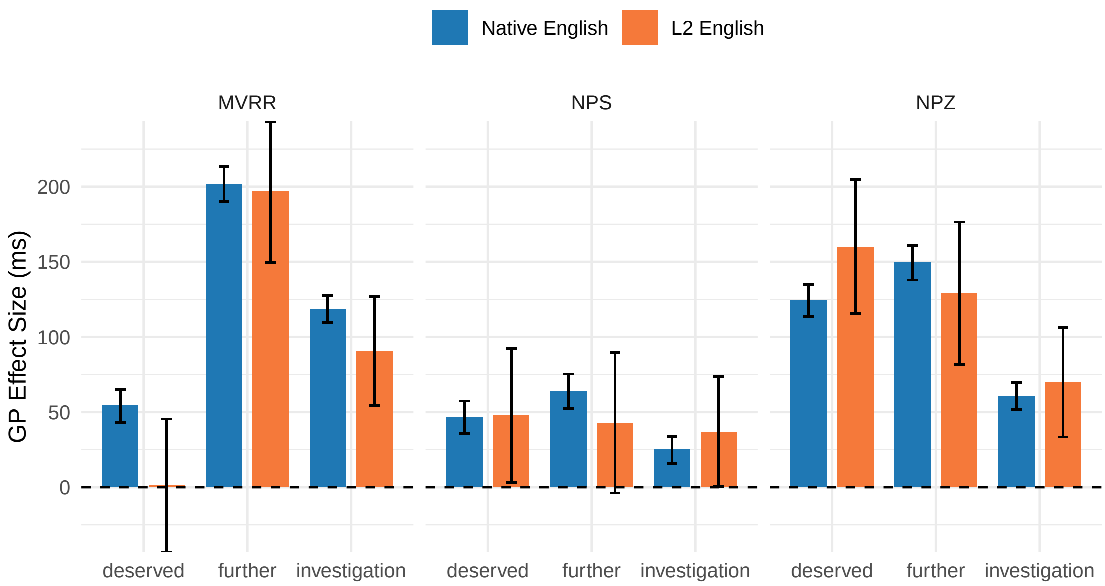


Figure 1: Estimated magnitude of garden path effects. Error bars are 95% confidence intervals.

[1] Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*.

[2] McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *JML*.

[3] Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: a visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636–643.

[4] Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210.

[5] Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *JML*, *40*(1), 136–150.

[6] Huang, K.-J., Arehalli, S., Kugemoto, M., Muxica, C., Prasad, G., Dillon, B., & Linzen, T. (2024). Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. *JML*.

[7] Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).

[8] Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343.

Lexical stress precedes syllable structure during speech planning – evidence from EEG multivariate pattern analysis

Giulia Li Calzi[1,2], Antje S. Meyer[1,3], Constantijn L. van der Burght[1,4]
[1] Max Planck Institute for Psycholinguistics; [2] Centre for the Interdisciplinary Study of Language Evolution, University of Zurich; [3] Donders Institute for Brain, Cognition, and Behaviour, Radboud University; [4] Leiden University Centre for Linguistics, Leiden University
Stan.vanderBurght@mpi.nl

To produce a word, speakers need to decide which concept to express, select an appropriate item from the mental lexicon, and spell out its phonological form. The temporal dynamics of these processes remain a subject of debate. Concerning the phonological phase, (psycho)linguistic theories have long argued whether a word's stress pattern is retrieved independently, or whether this information is derived from a word's syllable structure. To address this issue, we investigated the time course of lexical access in picture naming with electroencephalography (EEG). Native speakers of Dutch (N=30; 23 female; age 23.4 ± 4.2 years) named 48 pictures using simple nouns. The picture names varied in semantic category (animate or inanimate), stress pattern (first or second syllable), and the structure of the first syllable (open or closed). We analysed the EEG data using time-resolved multivariate pattern analysis (MVPA), which quantifies the extent and time course in which neural representations support the discrimination of relevant stimuli dimensions. Using time-resolved MVPA, we decoded the availability of semantic category, stress pattern, and syllable structure during speech preparation. An additional, exploratory analysis was performed on the availability of the word-initial phoneme (plosive vs. fricative). The results demonstrated above-chance decoding of semantic category within 100 ms after picture onset, confirming early access to semantic information. This was followed by stress pattern and syllable structure, at around 150 and 250 ms after picture onset, respectively. These results suggest that a word's stress pattern can be retrieved before syllable structure information becomes available. The exploratory analysis demonstrated the availability of the word-initial phoneme within 100 ms after picture onset. This result, while not part of the a-priori analysis plan, tentatively supports the view that during picture naming semantic and phonological information are accessed rapidly and in parallel. In sum, our results offer novel insights for theories of language production, suggesting that the retrieval of lexical stress is independent from syllable structure. Importantly, this study demonstrates the successful application of MVPA on pre-articulation data from EEG, a widely available method, offering an accessible approach to address novel questions in speech production research.

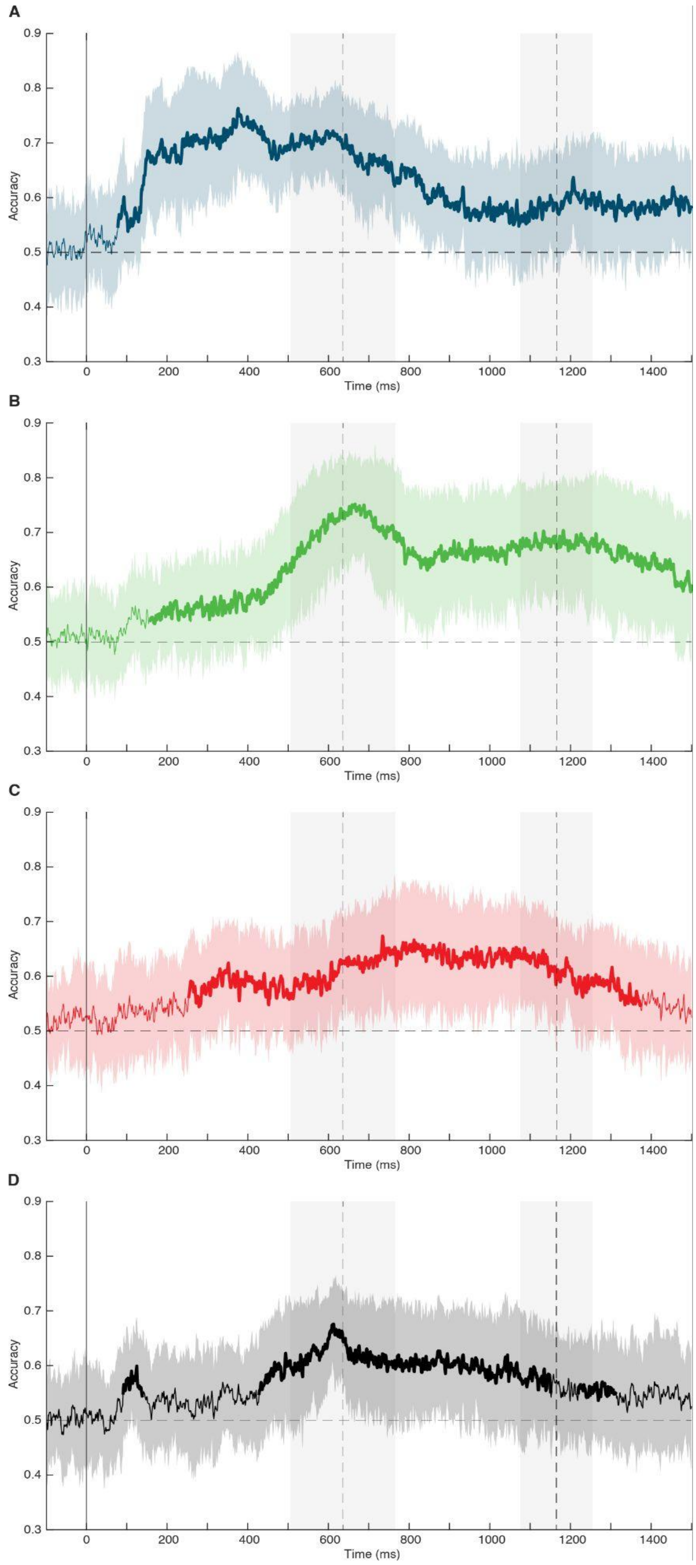


Figure 1: **The time course of phonological encoding of single words, depicting decoding accuracy for the experimental variables (A) animacy, (B) lexical stress, (C) syllable structure, and (D) word-initial phoneme.** Lines indicate mean classification accuracy across participants, with bold marking above-chance level decoding accuracy, and the shaded areas correspond to the standard error. As the bold marking shows, significant decoding of animacy and word-initial phoneme started around 100, followed by stress pattern and syllable structure at approx. 150 and 200 ms, respectively. The dashed horizontal line indicates chance-level decoding accuracy. Vertical lines indicate picture onset. Dashed vertical lines indicate the average onset and offset of speech, while shaded areas represent ±1 standard deviations around these onset and offset times.

A grammatical animacy agreement feature: evidence from processing in Polish

Zuzanna Fuchs & Anna Runova
University of Southern California
zfuchs@usc.edu

Background: In Polish, linguistic elements (ex. adjectives) agree with a head noun in grammatical gender (masc., fem., neut.), number, and case. Agreement with masc. nouns makes a further distinction in animacy: in the accusative, an adjective agreeing with a masculine animate (M.A) noun takes the suffix -*ego* (ex. *Kup zielonego żółwia*. 'Buy (the) green turtle$_{M.A}$.') while an adjective agreeing with a masculine inanimate (M.I) noun takes the suffix -*y* (ex. *Kup zielony długopis*. 'Buy (the) green pen$_{M.I}$.') [1]. Prior work has shown that processing -*ego* activates both masc. gender and semantic animacy [2]. However, this work did not address a special class of masc. nouns (M.X) in Polish that are semantically inanimate but take animate agreement marking: agreeing adjectives take the -*ego* suffix (ex. *Kup zielonego lizaka*. 'Buy (the) green lollipop$_{M.X}$.'). This raises the question of whether, in addition to semantic animacy, Polish agreement marking tracks abstract grammatical animacy analogous to grammatical gender. We test this by investigating facilitative processing of gender and animacy categorization features in Polish [2,3].
Method: Native Polish speakers (n=47) living in Warsaw, Poland completed a Visual World eye-tracking task. The distractor always had the opposite gender and animacy as the target; gender and animacy of the competitor were manipulated as in Table 1. We implemented the Covered Box Paradigm [4] [Figure 1] to ensure that gazes to the target in critical conditions were not over-inflated by non-linguistic preferences for looking at animate objects [5]. The study was fully counter-balanced.
Predictions: If -*ego* activates semantic animacy, we expect looks to M.A and M.X targets to be reduced relative to the different condition in the presence of a feminine animate (F.A) competitor and to be reduced for M.X targets in the presence of an M.A competitor. If -*ego* activates grammatical animacy, we expect looks to M.A targets to be reduced by M.X competitors over and above any reduction for M.I competitors.
Results: Cluster-based permutation analyses [6] [Figure 2] for looks to M.A targets revealed significant clusters for the contrasts between different (F.I comp.) and (a) same-animacy (F.A comp.) (2750-2850ms, p=0.02) and (b) same-abstract (M.X comp.) (2650-2950ms, p<0.001). An additional cluster was found for the contrast between same-gender (M.I comp.) and same-abstract (M.X comp). For M.X targets, clusters were found for the contrast between different (F.I comp.) and (a) same-abstract (M.A comp.) (2550-2950ms, p<0.001) but not (b) same-animacy (F.A comp.). Notably, there were less looks to M.X targets with M.A competitors than in the same condition (M.X comp.) (2500-2850ms, p<0.001).
Discussion: We find evidence that -*ego* activates semantic animacy, consistent with prior work [2]. We also find that -*ego* activates grammatical animacy (reduced looks to M.A targets by M.X competitors). This is novel evidence that agreement marking in animacy activates both a semantic feature and an abstract feature grammaticalized from it, though participants attend more to the semantic feature.

Figure 1 VWP sample trial; box disappears 200ms after onset of noun; targets were covered in half of trials, distractors in other half.

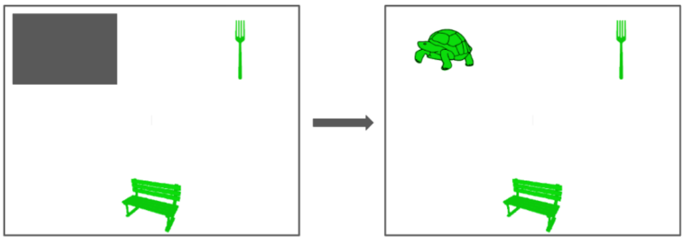


Table 1 Conditions for MA & MX targets; counter-balanced conditions not shown.

Condition	Targ.	Comp.	Dist.		Targ.	Comp.	Dist.
different	**M.A**	F.I	F.I		**M.X**	F.I	F.I
same-animacy	**M.A**	F.A	F.I		**M.X**	F.A	F.I
same-gender	**M.A**	M.I	F.I		**M.X**	M.I	F.I
same-abstract	**M.A**	M.X	F.I		**M.X**	M.A	F.I
same	**M.A**	M.A	F.I		**M.X**	M.X	F.I

Figure 2a Looks to M.A targets. Horizontal bars indicate significant clusters identified by CPA.

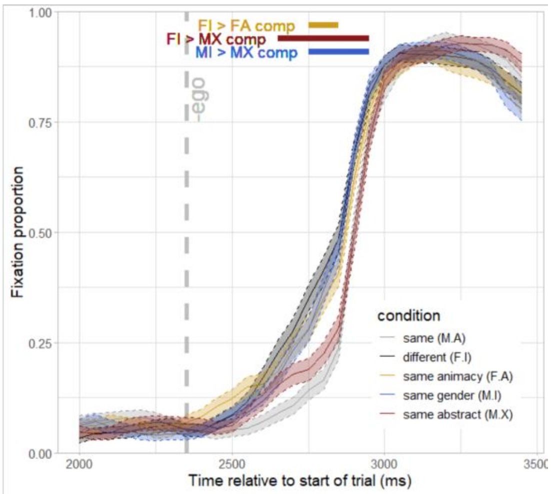
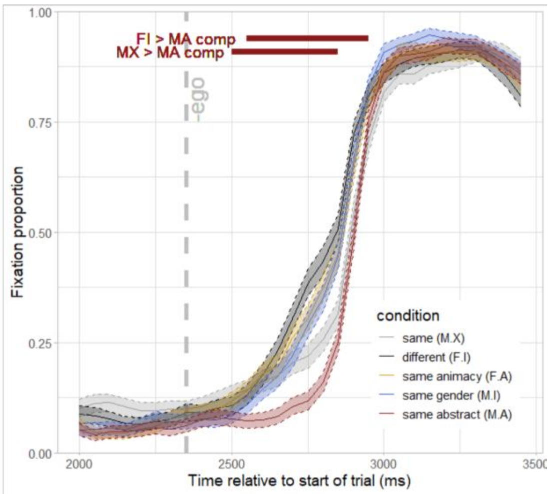


Figure 2b Looks M.X targets. Horizontal bars indicate significant clusters identified by CPA.



References

[1] Corbett, G. (1983). The Number of Genders in Polish. *Papers and Studies in Contrastive Linguistics*, 16, 83–89.

[2] Fuchs, Z. (2025). Independent processing of lexical agreement features: Evidence from Polish. *Language, Cognition & Neuroscience*, 1–15. https://doi.org/10.1080/23273798.2025.2489611

[3] Fuchs, Z. (2023). Processing of grammatical gender agreement morphemes in Polish: evidence from the Visual World Paradigm. *Morphology*, 34. 1-32.

[4] Schwarz, F., Bill, C., & Romoli, J. (2019). Reluctant Acceptance of the Literal Truth: Eye Tracking in the Covered Box Paradigm. *Proceedings of Sinn Und Bedeutung*, 20, 61–78.

[5] New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *PNAS*, 104(42), 16598–16603.

[6] Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.

Languages of Power: Metaphorical Grounding of Perceived Power in the Bilingual Mind

Yufen Wei and Guillaume Thierry

School of Psychology and Sport Science, Bangor University, UK

yfw21bgp@bangor.ac.uk

Abstract

Embodied cognition theories propose that abstract concepts are grounded in sensorimotor experiences (Barsalou, 2008), metaphorical mappings (Lakoff & Johnson, 1999), and social constructs (Pexman et al., 2023; Borghi et al., 2019). The grounding of perceived power, a culturally salient abstract concept, across languages remains underexplored. Our study examined the neural and behavioural manifestations of perceived power embodiment in Chinese-English bilinguals across three experiments, investigating how power-related concepts are metaphorically associated with vertical space and physical size, and how such grounding varies between first (L1, Chinese) and second (L2, English) languages.

Experiment 1 employed an auditory source localisation task with event-related potentials (ERPs) when participants reported whether power words (e.g., 'king', 'servant') and location words (e.g., 'sun', 'ground') originated from above or below their sitting position. High-power words produced faster responses when presented from above (congruent) versus below (incongruent), with no congruency effect for low-power words. Congruent high-power words elicited enhanced P3 and reduced N400 amplitudes in L1 only (Fig.1). English controls and location words showed no embodiment effects.

Experiment 2 investigated power-size association using a font-size judgement task. High-power words in bold-larger font prompted faster and more accurate responses than those in thin-smaller font, with stronger effects in L1 than L2 (Fig. 2). English controls showed robust congruency effects, with this effect twice as large when compared with bilinguals operated in their L2 English. Low-power words showed no congruency effects across groups.

Experiment 3 examined the neurophysiological basis of power-size embodiment, comparing high-power words to matched control words unrelated to power (e.g., 'daylight'). Larger fonts elicited enhanced P3 and late positive component (LPC) amplitudes for both word types, but font-size effects were significantly stronger for power than control words (Fig. 3). Unlike previous experiments, congruency effects were comparable between L1 and L2, with L2 proficiency correlating positively with effect magnitude across all late ERP components.

Differential ERP patterns across experiments reflect distinct processing mechanisms: N400 modulation in Experiment 1 indicated semantic integration difficulty when power-verticality metaphors were violated, while P3/LPC modulation in Experiment 3 reflected enhanced attention for metaphorically congruent power-size mappings. Perceived power embodiment varies systematically across metaphorical domains and languages, with physical size representing a more robust and universal grounding mechanism than vertical space. The asymmetric pattern between high- and low-power words supports a self-relevance hypothesis we proposed whereby concepts with greater personal relevance engage embodied representations more strongly than concepts with ambiguous self-relevance, extending multiple representation theories by incorporating individual relevance as a key modulator of abstract concept grounding.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, *29*, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Pexman, P. M., Diveica, V., & Binney, R. J. (2023). Social semantics: The organization and grounding of abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 378*(1870), 20210363. https://doi.org/10.1098/rstb.2021.0363

Figures

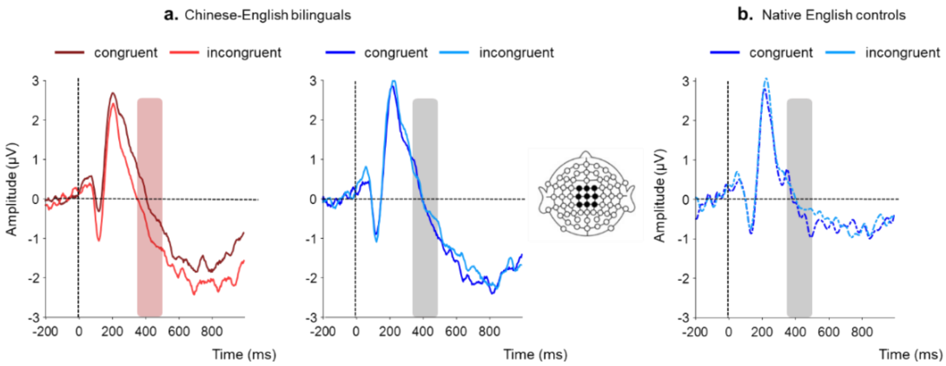


Fig. 1 (a) Language × Congruency interaction on N400 mean amplitudes on power words only in Chinese-English bilinguals; (b) ERPs elicited in native English controls by congruent and incongruent power words for visual comparison.

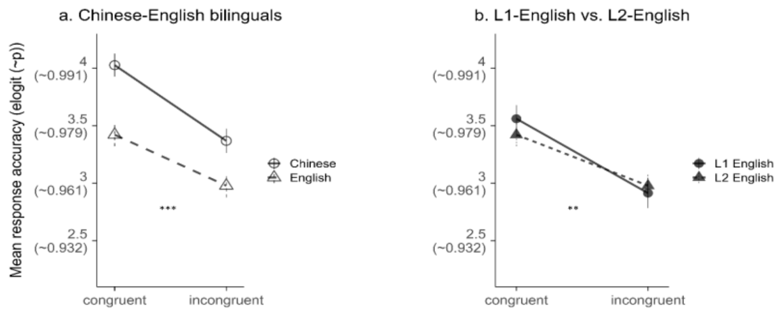


Fig. 2 Mean accuracy by Congruency condition for a) Chinese-English bilinguals in each language and for b) the group comparison of bilinguals in English (L2) and native English controls in English (L1, right panel).

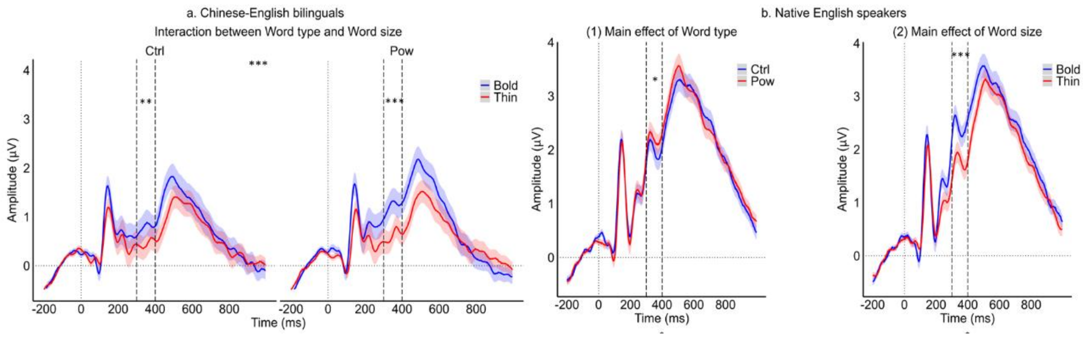


Fig. 3 Event-related potentials elicited by words at centroparietal electrodes (a) Word type × Word size interaction on P3 amplitude in Chinese-English bilinguals; (b) Main effects of Word type and Word size in the control group.

Resumptive pronouns are grammatical in English

Yourdanis Sedarous and Savithry Namboodiripad
University of Michigan
sedarous@umich.edu | savithry@umich.edu

BACKGROUND: Resumptive pronouns (RPs) refer to overt elements that are obligatorily bound by a co-indexed element in a structurally higher A-bar position, as schematized here: $[_{CP}\ XP_i\ [_{C'} \ldots RP_i \ldots]]$. Their occurrence is often attributed to one of three factors: base generation, syntactic movement, or so-called *intrusion* [1]. In this typology, both base-generated and movement RPs are regarded as *grammatical* elements [2], while "intrusive" RPs are labeled as a ungrammatical production epiphenomenon, and analyzed as an artifact of parsing and production, rather than as a grammatical element [3, 4]. In this paper we argue that "intrusive" RPs are actually grammatical elements and they do not affect a sentence's grammaticality status. We examine the interaction between RPs and syntactic satiation [5] in a language argued to exhibit "intrusive" RPs, English, and show that presence of an RP neither prevents structures that typically satiate from doing so, nor causes structures that resist satiation to become more acceptable. **METHODS:** Participants were recruited through Prolific for two different experiments, testing two island types: subject islands (N=34) and that-trace islands (N=66). Experimental items were constructed on the basis of Chaves and Dery (2019) [6], and presented auditorily to participants, who rated 288 total items on a 1-7 scale. We fit cumulative link mixed models for each island type, predicting RESPONSE with fixed effects for CONDITION (GAP vs. RP), PRESENTATION ORDER, and their interaction, with random intercepts for PARTICIPANT and ITEM. **RESULTS (SUBJECT ISLANDS):** The interaction between CONDITION and PRESENTATION ORDER was non-significant (β = –0.003, p = .189), and a likelihood ratio test comparing models with and without the interaction confirmed that including the interaction did not significantly improve model fit ($\chi^2(1)$ = 1.70, p = .192). The additive model revealed a significant, positive effect of PRESENTATION ORDER (β = 0.006, p < .001), indicating a general increase in acceptability, consistent with satiation. These results suggest that subject-island violations showed similar rates of satiation, regardless of RP presence. **RESULTS (THAT-TRACE):** Once again, the interaction between CONDITION and PRESENTATION ORDER was non-significant (β = –0.0016, p = .369), and likelihood ratio tests showed no improvement in model fit ($\chi^2(1)$ = 0.80, p = .372). Additionally, removing CONDITION ($\chi^2(1)$ = 0.80, p = .370) or PRESENTATION ORDER ($\chi^2(1)$ = 0.03, p = .871) did not significantly reduce fit, suggesting that neither predictor reliably explained variance in acceptability ratings. These results suggest that regardless of RP presence, That-Trace violations showed no evidence of satiation across the course of the experiment. **CONCLUSION:** Our results indicate that the presence of an "intrusive" RP does not affect the overall acceptability trajectory of a syntactic structure. Because of this, we suggest that RP preferences, whether observed within a language or across languages, should not be treated as categorical but rather as a continuous variable, reflecting variability in linguistic processing rather than strict grammatical distinctions.

Figures:

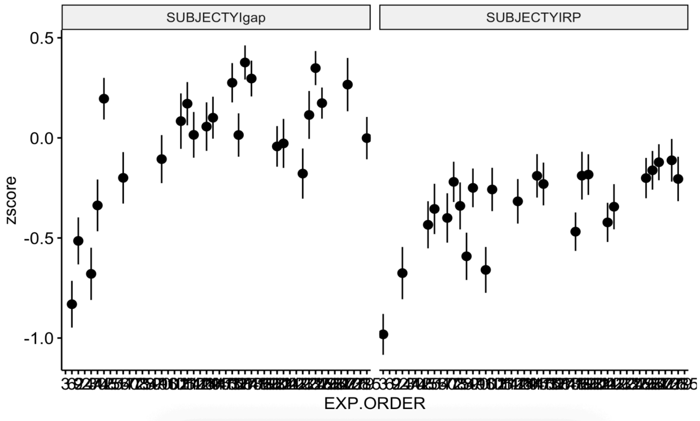


Figure 1. Ratings of SUBJECT-ISLAND extracted sentences with resumptive pronouns (right; *Which event did photos of it get international attention?*) and without (left; *Which event did photos of get international attention?*) resumptive pronouns, across the course of the experiment. Ratings transformed into by-subject z-scores for visualization purposes.

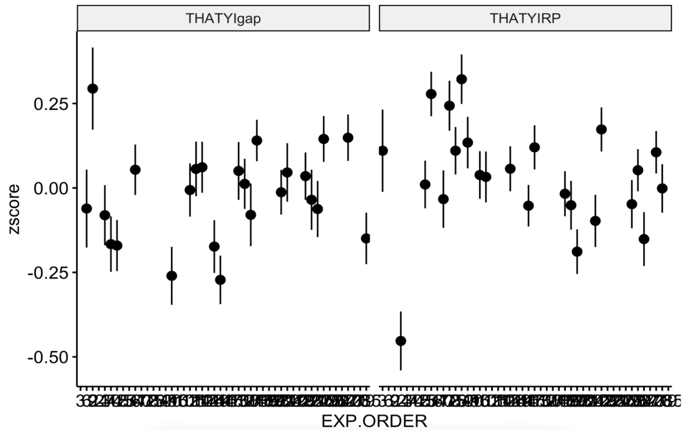


Figure 2. Ratings of THAT-TRACE extracted sentences with resumptive pronouns (right; *Which mistake did the senator fear that it would ruin the campaign?*) and without (left; *Which mistake did the senator fear that would ruin the campaign?*), across the course of the experiment. Ratings transformed into by-subject z-scores for visualization purposes.

Selected references:

[1] Rouveret, A. (2011). Some issues in the theory of resumption: A perspective on early and recent research. In A. Rouveret (Ed.), *Resumptive pronouns at the interfaces* (pp. 1–62). John Benjamins Publishing Company. https://doi.org/10.1075/lfab.5.01rou

[2] McCloskey, J. (2006). Resumption. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 94–117). Blackwell Publishing.

[3] Asudeh, A. (2011). Towards a unified theory of resumption. In A. Rouveret (Ed.), *Resumptive pronouns at the interfaces*(pp. 121–187). John Benjamins Publishing Company.

[4] Morgan, A. M., & Wagers, M. W. (2018). English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry*, *49*(4), 861-876.

[5] Snyder, W. (2022). On the nature of syntactic satiation. *Languages, 7*(1), 38. https://doi.org/10.3390/languages7010038

[6] Chaves, R. P., & Dery, J. E. (2019). Frequency effects in subject islands. *Journal of Linguistics*, 55(3), 475–521. DOI: http://doi.org/10.1017/S0022226718000294

Dissociating sublexical and lexical masked priming effects: Morphological decomposition interacts with prime lexicality.

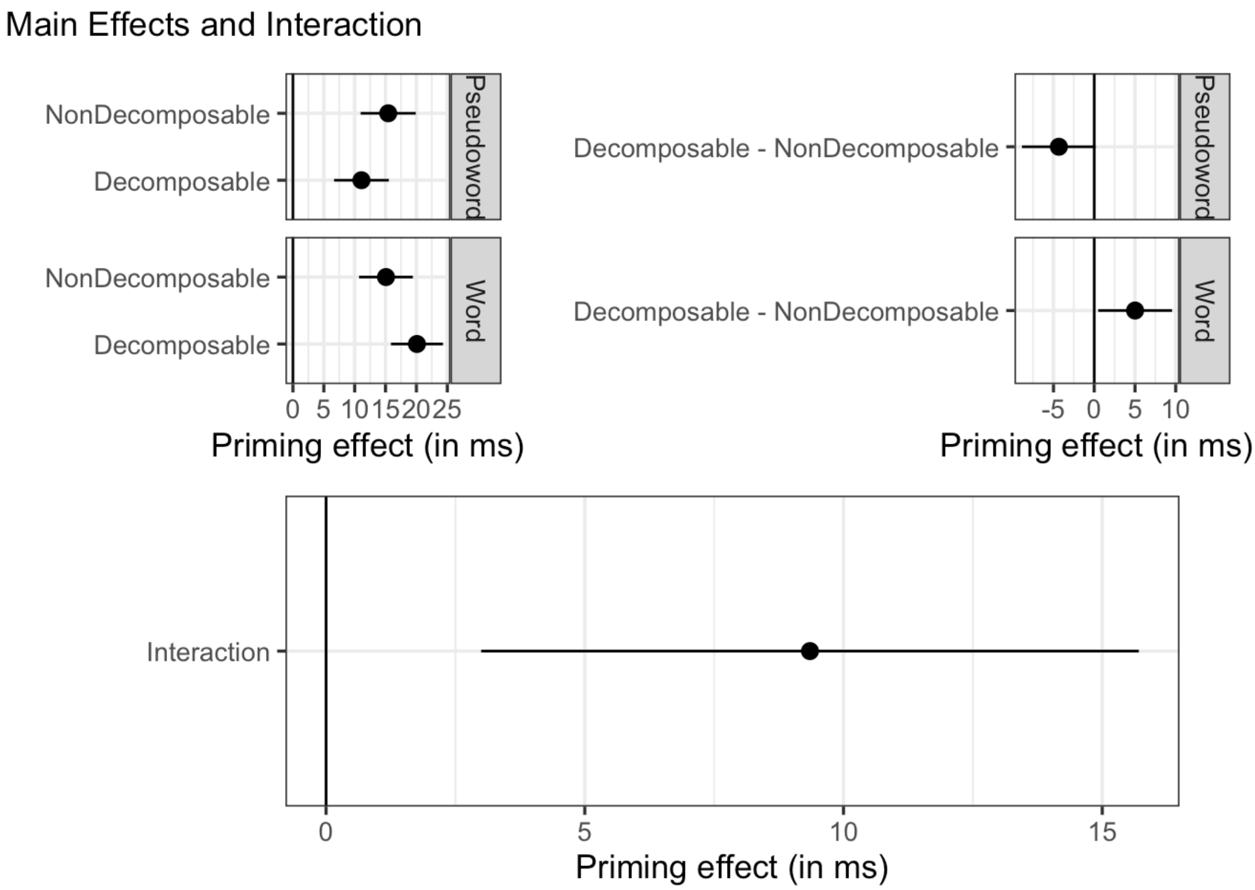
Crystal Jemy[1], Roberto Petrosino[1], Diogo Almeida[1]
[1] Psychology Program, Science Division, New York University Abu Dhabi
diogo@nyu.edu

Background: Substantial cross-linguistic masked priming evidence shows that [a] bimorphemic (***bone****less*) and [b] pseudo-affixed, monomorphemic words (***corn****er*) often trigger similar masked priming effects on their stems (*BONE* and *CORN,* respectively). In contrast, [c] monomorphemic words (e.g., ***cash****ew*) trigger negligible effects on the embedded stem (*CASH*). The dissociation between [a, b] and [c] has been taken as evidence that lexical access goes through an obligatory decomposition procedure based on a *visual* process informed by morpho-orthographic statistical regularities [1]. This procedure arguably occurs *before* accessing any lexical (morpho-syntactic, or semantic) information so that, at early stages, [b], but not [c], can be provisionally visually decomposed similarly to [a]. However, subsequent studies employing pseudoword primes have challenged this conclusion, finding that pseudowords containing embedded stems produce substantial priming regardless of their apparent morphological structure (e.g., *maskity* and *maskond* equally priming *mask*) [2-3]. This discrepancy has been explained via lexical inhibition, which may occur when primes are words, but not when they are pseudowords. Unfortunately, no study to date has fully crossed morphological structure and lexical status of the prime, and thus the apparent discrepancy observed in the literature relies on comparisons across studies, with different sets of conditions.
Method: We conducted an English masked priming experiment online (N_{total}=518). Four conditions were tested, crossing the lexical status and morphological structure of the prime: A morphologically complex, word prime condition (*appending-APPEND*), a morphologically simple word prime condition (*appendix-APPEND*), a morphologically complex, pseudoword prime condition (*appendity-APPEND*) and a morphologically simple pseudoword prime condition (*appendil-APPEND*), which were compared to an unrelated word prime condition (*bistro-APPEND*). Trials consisted of a 50ms-long prime preceded by a 500ms-long forward mask, and followed by a target, on which participants performed a lexical decision task. After excluding subjects and items with high error rates, and trials with outlying prime durations or extreme RTs, we performed a 2x2 repeated measures ANOVA with subjects as random effects [4].
Results: see Table 1 and Figure 1. Every condition showed priming effects. Crucially, morphological structure interacted with the lexical status of the prime, with the morphologically complex, word prime condition showing larger priming effects than its morphological simple counterpart ($M_{\text{complex\_word}}$=20 ms, 95% CI [16 24]; $M_{\text{simple\_word}}$=15 ms, 95% CI [11 19]), with the opposite effect for pseudoword primes ($M_{\text{complex\_pseudoword}}$=11 ms, 95% CI [7 15]; $M_{\text{simple\_pseudoword}}$=15 ms, 95% CI [11 20]).
Discussion: These results are incompatible with the early, visual morphological decomposition model. Contrary to its predictions, apparent morphological structure on strings only drives morphological decomposition when the prime is a morphologically complex word. Crucially, morphologically complex pseudowords do exhibit priming effects, but they are either smaller than or indistinguishable from purely orthographic priming effects. The results support a model in which there are two distinct processes at play in masked priming: an early prelexical priming process based on orthographic similarity, and a *lexical* morphological decomposition priming process.

Table 1.

lexicality	decompose	Mean priming	lower 95% ci	upper 95% ci
Pseudoword	Decomposable	11.1	6.7	15.5
Word	Decomposable	20.1	15.9	24.3
Pseudoword	NonDecomposable	15.4	11	19.9
Word	NonDecomposable	15.1	10.7	19.4

Figure 1.



References

[1] Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and cognitive processes*, *23*(7-8), 942-971.

[2] Morris, J., Porter, J. H., Grainger, J., & Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language and cognitive processes*, *26*(4-6), 558-599.

[3] Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded Stem Priming Effects in Prefixed and Suffixed Pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230.

[4] Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to Deal with 'the Language-as-Fixed-Effect Fallacy': Common Misconceptions and Alternative Solutions. *Journal of Memory and Language* 41 (3). Elsevier: 416–426.

Complement coercion revisited: Reassessing the psycholinguistic and the information-theoretic approach

Panagiota Rassia[1], Natalja Peiseler[1], Torgrim Solstad[1,2], Oliver Bott[1,2]
[1] CRC 1646, Project B01, Bielefeld University, [2] Department of Linguistics, Bielefeld University
panagiota.rassia@uni-bielefeld.de

Background: Coerced expressions like *Ingrid began the book* are typically analysed as requiring a type-shifting process whereby the entity-denoting complement noun (i.e., *the book*) is interpreted as an activity involving the noun in order to satisfy the matrix verb's (i.e., *began*) selectional restrictions. Classic psycholinguistic accounts on complement coercion ([1] i.a.) suggest that such expressions are cognitively taxing even when plausibility has been accounted for. Experimentally observed processing costs have been attributed to a coercion-specific mechanism that remedies the type mismatch between the verb and its complement noun. These studies compare a coerced expression to i) a preferred one which features the sentence's implicit reading (i.e., *read the book*), and ii) a non-preferred one that matches the coerced in terms of plausibility (i.e., *bought the book*). More recently, [2] approached the phenomenon from an information-theoretic angle, suggesting that any processing costs associated with the resolution of a type mismatch are in fact due to the complement noun's high surprisal. Indeed, the results of their reading eye-tracking and ERP experiments pattern with the cloze probability of a given noun preceded by either a coercion, a neutral, or a preferred verb (see Table 1). Crucially, however, this effect was limited to the regression-path times on the post-noun prepositional phrase (PP). [2] do not report acceptability norming results for their items, which raises the possibility that the increased processing cost they observe for coerced expressions may be due to acceptability differences among conditions.

Method: In an acceptability rating study (n = 60) using a slider (range = 0–100) we compared [2]'s items to a minimally reduced version where we omitted the PP following the object noun (see Table 1 for experimental design).

Results: We performed a linear mixed-effects regression analysis modelling acceptability as a function of the interacting factors VERB TYPE, NOUN FREQUENCY and CONSTRUCTION, including by-participant and by-item varying intercepts. The analysis revealed that coerced and neutral expressions received significantly lower ratings than preferred expressions (t = -3.06, p < .001 and t = -4.87, p < .001 respectively). While our results were qualitatively similar to [2]'s, sentence acceptability may still have affected processing. Our analysis further revealed a significant effect of CONSTRUCTION with items featuring a PP receiving lower ratings than those without a PP (t = 3.81, p < .001). This effect was significant in the coercion condition (t = -3.65, p < .001), which points to an acceptability-driven cost and the coercion effect reported in [2]. Figure 1 illustrates our results.

Discussion: Our findings point towards an acceptability-driven processing cost of coerced expressions instead of or in addition to [2]'s surprisal-induced effect. We are currently working on a conceptual replication of [2]'s study, where we seek to re-examine the validity of the information-theoretic explanation to complement coercion while accounting for the limitations in [2]'s study. In particular, we are developing new stimuli for a reading eye-tracking experiment that are normed for surprisal as well as for acceptability. In so doing, we tease these concepts apart. Moreover, in addition to tri-gram surprisal estimates, we will also use LLM-derived estimates (see [3]) which better capture the surprisal of the entire sentence and not just that of the verb phrase.

Condition	Probe sentence	Noun frequency		Mean cloze probability		PP
		High	Low	Hi-freq N	Lo-freq N	
coercion	Sie begann das	Buch	Exposé	.011	.001	im Urlaub
neutral	Sie kaufte das	Buch	Exposé	.013	.002	im Urlaub
preferred	Sie las das	Buch	Exposé	.18	.02	im Urlaub

Table 1: Experimental design and mean cloze probabilities as reported in [2]. The sentences in each of the conditions translate to (*coercion / neutral / preferred*): "She *began / bought / read* the book on vacation".

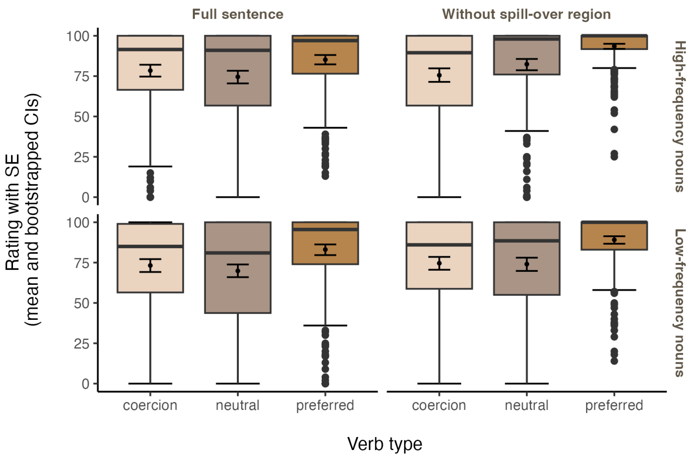


Figure 1: Acceptability ratings for the items used in [2]. The y-axis depicts the rating scale. VERB TYPE (coercion, neutral, preferred) is plotted on the x-axis. The upper half of the graph corresponds to high-frequency nouns, whereas low-frequency nouns appear on the lower half of the graph. Ratings for the full items appear on the left side of the graph, whereas ratings for items without a PP are displayed on the right side.

References

[1] McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*(1), B17–B25.

[2] Delogu, F., Crocker, M. W., & Drenhaus, H. (2017). Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs. *Cognition*, *161*, 46–59.

[3] Oh, B.-D., & Schuler, W. (2023). Transformer-Based Language Model Surprisal Predicts Human Reading Times Best with About Two Billion Training Tokens. *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Incremental processing of context during metaphor interpretation in preschool children: Evidence from a visual world eye-tracking study

Işın Tekin, Duygu Özge Sarısoy
Middle East Technical University
isin.tekin@metu.edu.tr, duyguo@metu.edu.tr

Background: Previous research shows that young children can interpret metaphorical expressions [1], however, less is known about time-course mechanisms underlying metaphor interpretation and how these mechanisms are modulated. Investigating this gap is especially important because the main models of metaphor processing diverge in their accounts of how processing unfolds. The indirect model proposes that metaphors are interpreted literally first, and after a reanalysis metaphorical meaning can be accessed [2, 3]. Direct access model suggests that metaphorical meaning is reached effortlessly [4]. In addition to these accounts, studies have shown that linguistic and non-linguistic factors influence metaphor processing [5, 6]. The present study investigates how the location of context, whether it appears before or after metaphorical expression, influences the metaphor processing in children and adults. We predict that context location may constrain metaphor processing, particularly for children, given that they are known to have difficulties integrating late-coming cues [7].

Method: 29 monolingual Turkish-speaking children aged 4;00-4;07 (M = 4;04, 15 females) and 17 adult native Turkish speakers (M = 21.7, 11 females) participated in a visual world paradigm experiment. Our stimuli included six verbal metaphors and their literal counterparts with the same verb presented in two lists for each of the following conditions, context-first (a), and context-last (b):

a. Kuzeni oyuncağı kırdı bu yüzden Ceren surat astı.
Cousin-Poss.3sg. toy-Acc. break-Past so Ceren face hang-Past
"Her cousin broke the toy so Ceren sulked."

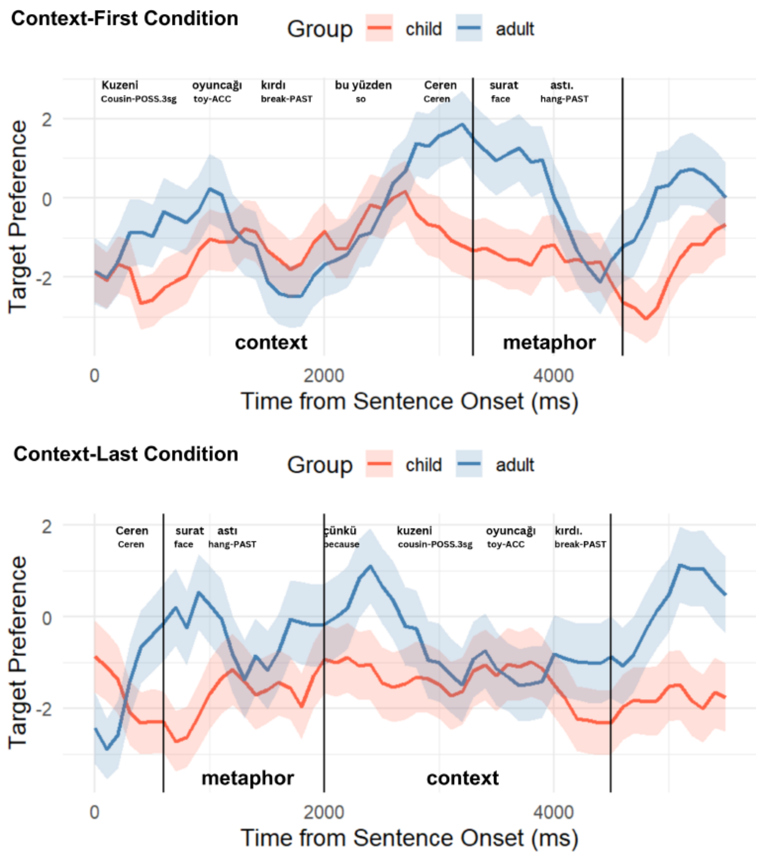
b. Ceren surat astı çünkü kuzeni oyuncağı kırdı.
Ceren face hang-Past because cousin-Poss.3.sg. toy-Acc. break-Past
"Ceren sulked because her cousin broke the toy."

Each sentence was accompanied by two illustrations depicting either the metaphorical (target) or the literal meaning of the expression (Fig. 1). While participants listened to the sentences, we recorded their eye movements and forced-choice picture selections.

Results & Conclusion: Forced-choice response analysis revealed a significant interaction between expression type and context location ($b = 0.403$, $SE = 0.140$, $p = .004$) with higher metaphor comprehension in the context-first condition (Children: $M = 0.45$, $SE = 0.05$; Adults: $M = 0.92$, $SE = 0.03$) than in context-last (Children: $M = 0.24$, $SE = 0.05$; Adults: $M = 0.70$, $SE = 0.05$). Growth curve analysis revealed that in the CF condition (Fig. 2), children's looks to the metaphorical image changed over time in a non-linear pattern in context-region ($Time^3 \times group$; $\beta = -21.62$, $SE = 5.45$, $p < .001$) and children were able to integrate metaphorical meaning by the end of the utterance ($Time^2 \times group$; $\beta = 15.67$, $SE = 6.26$, $p = .02$). In the CL condition (Fig. 2), children were less affected by the late-arriving context and showed flatter gaze patterns throughout the utterance ($\beta = -21.94$, $SE = 11.06$, $p = .06$). Our findings demonstrate that prior contextual information is a critical constraint in metaphor interpretation, especially in children, and that the presence of context prior to a metaphorical expression has a potential to shift the access to metaphorical meaning from indirect to direct.

Figure 1. Sample visual stimuli for metaphorical sentences (literal depiction/competitor on the left, metaphorical depiction/target on the right).

Figure 2. Gaze patterns for target (metaphorical depiction) preference for adults and children in context-first and context-last conditions.



References

[1] Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160-167.

[2] Grice, H.P. (1975). Logic and conversation. In: Cole, P., Morgan, J. (Eds.), *Syntax and Semantics*, vol. 3. Academic Press

[3] Searle, J. (1979). Metaphor. In: Ortony, A. (Ed.), *Metaphor and Thought*. Cambridge University Press, New York, pp. 92–123

[4] Gibbs, R. W. (1994). *The poetics of mind: figurative thought, language, and understanding.* Cambridge University Press.

[5] Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis.

[6] Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological bulletin,* 144(6), 641.

[7] Choi, Y., & Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of experimental child psychology*, 106(1), 41-61.

To thine native self be true: Exploring the link between self, emotion and language

Harrison Paff[1,2], Alissa Melinger[1], Sheila Cunningham[3], Josephine Ross[1]
[1] School of Humanities, Arts and Social Sciences, University of Dundee; [2] School of Psychology, University of Queensland; [3] Department of Sociological and Psychological Sciences, Abertay University
hpaff001@dundee.ac.uk

Language has a critical role in the development and experience of the self [1]. Bilingual speakers, for example, report feeling less *themselves* when conversing in a foreign than in a native language [2]. This foreign language effect (FLE) has been attributed to reduced emotionality in a second language (L2) [3], but it could also arise due to weaker links between the L2 and the self-system which develops in early childhood; the home L1 learning environment offers plentiful opportunities for emotional associations and autobiographical connections. In contrast, L2 is often acquired in an emotionally and autobiographically sparse academic setting [4]. The self-system is studied using paradigms that compare the processing of self and other-relevant stimuli; self-relevant stimuli are attended to faster [5], processed quicker and remembered better [6] than other-relevant stimuli and this self-bias effect is modulated by positive affect [7]. These paradigms, however, are limited to a monolingual context. Extending these findings to the bilingual context, our study asks whether self-bias effects are also modulated in the L2. The reduced emotional hypothesis [3] predicts they are, but only for emotional stimuli. Alternatively, our novel Dual Selves Hypothesis predicts a reduced self-bias effect across all stimuli, irrespective of emotional valance.

To test if self-processing is reduced in the L2 and discriminate between an account underpinned by emotional vs self-processing, we employed two self-bias tasks that allow systematic manipulation of emotionality and self-reference; the incidental and evaluative self-reference effect (e/iSRE) tasks, which show better recall of self- than other-referenced stimuli [6, 8]. The eSRE requires stimuli to be encoded in relation to autobiographical knowledge, whereas the iSRE does not. Any language or emotion modulations should therefore be observed with the eSRE but not the iSRE. In our version of these tasks, participants were presented with trait adjectives paired either with their own or someone else's name (Harry Potter). As well as the self-other manipulation, we varied trait adjective language, presenting words in German (L1) or English (L2). We also varied trait adjective emotional valance with positive, neutral and negative words. In Exp 1 ($N = 65$), participants identified the trait adjective's position relative to the name (incidental processing) and, in Exp 2 ($N = 65$), participants decided if the trait adjective described the person named (evaluative processing, e.g., are you stingy?). After the encoding phase, a surprise source memory test asked participants to indicate whether a word had been presented in the previous task, and if so, with which name. Participants were German-English unbalanced bilinguals who had learned English in an educational setting.

In Exp 1, we found a source memory iSRE that did not vary via valence or language. In Exp 2, we found a referent x valence interaction showing a source memory eSRE only with positive words collapsed across language. We also found a referent x language interaction showing a source memory eSRE only with German words collapsed across valence. Our findings show that SRE emotional modulation depends on evaluating the word against the self-schema. Critically, however, language did not attenuate the eSRE positivity bias. Thus, reduced L2 self-processing is likely not driven by reduced emotion processing [3]. Rather, the results are more consistent with the Dual Selves Hypothesis which links reduced L2 self-processing to the context and developmental timeline of L2 acquisition.

References

[1] Mead, G. H. (1934). Mind, Self, and Society from the Standpoint of Social Behaviorist. University of Chicago Press, Chicago, IL, USA.

[2] Pavlenko, A. 2006. “Bilingual Selves.” In *Bilingual Minds: Emotional Experience, Expression, and Representation*, edited by A. Pavlenko, 1–33. Bristol: Multilingual Matters.

[3] Ivaz, L., Costa, A., & Duñabeitia, J. A. (2016). The emotional impact of being myself: Emotions and foreign-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 489 - 496. https://psycnet.apa.org/doi/10.1037/xlm0000179

[4] Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current opinion in Psychology*, *17*, 74-78. https://doi.org/10.1016/j.copsyc.2017.06.005

[5] Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human perception and performance*, *38*(5), 1105. https://psycnet.apa.org/doi/10.1037/a0029792

[6] Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of personality and social psychology*, *35*(9), 677. https://psycnet.apa.org/doi/10.1037/0022-3514.35.9.677

[7] Cunningham, S. J., & Turk, D. J. (2017). A review of self-processing biases in cognition. *Quarterly journal of experimental psychology*, *70*(6), 987-995. https://doi.org/10.1080/17470218.2016.1276609

[8] Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and cognition*, *17*(3), 1040-1045. https://doi.org/10.1016/j.concog.2008.02.004

Figure 1. Incidental task source memory performance as a function of Referent, Language and Valence. Error bars represent one standard error of the mean.

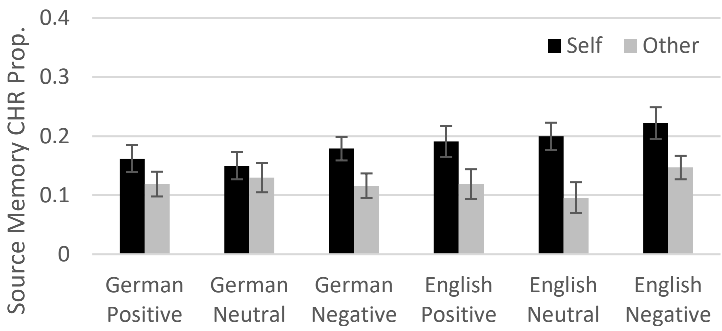
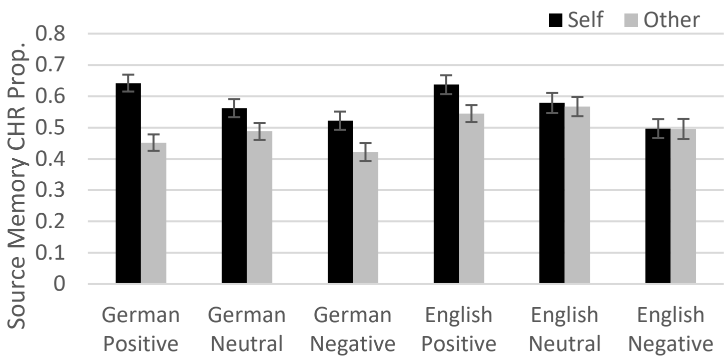


Figure 2. Evaluative task source memory performance as a function of Referent, Language and Valence. Error bars represent one standard error of the mean



Rethinking Reasoning: When Next-Token Prediction Mimics Thought

Jinbiao Yang
Max Planck Institute for Psycholinguistics
Jinbiao.Yang@mpi.nl

Background:

In cognitive science, reasoning is typically understood as a slow and structured “System 2” process. But does such reasoning require a dedicated mental mechanism, or could it instead emerge naturally from more general processes of language use? Large language models (LLMs) provide a new perspective on this question: these models simply predict the next word in a sequence; nevertheless, under certain training or prompting conditions, they can produce outputs that resemble reasoning—structured, multi-step progressions. In other words, behavior that appears to mimic reasoning can arise from the basic predictive mechanism. Since existing reasoning LLMs are trained with highly complex pipelines, here we design an extremely simplified approach to reproduce reasoning-like patterns, in order to ask whether what we call “reasoning” might be just a reorganization of predictive distributions.

Method:

We explored this question using GPT-Neo, a language model not explicitly trained for reasoning, as a case study. We post-trained the model on multi-step arithmetic samples using two formats: (A) direct-answer (e.g., "19 - 8 - 2 = ? [answer] = 9"), and (B) structured reasoning (e.g., "13 + 19 - 6 = ? [thinking] = 32 - 6 [answer] = 26"). By systematically varying the ratio of A and B in training, and evaluating the model on unseen problems (e.g., "12 + 7 - 19 = ?"), we examined how exposure to intermediate reasoning affected the model's behavior.

Results:

1. The model’s tendency to reason was directly linked to the proportion of training data with intermediate steps:

- Training predominantly on direct answers → the model almost never produced reasoning steps.
- Increasing structured reasoning examples → the model more frequently generated intermediate steps, with both accuracy and confidence improving.

2. Even models trained with minimal reasoning data could be prompted (e.g., "12 + 7 - 19 = ? [thinking]”) into generating reasoning steps

Discussion:

Our findings suggest that what appears as “reasoning” is not an entirely new capacity added onto LLMs, but something that can emerge naturally from a general-purpose language prediction mechanism under the right conditions. When training emphasizes direct-answer patterns, the model defaults to shortcut strategies. But when intermediate steps are made more frequent or more available, multi-step reasoning becomes the locally optimal generation path.

This mechanistic view offers an analogy for human reasoning. Traditionally, psychology has treated “System 2” as an independent system dedicated to slow, deliberate thought. Our results suggest an alternative: reasoning may not require a separate mental module at all. Instead, it may reflect the reorganization of predictive trajectories: When people are required to suppress fast, intuitive “System 1” responses and extend the unfolding of language and thought, intermediate steps naturally surface, forming the reasoning chains we recognize.

This perspective carries both engineering and cognitive significance. From an engineering standpoint, it implies that many recent methods for improving LLM reasoning (e.g., chain-of-thought prompting, reinforcement learning) work by reshaping the predictive landscape to make intermediate steps more attractive. From a cognitive standpoint, it suggests a reinterpretation of reasoning itself: not as a distinct mental faculty, but as a form of regulation over predictive processes shaped by context and experience.

Memory and Focus: How Bound Focus Affects Illusions of Plausibility

Samuele Bruzzese[1], Buhan Guo[1], Shayne Sloggett[1],
[1] University of York
sb2441@york.ac.uk

Background: Linguistic focus is a communicative tool believed to increase the salience in memory of affected constituents. Psycholinguistic research consistently show that comprehenders' processing focus experience (i) heightened working memory encoding, (ii) enhanced error detection, and (iii) improved long-term recall of focused constituents [1][2][3][4], which could all be due to a higher activation state in memory of focused constituents. These effects seem to align with cue-based models of memory in language processing [5], which predict that a constituent's retrieval is affected by its activation level in memory.

One attempt to investigate focus effects on retrieval was conducted by [6] in an eye-tracking study on Illusions of Plausibility (IoP). In Experiment 1, they examined sentences like (1), where the presence of a plausible but grammatically illicit distractor facilitated the licensing of the predicate "shatter" when the target was a grammatically licit but implausible referent (an IoP). In Experiment 2, they introduced focus to increase the saliency of the grammatically licit yet implausible target, which should reduce distractor interference and therefore the IoP. Their findings revealed the opposite pattern: an earlier IoP effect in Experiment 2 compared to Experiment 1.

Method: Building on recent work by [7], we propose that [6]' results may be due to their design not indicating whether focus was on the target or the distractor, and the absence of contrastive alternatives to the focused constituent [7] [8] [9]. If comprehenders interpreted the distractor as being in focus, the earlier facilitatory effects of Experiment 2 are unsurprising, while the lack of contrastive alternatives in the original items may have increased parsing difficulty of the focus itself [7]. This research test whether focus placement modulates retrieval interference by running two experiments: a maze task (n = 62) and an eye tracking experiment (desired n = 48, data collection in progress). To systematically control focus interpretation, our experimental design manipulates preceding contexts introducing contrastive alternatives, and questions establishing whether focus is on the target or the distractor in items based on [6]'s original design, as shown in (3).

Results: Preliminary results from the maze-task (Figure 1) show a strong plausibility effect in the distractor focus condition, and an attenuated - though not eliminated - plausibility effect in the target focus condition. There is also a distractor effect, such that implausible distractors cause longer response times than plausible distractors. The plausibility and distractor effects combined suggest that there is an IoP effect which is attenuated when the target is in focus, in line with the original predictions of [6].

Discussion: Our preliminary results suggest that focus placement influences target and distractor competition at retrieval. This provides important implications for focus theory and cue-based retrieval models. If focus were a simple activation boost, we would expect the target focus condition to cancel out the IoP, as an activation-based model [5] would predict that increased target activation blocks distractor interference. Our data show that the target focus condition attenuates the IoP, rather than eliminating it. This pattern may suggest that focus, rather than boosting activation, modulates retrieval by indicating contrast relationships between the target and distractor, influencing which element functions as the baseline for competition at retrieval.

Examples

1. Sue remembered the plate/letter that the butler with the cup/tie accidentally shattered…
2. What Sue remembered was the plate/letter that the butler with the cup/tie accidentally shattered…
3. **Context Question:**
   - What did the butler shatter? **(target focus)**
   - Which plate did the butler shatter? **(distractor focus)**

Target Sentence: It was the cup/napkin near the plate/letter that the butler accidentally shattered while...

Results plot

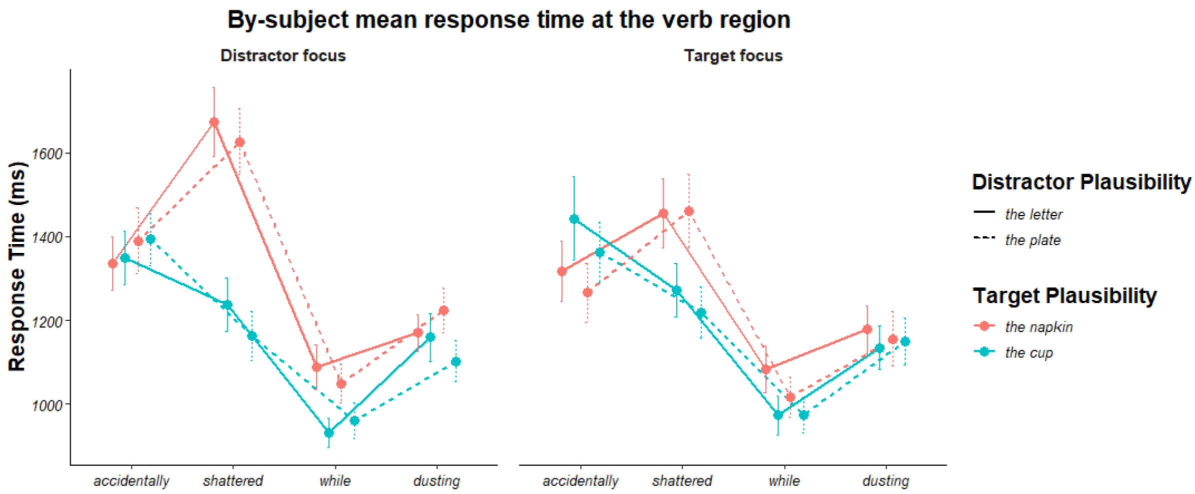


Figure 1: By subject mean response times for the maze-task (n=62). On the left: mean response times in the distractor focus conditions. On the right: mean response times in the target focus conditions.

References

[1] Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *J. Mem. Lang.*, *34*(2), 232–267.

[2] Brédart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*, 135–144.

[3] Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations□. *J. Mem. Lang.*, *56*(3), 357–383.

[4] Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Mem. Cognit.*, *35*(1), 73–86.

[5] Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.*, *29*(3), 375–419.

[6] Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *J. Mem. Lang.*, *102*, 16–27.

[7] Hoeks, M., Toosarvandani, M., & Rysling, A. (2023). Processing of linguistic focus depends on contrastive alternatives. *J. Mem. Lang.*, *132*, 104444.

[8] Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Lang. Cogn. Neurosci.*, *31*(2), 217–235.

[9] Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.*, *1*(1), 75–116.

ManyLanguages: A global network for Big Team Language Science

Erin Buchanan[1] on behalf of the ManyLanguages Consortium
[1]Harrisburg University of Science and Technology
many-languages-board@many-languages.com

ManyLanguages is a globally distributed network of language science laboratories that coordinates data collection for democratically selected studies. Our mission is to facilitate the connection between language science researchers to diversify the languages, participants, researchers, and projects represented in the language sciences. We will facilitate the collection of evidence across the language sciences by supporting a distributed laboratory network that is ongoing, diverse, and inclusive. We embrace open science principles by sharing collected data, materials, translations, and other research outputs from the network. We strive to engage research across a broad spectrum of language sciences creating an inclusive and diverse environment for ideas, investigation, and participation.

Friday Morning Posters

The processing of Chinese reflexives as plain anaphors and intensifiers

Jun Lyu (Peking University)
junlyu@pku.edu.cn

Background It is generally agreed that, in neutral contexts, the two Chinese reflexives, *ziji* ('self') and *ta-ziji* ('s/he-self'), tend to be construed as 'plain anaphors,' adhering to Principle A (Chomsky'81) (e.g., Dillon et al.'14,'16; Wang'17; Chang et al.'20). However, two questions are open for further research. **First**, while previous research suggests that anaphors of diffent forms may exhibit differing sensitivities to syntactic and non-syntactic (e.g., semantic, discourse) information (e.g., Kaiser et al.'09), insufficient attention has been paid to the distinct processing patterns of *ziji* and *ta-ziji* from a comparative perspective (but see Dillon et al.'16; Wang'17). **Second**, it remains unclear whether the intensified use of Chinese reflexives results in different processing and interpretation patterns compared to their anaphoric use. To address these two questions, this study employs self-paced reading (SPR) and compares the processing of *ziji* and *ta-ziji* as plain anaphors and as intensifiers.

Methods In **Exp.1** on plain anaphors ($N_{\text{participants}}$ = 49), we manipulated the factors Reflexive (*ziji*/*ta-ziji*) and Distance (local/non-local) in a 2x2 factorial design. We used animacy to establish local vs. non-local antecedent-reflexive dependencies, shown in example (1). **Exp.2** (*N* = 50) on intensifiers also controls these two factors, except that subordinate clauses are always in a *it*-cleft construction (*shi*…), shown in example (2). In both experiments, the reflexive is in subject position of the subordinate clause, a position argued to be in the same binding domain as its closest antecedent in the matrix clause (Huang & Liu'01). Native Chinese participants read 24 target sentences and 36 fillers at their own pace (**SPR**) and indicated their acceptability judgment (**AJ**) of the sentences on a 1-7 Likert scale.

Results In **Exp.1** on plain anaphors (see **Fig.1-2**), we found clear locality bias effects in offline acceptability judgments due to a main effect of Distance ($p < .01$). In the self-paced reading task, we found a main effect of Distance at the critical (reflexive) region ($p <.05$), in addition to a main effect of Reflexive ($p < .005$, as *ta-ziji* is syllabically longer). However, pairwise comparisons indicate that the locality bias is only significant for the processing of *ta-ziji* ($p < 0.001$) but not for *ziji* ($p > .1$). Crucially, we discovered a Reflexive x Distance interaction ($p < .05$) at the second spillover region, because the locality bias effect only exists for *ta-ziji* ($p < .05$). Only at the final region does *ziji* show a delayed, strong locality bias effect ($p < .001$). In **Exp.2** on intensifiers (see **Fig.3-4**), no effect has been detected in the offline judgment task. In the self-paced reading task, from the reflexive to the pre-final region, we only found a main effect of Reflexive at the reflexive region ($p < .001$) due to longer reading times of *ta-ziji* compared to *ziji*. At the final region, we discovered a reversed main effect of Distance ($p < .05$) as local binding leads to slower reading times compared to non-local binding.

Conclusions The present study has produced three key findings. **First**, we replicated in Exp.1 the locality bias of *ziji* and *ta-ziji* (e.g., Dillon et al.'14,'16). **Second**, unlike some previous studies, we found that *ta-ziji* exhibits a stronger locality bias than *ziji*, which fits with a form-specific approach to anaphora resolution (e.g., Kaiser et al.'09). **Third**, the intensified uses of *ziji* and *ta-ziji*, as examined in Exp.2, drastically altered the processing and interpretations of these reflexives, which suggests that discourse-level information can have an early impact on the processing of reflexives (at least for *ta-ziji*).

(1) An example set (24 in total) of target sentences in **Exp.1** on plain anaphors (sentences are translated into English; word order kept intact). Square backets indicate local binding domains.

a. **Local binding** (<u>dependency length created using animacy (mis)match; BA is a light verb that introduces an SOV word order; subscripts indicate presentation regions in SPR</u>)

Company$_1$ *stated*$_2$ *[**Xiaoli**$_3$ thought$_4$ {**ziji/ta-ziji**}$_5$ BA$_6$ research$_7$ plan$_8$ screwed up.$_9$]*

b. **Non-local binding**

Xiaoli$_1$ *stated*$_2$ *[**company**$_3$ thought$_4$ {**ziji/ta-ziji**}$_5$ BA $_6$ research$_7$ plan$_8$ screwed up.$_9$]*

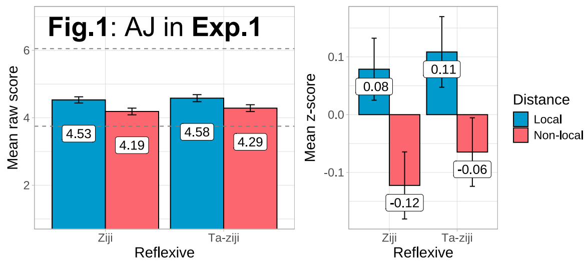
(2) An example set (24 in total) of target sentences in **Exp.2** on intensifiers (**SHI** introduces an *it*-cleft ***focus*** construction in Chinese, thus forcing an intensifier reading)

a. **Local binding**

Company$_1$ *stated*$_2$ *[**Xiaoli**$_3$ thought$_4$ SHI$_5$ {**ziji/ta-ziji**}$_6$ BA $_7$ research$_8$ plan$_9$ screwed up.$_{10}$]*

b. **Non-local binding**

Xiaoli$_1$ *stated*$_2$ *[**company**$_3$ thought$_4$ SHI$_5$ {**ziji/ta-ziji**}$_6$ BA $_7$ research$_8$ plan$_9$ screwed up.$_{10}$]*



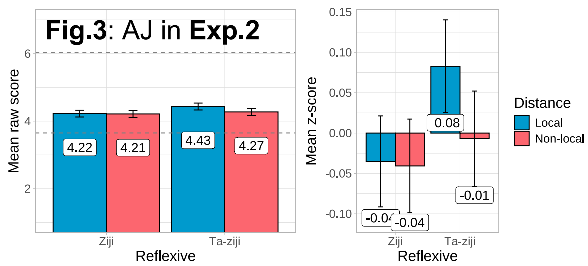
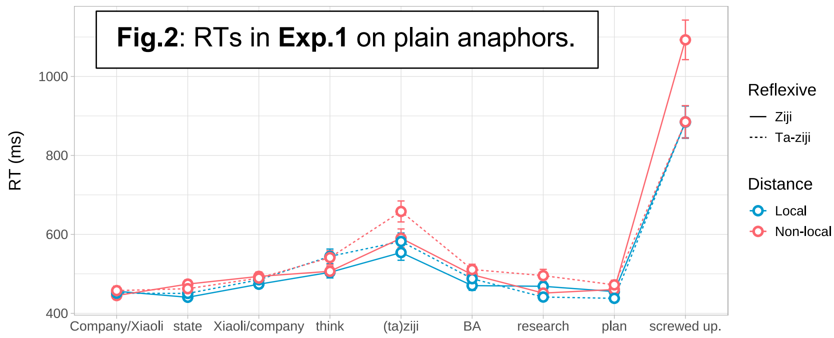


Fig.1 (left) and **Fig.3** (right): Acceptability judgment results from **Exp.1** and **Exp.2**, respectively. The left panel in each figure uses a raw rating scale; the right panel uses a z-score scale. Dashed lines at upper and lower parts of the graphs indicate 'acceptable' and 'unacceptable' fillers.



Bilinguals' Neurocognitive Profiles in L1 and L2: N400 vs. P600 Dominance Reflects Divergent Processing of Filler-Gap Dependencies

Ioannis Iliopoulos
University of Potsdam

Prior ERP research suggests that the same linguistic violations can elicit a P600 response in some participants while triggering an N400 in others, enabling their categorization into positivity- (P600) or negativity- (N400) dominant subgroups in both L1 and L2 [1–4]. N400 dominance reflects a reliance on lexical-semantic cues, while P600 dominance indicates engagement of combinatorial morphosyntactic mechanisms [1].
Filler-gap dependencies (FGDs) provide a well-established paradigm for probing syntax–semantics integration [5]. Given that P600/N400 subgroup performance has not yet been tested in a within-subject design examining how bilinguals process such dependencies in both their L1 and L2, we ask: Do ERP dominance profiles reflect stable processing strategies across languages and FGD-related tasks in bilinguals?
We tested 106 L1 German/L2 English bilinguals (M_{age} = 25.54, SD = 4.2; M_{AoA} = 8.02, SD = 2.1; CEFR B1–C2) in four EEG experiments (two per language) examining filler plausibility (lexical-semantic) and gap availability (syntactic) violations. Stimuli were presented word by word. ERPs were measured at 350–600 ms and 600–1000 ms from the onset of the underlined words (*Tables 1 & 2*). Subgroups were established separately for each experiment via LMM-derived random slopes comparing N400 (350–600 ms) and P600 (600–1000 ms) mean amplitudes. Despite this experiment-specific categorization, correlation tests revealed significant positive correlations between participants' N400-P600 slope differences across experiments (all $r > 0.315$, $p < 0.01$), indicating stable individual tendencies in ERP dominance.
At the group level, participants exhibited a biphasic N400–P600 response to lexical-semantic violations and a P600 to syntactic violations in both languages. However, subgroup analyses showed divergent patterns:

N400-Dominant Subgroups:

- Lexical-semantic violations elicited a significant N400 in the L1, followed by a non-significant negativity in the later time window. In the L2, a significant negativity was observed in both time windows.
- Syntactic violations elicited a biphasic N400-P600 response in the L1. In the L2, the N400 did not reach significance, while the P600 was significant.

P600-Dominant Subgroups:

- Lexical-semantic violations elicited a non-significant positivity in the early time window that was followed by a significant P600 in the later time window in both languages.
- Syntactic violations elicited significant positivities in both time windows in both the L1 and L2.

These findings demonstrate, first, that individual neurocognitive profiles remain stable across languages and violation types; and second, that ERP response patterns for each task are highly similar across languages within each subgroup, although specific ERP responses varied depending on violation type. Notably, syntactic violations elicited structural repair responses regardless of dominance. Taken together, these results suggest that filler-gap dependency processing in bilinguals is driven primarily by an individual's ERP dominance and task-specific demands, rather than by language status.

Table 1. Example Stimuli for the Filler Plausibility Experiment (Plausible vs. Implausible)

German:

Tim liebte den {Kuchen/#Kunden}, den die Bäckerin dekoriert hatte mit großer Mühe.

Tim loved the {cake/#customer}, that the backer decorated had with great effort.

Tim loved the {*cake/#customer*} *that the baker had decorated with great effort.*

English:

Sharon photographed the {bottle/#scientist} that the kind waiter opened during his shift.

Table 2. Example Stimuli for the Filled Gap Experiment (Grammatical vs. Ungrammatical)

German:

Michelle kannte den Lernstoff, {bevor/*den} die Dozentin ihrem Kurs diesen vermittelt hatte.

Michelle knew the material {before/*that} the professor her course this taught had.

Michelle knew the material {*before/*that*} *the professor had taught it to her course.*

English:

Michael watched the boy, {while/*for whom} Susan was playing some music for him at the party.

References

[1] Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia, 56, 289-301.

[2] Ciaccio, L. A., Bürki, A., & Clahsen, H. (2023). Inter-individual variability in morphological processing: An ERP study on German plurals. Journal of Neurolinguistics, 67, 101138.

[3] Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. Bilingualism: Language and Cognition, 17(2), 277-293.

[4] Grey, S. (2023). Variability in native and nonnative language: An ERP study of semantic and grammar processing. Studies in Second Language Acquisition, 45(1), 137-166.

[5] Felser, C., & Jessen, A. (2020). Brain responses elicited by implausible fillers and filled object gaps in German. In *Typical and Impaired Processing in Morphosyntax* (pp. 75-90). John Benjamins Publishing Company.

Sentential negation causes both NPI and NCI illusions in Czech

Radim Lacina[1,2], Mojmír Dočekal[2]
[1] Department of Czech Language, Faculty of Arts, Masaryk University; [2] Department of Linguistics and Baltic Languages, Faculty of Arts, Masaryk University
rlacina@mail.muni.cz

Background: When comprehenders establish dependencies online, they can be led astray by potentially licensing yet grammatically irrelevant elements, giving rise to linguistic illusions [1]. Among these are NPI illusions such as **The student who no teacher likes has ever got an A* [2]. A crucial result has been that while sentential negation is an NPI licensor in grammatical sentences, it is ineffective as an illusory licensor [3]. This has been argued to be incompatible with the cue-based retrieval approach to NPI illusions [4] and as supporting the active scope [3,5] and scalar theories [6,7]. Recently, an illusion with another type of negative dependency, negative concord items (NCIs), has been discovered in Czech [8]. In such strict negative concord languages, NCIs (e.g., the Czech quantifier *žádný* 'no') require the presence of sentential negation in the same clause. As NCIs are said to be licensed in the syntax, similarly to agreement [9], sentential negation should be effective in causing illusions, unlike with NPIs.

Current study and methods: We tested the effectiveness of sentential negation as an illusory licensor for both NPIs (Exp1) and NCIs (Exp2) in two speeded acceptability judgement experiments with native speakers of Czech ($N_1 = 143$, $N_2 = 179$, $\text{Items}_{1,2} = 32$). As for NPIs illusions, which have not been documented in Czech (or any Slavic language), we expected negation to be ineffective. On the other hand, NCIs were predicted to be vulnerable to illusory licensing from sentential negation. We presented sentences such as those in (1), which included either the strong NPI *ani jeden* ('even one') in Exp1 or the quantifier NCI *žádný* ('no'), using RSVP. The presence or absence of the negation morpheme *ne-* manipulated grammaticality in the matrix clause and the potential for illusory licensing in the relative clause.

1) Minulý měsíc (ne)vyhrál karbaník, který (ne)hledá dlouhodobou práci, ani jednu / žádnou sázku.
   Last month (NEG)won gambler, who (NEG)seeks long-term job even one / no bet
   'Last month the gambler, who is/isn't looking for a long-term job, won/didn't win even one / no bet.'

Results: The acceptability data in Figure 1 show a clear illusion effect with both NPIs and NCIs, which was statistically tested for with a mixed effects logistical regression specified with the maximal random effects structure. The crucial effect of sentential negation within the RC nested within the ungrammatical condition was significant in the NPI Exp1 ($\beta = 0.79$, $SE = 0.08$, $z = 10.03$, $p < 0.001$***) as well as in the NCI Exp2 ($\beta = 0.65$, $SE = 0.08$, $z = 8.08$, $p < 0.001$***).

Discussion: We find that that Czech comprehenders are vulnerable to grammatical illusions both with NCIs and NPIs. Crucially, these illusions were in both cases caused by the presence of sentential negation in an intervening RC. While this was predicted in the case of NCIs, due to them functioning akin to syntactic agreement, the surfacing of the NPI illusion was contrary to both our predictions, as well as the previous findings on English [3]. This result thus goes against both the active scope [3,5] and scalar theories [6,7], while being compatible with the cue-based retrieval theory [4]. Our study further replicates previous findings on Czech NCI illusions [8] and highlights the value of broadening the range of languages studied within the domain of linguistic illusions where differences in vulnerability profiles might be linked to the particular structures of the languages in question and their interaction with the human language processor.

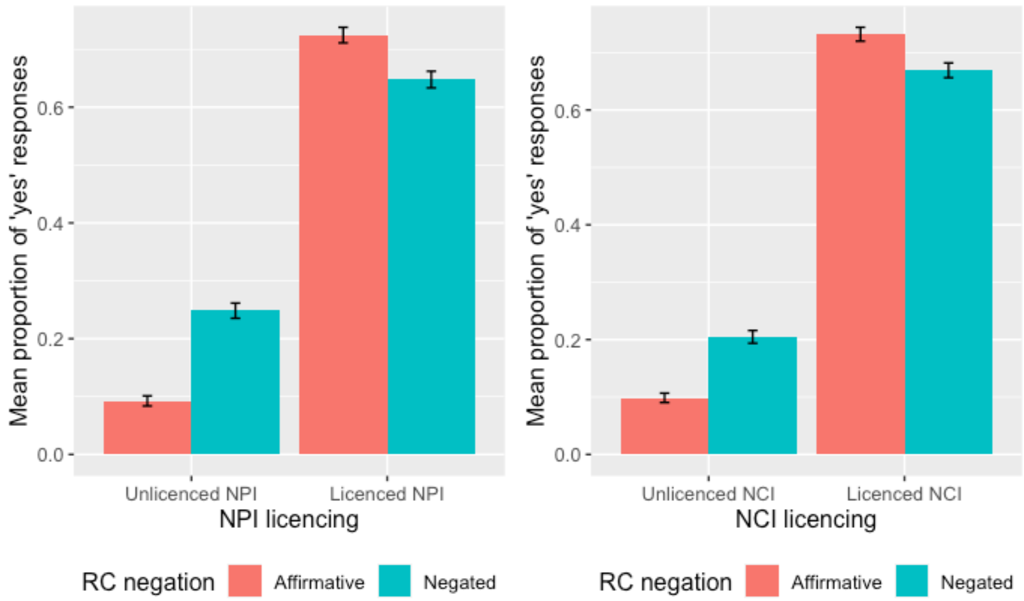


Figure 1: The mean proportion of 'yes' responses per condition in Exp1 (NPIs, left) and Exp2 (NCIs, right) together with their associated standard errors.

References

[1] Phillips, C., Wagers, M., & Lau, E. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the Interfaces* (pp. 147–180). Brill.

[2] Muller, H., & Phillips, C. (2020, 03). Negative Polarity Illusions. In *The Oxford Handbook of Negation*. Oxford University Press.

[3] Orth, W., Yoshida, M., & Sloggett, S. (2021). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47 (6), 906.

[4] Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32 (4), 685-712.

[5] Orth, W., Sloggett, S., & Yoshida, M. (2025). Positive polarity items: an illusion of ungrammaticality. *Language, Cognition and Neuroscience*, 0 (0), 1–25.

[6] Schwab, J. (2022). Lexical variation in NPI illusions–NPI illusions as a scalar phenomenon. *Glossa Psycholinguistics*, 1 (1), 21.

[7] Muller, H. (2022). *What could go wrong? linguistic illusions and incremental interpretation* (Unpublished doctoral dissertation). University of Maryland, College Park.

[8] Lacina, R. (2025). The negative concord illusion: An acceptability study with Czech neg-words. *Linguistics Vanguard*.

[9] Zeijlstra, H. (2008). *Negative concord is syntactic agreement*. Manuscript, University of Amsterdam, 5, 113.

Say what you mean: Linguistic vividness and information theory

Alexander Kilpatrick[1], Rikke Bundgaard Nielsen[2]
[1]Univeristy of Aizu, [2]University of Melbourne
alex@u-aizu.ac.jp

Background: This paper tests the hypothesis that words that are highly specific, concrete, and/or imaginable—here grouped under the term linguistic vividness—carry more information than words that are abstract, general and harder to imagine. We test this hypothesis using the measure of surprisal, a concept derived from information theory [1]. Surprisal quantifies the unexpectedness or unpredictability of a linguistic element, such as a phoneme or word, within a given context. According to information theory, improbable events convey more information than highly probable events. This means that improbable events are more cognitively demanding but also better retained in long-term memory. For instance, [2] demonstrates that words composed of improbable phoneme sequences not only carry more information but also require greater cognitive effort to process, which paradoxically improves their memorability.

Method: We tested our hypothesis with a dataset comprised of a corpus of ~50 million words of spoken American English [3] which was cross-referenced with a pronouncing dictionary [4] from which we calculated bigram Surprisal—the negative log probability of one phoneme following another—which was then averaged for each word. Three dependent variables—Specificity [5], Concreteness [6], and Imaginability [7]—were taken from existing psycholinguistic experiments while six independent variables—Phoneme Length [4], Iconicity [8], Morpheme Count [9], parts of speech (POS) [10], age of acquisition (AOA) [11], and prevalence [12] were cross referenced from other datasets and included in linear mixed effects models. Anonymised data and code here: https://tinyurl.com/3pz9ezfs

Results: The results (Table 1) revealed a significant relationship between linguistic vividness and phonemic surprisal. In this datae, words rated higher in specificity, concreteness, and imaginability exhibited greater average phonemic bigram surprisal, even when controlling for the independent variables.

Discussion: This pattern points to an unconscious, cooperative tendency in speech communities—at least in American English—to assign higher-information forms to more vivid meanings. We do not argue that the results arise from the implementation of a deliberate strategy, but rather that this alignment may reflect a shared communicative instinct for precision: By using words characterised by high levels of surprisal and linguistic vividness, a speaker can effectively direct the cognitive resources of the listener, guiding their attention to important information and ensuring optimal retention. These findings align with prior work on negative valence [14], where high-surprisal words were also better remembered, supporting the Attentional Optimization Hypothesis (AOH: [14]). According to AOH, phonologically marked forms help direct attention toward meanings that merit deeper cognitive investment. Crucially, vividness predicted surprisal even after controlling for variables like phoneme length, morpheme count, AoA, prevalence, and iconicity, suggesting this relationship is not a by-product of lexical frequency or familiarity. Instead, it may reflect a broader tendency in language to highlight conceptually rich meanings through phonological form. This unconscious alignment between sound and meaning may serve to enhance both precision and memorability in communication. Future work should test whether similar patterns appear across languages, particularly those with different phonological or morphological systems.

	Specificity	Concreteness	Imaginability
(Intercept)	4.222 ***	25.067 ***	2.786 **
Average Surprisal	**2.823 ****	**5.13 *****	**5.454 *****
Iconicity	3.305 ***	2.026 *	7.326 ***
Phoneme Length	2.626 **	0.625	6.234 ***
Morpheme Length	-2.473 *	-5.789 ***	-7.372 ***
AoA	-2.185 *	-25.814 ***	-37.638 ***
Prevalence	-7.461 ***	-8.524 ***	7.434 ***
POS_Adjective		-1.72 .	2.213 *
POS_Adverb	-0.535	-0.785	0.709
POS_Article			-0.617
POS_Conjunction			-0.471
POS_Determiner		-0.827	0.227
POS_Interjection		-1.098	0.677
POS_Name	1.841 .	11.105 ***	3.198 **
POS_Noun	1.69 .	33.03 ***	3.178 **
POS_Number		-1.474	2.229 *
POS_Preposition		-1.478	0.624
POS_Pronoun		-1.089	0.897
POS_Unclassified			1.906 .
POS_Verb	-1.496	4.031 ***	2.157 *

Table 1: Results of the three linear mixed effects models. Significance codes: < 0.001 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379-423.

[2] Kilpatrick, A., & Bundgaard-Nielsen, R. L. (2025). Exploring the dynamics of Shannon's information and iconicity in language processing and lexeme evolution. *PLoS One*, *20*(4), e0321294.

[3] Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

[4] Carnegie Mellon University. (n.d.). *CMU Pronouncing Dictionary*. Retrieved April 1, 2025, from https://www.speech.cs.cmu.edu/cgi-bin/cmudictCMU Speech

[5] Ravelli, A. A., Bolognesi, M. M., & Caselli, T. (2024). Specificity ratings for English data. Cognitive Processing, 1-20.

[6] Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. Behavior research methods, 51, 1258-1270.

[7] Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: a multiparametric approach. Cerebral cortex, 20(8), 1799-1815.

[8] Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2024). Iconicity ratings for 14,000+ English words. Behavior research methods, 56(3), 1640-1655.

[9] Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. Behavior research methods, 50, 1568-1580.

[10] Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

[11] Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav Res,* 44, 978-990.

[12] Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behav Res, 51*, 467-479.

[13] Kilpatrick, A. (2025). The negativity bias is encoded in language. JASA Express Letters, 5(2).

[14] Kilpatrick, A., & Bundgaard-Nielsen, R. L. (Under Review). Say it like you mean it: Linguistic vividness and the attentional optimization hypothesis

The role of Morphological Skills as a Compensatory Mechanism in adult Developmental Dyslexia

Katerina Stoumpou[1], Prof Ghada Khattab[1], Dr Faye Smith[1]
[1]Speech & Language Sciences, Newcastle University
K.Stoumpou2@newcastle.ac.uk, ghada.khattab@newcastle.ac.uk, faye.smith@newcastle.ac.uk

Introduction: University students with dyslexia are often described as 'high-functioning' or 'compensated', suggesting that they have developed mechanisms to overcome difficulties associated with dyslexia and achieve a high level of reading necessary to engage with academic texts [1]. Morphological awareness, the ability to manipulate word parts that carry meaning [2], is a relatively well-preserved skill in dyslexia [3] and could be such a mechanism. Morphological awareness may support word reading and reading comprehension processes by providing an alternative reading unit, the morpheme, which combines information on form, pronunciation and meaning [4], thereby reducing reliance on phonology, a known weakness in dyslexia [5]. While research on morphology in dyslexia remains limited and findings are still inconclusive, this paper offers an in-depth investigation of morphological compensation across different literacy abilities as well as the role of morphology in skilled reading.
Methods: A cross-sectional behavioural study assessed university students and graduates with (n = 53) and without (n = 68) dyslexia on various psycholinguistic skills, including a comprehensive battery of morphological tasks adapted to minimise phonological load. Hierarchical regressions examined the contribution of morphology to word reading, spelling and reading comprehension, while controlling for phonological, orthographic, vocabulary and reasoning skills. A central concept in this study is reading resilience, the ability to comprehend text despite decoding difficulties [6, 7]. Quantile regressions explored the link between morphology and varying levels of reading resilience, using a refined index of reading comprehension that accounted for phonological and decoding skills, thereby considering participants' relative strengths and weaknesses in those domains.
Results: Morphological skills explained unique variance in reading comprehension in the dyslexia group, beyond phonological, orthographic, vocabulary and reasoning skills ($\Delta R^2 = .11$, $b = .44$, $t(48) = 2.71$, $p = .01$); however, they were not predictive of word reading or spelling. Participants with dyslexia who exhibited high reading resilience performed comparably to controls in most morphological tasks, despite their lower performance in phonological and decoding tasks. Furthermore, morphological skills were the only consistent predictor of resilience in dyslexia across the lower and middle quantiles, contributing substantially beyond other higher-level skills. Conversely, for controls, vocabulary was the key predictor of resilience in the same quantiles, suggesting a double dissociation in the skills underpinning resilience between the two groups (Figure 1). Ongoing analysis will further clarify the role of morphology in adult dyslexia with more detailed findings presented at the conference.
Conclusion: Preliminary results support the idea of morphological compensation in dyslexia, aligning with theories viewing morphology as a central component of word reading, spelling and reading comprehension processes [8-10]. These findings will help inform educational practices and interventions for adolescents and adults with dyslexia.

References

[1] Law, J. M., Wouters, J., & Ghesquière, P. (2015). Morphological Awareness and Its Role in Compensation in Adults with Dyslexia. Dyslexia, 21(3), 254–272.
[2] Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 189–209). Lawrence Erlbaum Associates, Inc.
[3] Martin, J., Frauenfelder, U. H., & Cole, P. (2014). Morphological awareness in dyslexic university students. *Applied Psycholinguistics*, *35*(6), 1213.
[4] Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological science in the public interest*, *19*(1), 5-51.
[5] Cavalli, E., Duncan, L. G., Elbro, C., El Ahmadi, A., & Colé, P. (2017). Phonemic—Morphemic dissociation in university students with dyslexia: an index of reading compensation? *Annals of Dyslexia*, *67*(1), 63–84.
[6] Farris, E. A., Cristan, T., 2nd, Bernstein, S. E., & Odegard, T. N. (2021). Morphological awareness and vocabulary predict reading resilience in adults. *Annals of Dyslexia*, *71*(2), 347–371.
[7] Patael, S. Z., Farris, E. A., Black, J. M., Hancock, R., Gabrieli, J. D. E., Cutting, L. E., & Hoeft, F. (2018). Brain basis of cognitive resilience: Prefrontal cortex predicts better reading comprehension in relation to decoding. PloS One, 13(6), e0198791.
[8] Kim, Y. S. G. (2020). Toward integrative reading science: The direct and indirect effects model of reading. *Journal of Learning Disabilities*, *53*(6), 469-491.
[9] Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. Journal of Research in Reading, 44(1), 10-26.
[10] Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific studies of Reading*, *18*(1), 22-37.

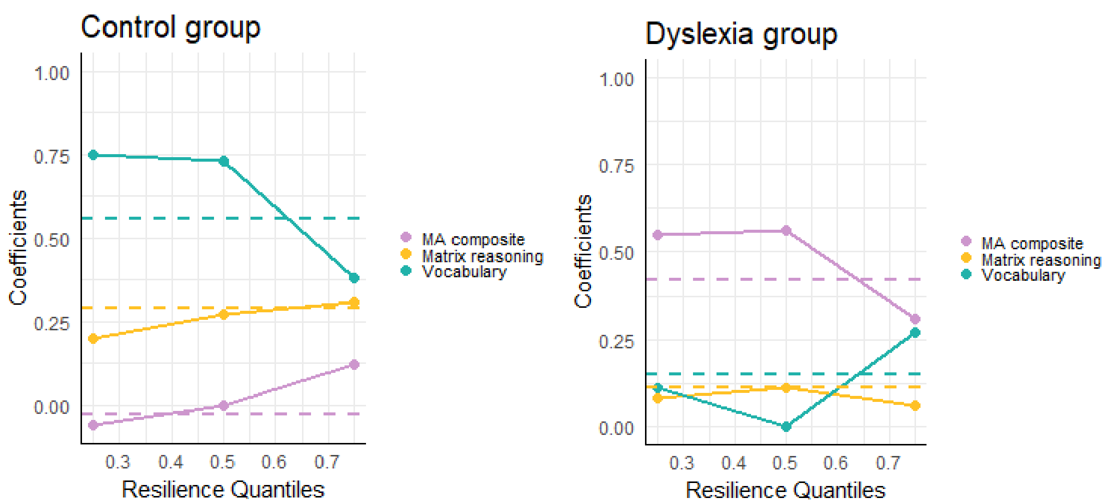


Figure 1. *Quantile process plots for Morphological Awareness composite, Vocabulary and Matrix Reasoning across reading resilience in Controls and Dyslexia. The dashed lines represent the coefficients in Ordinary Least Squares Regression while the solid lines represent the coefficients in the Quantile Regression.*

Does multimodal pre-activation influence linguistic expectations in LLMs and humans?

Sasha Kenjeeva, Giovanni Cassani, Noortje J. Venhuizen, Afra Alishahi
Department of Cognitive Science and Artificial Intelligence, Tilburg University
a.b.kenjeeva@tilburguniversity.edu

Background. The meaning representations that humans construct for words capture both linguistic and multimodal sensorimotor information [1]. Rommers et al. [2] showed that a word's sensorimotor components might also be pre-activated during online sentence processing: a context in which "moon" is a highly predictable continuation resulted in facilitated processing for "tomato" (similar shape) compared to "rice". To investigate to what extent multimodal pre-activation influences linguistic expectations during sentence processing, we here describe a data-driven experimental setup—with materials normed for plausibility, visual and co-occurrence similarity—that orthogonally manipulates multimodality (sensorimotor similarity; see below) and linguistic predictability (Cloze probability). We hypothesize that high sensorimotor similarity to the likeliest Cloze completion should result in decreased processing effort, even when a word is not predictable from the linguistic context. We report processing effort in terms of LLM surprisal and additionally plan to conduct a human self-paced reading (SPR) study, to shed light on processing behaviour of humans and (multimodal) language models.

Stimuli design. We designed 37 context frames with 5 plausible continuations each (n=185), manipulating linguistic predictability in context (L dimension) using Cloze data [3], and multimodal similarity (MM dimension) to the likeliest Cloze completion – the target – according to word vectors derived from the Lancaster sensorimotor norms [4] and ViSPA, a psychologically-validated computer vision model [5]. Refer to Table 1 for an example of a context: the critical prediction is that the pre-activation of the visual features of the referent of the likeliest completion ("watch") will result in decreased processing difficulty of "compass" (L-MM+) compared to "dog" (L-MM-), due to the visual similarity between a compass and a watch. Figure 1a shows that MM+/- is reflected by high and low visual similarity to the target word, respectively, in both Lancaster norms and ViSPA. As shown in Figure 2a, this effect is partially confounded by simple linguistic similarity, as measured by word2vec [6].

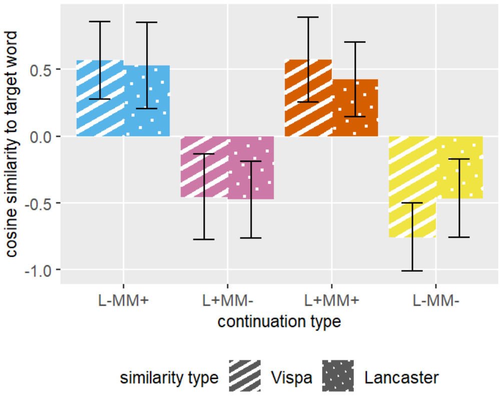
Norming studies. To validate our stimuli design, we collected plausibility, visual similarity and co-occurrence similarity ratings. For each study, we collected $\sim$20 data points per sentence. Participants (n=97) rated most sentences (170/185) as plausible. In the visual (n=145) and co-occurence similarity (n=229) studies, participants viewed the context frames with the target word and rated which of two words on a slider was more visually similar/more likely to appear in similar sentences to the target word. Results from both studies mirrored the similarity patterns of our data-driven measures (see Figures 1b and 2b), thus validating our stimuli design.

LLMs experiment. To test whether visual similarity might lower surprisal in a model trained on text and images, we fitted human ratings for visual and co-occurrence similarity to predict surprisal in two auto-regressive LLMs: GPT2 (language-only [7]) and QWEN2-VL (dual-stream vision-language [8]). Contrary to our prediction, visual similarity did not significantly predict surprisal in either model for the critical comparison (L-MM+ vs L-MM-). Co-occurrence similarity also had no effect on surprisal, indicating that the surprisal estimates of both LLMs seem to follow Cloze predictability patterns, regardless of being trained with visual information or not.

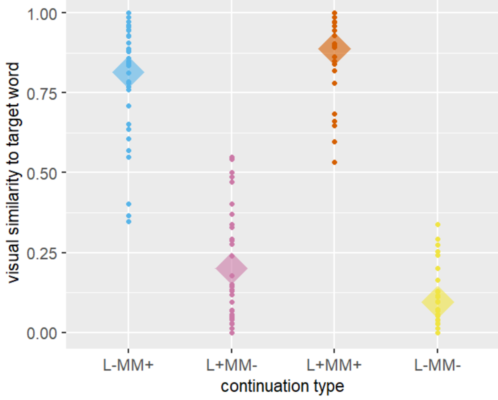
SPR experiment. We are currently conducting a self-paced reading study to investigate whether, contrary to the LLMs, visual similarity influences human reading times. We will then determine whether humans' online processing of plausible sentences entails a multimodal dimension that goes beyond Cloze predictability.

Context	Completion	Continuation Type	Manipulation
The impatient man kept glancing at his …	*watch*	target	highest Cloze
	compass	L-MM+	zero Cloze / high visual similarity
	wife	L+MM-	low Cloze / low visual similarity
	phone	L+MM+	low Cloze / high visual similarity
	dog	L-MM-	zero Cloze / low visual similarity

Table 1: Example of a context frame with its five continuations

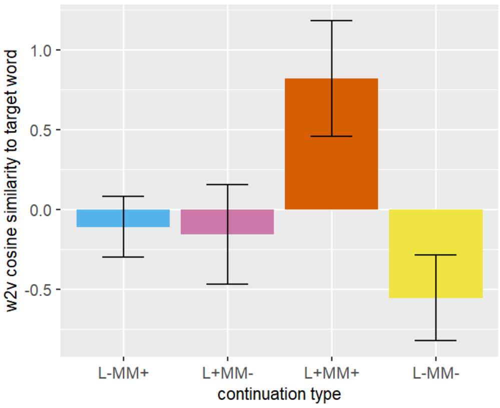


(a) Computational visual similarity

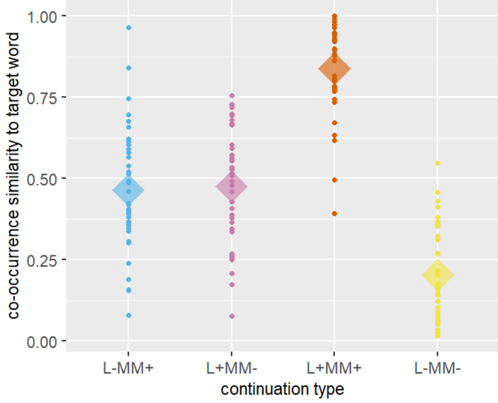


(b) Human-rated visual similarity (mean and per context)

Figure 1: Visual similarity of each continuation type to the target word



(a) Co-occurrence similarity from word2vec



(b) Human co-occurrence similarity (mean and per context)

Figure 2: Co-occurrence similarity of each continuation type to the target word

References. [1] Davis, C. P., & Yee, E. (2021). Building semantic memory from embodied and distributional language experience. *Wiley Interdiscip. Rev. Cogn. Sci., 12*(5), e1555.

[2] Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia, 51*(3), 437–447.

[3] Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behav. Res. Methods, 52*(4), 1795–1799.

[4] Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behav. Res. Methods, 52*(3), 1271–1291.

[5] Günther, F., Marelli, M., Tureski, S., & Petilli, M. A. (2023). ViSpa (Vision Spaces): A computer-vision-based representation system for individual images and concept prototypes, with large-scale evaluation. *Psych. Rev., 130*(4)

[6] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space.*arXiv:1301.3781.*

[7] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

[8] Wang, P. et al. (2024). Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv:2409.12191.*

Does overt production facilitate language prediction in challenging situations only?

Ana Bautista[1, 2] Francesca Branzi[3] & Clara Martin[1, 4]
[1] BCBL, Basque Center on Cognition, Brain & Language; [2] University of the Basque Country (UPV-EHU)
[3] University of Liverpool; [4] Ikerbasque, Basque foundation for science
a.bautista@bcbl.eu

Background: Previous research suggests that overt production facilitates predictive processing in native speakers. In a previous study [1], participants read highly and low constraining sentence contexts, followed by a picture of the final noun of the sentence, which they had to name as fast as possible. In two separate blocks, participants were asked to read the sentences aloud or silently. The authors found that reading aloud significantly increased picture naming speed for highly constraining sentences, supporting the hypothesis that language prediction involves the production system [2]. However, it remains unclear whether this mechanism extends to second language (L2) speakers, who may rely less on production due to lower proficiency. To adapt the paradigm for L2 populations, we increased word presentation times to accommodate slower processing in a less-proficient language. We first tested the new slower timing with first language (L1) speakers, expecting to replicate the original interaction between reading mode and sentence constraint.

Method: We tested 28 native speakers of Spanish online using jsPsych and JATOS. Participants read highly and low constraining sentences in Spanish, aloud and silently, while we measured the naming times locked to the presentation of the final picture. Importantly, we increased the durations of several word and picture presentations relative to those used in the original study (Figure 1).

Results: A mixed-effect model indicated a significant effect of reading mode (faster naming times when reading aloud vs. silently) and of sentence constraint (faster naming times when reading highly vs. low constraining sentences) on naming times. However, no significant interaction effect was found.

Discussion: In contrast with the original experiment [1], in our study language production did not significantly increase picture naming speed for highly constraining sentences (Figure 2). We interpret this null effect as a consequence of having increased word presentation times. In this sense, we hypothesize that the facilitatory effect of overt production might be reserved for highly challenging conditions (e.g., speeded reading time). Longer presentation times might reduce task difficulty, thereby diminishing the need for anticipatory processing via the production system. To test this, we are currently replicating the original (faster) version of the task, with L1 speakers. Results from this replication will be presented at the conference. This line of research aims to identify the boundary conditions under which production supports prediction, which is a crucial step in understanding how and when prediction operates in language comprehension.

References

[1] Lelonkiewicz, J. R., Rabagliati, H., & Pickering, M. J. (2021). The role of language production in making predictions during comprehension. Quarterly Journal of Experimental Psychology, 74(12), 2193-2209.

[2] Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. Psychological Bulletin, 144(10), 1002-1044.

Figures

Figure 1. Trial presentation times used in the original study (left) and in the current experiment (right). Left figure adapted from Lelonkiewitz et al. (2021) [1].

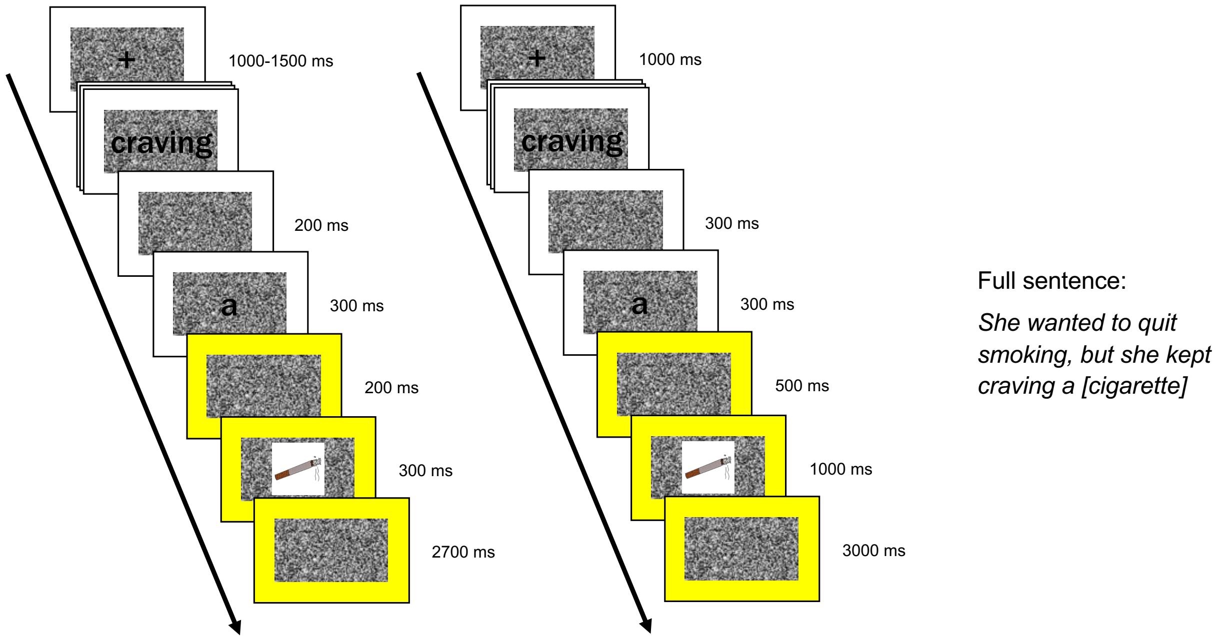
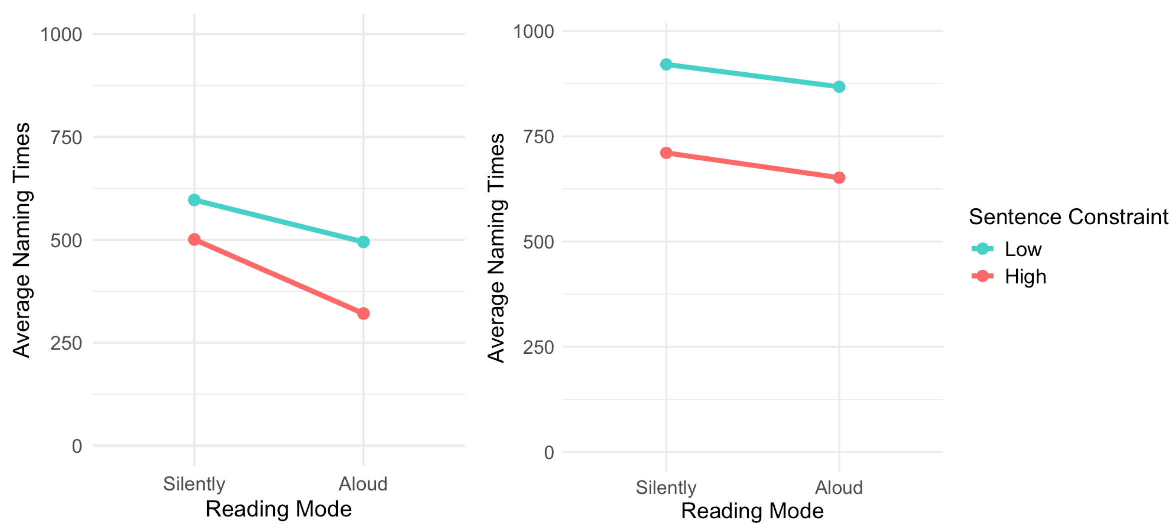


Figure 2. Results of the original study (left) and of the current experiment (right). Left figure adapted from Lelonkiewitz et al. (2021) [1].



Processing morphologically complex words: Insights from Russian

Anna Gupta, Carsten Eulitz
University of Konstanz
anna.gupta@uni-konstanz.de

Background

Complex words are decomposed into morphological constituents in the process of word recognition. Experimental research has focused on whether decomposition occurs or not, but what affects decomposition remains a pertinent question. It is well established in theoretical linguistics that complex words are arranged in a hierarchical fashion. Studies considering that word recognition is governed by hierarchical morphological structure tested this idea experimentally by investigating the concept of derivational depth and its effects on the processing of English and German words [1, 2, 3, 4]. Derivational depth is the degree of complexity of a word internal structure, i.e., the number of steps that are required to derive a complex word from its base. The primary goal of the current study is to expand on previous research and to explore the extent to which derivational depth affects the processing of complex words with overt affixation in Russian. Our secondary goal is to investigate whether the presence of inflectional endings in derived words interferes with the processing of derivational depth.

Method

We will conduct a delayed (long-lag) repetition priming experiment where primes and targets are separated by 5-7 intervening items. Participants will perform a visual lexical decision task. Stimuli include three sets of prime-target pairs in which derivational depth of primes has been manipulated: (1) one-step nouns (*1StepN*) derived from nouns in one step by adding a derivational suffix, *put-nik* ‘traveller’ ~ *put’* ‘way’; (2) one-step adjectives (*1StepA*) derived from nouns in one step by attaching a derivational suffix and an inflectional case/number marking ending, *shum-n-iy* ‘noisy’ ~ *shum* ‘noise’, and (3) two-step nouns (*2StepN*) formed from nouns via adjectives in two steps by means of two derivational suffixes, *chest-n-ost* ‘honesty’ ~ *chest’* ‘honour’ (see Table 1 for other examples). We included the *1StepN-2StepN* comparison to investigate the effects of derivational depth in words within the same part of speech. For this comparison, however, base forms used as targets are different. In the other comparison (*1StepA-2StepN*), base forms and thus targets are identical. This gives us the opportunity to explore the effects of depth but to control for potential effects of base properties such as frequency, length, and others.

Results

Our data collection will be completed in August 2025. Based on previous research, we expect larger processing costs reflected in reduced priming effects for two-step nouns compared to one-step nouns and one-step adjectives because two morphological operations (steps) are performed while decomposing two-step words resulting in lower activation of their bases.

Conclusion

We surmise that the processing of Russian complex words in our study will be affected by derivational depth (one-step vs. two-step), thus providing evidence for the role of hierarchical morphological structure. The findings from the study will have important implications for cross-linguistic generalizations.

References

[1] Meinzer, M., Lahiri, A., Flaisch, T., Hannemann, R., & Eulitz, C. (2009). Opaque for the reader but transparent for the brain: Neural signatures of morphological complexity. *Neuropsychologia*, *47*(8), 1964–1971.

[2] Pliatsikas, C., Wheeldon, L., Lahiri, A., & Hansen, P. C. (2014). Processing of zero-derived words in English: An fMRI investigation. *Neuropsychologia*, *53*, 47–53.

[3] Darby, J., & Lahiri, A. (2016). Covert morphological structure and the processing of zero-derived words. *The Mental Lexicon*, *11*(2), 186–215.

[4] Wheeldon, L., Schuster, S., Pliatsikas, C., Malpass, D., & Lahiri, A. (2019). Beyond decomposition: Processing zero-derivations in English visual word recognition. *Cortex*, *116*, 176–191.

Table 1

Stimuli Examples

	Condition	Target	Related Prime	Unrelated Prime
(1)	1StepN	школ-а *shkol-a* school-SG.F.NOM ‘school’	школь-ник *shkol-nik* school-er.SG.M.NOM ‘schoolboy’	лес-ник *les-nik* forest-er.SG.M.NOM ‘forester’
(2)	1StepA	вред *vred* harm.SG.M.NOM ‘harm’	вред-н-ый *vred-n-iy* harm-ful-ADJ.SG.M.NOM ‘harmful’	нерв-н-ый *nerv-n-iy* nerv-ous-ADJ.SG.M.NOM ‘nervous’
(3)	2StepN	вред *vred* harm.SG.M.NOM ‘harm’	вред-н-ость *vred-n-ost* harm-ful-ness.SG.F.NOM ‘harmfulness’	нерв-н-ость *nerv-n-ost* nerv-ous-ness.SG.F.NOM ‘nervousness’

The markedness effect on form-based predictions of sound and number: Evidence from a visual-world study

Daiwen Gong[1], Aine Ito[1]
[1] National University of Singapore
daiwengong@gmail.com; daiwen_gong@u.nus.edu

Background: Prediction plays a key role in comprehension, but it remains unclear whether listeners can predict the sound or number of upcoming words based on phonological/number agreement (e.g., “a/an … banana/apple” and “there is/are the … bowl/plates”). Prior research has yielded mixed findings on such predictions [1–3]. It has largely overlooked the role of markedness, which may influence predictive processing. In binary contrasts, the unmarked item lacks a distinctive property that is meanwhile present in its marked counterpart [4]. For instance, *an* carries an additional /n/ compared to *a*, so *an* is phonologically marked, and *a* is unmarked. Likewise, in the present tense, the copula verb *be* inflects for number: *is* is marked for singular number, while *are* is marked for plural. Marked forms, due to their specified features, tend to be better maintained in the focus of attention [5,6]. For example, comprehenders can identify agreement violations involving marked forms faster than those involving unmarked ones [7]. To investigate whether the markedness of lexical cues modulates predictive processing, we conducted a visual-world eye-tracking study focusing on native English speakers’ predictions of sound and number information.

Method: Forty participants viewed scenes with two images depicting nouns that varied in either initial sound (consonant vs. vowel) or number (singular vs. plural) while listening to sentences in which the indefinite article (*a*/*an*) or the copula verb (*is*/*are*) served as reliable cues to the upcoming noun (see Figure 1 for examples). In sound prediction trials, the unmarked (*a*) and marked (*an*) indefinite articles can be used to predict consonant-initial and vowel-initial nouns, respectively. In number prediction trials, the copula verbs, *is* and *are*, both marked for number, can be used to predict singular and plural nouns in there-be constructions. We conducted cluster-based permutation analyses (CPA) on fixation data (Figure 2). Due to an unexpected bias toward plural (vs. singular) images, a post hoc analysis was conducted to examine whether participants were more likely to fixate on plural images after hearing *are* compared to *is*, and vice versa (Figure 3). This analysis allowed us to test the effect of markedness while keeping the images constant.

Results: Results showed that participants were more likely to fixate on the target image than the competitor image before the target word was mentioned when the cue was marked (*an*) but not when it was unmarked (*a*) in sound prediction trials. In number prediction trials, the predictability effect was found for both *is* and *are*, with no significant difference between them. These findings suggest that cue markedness influences form-based predictions, with marked cues generating stronger predictive effects due to their specified phonological or morphological properties.

Conclusion: Our findings demonstrate that marked cues (*an* in sound prediction) are more likely to elicit predictions than unmarked cues (*a* in sound prediction). When two marked forms are equally specified (*is* and *are* in number prediction), both can facilitate prediction. Marked forms that carry explicit surface features (e.g., an additional /n/) may be more easily detected and maintained in attentional focus, hence triggering stronger prediction.

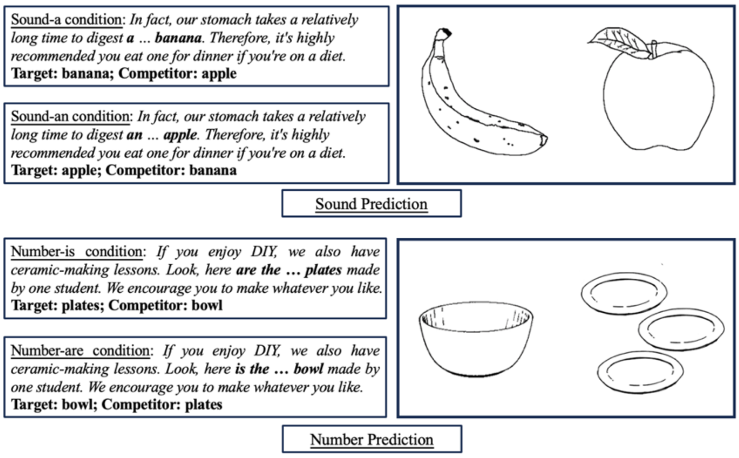


Figure 1 | Examples of sentences and images used in Sound-Prediction and Number-Prediction.

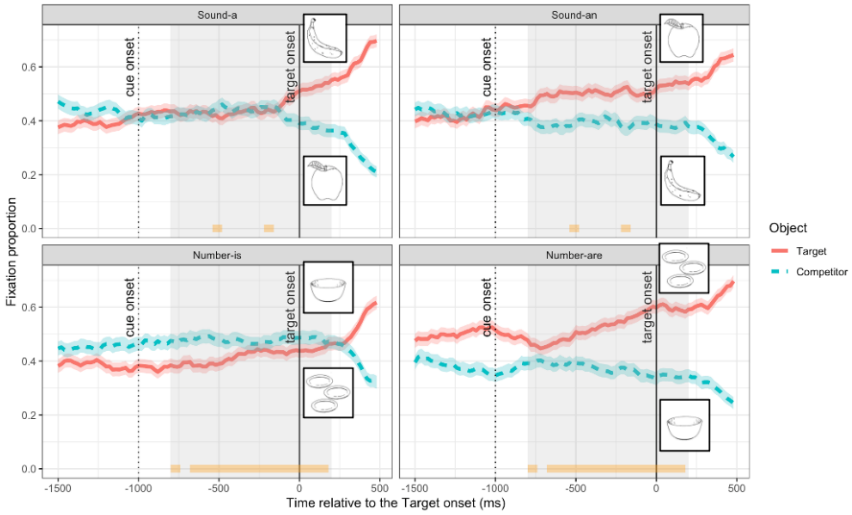


Figure 2 | Proportion of fixation to target and competitor images over time for each condition. The gray shaded area indicates the critical time window. Time 0 indicates the onset of the target word. The yellow lines at the bottom indicate clusters in which the target image attracted significantly more fixations in the an/are conditions than in the a/is conditions.

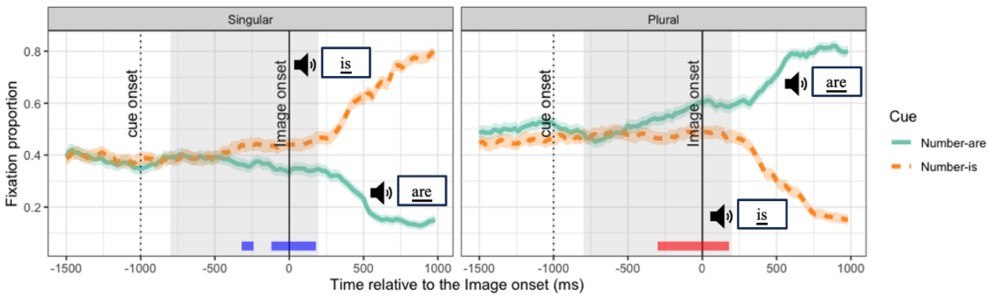


Figure 3 | Proportion of fixation to target and competitor images over time for each condition. The gray shaded area marks the critical time window, with Time 0 indicating the onset of the target word. The blue and red lines at the bottom represent significant negative and positive clusters, respectively.

References

[1] Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. Journal of Experimental Child Psychology, 173, 351–370.

[2] Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. Cognition, 146, 349–370.

[3] Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. Frontiers in Psychology, 6.

[4] Jakobson, R. (1957). Shifters, verbal categories, and the Russian verb. Russian Language Project, Department of Slavic Languages and Literatures, Harvard University.

[5] Wagers, M., & McElree, B. (2022). Memory for linguistic features and the focus of attention: Evidence from the dynamics of agreement inside DP. Language, Cognition and Neuroscience, 37(9), 1191–1206.

[6] Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. Quarterly Journal of Experimental Psychology, 67(7), 1274–1304.

[7] Alemán Bañón, J., & Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: An event-related potential investigation. Language, Cognition and Neuroscience, 31(10), 1273–1298.

Tacit knowledge of stylistic variation: Evidence from (ING) perception in native and non-native listeners

Aini Li[1] and Lacey Wade[2]

1: Department of Linguistics and Translation, City University of Hong Kong; **2**: Department of Linguistics, University of Kansas

Corresponding author: Aini Li (ainili@cityu.edu.hk)

In quantitative sociolinguistics, it is well-established that language variation is stylistically conditioned, with prestige variants typically favored in formal speech and vernacular features in casual styles [1, 2]. However, it remains unclear the extent to which native and non-native speakers are aware of the stylistic conditioning on variation, and whether/how they utilize such knowledge when processing variation in real time. This study probes listeners' tacit knowledge of stylistic conditioning on variation by examining their behaviors of uncertainty resolution in speech perception, using the sociolinguistic variable (ING)—the alternation between *-ing* and *-in* as in *thinking ~ thinkin'*. We compare native American English listeners with Cantonese-speaking second-language English (non-native) listeners to explore how language experience influences sociolinguistic processing. The findings can shed light on how different listener groups process variation in real-time language use and the mechanisms underlying the acquisition of sociolinguistic competence.

Two variant identification experiments (within-subjects design) compared English L1 (Exp 1, N =60) and L2 listeners (Exp 2, N=60). Both groups reported whether they heard *-in'* or *-ing* when the actual variant was masked. 24 critical and 72 filler sentences were constructed and recorded by a female native American English speaker in both casual and formal sentence frames, matched for duration. For critical items, the ING portion in all targets (e.g., carvING) was replaced with white noise of identical length and amplitude, to create uncertainty and to prevent listeners from relying on variant relative durational cues in their uncertainty resolution; noise bursts were also randomly inserted throughout the sentence to mask the manipulation. For filler items, similar manipulation was performed except that no ING-suffixed words were included. In a norming study, native listeners (N=10) indicated on a slider how likely the sentence they heard was produced under a formal style to verify the effectiveness of style manipulation. In Exp 1, native listeners (N=60) heard sentences in both styles (one guise per sentence and counterbalanced across participants). After hearing each sentence, listeners were asked to choose out of two written word options the one they thought they heard (e.g., *sleepin*' vs. *sleeping*). Exp 2 had the same procedure except the non-native listeners (N=60) first completed the variant identification task and then rated sentence formality, with an additional LEAP-Questionnaire [3] to assess their Cantonese and English proficiency.

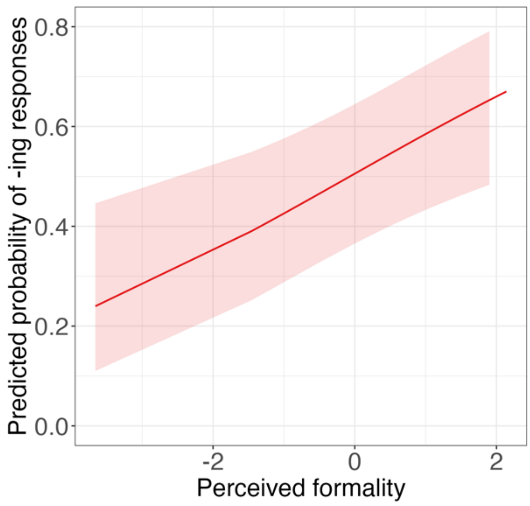
Results based on mixed-effects regression modeling found that for native listeners, there was a significant effect of style rating ($p < 0.01$): Native listeners were significantly less likely to perceive the uncertain white noise as *-ing* when the utterance was perceived as casual (i.e. lower formality rating) (Figure 1). This suggests that native listeners utilize co-varying stylistic cues to infer their variant choices in perception. In comparison, non-native listeners did not show a significantly stronger bias towards perceiving *-ing* when the sentence was rated as more formal ($p = 0.77$) (Figure 3). Analysis based on their English proficiency found that while English proficiency correlates with their sensitivity to stylistic differences, it does not correlate with their sensitivity to the stylistic covariation of (ING). As a result, both groups can distinguish a formal sentence from a causal sentence, likely due to presumably universal cues to style, like vowel reduction (Figure 2 vs. Figure 4), but only native listeners possess the tacit sociolinguistic knowledge of the stylistic conditioning on (ING) variation. These findings indicate that even advanced L2 learners may not automatically develop native-like sociolinguistic competence and bring up questions about the time course of developing sociolinguistic competence in an L2.

References

[1] Labov, W. (2010). Principles of linguistic change, cognitive and cultural factors (volume3). *Malden, MA: John Wiley & Sons*.

[2] Schilling, N. (2013). Investigating stylistic variation. *The handbook of language variation and change*, 325-349.

[3] Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2020). Language Experience and Proficiency Questionnaire. *Bilingualism: Language and Cognition, 23(5), 945-950.*



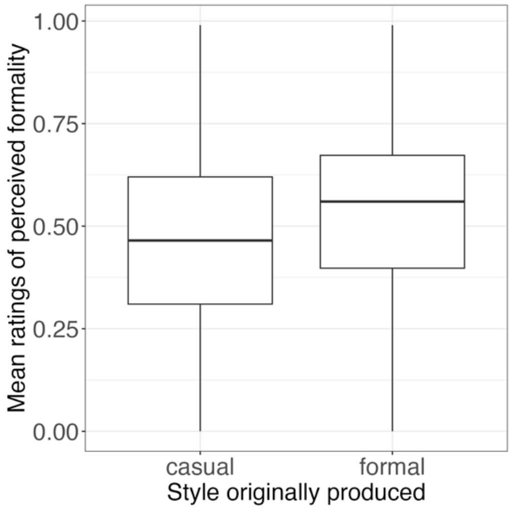
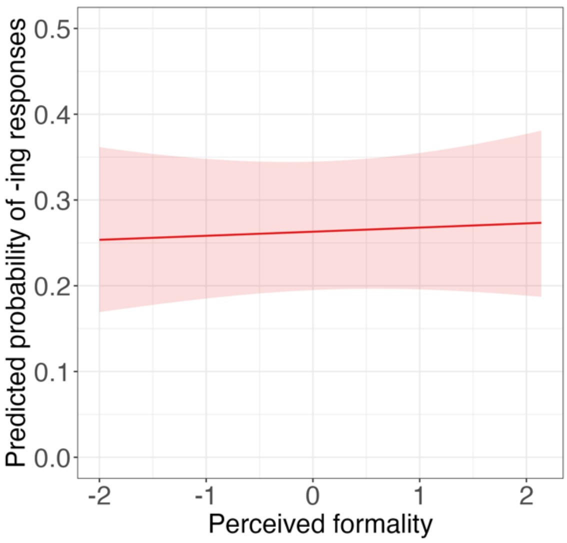


Figure 1

Figure 2

Figure 1 (left): Native American English listeners: Results of variant identification
(note: the higher the rating, the more likely listeners judged the guise as formal)
Figure 2 (right): Native American English listeners: Results of the style rating task



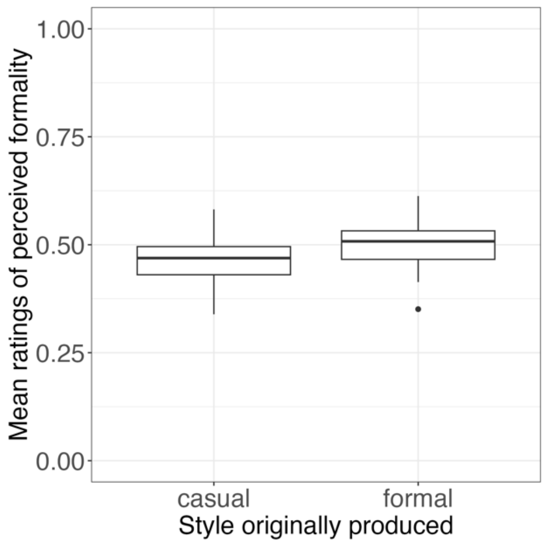


Figure 3

Figure 4

Figure 3 (left): Non-native listeners: Results of style rating
Figure 4 (right): Non-native listeners: Results of variant identification
(note: the higher the rating, the more likely listeners judged the guise as formal)

Do code-switching and sociolinguistic environment modulate the processing of ambiguous pronouns? Insights from Italian-English bilinguals

Camilla Masullo[1] & Beatrice Giustolisi[1]

[1]University of Milano-Bicocca

Corresponding author's email: camilla.masullo@unimib.it

Code-switching (CS) is the bilingual practice of alternating between languages within a single interaction or utterance [1]. While CS was found to be employed for socio-pragmatic reasons [2] and to facilitate low-frequency word prediction [3], as far as we know, no study has explored its effects on critical structures in bilingual processing, such as syntax-pragmatics interface phenomena [4]. We aim to do so by examining whether the processing of ambiguous pronouns in sentences presented in Italian and sentences that involve CS from English to Italian differs in terms of referent selection and RTs. Contrary to English, which tends to refer ambiguous anaphoric pronouns to the subject of the preceding sentence, Italian, a pro-drop language, presents different tendencies for null and overt ambiguous pronouns, with null pronouns more frequently referring to the subject of the preceding sentences and overt pronouns more often referred to the object [5]. Furthermore, given the role of sociolinguistic context in bilingual language processing [6] and the susceptibility of ambiguous pronouns to language attrition [7], we seek to explore whether linguistic environment (i.e., living in Italy VS. in an anglophone country) influences ambiguous pronouns resolution, both jointly and independently of CS.

Participants were recruited on Prolific and were pre-screened using its built-in screeners and an ad-hoc questionnaire according to the following criteria: (i) being Italian-English bilinguals, with Italian as a native language and an estimated level of at least B2 CEFRL English proficiency and residing (ii) either in Italy (Group 1, N = 46) or in an anglophone country (Group 2, N = 47). English proficiency was checked using Cambridge's online English test. In the experiment, participants in both groups listened to the same auditory stimuli: each test item began with a sentence introducing two animate characters of the same grammatical gender and with comparable lexical frequency [8, 9], followed by a second sentence containing an ambiguous pronoun referring to the first part of the test item. Each test item occurred in one of four different conditions (see Table 1), for a total of 240 test items. Half of the test items were entirely in Italian, while the other half involved CS. Both conditions included an equal number of null and overt ambiguous pronouns. The 240 sentences were equally distributed in 4 lists. Each list contained an equal number of sentences with/without CS, and with a null/overt pronoun. Each participant received 1 of the 4 lists and 60 additional fillers.

Results from GLME and LME models show no effect of CS on antecedent selection. We found that the ambiguous pronoun condition significantly influenced antecedent selection, with null pronouns more often resolved to the subject and overt pronouns to the object (t = 12.03). As Figure 1 shows, this tendency was significantly more pronounced for participants in Italy than those in an anglophone country (t = - 6.69). Additionally, ambiguous pronouns resolved to the object led to longer RTs than those referring to the subject, especially in the case of null pronouns (t = 4.52). Our results reveal no effect of CS on resolving ambiguous pronouns but highlight the role of the sociolinguistic environment in bilingual language processing. Compared to Italian-English bilinguals residing in Italy, Italian-English bilinguals living in an anglophone context showed a weaker tendency to link ambiguous null pronouns to the subject, which may point to potential syntactic attrition [7].

References

[1] Myers Scotton, C., & Ury, W. (1977). Bilingual Strategies: The Social Functions of Code-Switching. Int. J. Sociol. Lang., 1977(13).

[2] Myslín, M., & Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language,* 91(4).

[3] Tomić, A., & Valdés Kroff, J. R. (2022). Expecting the unexpected: Code-switching as a facilitatory cue in online sentence processing. *Biling.: Lang. Cogn.,* 25(1).

[4] Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguist. Approaches Biling.*,

[5] Carminati, M.N. (2002). *The processing of Italian subject pronouns*. University of Massachusetts at Amherst dissertation.

[6] Beatty-Martínez, A.L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M.T., Guzzardo Tamargo, R.E., & Kroll, J.F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *J. Exp. Psychol.*: Learn. Mem. Cogn., *46*(6).

[7] Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *Int. J. Biling.*, *8*(3).

[8] Crepaldi, D., Amenta, S., Pawel, M., Keuleers, E., & Brysbaert, M. (2015, September). SUBTLEX-IT. Subtitle-based word frequency estimates for Italian. In *Proc. Ann. Meet. Italian Assoc. Exp. Psychol*.

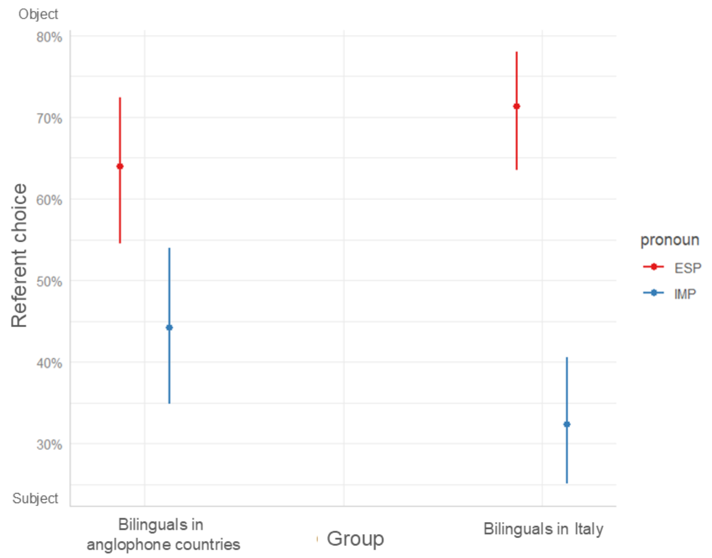
[9] van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Q. J. Exp. Psychol.*, 6.

Table 1. Experimental manipulation and stimuli examples.

		Code-switching	
		Yes	No
Pronoun	Null	N=60 (see 1a)	N=60 (see 1b*)
	Overt	N=60 (see 1c**)	N=60 (see 1d)

(1a/c)** *The woman greets the girl solo quando **[lei]**** attraversa la strada*
The woman.NOUN.SBJ. greets the girl.NOUN.OBJ only when **[she.3S.SUBJ.]**** crosses the road
(1d/b*) *La donna saluta la ragazza solo quando **[lei]**** attraversa la strada*
The woman.NOUN.SBJ. greets the girl.NOUN.OBJ only when **[she.3S.SUBJ.]*** crosses the road

Figure 1. Predicted probabilities of referent choices for bilinguals residing in Italy vs. in an anglophone country.



Language of Study

The experiment contained stimuli with ambiguous referential pronouns, both in Italian and with CS from English to Italian. Italian is a pro-drop language where ambiguous referential pronouns can be either overt or null pronouns (examples 1a, 1b, and 1c).

(1a) Marco chiama Luca appena ***lui*** entra nel giardino
Marco.NOUN.M calls Luca.NOUN.M as soon as **he.3SG.M** enters the garden
'Marco calls Luca as soon as he enters the garden'

(1b) Marco chiama Maria appena ***lui*** entra nel giardino
Marco.NOUN.M calls Maria.NOUN.F as soon as **he.3SG.M** enters the garden
'Marco calls Maria as soon as she enters the garden'

(1c) Marco chiama Luca appena entra nel giardino
Marco.NOUN.M calls Luca.NOUN.M as soon as **Ø** enters the garden
'Marco calls Luca as soon as (she/he) enters the garden'

In 1a, the pronoun "lui" (i.e., "he") is ambiguous as it can refer to both the subject and the object of the preceding sentence, as the two referents share the same grammatical gender. The ambiguous pronoun could also refer to a potential third grammatically masculine referent not mentioned in the preceding context. Ambiguity might be reduced in cases like (1b), where only one of the antecedents presents the same grammatical gender as the ambiguous pronouns (i.e., "Marco", masculine). Nevertheless, the option for a potential third grammatical masculine referent remains present. In 1c, the ambiguous pronoun is null, and it can refer to either one of the preceding antecedents or a third unmentioned referent of any grammatical gender.

Linguistic conditioning to change the emotional and gustatory meanings of new words

Mizuki Yoshio & Toshimune Kambara
Department of Psychology, Hiroshima University
m252758@hiroshima-u.ac.jp

Background: Previous research has suggested that some of the figurative languages are rooted in embodied cognition. For example, taste-related words are often used to convey emotional valence (e.g., “sweet talk” for positive expressions; “salty remark” for negative ones). Prior studies have examined the linkage between valence and taste judgements of words [1, 2]. However, no experimental studies have yet examined the transfer of gustatory meanings of real valenced (i.e., positive or negative) words to pseudowords using linguistic conditioning. This study examined whether words with positive and negative valence can influence affective and gustatory meanings through semantic conditioning in two experiments.
Method: The methodology was modeled after several previous studies on evaluative conditioning that used verbal stimuli [3, 4]. The first experiment (N = 126) consisted of five phases: The first rating phase, a conditioning phase, the second rating phase, a contingency awareness questionnaire [5, 6], and a contingency memory task [6, 7]. In the rating phases, participants evaluated pseudowords in terms of emotional and gustatory meanings (familiarity, valence, arousal, sweetness, saltiness, sourness, bitterness, umami, and spiciness). During the conditioning phase, a forward conditioning procedure was used: participants were visually presented with the pseudowords (conditioned stimuli: CS), followed by auditorily presented positive or negative real words (unconditioned stimuli: US). To strengthen CS-US associations, participants were also asked to repeat the US aloud. In the second experiment (N = 128), we additionally assessed participants’ preference toward each taste (i.e., sweetness, saltiness, sourness, bitterness, umami, and spiciness).
Results: The analyses were conducted using the cumulative link mixed model (CLMM). Each model included condition (PW vs. NW) and phase (the first vs. second rating phase) as fixed effects, with random intercepts for participants and CSs, and random slopes for participants. In both experiments, significant main effects of phase and significant interactions between phase and condition were observed (all ps < .05). Post-hoc test showed that pseudowords paired with positive USs were rated as more positive, sweeter, and savorier (higher umami) in the second rating phase compared to the first rating phase (all ps < .001). In contrast, pseudowords paired with negative USs were rated as saltier, sourer, bitterer, and spicier (all ps < .001). In the second experiment, participants' taste preferences showed the following patterns: fewer than 20% reported disliking sweetness, saltiness, or umami; about half reported disliking sourness or spiciness; and approximately 80% reported disliking bitterness. Descriptive statistics on contingency awareness indicated that, in each experiment, over 40% of participants recognized the CS-US contingencies. Additionally, the accuracy of the memory task was greater than 60%.
Discussion: The results revealed that perceived gustatory meanings of valenced words can change those of pseudowords through semantic conditioning. Additionally, the results of the second experiment suggest that personal likings toward a specific taste did not seem to influence the gustatory-meaning transference. Taken together, emotional words can elicit certain taste impressions, regardless of people’s preference. Metaphors using taste words may be observed because people share embodied representations of emotions.

References

[1] Chan, K. Q., Tong, E. M. W., Tan, D. H., & Koh, A. H. Q. (2013). What do love and jealousy taste like? *Emotion, 13*(6), 1142–1149. http://dx.doi.org/10.1037/a0033758

[2] Zhou, Y., & Tse, C.-S. (2020). The taste of emotion: Metaphoric association between taste words and emotion/emotion-laden words. *Frontiers in Psychology, 11*, 986. https://psycnet.apa.org/doi/10.3389/fpsyg.2020.00986

[3] Ando, M., & Kambara, T. (2023). Japanese written pseudowords can be conditioned to Japanese spoken words with positive, negative, and active emotions. *Cognitive Processing, 24*(3), 387–413. https://doi.org/10.1007/s10339-023-01138-0

[4] Staats, C. K., & Staats, A. W. (1957). Meaning established by classical conditioning. *Journal of Experimental Psychology, 54*(1), 74–80. https://doi.org/10.1037/h0047716

[5] Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *The Quarterly Journal of Experimental Psychology, 63*(12), 2313–2335. https://doi.org/10.1080/17470211003802442

[6] Richter, J., & Gast, A. (2025). Simultaneous pairing increases evaluative conditioning: Evidence for the role of temporal overlap but not of onset synchrony. *Journal of Experimental Social Psychology, 116*, 1–15. https://doi.org/10.1016/j.jesp.2024.104689

[7] Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology, 97*(3), 404–420. https://doi.org/10.1037/a0016196

Notes: Linguistic features of Japanese and stimuli.

- Japanese has a vocabulary to express positive or negative contents by taste-related words. Specifically, positive contents can be metaphorically expressed as “甘い話 (“sweet talk”)” and “うまい話 (“savory talk”).” In contrast, negative content can be expressed as “辛口の意見 (“spicy remark”),” and “口を酸っぱくして言う (“to say something with one’s sour mouth”; to tell someone the same thing over and over again).”
- Conditioned stimuli were the following four pseudo words, which consist two moras and are written in Katakana character (one of the Japanese phonograms): ワユ (*wayu*), ヌヨ (*nuyo*), レヘ (*rehe*), and ソヒ (*sohi*).
- Unconditioned stimuli were all real Japanese words including emotional-evocative (e.g., 健康 means “healthy” in English; 汚い means “dirty” in English) and emotional-expressive words (e.g., 幸せ means “happiness” in English; 悲しみ means “sadness” in English).

Do Mandarin Speakers Retain Categories for Unaccusativity?

Kaiying K. Lin
Assistant Research Fellow, Institute of Linguistics, Academia Sinica (Taiwan)
limkhaiin@as.edu.tw

The Unaccusativity Hypothesis (Perlmutter, 1978) proposes a binary division of intransitive verbs into unergative and unaccusative classes, depending on whether their subjects behave like agents or patients syntactically. Later proposals (e.g., Sorace, 2000) suggest that verb behavior is not strictly binary but follows a semantic gradient, with event types such as change-of-location and motion predicting syntactic realizations probabilistically. This study tests whether Mandarin speakers (1) maintain binary syntactic classes, or (2) retain multiple semantic event type categories and associate these with syntactic classes. If such categories are psychologically real, speakers should be able to quickly retrieve and apply them to novel verbs in a sentence after brief exposure (Rosch & Mervis, 1975; Rosch, 1975; Murphy, 2002). However, our results suggest that Mandarin speakers do not robustly retrieve either (1) or (2), which indicates that Mandarin speakers may not categorize intransitive verbs using criteria related to unaccusativity.

Experiment. We conducted two acceptability judgment tasks, each with around 30 Mandarin-speaking adults. Participants listened to a description of a novel verb's meaning during the exposure phase, and during the test phase, they rated the verbs within diagnostic sentences (indicating categories) on a 0-5 scale. Sixteen nonce verbs (four for each category) were selected using a norming test, to approximate Sorace's (2000) semantics-based event types and prototypicality (core) levels: change-of-location (core unaccusative), change-of-state (non-core unaccusative), motional (non-core unergative), and non-motional (core unergative) verbs. Each verb was paired with two widely accepted diagnostic sentences (durative and postverbal-subject diagnostics) in Mandarin, yielding 32 critical items. The second experiment replicated the first but included an additional introduction during the test phase. We hypothesized that if speakers retain only binary syntactic classes, both unaccusative and unergative verbs should be easily retrieved within their respective groups, resulting in high accuracy in a diagnostic. On the other hand, if speakers retain multiple semantic event type categories, ratings for core verbs should be distinct from those for non-core verbs, and align with expected patterns in Mervis and Rosch (1981) that novel prototypical (core) items are retrieved more accurately and learned more readily than novel non-prototypical (non-core) items.

Results. We normalized ratings into z-scores and analyzed the data using Mixed Linear Regression models with verb classes (or categories) and diagnostics as independent variables. Neither experiment revealed sharp rating differences between core and non-core verbs within the unergative and unaccusative classes (Figures 1 and 2), suggesting that speakers do not robustly retain four semantic categories. Although statistically significant interactions between binary syntactic categories and diagnostics were observed (Experiment 1: $\beta = 0.8677$, $SE = 0.124$, $p < .001$; Experiment 2: $\beta = 1.238$, $SE = 0.136$, $p < .001$), accuracy based on ± 0.5 z-score thresholds remained nearly random (40–60%), offering only weak evidence for a potential unergative category. Overall, our findings suggest that while unaccusativity remains theoretically motivated for structural derivation, Mandarin speakers may not internally categorize intransitive verbs along these unergative–unaccusative or core–non-core lines. (489 words)

Figure 1, Verb ratings from Experiment 1. (N=29, The durative diagnostic should be rejected with unaccusative, but accepted with unergative-targeted verbs, while the postverbal-subject diagnostic should be accepted with unaccusative but rejected with unergative-targeted verbs)

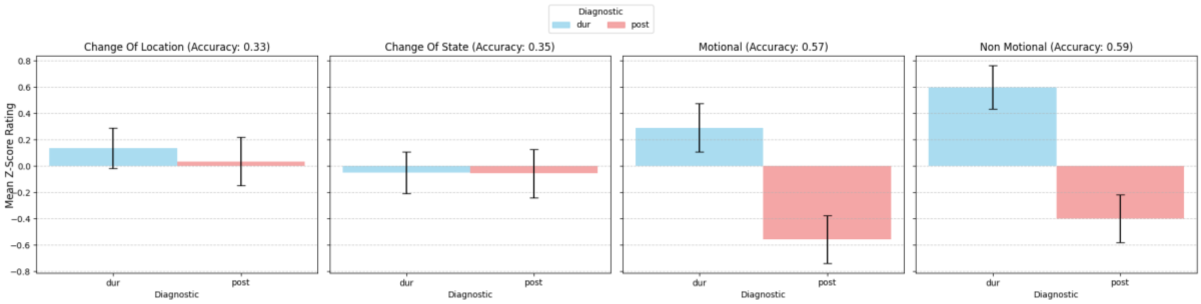
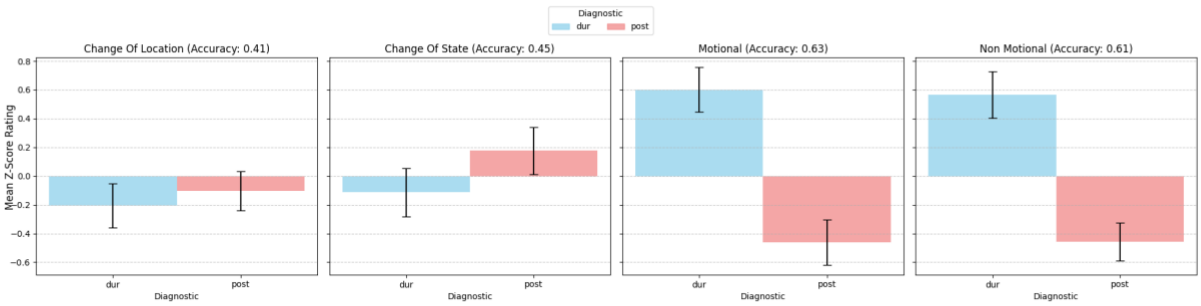


Figure 2, Verb ratings from Experiment 2 (N=31).



(1) (Sorace, 2000)

Core unaccusative verb	change of location	(unaccusative)
	change of state	(unaccusative)
	existence of state	
	uncontrolled process	
	controlled motional process	(unergative)
Core unergative verb	controlled non-motional process	(unergative)

(2) Samples of a non-motional unergative-targeted verb Si4 (in Mandarin)

Experiment	**Meaning**	**Exposure Phase**	**Test Phase**
1	to emit light / to glitter	Participants heard: *"This new verb is called **si4**. It means there is a person. The person sends out some light. The person glitters."*	Participants were asked to judge the acceptability of a diagnostic sentence without introduction: *"Si4-le yi-ge ren"* ('One person si4-ed', postverbal-subject diagnostic).

Selected Reference: Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology,* 32, 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513 Murphy, G. L. (2002). The big book of concepts. Boston Review. https://doi.org/10.7551/mitpress/1602.001.0001 Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In J. Jaeger (Eds.), *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society* (pp.157-190). Berkeley Linguistics Society. Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology*: *General,* 104, 192–233. doi: 10.1037/0096-3445.104.3.192 Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605. doi:10.1016/0010-0285(75)90024-9 Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76(4), 859–890. DOI: doi.org/10.2307/417202.

2	to emit light / to glitter	Participants heard: *"This new verb is called **si4**. It means there is a person. The person sends out some light. The person glitters."*	Participants heard: '*We can imagine a boy drinking magic water and his body is different now. Can we say: Si4-le yi-ge ren?'*

A Corpus of Joint EEG and Self-Paced Reading of Natural Dutch Texts

Sara Møller Østergaard, Bruno Nicenboim
Department of Cognitive Science and Artificial Intelligence, Tilburg University
s.m.ostergaard@tilburguniversity.edu

Background: There is an increasing interest in natural stimuli and experimental setups in psycholinguistic research, e.g., studies of reading times on natural texts [1], electroencephalographic (EEG) studies using all words in the sentence rather than single target words in hand-crafted contexts [2], and coregistration of EEG and eye-tracking [3, 4]. We present an open-access corpus of joint EEG and self-paced reading (SPR) of natural, medium-length, Dutch texts. The corpus contributes to a small collection of corpora with simultaneous recording of reading times and EEG. While this is often achieved using eye-tracking, the novel combination of SPR and EEG offers methodological advantages, particularly in aligning neural signals with word-level processing. Using SPR instead of eye-tracking eliminates parafoveal effects and allows EEG signals to be time-locked to word onsets, making it compatible with classical event-related potential (ERP) analyses. Furthermore, the inclusion of longer, natural texts enables novel analyses of reading and sentence processing.

Method: The corpus contains joint SPR and EEG recordings from 71 participants (47 female, mean age = 20.31) reading eight Dutch texts of approx. 800 words each. The linguistic stimuli are naturally occurring texts of different genres that were chosen based on overall fluency and comprehensibility. Every participant read seven texts using a SPR paradigm with central presentation and a single text in rapid serial visual presentation (RSVP). Bayesian hierarchical models were fit to validate the corpus by replicating surprisal effects on reading times and the N400 ERP component using log-probability and frequency of the given word as the predictors. The N400 was quantified as the average amplitude in centroparietal channels between 300 ms and 500 ms after the word onset. Log-probabilities of the words were obtained using unidirectional language models provided with the entire context of the texts preceding the word. The probabilities were extracted from four distinct models (see Table 1), and the average log-probability was used as the independent variable in the models. Word frequencies were obtained from the SUBTLEX-NL corpus [5]. All words were included in the statistical analysis.

Results: Both log-probability and word frequency had a negative effect on reading times, meaning that more probable and more frequent words elicited a faster reading pace, replicating previous findings [1, 4]. As expected [2, 4, 6], a positive effect of log-probability on N400 amplitude was found, indicating that less probable words elicit a more negative amplitude in the N400 time window. Previous studies have found either a positive or no N400 frequency effect [4, 6]. In contrast, we observed a negative word frequency effect on the N400 amplitude, with more frequent words eliciting a more negative N400 amplitude.

Use of the Corpus: Our corpus enables a range of future analyses. The simultaneous recording of reading times and EEG offers possibilities for exploring mechanisms underlying both behavioral and neurological responses during reading. The corpus allows for validation and exploration of EEG signal and ERPs for texts with longer linguistic dependencies. Additionally, the nature of the corpus makes it a useful asset in the development of computational cognitive models of reading. The corpus is available at DataverseNL [7].[1]

[1] Available from August 2025 at https://doi.org/10.34894/0O5XQ7

Table 1: Overview of Hugging Face models used for the analysis.

Hugging Face Reference
GroNLP/gpt2-small-dutch [8]
GroNLP/gpt2-medium-dutch-embeddings [8]
yhavinga/gpt2-large-dutch
yhavinga/gpt-neo-125M-dutch

Figure 1: Average SPR ERPs in centroparietal channels for content words with a log-probability <25% quantile, >75%, and between 25%-75%.

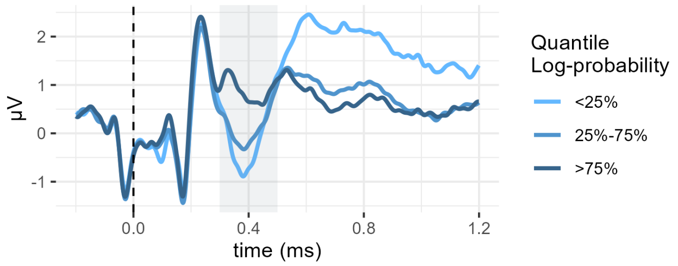


Table 2: Regression coefficients for models of reading times (only SPR), N400 amplitude (SPR and RSVP separately). All predictors were standardized. 95% credible intervals are reported in parentheses.

Coefficient	Reading Time	N400 (SPR)	N400 (RSVP)
Log-probability	$-0.02\ (-0.03, -0.01)$	$0.56\ (0.35, 0.77)$	$0.52\ (0.19, 0.85)$
Zipf Frequency	$-0.07\ (-0.08, -0.06)$	$-0.93\ (-1.22, -0.64)$	$-0.91\ (-1.20, -0.62)$

References

[1] Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., & Fedorenko, E. (2021). The Natural Stories corpus: A reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, *55*(1), 63–77. https://doi.org/10.1007/s10579-020-09503-7

[2] Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

[3] Hollenstein, N., Troendle, M., Zhang, C., & Langer, N. (2020). ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 138–146. https://aclanthology.org/2020.lrec-1.18/

[4] Frank, S. L., & Aumeistere, A. (2024). An eye-tracking-with-EEG coregistration corpus of narrative sentences. *Language Resources and Evaluation*, *58*(2), 641–657. https://doi.org/10.1007/s10579-023-09684-x

[5] Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

[6] Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1648–1662. https://doi.org/10.1037/xlm0000128 frequency effects.

[7] Østergaard, S., Lichtenberg, L., Boon, L., & Nicenboim, B. (2025). *EEG and Self-Paced Reading of Natural, Dutch Texts*. https://doi.org/10.34894/0O5XQ7

[8] de Vries, W., & Nissim, M. (2020). As good as new. how to successfully recycle english gpt-2 to make models for other languages.

How do Mandarin Chinese speakers prepare the form and content of their answers in turn-taking conversation.

Weijia Hu[1], Huanhuan Yin[1], Martin Pickering[2]
[1 2] Department of Psychology, University of Edinburgh
w.hu-11@sms.ed.ac.uk

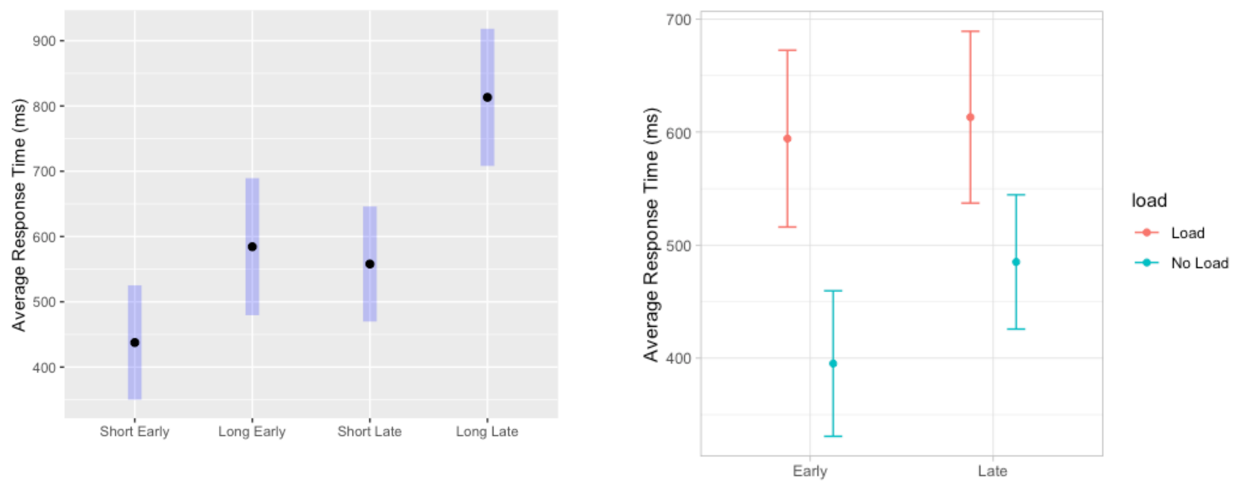
Background: During conversational turn-taking, interlocutors alternate between speaking and listening with gaps of around 200 ms—considerably shorter than the ~600 ms typically required to produce a single word [1, 2]. To achieve such rapid transitions, speakers likely begin planning responses while still comprehending [3]. However, it remains unclear whether this early planning involves only conceptual content or also phonological form, as existing findings are mixed [4, 5]. Given that Mandarin Chinese may select syllables as the first units of phonological planning while English primarily relies on individual phonemes, Mandarin speakers may engage in different phonological processing strategies due to this key difference [6]. This study investigates whether Mandarin Chinese speakers prepare both content and form early, and whether this preparation is constrained by verbal working memory. Previous research has shown that utterance length affects production time, with longer utterances typically requiring more planning [7]. Therefore, in Experiment 1, if Mandarin speakers prepare both content and form early, we expected an interaction where the disadvantage of producing longer answers would be reduced when critical information appears early, as speakers would have more time to prepare the phonological form in advance. Given that phonological planning is sensitive to working memory constraints [8, 9], Experiment 2 examined whether a concurrent verbal memory load disrupts early form preparation. We predicted an interaction where load would have a larger impact when critical information appears early, as maintaining early-prepared phonological representations in memory would be particularly demanding.
Method: In two question-answering experiments with native Mandarin Chinese speakers, we manipulated the position of critical information (early vs. late) within questions—for example, "*Which country has the Great Wall and is in Asia?*" vs. "*Which country is in Asia and has the Great Wall?*" , where "*the Great Wall*" serves as the critical information to the answer "*China*." Experiment 1 (n = 88) additionally manipulated answer length (short vs. long), with answers like "*China*" as short and "*The University of Edinburgh*" as long [5]. In Experiment 2 (n = 88), participants either answered questions normally (no-load condition) or while maintaining six unrelated Chinese characters in memory (load condition), which they recalled after responding. Response latencies were analysed in both experiments using linear mixed-effects models.
Results: Both experiments showed significant effects of critical information, with faster responses when it appeared early versus late. In Experiment 1, participants responded faster when producing short versus long answers. Crucially, we found a significant interaction: the answer length effect was larger when critical information appeared late. In Experiment 2, cognitive load increased response latencies, with an interaction where the load effect was significantly larger when critical information appeared early rather than late.
Discussion: Our findings demonstrate that Mandarin Chinese speakers prepare both the conceptual content and the phonological form of their responses early. However, early phonological planning is constrained by cognitive resources [8, 9]. When verbal working memory is limited, speakers do not always prepare the form early—even when critical information is available—suggesting that phonological planning is cognitively demanding. Therefore, speakers may plan their responses strategically, balancing the benefits and costs of early preparation. Thus, although Mandarin speakers may engage in earlier form preparation than English speakers, both groups require more time to prepare longer, phonologically complex responses [5].

References

1. Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G.,Rossano, F., De Ruiter, J. P., Yoon, K., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106
2. Indefrey, P., & Levelt, W. J. M. (2004b). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001
3. Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00731
4. Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*(1). https://doi.org/10.1038/srep12881
5. Corps, R. E., & Pickering, M. J. (2024). The role of answer content and length when preparing answers to questions. *Scientific Reports, 14*(1), 17110. https://doi.org/10.1038/s41598-024-68253-6)
6. O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282–302. https://doi.org/10.1016/j.cognition.2010.01.001
7. Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233. https://doi.org/10.1016/0749-596x(91)90004-4
8. Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*(1), 50–68. https://doi.org/10.1037/a0014411
9. Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *Quarterly Journal of Experimental Psychology, 70*(4), 811–831. https://doi.org/10.1080/17470218.2016.1167926

Figure

Mean response latency for Experiment 1 (left) and Experiment 2 (right). Error bars represent in purple standard error from the mean.



Polysemy and acoustic duration: Different senses come with different durations

Dominic Schmitz
English Language and Linguistics, Heinrich Heine University Düsseldorf, Germany
dominic.schmitz@uni-duesseldorf.de

Background: Research on lexical ambiguity and phonetic realisation has shown that morphological differences in homophonous elements influence fine-phonetic detail. For example, different types of word-final /s/ in English and German [e.g. 1, 2, 3] affect subphonemic acoustic duration, a finding accounted for by the discriminative lexicon model [DLM; 4, 5]. However, almost all studies in this area focus on homonymy, while polysemy has only recently been examined in this context. [6] found that in a reading task, speakers produced the word-final *-er* suffix (/ɐ/) significantly shorter when producing the gender-specific sense of a masculine role noun in German as opposed to its generic sense counterpart. The present study a) extends this research by finding similar durational differences using more natural speech showing that the effect is robust and not a mere artefact of the reading task, and b) analyses found durational differences using the DLM.

Method: The phonetic realisation of generic and gender-specific masculines was examined using a recall task. The 20 target items (10 female, 10 male stereotypicality; adopted from [6]) all ended in *-er* and appeared in context sentences. Each context was preceded by a sentence introducing a referent and followed by a question (cf. Example 1, referent in italics, target in bold). Gender-specificity was controlled via referent names: female-associated names for generic, male-associated for gender-specific masculines.

1. Das ist *Jenny* / *Jannis*. *Jenny* / *Jannis* ist **Kranführer** beim Bau. Was ist *Jenny* / *Jannis*?
   This is *Jenny* / *Jannis*. *Jenny* / *Jannis* is a **crane operator** in construction. *Jenny* / *Jannis* is what?

Participants first read the introduction and target sentence, accompanied by gender-matching comic-style portraits. After clicking *Ton aufnehmen* 'record sound', only the portrait remained and a question appeared, prompting recall. Figure 1 illustrates the trial structure. Trials were fully randomised, and the experiment was self-paced, with 210 native German speakers (70 speakers x 3 counterbalanced groups) recruited via Prolific. The experiment ran on *PennController for IBEX* [7], and participants received £8.5 per hour. The data ($n = 1790$) after excluding production errors, stuttering, and laughter were analysed using linear mixed-effects regression in R following standard procedures [8]. Models included either *sense* or DLM measures as the predictors of interest and control variables based on related homonymy research [1, 2, 3].

Results: The effect of *sense* showed that generic masculines come with significantly longer /ɐ/ durations than gender-specific masculines ($p < 0.001$, $\eta^2 = 0.32$ with 95% $CI = [0.29, 1.00]$). DLM measures show that the generic sense comes with lower levels of semantic co-activation, which in turn leads to longer acoustic durations ($p < 0.001$, $\eta^2 = 0.58$ with 95% $CI = [0.36, 1.00]$). See Figure 2 for both effects.

Discussion: In recalled as in read speech, generic and gender-specific masculines show significantly different /ɐ/ durations, with the suffix being longer in the generic sense. Since this pattern occurs across tasks, it cannot be attributed to reading or recall alone. The DLM offers a first explanation: generic masculines involve less semantic co-activation in the mental lexicon, resulting in longer acoustic durations. Importantly, not only homonymy but also polysemy appears to affect fine-phonetic detail.

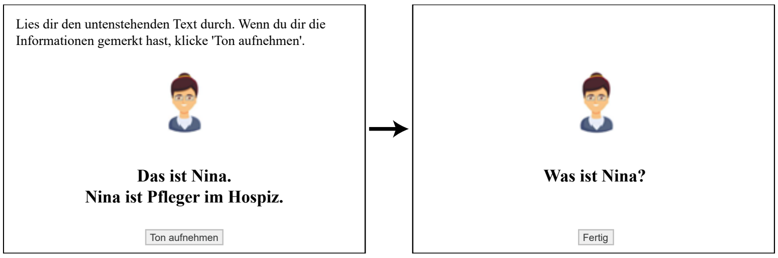


Figure 1: Trial structure: The left screen introduces Nina and provides information on her; the right screen asks about the given information. Left screen, top: Please read the text below. Once you memorised the information, please click on ‘record sound’. Left screen, bottom: This is Nina. Nina is a nurse at the hospice. Left screen, button: Record sound. Right screen: Nina is what? Right screen, button: Done.

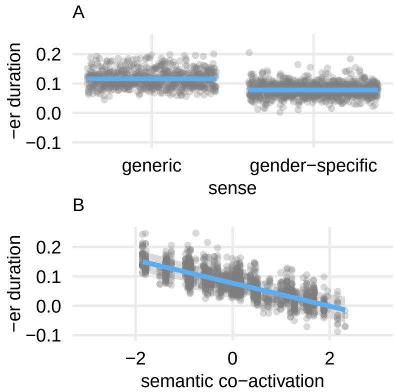


Figure 2: Partial effects of *sense* (A) and *semantic co-activation* (B) as predicted by the linear mixed-effects models.

References

[1] Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, *53*(1), 181–216. https://doi.org/10.1017/S0022226715000183

[2] Schmitz, D., Baer-Henney, D., & Plag, I. (2021). The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords. *Phonetica*, *78*(5-6), 571–616. https://doi.org/10.1515/phon-2021-2013

[3] Schmitz, D., & Baer-Henney, D. (2024). Morphology renders homophonous segments phonetically different: Word-final /s/ in German. *Proceedings of Speech Prosody 2024*, 587–591. https://doi.org/10.21437/SpeechProsody.2024-119

[4] Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, *2019*, 4895891. https://doi.org/10.1155/2019/4895891

[5] Schmitz, D., Plag, I., Baer-Henney, D., & Stein, S. D. (2021). Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.680889

[6] Schmitz, D. (forthcoming). Homophonous semantic minimal pairs differ in their subphonemic acoustic durations: The case of generic and specific masculines in German. *Proceedings of 20. Phonetik und Phonologie Tagung*.

[7] Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

[8] Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

It’s lunchtime, let’s eat restaurant! The influence of context on the processing of (a)typical thematic relations in Mandarin Chinese

Hsin-Ju Wu, Chia-Hsuan Liao
National Tsing Hua University, Taiwan
hsinjuwu@umich.edu

Background: Mandarin is known for its flexibility in the number and the types of arguments selected by verbs. Beyond the “canonical” theme or patient arguments (e.g., *eat hamburger* meaning “eat hamburger”), Mandarin allows verbs to take “non-canonical” arguments such as location or instrument (e.g., *eat restaurant* meaning “eat in a restaurant”) [1, 2]. A recent study has examined such computations at a phrase level, and shown a parsing preference for assigning a theme/patient role to the direct argument of a transitive verb [3]. In sentence processing, comprehenders have been shown to rapidly use verb information to predict upcoming arguments [4], but such findings are largely based on languages with relatively rigid verb-argument mappings. Mandarin thus provides a valuable testing ground for examining whether sentence context interacts with thematic relation computation. From a constraint-based perspective [5], comprehenders integrate multiple probabilistic cues—including lexical and contextual information—in real time. If so, the canonical theme/patient preference may be modulated, or even overridden, by context during processing.

Methods: The stimuli consisted of sentences with critical verb phrases placed in the final position. A 2×2 design was employed to examine whether there is an interaction between Sentence context (Supportive vs. Non-Supportive) and Argument canonicality (Canonical vs. Non-Canonical) (Table 1). While the arguments were either Canonical or Non-Canonical, they were equally unpredictable based on the verb alone, as confirmed by a cloze norming procedure. This allowed us to attribute processing differences to sentence context rather than verb-level biases. In terms of Sentence context manipulation, Supportive conditions provided a sentence context that biased the argument toward the Canonical or Non-Canonical ones. Non-Supportive contexts were created by intentionally mismatching the contexts and the arguments, resulting in implausible sentences. In total, 32 sets of the materials were created, which were counterbalanced into four experimental lists. Participants (N=32) read sentences via RSVP (600ms/word) and performed a lexical decision task at the end of each trial. Fillers ending with pseudowords were included to balance the frequency of button pressing. Reaction times were log-transformed and analyzed using a two-way repeated-measure ANOVA.

Results: The results revealed a significant interaction between Sentence context and Argument canonicality ($F(1,31) = 7.42$, $p < .05$). Post-hoc pairwise t-tests showed that in Supportive conditions, responses to Canonical arguments were significantly faster than those to Non-Canonical ones ($t(31) = -2.91$, $p < .01$), whereas in Non-Supportive contexts, the difference between Canonical and Non-Canonical arguments was not significant($t(31) = 0.81$, $p = .42$).

Discussion: Canonical arguments were processed more efficiently than non-canonical ones in supportive contexts, suggesting a default preference for a canonical role even when context favors a non-canonical one. However, this advantage disappeared in non-supportive contexts, where processing costs may override the default bias. The results support a constraint-based view of sentence processing, where context and argument structure of verbs jointly shape the interpretation of thematic relation. This highlights the flexibility of the comprehension system in weighing competing cues during real-time interpretation, rather than relying on rigid parsing preferences.

Table 1. Example Stimulus in Each Condition.

Context Type	Argument Type	Sentence Context							Verb	Argument
Supportive	Canonical (CA)	吃完 eat-complete	晚餐 dinner	後， after	除了 besides	碗盤 dishes	也要 also	記得 remember	**洗 wash**	**餐具。 tableware**
		"After dinner, besides the dishes, remember to wash tableware as well."								
Supportive	Non-Canonical (NCA)	今天 today	天氣 weather	非常 very	熱， hot	晚上 night	我 I	想要 want	**洗 wash**	**冷水。 cold-water**
		"Today's weather is very hot, I want to take a cold shower tonight."								
Non-Supportive	Canonical (CA)	今天 today	天氣 weather	非常 very	熱， hot	晚上 night	我 I	想要 want	**洗 wash**	**餐具。 tableware**
		"#Today's weather is very hot, I want to wash tableware tonight."								
Non-Supportive	Non-Canonical (NCA)	吃完 eat-complete	晚餐 dinner	後， after	除了 besides	碗盤 dishes	也要 also	記得 remember	**洗 wash**	**冷水。 cold-water**
		"#After dinner, besides the dishes, remember to wash cold water as well."								

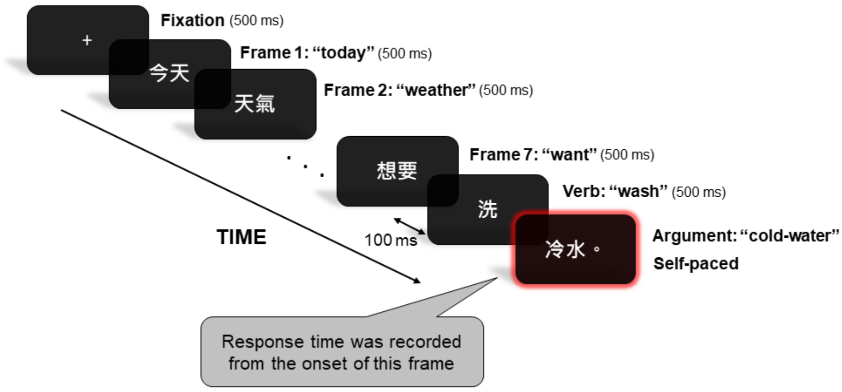


Figure 1. Presentation of Stimuli.

Table 2. Mean Reaction Times (ms) and SDs for Each Condition.

Context Type	Argument Type	Reaction Time	
		Mean	**SD**
Supportive	Canonical (CA)	668.75	171.18
Supportive	Non-Canonical (NCA)	721.55	193.35
Non-Supportive	Canonical (CA)	775.32	195.54
Non-Supportive	Non-Canonical (NCA)	755.35	183.96

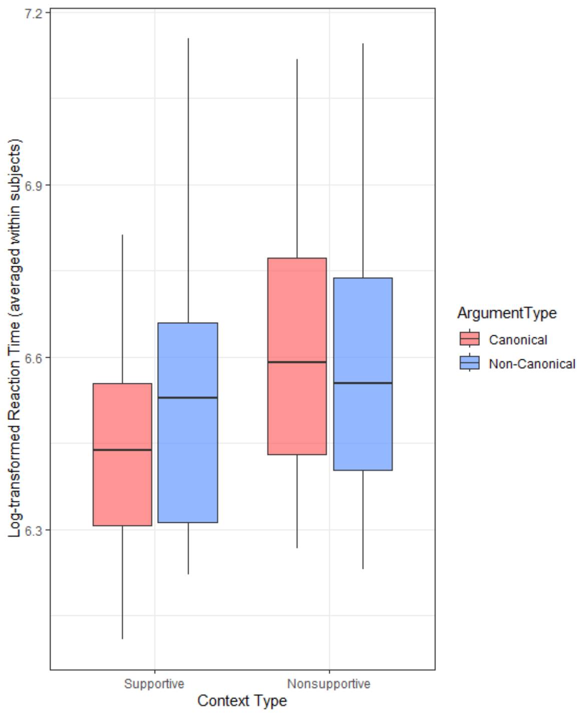


Figure 2. Log-transformed Reaction Times for Each Condition.

References

[1] Lin, T. H. (2001). Light Verb Syntax and the Theory of Phrase Structure (Doctoral dissertation). [University of California, Irvine].

[2] Li, Y. H. (2014). Thematic hierarchy and derivational economy. Language and Linguistics, 15(3), 295-339.

[3] Wu, H. J. (2024). The processing of canonical and non-canonical arguments in Mandarin Chinese (Master's thesis). [National Tsing Hua University, Taiwan].

[4] Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73(3), 247-264.

[5] Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. Language, Cognition and Neuroscience, 31(5), 602–616.

Phonology in morphological priming: Evidence from German complex verbs

Yiwei Si[1], Aditi Lahiri[1], Isabella Fritz[1]
[1]University of Oxford
yiwei.si@ling-phil.ox.ac.uk

Background: Compared to other Germanic languages, morphological priming experiments with German complex verbs have found that affix stripping occurs regardless of the semantic and form relatedness between the derived and base form [1]. The role of phonology in morphological priming, however, has largely remained underexplored, and no study has explicitly tested for phonological effects in German.

One phonological alternation in German complex verbs is stress placement, and consequently metrical patterns, which vary depending on the type of complex word, i.e., particle vs. prefix verbs (see Table 1). Specifically, while German prefix verbs (*verstéllen* 'misplace') carry the same stress as their base forms (s*téllen*, 'put'), stress must be alternated in particle verbs regardless of whether they consist of one (*áufstellen*, 'set up') or two particles (*hinéinstellen*, 'put into'). In German, main stress falls on the leftmost foot. Accordingly, particle verbs have two feet, whereas prefix verbs have only one.

Method: To test whether stress (and metrical) differences in derivational pairs impact word recognition, we conducted a cross-modal ERP priming study with a lexical decision task. 24 German L1 speakers were tested in Oxford, UK. Using auditorily presented primes and visually presented targets, we tested whether stress alternations between three types of complex verbs, i.e., single particle verbs (1PartVs, e.g., *áufstellen*), double particle verbs (2PartVs, e.g., *hinéinstellen*) and prefix verbs (PfxVs, e.g., *verstéllen*), and their bases (*stéllen*) affect morphological processing. The dominant meaning of all primes was semantically transparent. Participants heard a complex word (prime) followed by the base form (target) appearing on the screen. As control primes, participants were presented with complex verbs (1PartVs, 2PartVs and PfxVs) which were neither morphologically, semantically, nor formally related to the target (e.g., *zubinden~reden* 'tie up-speak').

Results: Behavioural results showed equal priming across all three verb types (see Figure 1A for RT priming effects). ERP data, however, showed that PfxVs, which retain the same stress placement as their bases, elicited a longer N400 effect (experimental – control items) compared to particle verbs, where stress placement differs between derived and base forms. Furthermore, P600 effects were observed only in particle verbs but not in the PfxV condition. Both effects were observed over the centro-parietal regions.

Discussion: We interpret the prolonged N400 effect in the PfxV condition as reflecting more efficient lexical access compared to the 1PartV and 2PartV conditions. Lexical access of PfxVs was facilitated by the phonological match between prime and target (both having one foot). In contrast, particle verb processing was effortful due to mismatches in phonological units and stress placement. The two particle verb conditions showed very similar processing patterns despite 2PartVs being morphologically more complex with the additional deictic morpheme *hin* 'hither'/*her* 'thither'. This suggests that the ease of affix stripping is influenced not by morphological complexity or word length alone (e.g., syllable count), but rather by metrical structure, i.e., the number of feet.

Whilst prime-target phonological alignment aided lexical access (N400), we hypothesise that the mismatch between prosodic (one foot) and morphological units (prefix morpheme and stem morpheme) was responsible for increased processing costs (P600) for PfxVs. This was not the case for 1PartVs where prosodic and morphological structures (2 feet, 2 morphemes) aligned. 2PartVs behaved the same as 1PartVs, suggesting that double particles are considered as a single morpheme rather than being processed into constituents. Overall, our study showed that the number of prosodic units impact decomposition processes and, together with stress alternations between complex and base form, added processing costs during word recognition.

Table 1: Phonological differences of 1PartVs, 2PartVs and PfxVs in German. The (x) marks the head of the foot and the (X) marks the syllable carrying main stress.

Condition	Example	Number of feet	Foot structure	Meaning
(a) 1PartV	*áufstellen*	two-feet	X (x) (x .) ˈaʊ̯f ˌʃtɛ ln̩	lit. up-put 'set up'
(b) 2PartV	*hinéinstellen*	two-feet	X (x) (x .) hɪn ˈaɪ̯n ˌʃtɛ ln̩	lit. hither-in-put 'put into'
(c) PfxV	*verstéllen*	one-foot	X (x .) fɛɐ ˈʃtɛ ln̩	lit. prefix-put 'misplace'

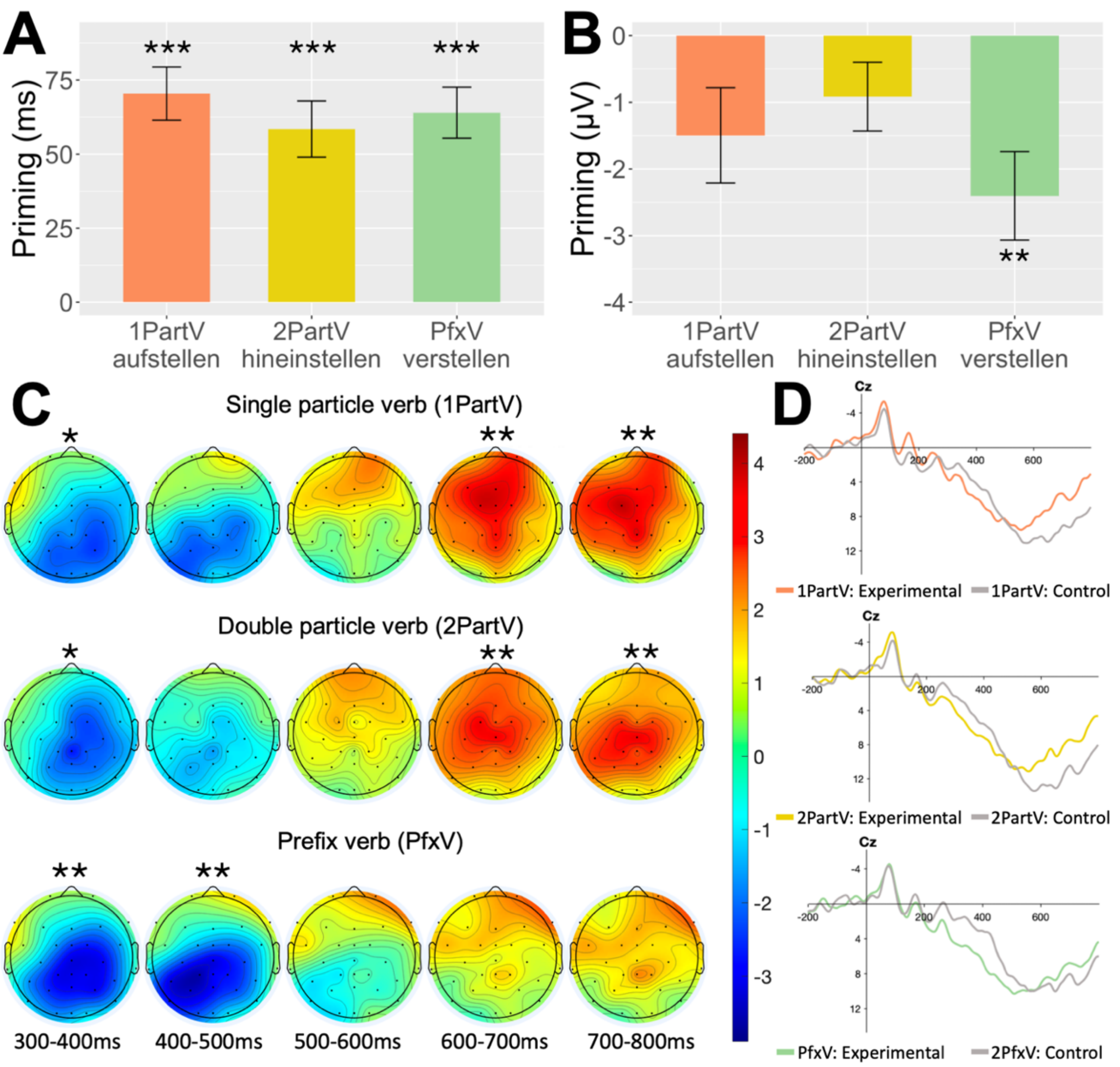


Figure 1. (A) Magnitude of the Priming Effect (experimental – control). (B) Difference (in µV) between experimental–control items in the 400-500ms time-window. (C) Topographical plots (experimental–control items). (D) Brainwaves of a representative electrode (Cz) illustrating N400 and P600 priming effects.

References

[1] Smolka, E., Libben, G., & Dressler, W. U. (2019). When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience*, *34*(5), 599–614.

Processing of Homonyms in Bilingual Children: A Visual World Eye-Tracking Study

Katrin Odermann, Renate Delucchi Danhier, Barbara Mertins

TU Dortmund University

katrin.odermann@tu-dortmund.de

Background: Semantic awareness is crucial for both word decoding and higher-level comprehension. Since the connection between word form and meaning is arbitrary, homonyms — words with multiple meanings — pose a special challenge for preschoolers, who often expect a one-to-one form-meaning match [1]. Homonyms are therefore particularly well suited to assess semantic awareness: distinguishing between multiple meanings of the same form requires children to distinguish surface structure from underlying meaning and flexibly consider context — a key metalinguistic skill. While young children begin to recognize homonyms, they tend to favor the more frequent, primary meaning, even when context suggests another [3]. Lexical access to homonyms involves co-activation of meanings, with secondary meanings typically accessed later [2]. In bilinguals, this process is complicated by cross-language co-activation, involving multiple forms and meanings.

Method: This study used a Visual World Paradigm and eye-tracking to explore the processing of homonyms in 44 German speaking children (23 monolingual, 21 bilingual), aged 4–6 (mean = 5,13). Three research questions guided the study: 1. Is the primary meaning activated even when context supports only the secondary? 2. Do monolingual and bilingual children differ in terms of fixation duration on the primary meaning? 3. Do fixation patterns on the primary meaning change over time? These questions are key to understanding not only how children resolve lexical ambiguity, but also how monolingual and bilingual language experience shapes this process. Focusing on online measures like eye fixations reveals differences in underlying processes that are not captured by final responses alone. Parental questionnaires provided background information, and children's vocabulary size was measured using the German version of the Peabody Picture Vocabulary Test (PPVT). The experimental task, conducted in German, included homonym and control trials within a visual-world paradigm. Each homonym trial paired an auditory stimulus with a visual stimulus to assess children's recognition of both meanings of a homonym and their sensitivity to potential semantic conflict between them. Control trials contained no lexical ambiguity and served as a baseline for comparison.

Results: PPVT results showed significantly lower vocabulary scores in bilinguals. A linear mixed model revealed significantly longer first fixation durations on both primary and secondary homonym meanings compared to distractors, indicating co-activation in both groups. Age significantly influenced fixation duration ($\beta = -2.37$, $t = -2.16$, $p = 0.037$), with durations decreasing with age. In monolinguals, first fixations on the secondary meaning were significantly longer than on distractors ($\beta = 200.80$, $t = 5.79$, $p < 0.001$). Bilinguals showed longer fixations for both meanings compared to distractors and semantic competitors, and longer fixations on the primary meaning than monolinguals ($\beta = 105.78$, $t = 2.11$, $p = 0.036$). SES and PPVT scores had no significant effects. A second analysis on fixation counts over five seconds revealed the primary meaning was fixated throughout, while fixations on the secondary meaning increased and remained high. No group differences emerged in the temporal course of visual attention.

Discussion: The results of the present study underline the complexity of homonym processing in young children. Both groups co-activated meanings, but only bilinguals showed longer fixations than monolinguals, likely due to smaller vocabulary and higher cognitive load resulting from dual-language access. Overall, older children processed meanings more efficiently as shown by shorter first fixations. Generally, the study emphasizes the dynamic nature of semantic processing and word retrieval in early language development, with age emerging as a significant predictor for the efficiency of semantic processing. The study provides novel insights into ambiguity resolution as a core challenge in early semantic awareness development, and into the strategies children employ when navigating lexical ambiguity, particularly in bilingual contexts. This is the first eye-tracking study examining semantic-level metalinguistic awareness in children.

References:

[1] Mazzocco, M. M. M. (1997). Children's interpretations of homonyms: A developmental study. *Journal of Child Language*, *24*(2), 441–467. https://doi.org/10.1017/S0305000997003103

[2] Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*(1), 203–214. https://doi.org/10.1017/S030500090300583X

[3] Beveridge, M., & Marsh, L. (1991). The influence of linguistic context on young children's understanding of homophonic words. *Journal of Child Language*, *18*(2), 459–467. https://doi.org/10.1017/S0305000900011168

Figures:

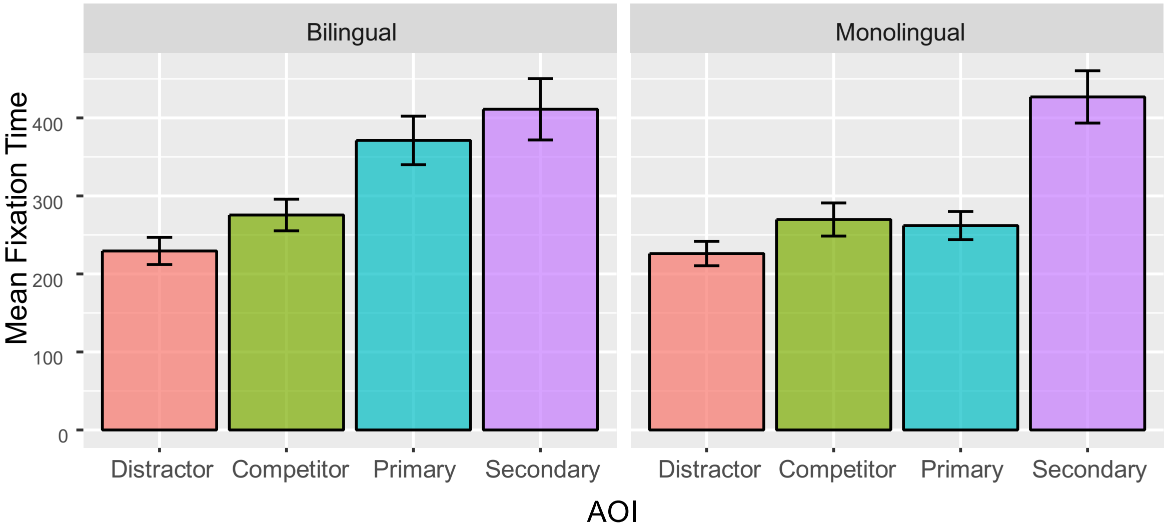


Figure 1: Mean first fixation duration per AOI and groups

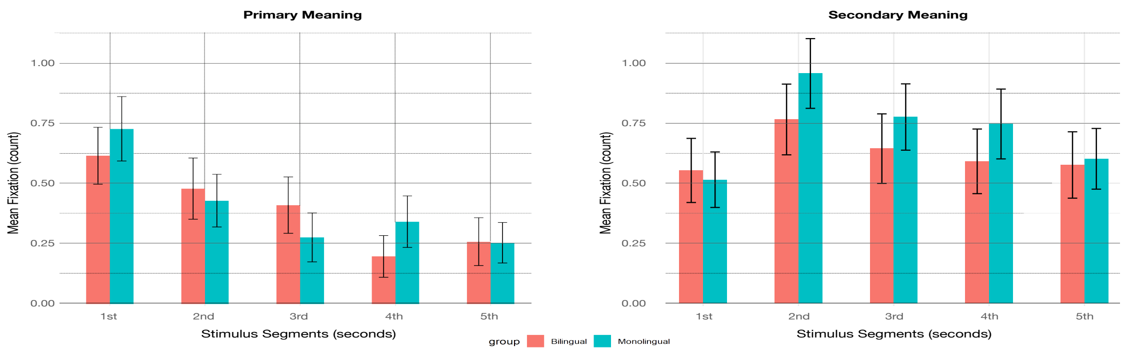


Figure 2: Average fixation number for primary and secondary AOIs over stimulus presentation time

A matter of time and meaning: a time-to-event analysis of response times in a semantic categorization task

Anastazja Rosanoff[1], Peter Hendrix[1]
[1]Tilburg University
a.rosanoff@tilburguniversity.edu, p.h.g.hendrix@tilburguniversity.edu

Background: The Calgary Semantic Decision Project (CSDP; [1]) provides response data from a semantic categorization task in which participants are asked to indicate if a word shown on the screen is a concrete word (e.g., “apple”) or an abstract word (e.g., “theory”). Previous studies have shown that orthographic and semantic properties of words influence response times in this dataset [see, e.g., 2]. However, these studies did not examine how the effects of predictors evolve over time. Here, we adopt a technique from time-to-event analysis - the piecewise exponential additive mixed model (PAMM; [3]) - to model how the instantaneous probability of a response evolves over time. This analysis technique allows us to investigate the temporal dynamics of predictors effects and thus offers insight into how lexical processing unfolds over time in the semantic categorization task.

Method: We fit a PAMM to the response time data for the 4,994 words in the CSDP for which lexical-distributional information was available in the British Lexicon Project [4]. We modeled time-varying, non-linear predictor effects using tensor product interactions between time and predictor variables [5]. In the interest of model interpretability, we restricted tensor product interactions to fourth-order non-linearities in both dimensions. No such restriction was applied to the smooth term for the main effect of time.

Results: The results of the PAMM analysis are presented in Figure 1. Consistent with previous findings [e.g., 1], we observed a facilitatory effect of word frequency, with high-frequency words associated with a higher instantaneous probability of a response. This effect was most prominent early in the response time window. We also found an effect of mean bigram frequency: words with lower average letter bigram frequency showed a higher instantaneous response probability. This effect was relatively stable across much of the analysis window, from approximately 720 to 1200 ms after stimulus onset.

In addition, we observed effects of two semantic predictors: concreteness and semantic neighborhood density [cf., 1]. The effect of concreteness emerged early, whereas the influence of semantic neighborhood density appeared later in the response time window. Finally, we found an early effect of orthography-to-semantics consistency [6]: words with more consistent mappings between form and meaning were associated with a higher instantaneous probability of a response.

Discussion: The PAMM analysis revealed early effects of form-based predictors, meaning-based predictors, and the consistency between form and meaning. These effects overlapped in time, suggesting that form and meaning jointly influence lexical decision-making from the start of the response time window. The early effect of concreteness, in particular, indicates that semantic information is available and used early in the decision making process. As such, the current findings support models of lexical processing that assume parallel activation and integration of form and meaning [7]. The later effect of semantic neighborhood density, on the other hand, suggests that competition among related meanings begins to influence the decision making process at a later integration stage. These findings underscore the value of time-sensitive modeling approaches for revealing the dynamic interplay of orthographic and semantic information in real-time word recognition.

Figures

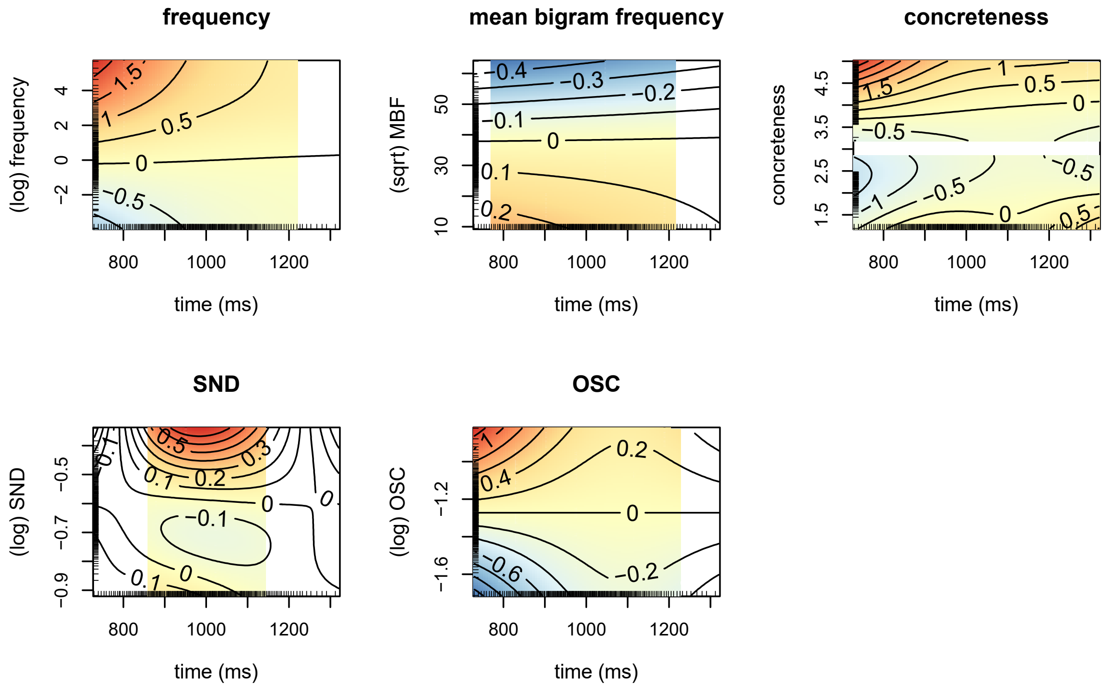


Figure 1. Results of the PAMM analysis on the Calgary semantic decision project data. Shown are predictor-related adjustments to the main effect of time. Warmer colors indicate higher hazard rates. SND refers to semantic neighborhood density; OSC to orthography-to-semantics consistency.

References

[1] Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: Concrete/abstract decision data for 10,000 English words. *Behavior Research Methods*, *49*(2), 407-417. https://doi.org/10.3758/s13428-016-0720-6.

[2] Pexman, P. M., & Yap, M. J. (2018). Individual differences in semantic processing: Insights from the Calgary semantic decision project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1091-1112. https://doi.org/10.1037/xlm0000499.

[3] Bender, A., Groll, A., & Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, *18*(3-4), 299–321. https://doi.org/10.1177/1471082X17748083.

[4] Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287-304. https://doi.org/10.3758/s13428-011-0118-4.

[5] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R (2nd ed.).* Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

[6] Marelli, M., Amenta, S., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Quarterly Journal of Experimental Psychology*, *68*(8), 1571–1583. https://doi.org/10.1080/17470218.2014.988736

[7] McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375.

Patterns fast and slow: The structure and statistics of language shape high- and low frequency neural signals

Sophie Slaats[1*], Alexis Hervais-Adelman[1,2]

[1] Dynamics of Brain and Language Lab, Université de Genève, Genève, Suisse; [2] LiZZ, Universität Zürich, Zürich, Schweiz

*Sophie.Slaats@unige.ch

The neural mechanisms that govern the combination of words to form phrases and sentences remain an important open problem in the neurobiology of language. Recent studies of syntactic structure building suggest a role for both low- and high frequency oscillatory activity: delta band (<4 Hz) activity and high gamma power (>70 Hz) (HGP) are found to be sensitive to the syntactic structure of the input [1,2]. Interestingly, these frequency bands also encode information about statistical regularities at the word level, such as surprisal and entropy [2,3]. These regularities can function as a *cue* to the underlying syntactic structure, for example, by indicating the presence of a phrase boundary [4]. Indeed, previous work suggests that delta-band neural correlates of structure building are affected by lexical probability [5]. In this study, we aim to disentangle neural sources for structure building and lexical probability in delta band activity and HGP using spatiotemporally precise intracranial recordings. In addition, we investigate whether the previously reported delta-band effect of probability on structure building can be replicated, and test whether a similar effect exists in HGP.

To this end, we analyze an open dataset of intracranial EEG recordings (ECoG) from native speakers of Dutch who watched a short film [6]. We analyze data from 18 adults (18-52 years old, 12 women). All participants are right-handed and were implanted with left-lateralized ECoG grids (m. 76 channels, sd 16). We use Temporal Response Functions (TRFs) with a cumulative model comparison approach to measure the neural encoding of syntactic and lexical probabilistic features in HGP (70-150 Hz) and delta- (0.5-4 Hz) activity. In addition, we test whether neural correlates of structure are affected by lexical probability by splitting the syntactic features into two based on lexical probability. The syntactic features are bottom-up and top-down node counts, obtained from a manual constituency parse of the stimulus (adapted minimalist framework). The lexical probabilistic features are surprisal and entropy extracted from GPT2 for Dutch [7].

Preliminary results suggest that HGP and delta band signals are differently modulated by the structure and statistics of language, and that these have (partially) separable neural sources. To wit, HGP and delta band activity are modulated by all structural and statistical features in a widespread frontotemporal network, which is spatially captured by surprisal (see Figure 1A and B). The other features contribute to variance in the signal mainly, but not exclusively, in subsets of this network. At the interface between statistics and structure, our results suggest that surprisal and structure building interact in both frequency bands. In the delta band, surprisal determines the latency of (top-down) structure building in superior temporal gyrus (see Figure 2). In HGP, the difference appears to be one of amplitude rather than time. HGP increases more as a function of node count for high- than for low surprisal words. These effects are most salient after ~300ms.

These findings suggest that structural and lexical probabilistic features jointly shape high- and low frequency neural signals [3], not only additively, but also in interaction. The interaction between probability and syntactic structure appears to be one of the *latency* in the delta-band, and one of *amplitude* in HGP. This supports a role for delta-band activity and HGP in representing probabilistic and syntactic information [8,9], and is in line with a role for lexical probability as a cue for structure [5]. The results of this study will speak to neurobiological models of language comprehension with a focus on oscillatory activity.

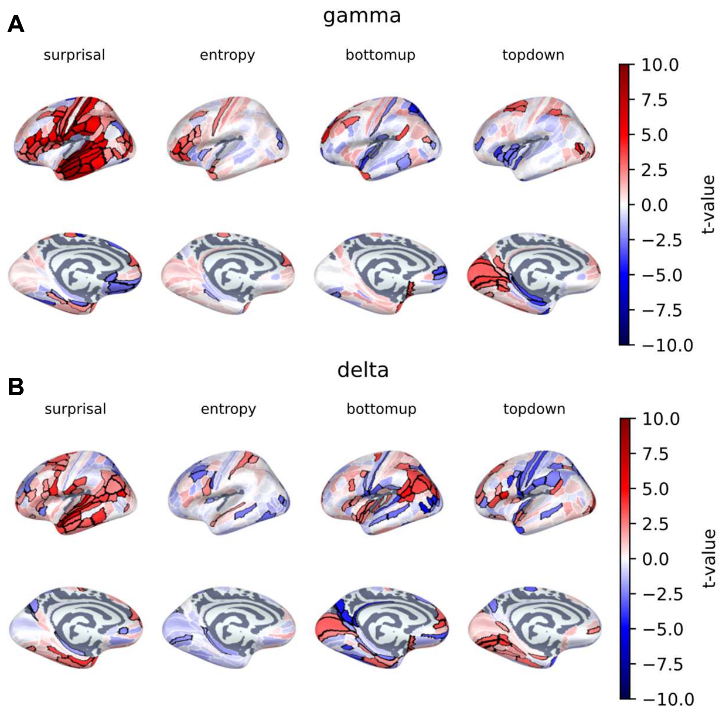


Figure 1. The effect of each of the features of interest on the reconstruction accuracy of the models in **(A)** high gamma power (70-150 Hz) and **(B)** the delta band (0.5-4 Hz). Colors indicate t-values of the coefficients from linear mixed models. Significance was calculated with permutation. Positive t-values (red) indicate that a feature positively contributes to model fit. Opaque parcels with black borders reached significance ($p < 0.05$); parcels with white borders did not. Due to different ECoG grids, not all parcels contain data from all participants. The medial view is slightly tilted. The parcellation used is HCPPPM1 [10].

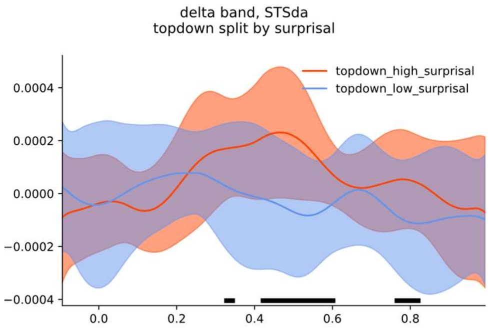


Figure 2. The grand average TRF for top-down node counts in the superior temporal sulcus split for high- and low surprisal values. The shaded ribbon indicates standard deviation. The black line at the bottom shows the time-window where the difference (t-value) between high- and low surprisal structure building is larger than 97.5 percent of models with a randomly split feature.

References

1. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **19**, 158–164 (2016).
2. Nelson, M. J. *et al.* Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E3669–E3678 (2017).
3. Weissbart, H. & Martin, A. E. The structure and statistics of language jointly shape cross-frequency neural dynamics during spoken language comprehension. *Nat. Commun.* **15**, 8850 (2024).
4. McCauley, S. M. & Christiansen, M. H. Language learning as language use: A cross-linguistic model of child language development. *Psychol. Rev.* **126**, 1–51 (2019).
5. Slaats, S., Meyer, A. S. & Martin, A. E. Lexical surprisal shapes the time course of syntactic structure building. *Neurobiol. Lang.* **5**, 942–980 (2024).
6. Berezutskaya, J. *et al.* Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. *Sci. Data 2022 91* **9**, 1–13 (2022).
7. de Vries, W. & Nissim, M. As Good as New. How to Successfully Recycle English GPT-2 to Make Models for Other Languages. in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* 836–846 (2021). doi:10.18653/v1/2021.findings-acl.74.
8. Donhauser, P. W. & Baillet, S. Two distinct neural timescales for predictive speech processing. *Neuron* **105**, 385-393.e9 (2020).
9. Martin, A. E. A compositional neural architecture for language. *J. Cogn. Neurosci.* **32**, 1407–1427 (2020).
10. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

Features all the way down: visual masking interacts with age of acquisition and iconicity in picture naming

Giulio Massari[1,*], Fanny Meunier[1] & Raphaël Fargier[1]

[1]Université Côte d'Azur, CNRS, BCL, France

*Corresponding author: Giulio.MASSARI@univ-cotedazur.fr

Background: Traditionally, models of language production saw word retrieval as a discrete sequence of processes mapping different linguistic features onto different levels of processing [1]. Consequently, factors modulating retrieval in tasks like picture naming were also assumed to have a very localized area of influence [2]. However, there is now evidence that information can flow more flexibly across processing levels [3], and that many factors impact production rather at the interface between levels. For example, while age of acquisition has long been considered a purely lexical variable, empirical results show that its impact actually extends to visuo-conceptual and phonological stages [4-5]. This is in line with the observation that properties connecting input and output levels of production, thus only indirectly communicating in most processing models, modulate retrieval as well. For instance, iconicity, i.e., the property of having a non-arbitrary connection between visuo-conceptual and phonological features of words (Fig. 1), has recently been observed to speed up retrieval in interaction with age of acquisition [6]. This suggests that linguistic factors can display distributed, multiple-level influence on retrieval processes, but the dynamics of such influence across processing levels remain unknown. To investigate this, we tested whether a factor which *in principle* should have a purely visual locus, i.e., masking of a prompt image, impacts word production in interaction with variables having a strong (i.e., iconicity) or more marginal (i.e., age of acquisition) visual component. **Method:** We employed a picture naming task in French, comprising 200 images. Half of the images were associated to words acquired early in life, while the other half to words acquired later. Within the 200 stimuli, we also embedded a smaller selection of 40 iconic vs. 40 non-iconic words. Items were assigned to a low-masked or high-masked condition, balanced upon critical variables and counterbalanced across participants (Fig. 1). **Results:** A linear mixed-effects model with Participant and Item random effects, run on production latencies (n=28 participants), revealed main effects of Masking and AoA (both $p<.001$), as well an interaction between them ($p=.027$): i.e., a stronger masking disadvantage was observed among late acquired than among early acquired items (Fig. 2). A second analysis, restricted to the iconicity selection, also revealed a main effect of Masking ($p<.001$), as well as an interaction with Iconicity ($p=.013$): i.e., a much stronger masking effect was observed among iconic than among non-iconic items; conversely, iconic words were retrieved *faster* than non-iconic words under low-masked conditions, but *slower* under high-masked ones (Fig. 3). **Discussion:** These results suggest that even a visual factor such as masking displays consequences all the way down through production, while interacting with different linguistic variables. The nature of this interaction is, however, dependent on the variable at stake. Masking and age of acquisition effects interacted but maintained a consistent direction, as compatible with the view that age of acquisition impacts the overall strength of connections between features in a word's network [7]. The iconicity effect, on the contrary, changed direction depending on masking intensity, suggesting that this factor stems back from a specific visual-phonological connection, hampered under strongly masked conditions [8]. Overall, the present results support the hypothesis that age of acquisition and iconicity impact the connection between (different) levels of processing, rather than processing at a specific stage.

References

[1] Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci., 22(1),* 1-75.

[2] Alario, F.X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U.H., & Segui, J. (2004). Predictors of picture naming speed. *Beh. Res. Meth. Instr. Comp., 36(1)*, 140–155.

[3] Mascelloni, M., McMahon, K., Piai, V., Kleinman, D., & de Zubicaray, G. (2021). Mediated phonological-semantic priming in spoken word production: Evidence for cascaded processing from picture-word interference. *Q. J. Exp. Psychol., 74(7),* 1284-1294.

[4] Catling, J.K., Dent, K., & Williamson, S. (2008). Age of acquisition, not word frequency affects object recognition: Evidence from the effect of visual degradation. *Acta Psychol., 129,* 130-137.

[5] Perret, C., Bonin, B., & Laganaro, M. (2014). Exploring the multiple-level hypothesis of AoA effects in spoken and written object naming using a topographic ERP analysis. *Brain Lang., 135,* 20-31.

[6] Massari, G., Meunier, F., & Fargier, R. (2025, under review). Sharp as a tack, round as a ball: Iconic mappings facilitate word retrieval in spoken language production. *Lang. Cogn. Neurosci.*

[7] Elsherif, M.M., Preece, E., & Catling, J.C. (2023). Age-of-acquisition effects: A literature review. *J. Exp. Psychol. Learn. Mem. Cogn.*, *49(5),* 812–847.

[8] Emmorey, K. (2014). Iconicity as structure mapping. *Phil. Trans. R. Soc. B, 369: 20130301.*

Figures

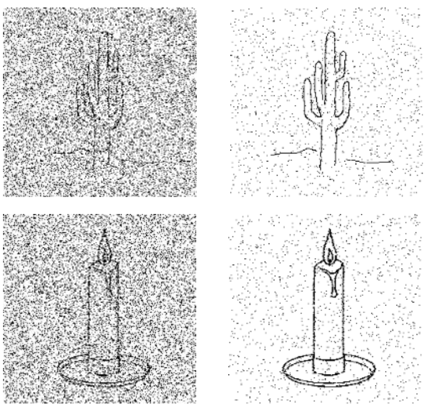


Fig. 1. Examples of items comprising conditions. Following [4], leftmost items were assigned to the high-masked and rightmost items to the low-masked condition, with counterbalanced assignment across participants. Following [6], the uppermost item, ‘*cactus’* (‘cactus’), was assigned to the iconic condition, as the word’s phonology raises an association with a spiky shape, and its referent is indeed spiky. The lowermost item, *‘bougie’* (‘candle’), was assigned to the non-iconic condition, as the word’s phonology does not bias into thinking about any visual shape.

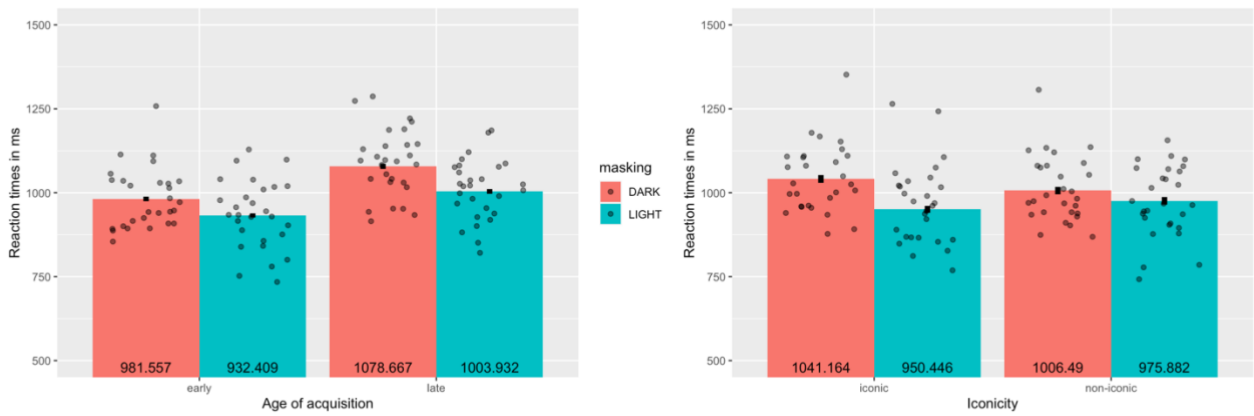


Fig. 2 and 3. Mean reaction times across conditions (left: Masking X AoA; right: Masking X Iconicity). Bars represent condition means overall; scatter points represent individual performances, averaged across items.

What drives incremental sequence learning?

Jessie S. Nixon[1], Erdin Mujezinović[2], Ruben van de Vijver[2]
[1]University of Bielefeld; [2]Heinrich Heine University of Düsseldorf
jessie.nixon@uni-bielefeld.de

Spoken language unfolds over time, leading to structural relations in language at multiple levels. This raises the question of how such sequences in language are learned. In seminal work, [1] showed that listeners learn from statistical information syllable sequences. This demonstrated a mechanism by which language could be learned and led to an explosion of follow up work. Yet few studies have addressed the question of which learning mechanisms are involved. An exception is e.g. [2], which investigated sequence learning in typing. The present study aims to test a) whether we can find evidence of *incremental learning during exposure* and b) which *learning mechanisms* best predict response accuracy during learning of spoken syllable sequences.

Our experiment was based on [1] and used the same stimuli: 6 trisyllabic 'words' made up of 11 spoken CV syllables; 300 tokens of the 6 words were pseudorandomised and combined into a 21-minute running speech stream without pauses between words. While [1] carried out a forced-choice test *after training*, we were interested in incremental learning. Therefore, we asked participants (16, German speaking) to type the *upcoming syllable* at 167 pseudorandom intervals. Intervals occurred both within and between words and had varying predictability. A response was correct if it matched the next syllable when the speech stream restarted.

We expected, firstly, that learning would occur throughout the experiment, such that before the end of training, response accuracies would be above chance. Secondly, we expected that accuracy would depend on the predictability of the upcoming syllable in the sequence. To measure predictability, we used two measures corresponding to different proposals for the learning mechanisms underlying sequence learning: transitional probability (TP) and error-driven learning (EDL). TP was the mechanism originally proposed by [1] (TP of Y given X = frequency of bigram XY / frequency of X). EDL used the Rescorla-Wagner equations implemented in the `edl` R package [3] with preceding syllables as *cues* and the upcoming syllable as the *outcome* to be predicted.

Determining chance level in the present study is not straightforward as responses were open. We calculated chance as 9.1%. This calculation assumes that participants were drawing from the 11-syllable set, which was not always the case, so this estimate is conservative with respect to our hypothesis. Mean accuracy across all participants and trials was 22%. A one-tailed t-test showed this to be significantly above chance ($t=16.4$, $p<.001$). Accuracy was above chance for 14/16 participants (largest $p<.003$). Two participants were not above chance ($p=.74$; $p=.18$). This demonstrates that incremental learning took place during exposure. This is a strong test of learning, as it includes even early trials, where presumably less learning had yet taken place.

To determine which learning mechanism best predicts accuracy we generated the two measures of predictability (TP and EDL activation) for each individual participant on each trial. In two separate binomial GAMM models (Wood, 2011), we included a smooth for the predicability measure (TP or activation) and random effects for participants, with response accuracy as the dependent variable (Fig. 1). Model comparisons using `compareML()` found that activation predicted accuracy better than TP (fREML difference score 12.4). The large fluctuations in accuracy for TP (Fig. 1) are *not expected theoretically*. It should be noted that with [1]'s design and stimuli, there are TP ranges with sparse data (see Fig. 1, right, e.g. between TP 0.8-0.99), reflected in large confidence intervals. Nevertheless, TP does not seem to predict accuracy well: a positive correlation is expected, but a check of the raw data confirms that accuracy was actually higher (38%) with TP=0.5 than with TP=1 (18%). Sparse data is not an issue here as these were among the most frequent TP values. This result reflects the fact that high TP does not always lead to high accuracy; e.g. in early trials. Although TP may provide a good overall prediction on relatively large amounts of summed TP data, as in [1], it does not appear to be a good predictor of incremental learning data.

The present results provide insights into the mechanisms that enable learning of sequential information in language. The results extend the work of [1] in a number of ways. Rather than a word-nonword forced choice as in [1], participants were able to make open predictions about upcoming syllables with an above-chance success rate. We found that participants' prediction success was better estimated by an error-driven learning model that used previous syllables as cues to the upcoming syllable compared to a model based on transitional probability. While TP makes intuitive sense as a driver of learning, it seems to overestimate predictability at certain points, especially early in learning. TP is outperformed here by EDL, which also takes into account other important factors, such as the history of learning (which here includes effects similar to item frequency) and cue competition. The theoretical implications of the different learning mechanisms will be discussed.

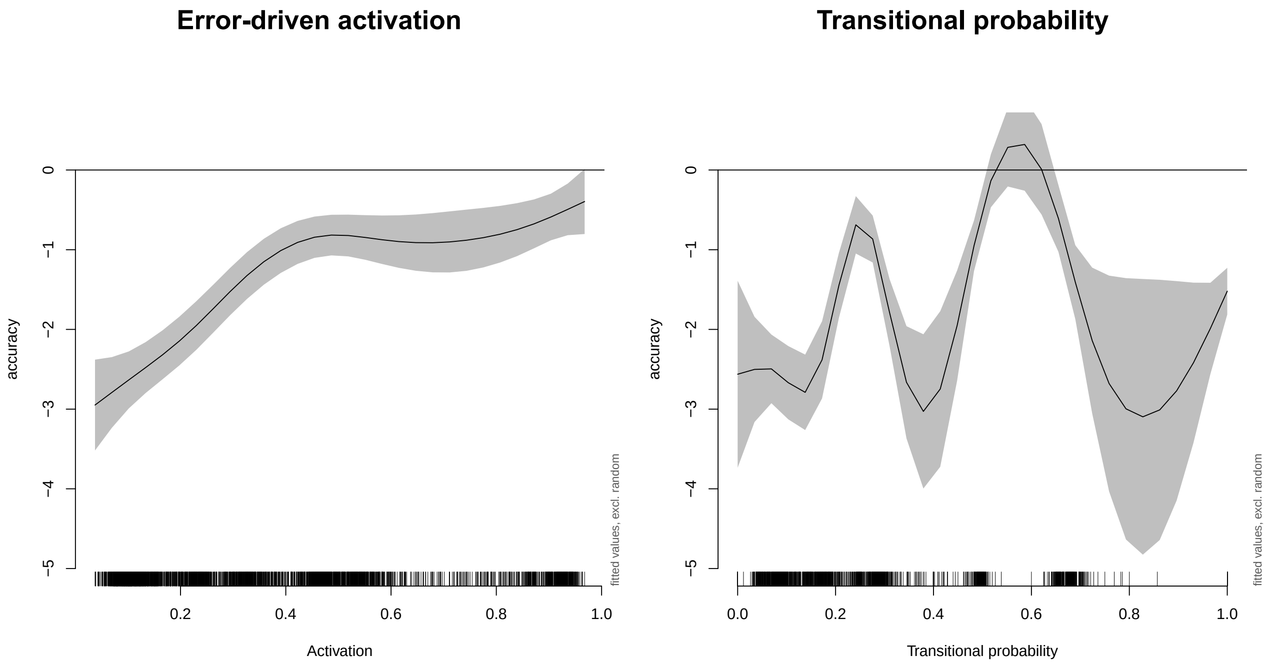


Figure 1: Results of binomial GAMM models of response accuracy with smooths for error-driven (EDL) activation (left) and transitional probability (right). The predictability measure (Activation or Transitional probability) is on the x-axis; accuracy is on the y-axis (log odds scale). Data points are represented by tick marks along the x-axis. Shaded areas represent confidence intervals. Random effects are excluded from the plots.

References

[1] Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, 35(4):606–621.

[2] Tomaschek, F., Ramscar, M., and Nixon, J. S. (2024). The keys to the future? an examination of associative versus discriminative accounts of serial pattern learning. *Cognitive Science*.

[3] van Rij, J. and Hoppe, D. (2020). edl: Toolbox for error-driven learning simulations with two-layer networks. R package version 0.3.

Age differences in spoken language comprehension: verb-argument and formality-register congruence influence real-time sentence processing

Katja Maquate[1], Angela Patarroyo[1], Angelina Ioannidou-Tsiomou[1], Pia Knoeferle[1,2,3]
[1] Humboldt-Universität zu Berlin; [2] Einstein Center for Neurosciences Berlin; [3] Berlin School of Mind and Brain
katja.maquate@hu-berlin.de

Semantic and social contextual constraints are known to mediate language comprehension, but it remains unclear whether they exert equal influence. While some studies suggest parity [1, 2], others indicate that individual differences, such as age [3], differ in how they affect processing social and semantic information. Formality and its associated register offer a means to jointly examine social context, semantic effects, and individual differences. Using the Visual World Paradigm, we tested German participants' processing of formality register and verb-argument (in)congruence. To assess the role of individual differences, we examined age (18–45 years) as a factor, hypothesizing that with advancing in age, one is more established in professional settings and may have greater experience navigating formal and informal contexts.

Method: 66 participants (18-45 years of age, *N*=32 > 32 years) first listened to a context sentence that set up a formal or informal situation (Table 1) while looking at images associated either with a formal (e.g. fancy shoes and chic clothes) or informal (e.g. old shoes and casual clothes) context (Figure 1). Continuing to look at the images, they then listened to a target sentence (1, e.g. *I'm soon tying my* $\textit{shoes}_{\textit{colloquial}}$, *N*=32 critical items, *N*=64 fillers). The verb in the target sentence semantically (mis)matched its arguments (e.g. tie shoes vs. clothes). The images represented candidate post-verbal referents (e.g. shoes or clothes), creating semantic congruence between the verb and two out of four candidate referents. This verb-argument congruence factor was crossed with congruence between the formality of the context sentence and the (informal vs. more formal) register of the verb's argument (e.g. shoes $_{\text{standard}}$ vs. $\text{shoes}_{\text{colloquial}}$, Table 1).

Hypotheses: During the context sentence, all participants should fixate the formality matching (vs. mismatching) pictures more. During the verb and/or object region, we predicted more fixations towards the formality matching (vs. mismatching) target picture with advancing age. We predicted anticipatory / integrative effects in the verb & object region for semantically matching (vs. mismatching) targets regardless of age. Finally, formality and semantic congruence could also interact in the verb and/or object region.

Results: Participants, as predicted, took the context formality into account to anticipate matching images ($p < .01$). Unexpectedly, the older the participants were, the more they took the formality of the context into account ($p < .01$, Figure 2). As predicted, participants anticipated ($p < .01$) and integrated ($p < .01$) the named object noun. We observed an unpredicted 3-way interaction between context formality, verb congruence, and age in the verb period ($p < .05$). Post-hoc comparisons indicated that in a formality mismatching situation, middle-aged (but not younger) adults were significantly more likely to fixate the target image when verb congruency matched (vs. mismatched), see green line in right pane in Figure 3: E.g., after an informal context, middle-aged participants looked more at the fancy shoes when the verb constraints matched versus mismatched ($p < .01$).

Discussion: Middle-aged adults showed competitor effects based on the verb's constraints even though context formality had already ruled one of the two potential candidate images out. Our findings indicate that individual differences affect how semantic and social contextual constraints mediate language comprehension.

Figure 1: Example of a four-image visual display. One pair of shoes denotes a formal and one an informal situation, likewise one depiction of clothes is associated with a formal and one with an informal context.

Condition	Context / Formality	Target Sentence
(a) Full match	formal / match	Ich binde gleich meine Schuhe$_{\text{standard}}$
(b) V-A mismatch	formal / match	Ich binde gleich meine #Kleidung$_{\text{standard}}$
(c) Register mismatch	formal / mismatch	Ich binde gleich meine Latschen$_{\text{colloquial}}$
(d) Double mismatch	formal / mismatch	Ich binde gleich meine #Klamotten$_{\text{colloquial}}$
(a) Full match	informal / match	Ich binde gleich meine Latschen$_{\text{colloquial}}$
(b) V-A mismatch	informal / match	Ich binde gleich meine #Klamotten$_{\text{colloquial}}$
(c) Register mismatch	informal / mismatch	Ich binde gleich meine Schuhe$_{\text{standard}}$
(d) Double mismatch	informal / mismatch	Ich binde gleich meine #Kleidung$_{\text{standard}}$

Table 1: Illustration of the experimental conditions. Example context: *Elegant*$_{formal}$*/schlampig*$_{informal}$ *angezogen spricht*$_{formal}$*/labert*$_{informal}$ *Peter.* Translation context sentence (not shown in table): *Elegantly/sloppily dressed says/ramples Peter.* Translation target sentence: *I'm soon tying my shoes*$_{colloquial\ or\ standard/formal}$ */ clothes*$_{colloquial\ or\ standard/formal}$. The # marks semantic mismatch. V-A mistmatch = Verb-Argument mismatch.

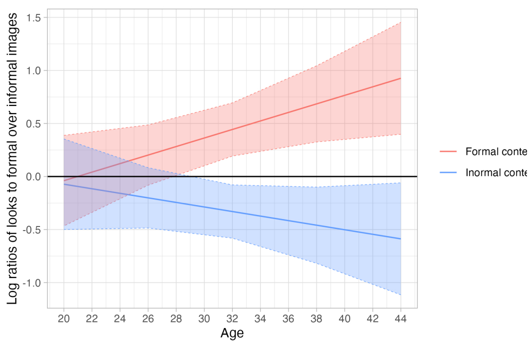


Figure 2: Significant context formality × age interaction for the context-sentence time period. Log ratios above 0 = looks to formal image, below 0 = informal. Shading = 95% CIs.

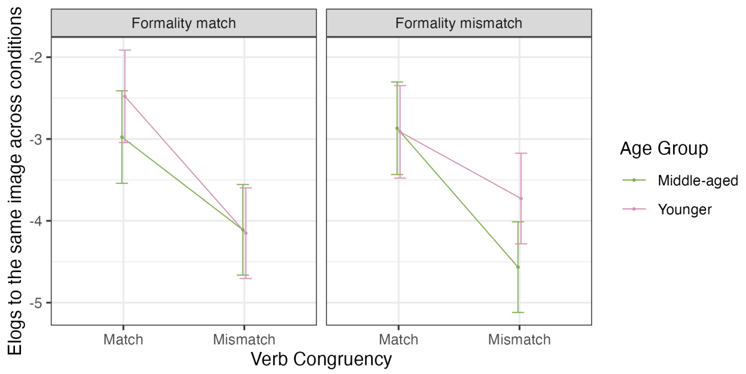


Figure 3: Significant three-way interaction: context formality, verb congruency, and age for the verb time period. Elogs to same image across conditions, e.g., the object noun referent in the full match condition. Error bars = 95% CIs.

References

[1] Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

[2] Yurchenko, A., Arutiunian, V., Shitova, N. M., Bergelson, M., & Dragoy, O. (2023). Register switching involving lexical-semantic processing in Russian: An ERP study. *Journal of Neurolinguistics*, *65*, 101111.

[3] Maquate, K., & Knoeferle, P. (2021). Referential vs. non-referential world-language relations: How do they modulate language comprehension in 4 to 5-year-olds, younger, and older adults? *Frontiers in Psychology*, *11*, 542091.

Semantics in reading-time corpora

Laia Colina Fortuny[1], Li Kloostra[1], Johan Bos[2], Jakub Dotlačil[1]

[1] Utrecht University, Utrecht, The Netherlands
[2] University of Groningen, Groningen, The Netherlands
j.dotlacil@uu.nl

Background: In the last years, it has become standard to evaluate processing models not just on individual experiments but on psycholinguistic corpora, which include neural and behavioral measures for texts. Crucially, the corpora need to be endowed with linguistic annotations if their measures are to be connected to theories of language. Nowadays, it is a common practice to provide at least morphological information and possibly syntactic constituency and dependency information for such corpora. Such enrichment was important for testing syntactic theories of processing (e.g., [1]). However, we lack a psycholinguistic corpus that would also include semantic annotations. This arguably affects the current state of affairs of theorizing on processing, which heavily leans on syntactic theories, lacking data and models on which semantic processing theories could be evaluated. Providing a reading-time corpus with rich (discourse) semantic annotations, developed independently in formal semantics, should help close this gap.

Method: We chose an existing psycholinguistic corpus (UCL corpus [2]) and enriched it with annotations for semantics. The annotations were carried out within the Parallel Meaning Bank, PMB [3], which provides meaning representations for all language levels - from words to discourses. The annotations are compatible with independently developed databases/frameworks: wordnet for the word senses, verbnet for thematic roles, CCG for syntax and Discourse Representation Theory (DRT) for discourse. We used the automatic parser+BOXER for the first round of annotations and manually checked and cleaned all sentences.

Results+Discussion: An example of a fully annotated sentence is given in (1), which shows two meaning representations: a sequence notation, and a box notation, common in DRT. Focusing on the former annotation, we note that the semantic information carries the wordnet synsets (listed/displayed in the first column, starting with person.n.01), the verbnet roles and discourse operators (in the second column) and the word strings that introduce operators and synsets (in the third column). Inferred meaning (e.g., the hearer in imperatives) is also represented. The information can be used to assess semantics on processing, as showcased here on two case studies: **(A)** Can we find encoding interference due to semantic similarity? To address this question, we checked in the annotated corpus whether a word that introduces a male is processed slower if there is already another male present in the same discourse. We used Bayesian hierarchical models with log-frequency, word position and word length covariates and the maximal random effect structure (subjects and words being random effects). We observed indeed that there is a clear increased difficulty when the word introduces an extra male in discourse, (see (2)). This provides novel support for theories on encoding interference [4]. **(B)** Can we find integration cost due to the distance between a role assigner (e.g., a verb) and its arguments, as proposed by Dependency Locality Theory? Following the methodology just described, we found integration cost in the data (see (3)), but only if we exclude cases in which the role assigner is right next to the argument. This is in line with a previous corpus study of [5], but for the first time evaluated on semantic corpus. The case studies are just a few examples showing that the annotated corpus should be useful to enhance our understanding of semantic processing and make the incorporation of semantic theories into processing models more testable in future research.

(1) Come within the walls to escape the enemy.

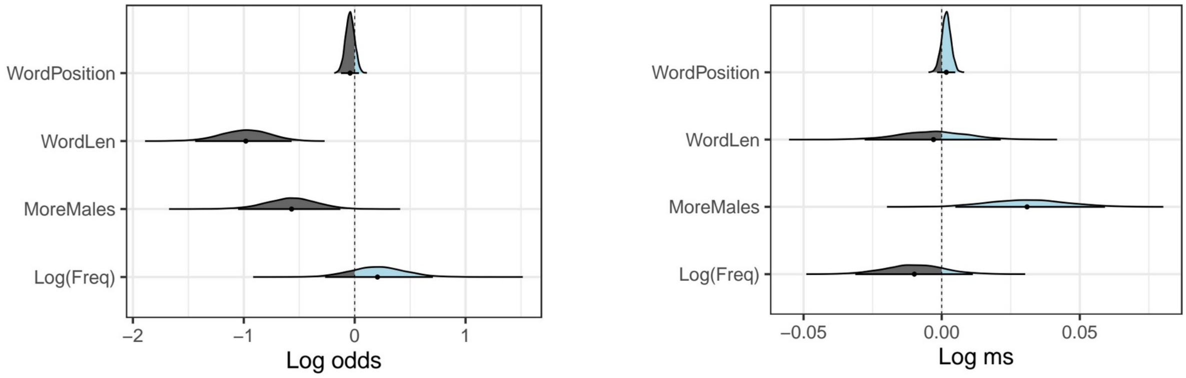
Meaning Representation (sequence notation)

```
person.n.01 EQU hearer                        %
come.v.01   Theme -1 Location +1 Result +2 % Come within  [0-11]
wall.n.01                                   % the walls to [12-24]
escape.v.01 Theme -3 Source +1             % escape       [25-31]
person.n.01 Role +1                         % the enemy.   [32-42]
enemy.n.03                                  %
```

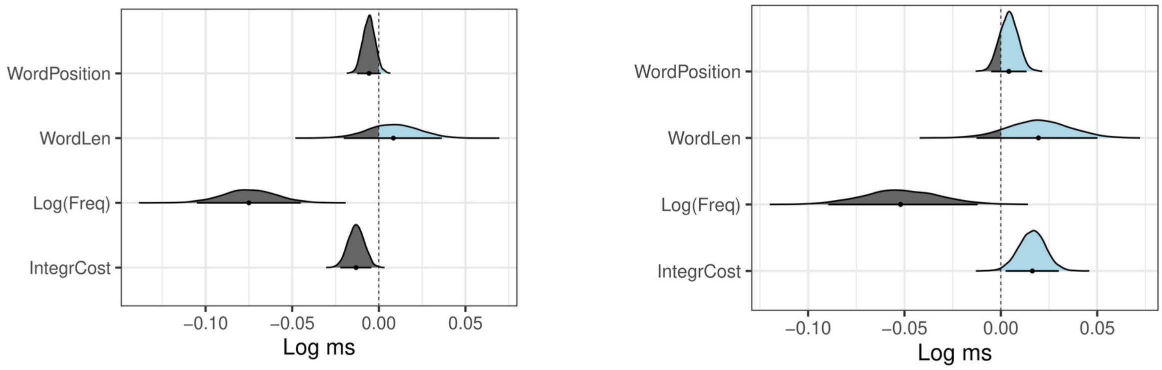
Meaning Representation (box notation)

```
b1: x1 e1 x2 e2 x3 x4
- - - - - - - - - - -
person.n.01(x1)
x1 = hearer
come.v.01(e1)
e1 Theme x1
e1 Location x2
e1 Result e2
wall.n.01(x2)
escape.v.01(e2)
e2 Theme x1
e2 Source x3
person.n.01(x3)
x3 Role x4
enemy.n.03(x4)
```

(2) Posteriors for reading measures on words introducing a male. The thick horizontal lines represent 95% credible intervals. MoreMales measures how many males appeared in the discourse up to then. Left graph: skipping rate of the target word in the eye-tracking data, showing a decrease in skipping when more males already present. Right graph: reading times on the spillover word in the self-paced reading data, showing an increase in reading times (reading times on eye tracking measures show similar effects).



(3) Posteriors for early eye-tracking reading measures on words introducing thematic roles. IntegrCost is a measure of distance between the word and its most remote argument. The left graph shows posteriors if we include all data. The right graph excludes cases in which the most remote argument is the previous word.



[1] Shain, C. et al. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience*, *42*(39), 7412-7430.

[2] Frank, S. L. et al. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior research methods*, *45*, 1182-1190.

[3] Abzianidze, et al. (2017). The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. *arXiv preprint arXiv:1702.03964*.

[4] Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of memory and language*, *55*(4), 601-626.

[5] Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193-210.

Event construal through social verbs in English, German, and Hungarian: The LISADA corpus

Tiziana Srdoc*, Elena Marx, Anna Viola Sáfrány, Eva Wittenberg
Central European University, Vienna
*SrdocT@ceu.edu

Humans navigate complex social environments by rapidly making sense of various social cues and events, but how we talk about these events matters. The same event can be framed differently, for example, as “buying” or “selling,” and an action like “kissing” can be construed as mutual, joint, or both, depending on the linguistic context [1]. While previous research has shown that different lexical and grammatical cues influence the mental representation of events [2–5], the specific impact of language on the construal of social events remains relatively unexplored. Thus, in this study, we investigated how people interpret linguistic descriptions of social events that are inherently social (e.g., “*to meet”)* or potentially social (e.g., “*to dance”*), and how these interpretations align or differ across languages.

To answer these questions, we created the *Linguistic Impact on Social Action Construal Database* (LISADA), a corpus of social action comprehension in English, German, and Hungarian [6]. The corpus consists of native speakers’ ratings on two key semantic dimensions: mutuality (doing something to each other) and jointness (doing something together). Ratings were collected for 240 infinitive verbs selected for their ability to express social actions involving multiple participants [7, 8]. Participants (N = 60 each for English and German, N = 30 for Hungarian) completed two between-subjects forced-choice tasks: indicating whether the presented verb expressed a mutual or a self-directed action (Mutuality Task), or a joint or an individual action (Jointness Task; Fig. 1). We used hierarchical cluster analyses and zero-order correlations to identify verb groupings and assess the relationship between mutuality and jointness across languages. To illustrate a case of application, we also examined whether morphosyntactic differences between German and English affect social event construal along these dimensions. Specifically, we tested whether verbs that occur with the German marker *sich*, which is ambiguous between reciprocal and reflexive readings, introduce greater response variability (i.e., CI size) within and across languages. We compared nested mixed-effects linear models that incrementally added language and occurrence with *sich* and their interaction as predictors.

Our findings revealed both cross-linguistic similarities and differences. All three languages exhibited a strong positive correlation between mutuality and jointness (English: $\tau = 0.75$; German: $\tau = 0.65$; Hungarian: $\tau = 0.64$). However, the number and structure of verb clusters varied: four in English, two in German, and three in Hungarian (Fig. 2). The data also revealed subtle differences in the construal of social actions, with some verb clusters showing disagreement between mutuality and jointness, suggesting that the dimensions do not map directly onto each other. Furthermore, the cross-linguistic comparison revealed greater inter-participant disagreement for *sich* verbs, as well as equivalent English verbs, in the Jointness Task ($df = 1$, $\chi^2 = 24.30$, $p <.001$; no effect for the Mutuality Task, $df = 1$, $\chi^2 = 2.67$, $p = 0.10$; Fig. 3), suggesting that the observed variance is rooted in conceptual rather than morphosyntactic factors. That verbs, even without additional morphosyntactic properties, induce similar social judgments cross-linguistically aligns with previous research showing that verbs rapidly activate conceptual knowledge about agents, patients, and instruments [9]. In sum, our LISADA corpus offers novel insights into how social events are linguistically and conceptually construed across languages and as illustrated by the exemplary test case, provides a valuable resource for further research at the interface of language and social event cognition.

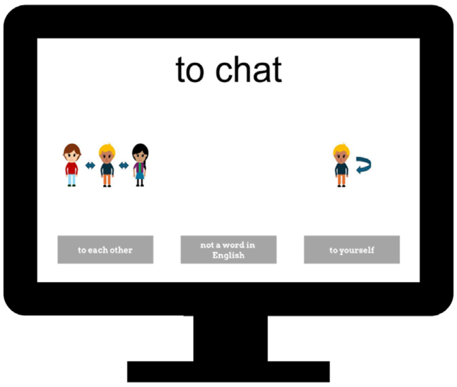


Figure 1 Example of a trial in the mutual condition

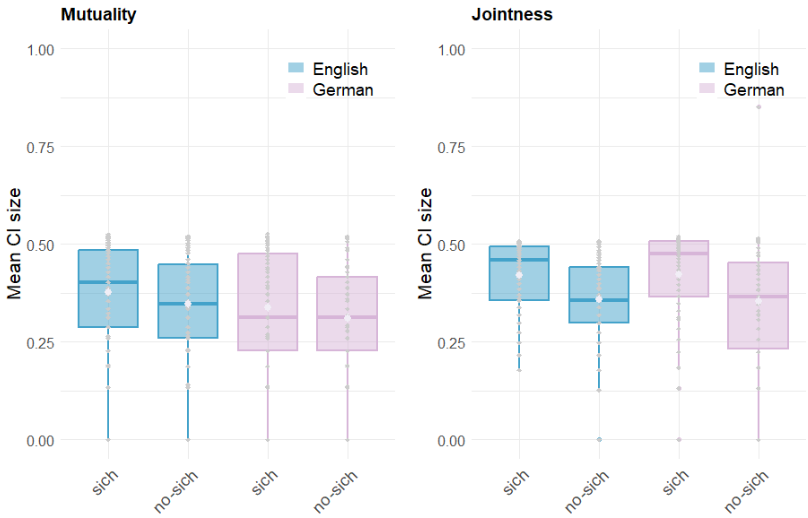


Figure 3 Comparison of CI sizes for mutuality and jointness scores between English and German. Box plots represent the distribution of CI sizes for individual verbs; dots represent overall mean scores

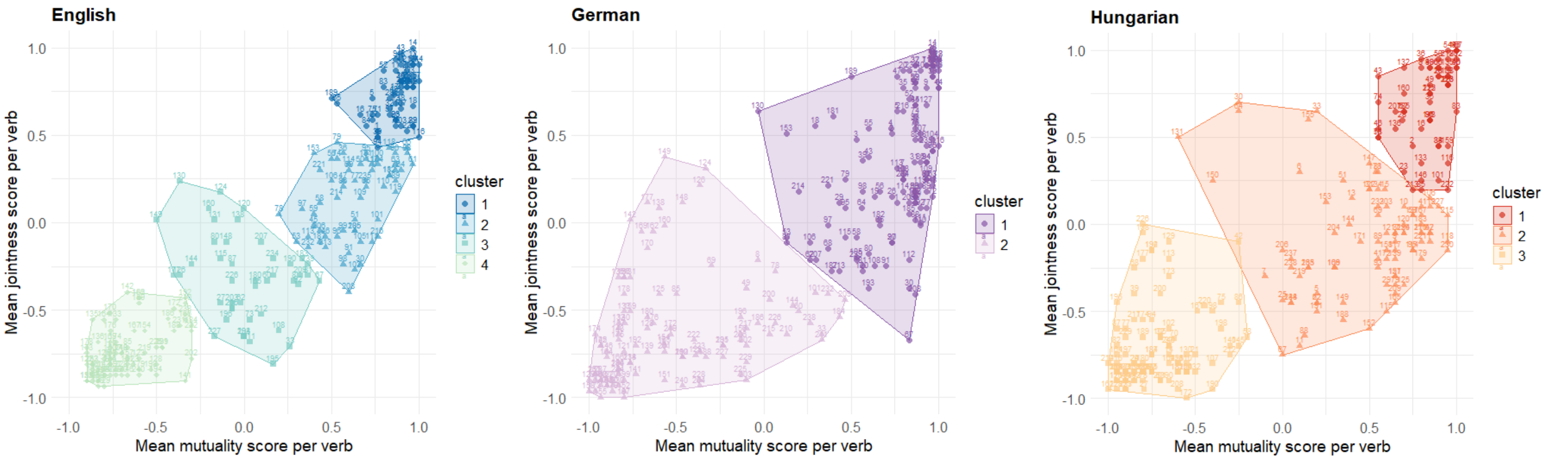


Figure 2 Scatterplots of cluster solutions for English, German and Hungarian verbs based on mutuality and jointness scores. Individual verbs are numbered from 1 to 240 correspondingly across languages; values range from -1 (indicating low mutuality or jointness) to 1 (indicating high mutuality or jointness)

References

[1] Wittenberg, E., Khan, M. & Snedeker, J. (2017). Investigating thematic roles through implicit learning: Evidence from light verb constructions. *Frontiers in Psychology, 8*, 1089.

[2] Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition, 214*, 104744.

[3] Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*(1), 155-184.

[4] Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition, 206*, 104516.

[5] Slivac, K., Hervais-Adelman, A., Hagoort, P., & Flecken, M. (2021). Linguistic labels cue biological motion perception and misperception. *Scientific reports, 11*(1), 1-14.

[6] Srdoc, T., Marx, E., Sáfrány, A. V., Balla, A., & Wittenberg, E. (2025). *Linguistic Impact on Social Action Construal Database (LISADA).* https://osf.io/xzdqc/

[7] Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press

[8] Nedjalkov, V. P. & Geniušienė, M. (2007). Questionnaire on reciprocals. In V. P. Nedjalkov (Ed.), *Reciprocal constructions* (p. 3-113). Amsterdam: John Benjamins.

[9] Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*(4), 516–547.

When focus overrides form: Prosodic rephrasing in Mandarin complex nominals

Yu-yin Hsu[1], Anqi Xu[2]

The Hong Kong Polytechnic University[1] ; Harbin Institute of Technology, Shenzhen[2]

yu-yin.hsu@polyu.edu.hk, xuanqi@hit.edu.cn

Background: Extensive research has established that focus in Mandarin is marked by increased duration and expanded F_0 range on the focused constituent, followed by post-focus compression (PFC), a systematic reduction in these acoustic dimensions [1,2]. However, prior investigations have primarily examined focus marking in simple words, leaving it unclear whether focus also affects the prosodic structuring within more complex phrases. Meanwhile, research on prosodic phrasing in Mandarin has either been predominantly theoretical [3], lacking empirical validation, or has not explicitly examined how prosody is realized under different focus conditions [4]. Given that Mandarin phrases exhibit consistent durational patterns that reflect their prosodic grouping (more reliable than f_0) [5], deviations from these patterns under focus may reflect changes in how the elements of a phrase are prosodically grouped and organized. This study investigates whether focus can reshape the default prosodic structure of Mandarin numeral-classifier-noun phrases. We hypothesize that focus may alter the typical prosodic grouping of these complex nominals.

Method: We conducted two speech production experiments with native Mandarin speakers. Numeral-classifier-noun phrases were embedded in carrier sentences, with a contrastive focus elicited on either the numeral (CNUM) or entire noun phrase (CNP), alongside a baseline condition with no focus (ODNP), presented in a mini-dialogue paradigm (Table 1). Experiment 1 used monosyllabic numerals (14 speakers), and Experiment 2 used disyllabic numerals (14 speakers). All syllables in the target phrase bore the same tone (e.g., 三枝花 sān zhī huā 'three flowers', and 三千枝花 sān qiān zhī huā 'three thousand flowers'). Durational measurements were analyzed using linear mixed-effects models, with Syllable Location, Focus Type, Tone, and their interaction as fixed factors.

Results: In Experiment 1, there was a significant interaction between Focus Type and Syllable Location ($\chi^2(6)=221.62$, $p<.001$). As shown in Figure 1, phrases under CNP resembled those in ODNP, whereas phrases under CNUM displayed a distinct durational pattern, suggesting prosodic restructuring when the focus span (numeral) diverged from the default prosodic unit (NP). Post-hoc analyses confirmed that in ODNP and CNP, nouns were significantly longer than numerals ($p<.001$), but in CNUM, no significant duration difference was observed between numerals and nouns ($p=.998$). Similarly, Experiment 2 revealed a significant interaction between Focus Type and Syllable Location ($\chi^2(8)=185.36$, $p<.001$). The duration ratio of the first numeral did not differ between ODNP and CNP ($p>.050$), but a significant difference emerged in CNUM ($p<.001$), indicating that numeral focus induced reorganization of prosodic structure even with disyllabic numerals.

Discussion: In both the monosyllabic and disyllabic numeral conditions, phrases with old information (ODNP) or focus on the entire NP (CNP) preserved the default temporal patterning. This indicates that when the focus domain coincides with the default prosodic unit (the NP), focus does not induce internal prosodic restructuring. However, under corrective focus on the numeral (CNUM), the durational profiles diverged from the expected rhythmic patterns reported for trisyllabic compounds [5], likely due to on-focus lengthening [1–2]. These findings suggest that prosodic phrasing in Mandarin is influenced not only by morphosyntactic structure (e.g., compounds vs. phrases) but also by pragmatic factors such as focus. Focus marking, therefore, can reorganize prosodic structures, highlighting the intricate interaction among prosodic organization and syntax in Mandarin. Future studies should explore whether the observed prosodic restructuring generalizes to other complex nominals both within Mandarin and cross-linguistically.

Figure 1. Effect of focus condition on the duration ratio of each syllable over the whole monosyllabic/disyllabic numeral-classifier-noun phrases.

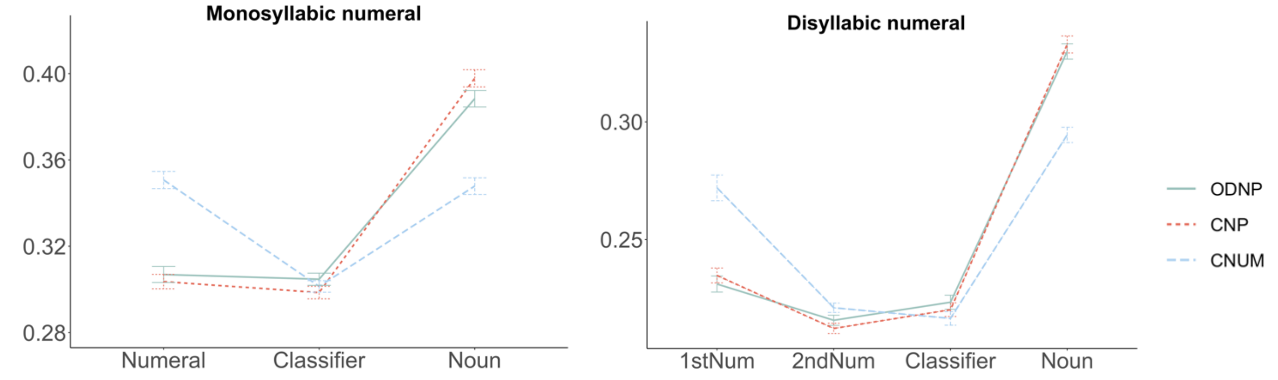


Table 1. Mini-dialogue illustrating three focus conditions with Mandarin numeral-classifier-noun phrases. The words bearing corrective focus are in bold, and the corresponding words being corrected in the leading sentence are underlined.

Information	**Leading sentence**	**Target sentence**	**Pinyin & Characters**
Old NP (ODNP)	What happened to <u>five bowls of liquor</u>?	**Five bowls of liquor** spilled all over the table.	
Contrastive NP (CNP)	<u>A bottle of water</u> spilled all over the table.	No, **five bowls of liquor** spilled all over the table.	wǔ wǎn jiǔ sǎ le yī zhuō zi 五 碗 酒 洒了一 桌 子
Contrastive numeral (CNUM)	<u>Two</u> bowls of liquor spilled all over the table.	No, **five** bowls of liquor spilled all over the table.	

References

[1] Chen, Y. (2010). Post-focus F0 compression—Now you see it, now you don’t. Journal of Phonetics, 38(4), 517–525.

[2] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. Journal of Phonetics, 27(1), 55–105.

[3] Duanmu, S. (1999). Zhongyin lilun he Hanyu de ci chang xuanze [Metrical theory and word length choices in Chinese], Zhongguo Yuwen, 4(271), 246–254.

[4] Ni, C., & Liu W., (2008). Durational Characteristics and Pitch Characteristics of the Prosodic Phrase in Mandarin Chinese, in *2008 Chinese Conference on Pattern Recognition*, 1–5.

[5] Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. Journal of Phonetics, 37(4), 502–520.

The tonal system of Mandarin

Mandarin Chinese is a lexical tone language, where each syllable is associated with one of four canonical tones, and a neutral tone. The tonal inventory is as follows:

- Tone 1 (High-Level, ¯): A steady high pitch (e.g., 妈 /mā/ ‘mother’).
- Tone 2 (Rising, ˊ): A dynamic pitch rise from mid to high (e.g., 麻 /má/ ‘hemp’).
- Tone 3 (Low-Dipping, ˇ): A falling-rising contour (mid-low-high) (e.g., 马 /mǎ/ ‘horse’).
- Tone 4 (Falling, ˋ): A sharp fall from high to low pitch (e.g., 骂 /mà/ ‘to scold’).
- Neutral tone: An unstressed, reduced tone, lacking a fixed contour (e.g., 吗 /ma/ — a question particle).

Focus marking in Mandarin

In Mandarin Chinese, on-focus words are typically marked by increases in both syllable duration and fundamental frequency (F_0), followed by a sharp reduction in these parameters. This pattern of post-focal reduction is known as post-focus compression (PFC) [1,2]. However, tone type influences the realization of focus. Only words with high tone (Tone 1) exhibit the typical raised on-focus f_o and reduced post-focus f_o, whereas other tonal conditions rely more on pitch expansion, rather than voice pitch. For example, the focused words with a falling tone (Tone 4) and a low tone (Tone 3) may even show decreased f_o [2]. Moreover, the post-focus regions exhibit varied prosodic patterns, featuring compressed F0 range and occasionally reversed patterns. Some studies suggest that some tonal contexts may not necessarily trigger a f_o range reduction in the post-focus words [1]; for instance, a post-focus rising tone (Tone 2) shows compression after a high tone but expansion after a low tone, while falling tone after focus shows expansion after a high tone but remains unchanged after a low tone.

Prosodic phrasing in Mandarin

Duanmu [6] and Feng [7] posit that disyllabic footing functions as a default rhythmic mechanism in Mandarin, accounting for the predominance of disyllabic words and the systematic prosodic division of trisyllabic structures into "1+2" or "2+1" patterns, and quadrisyllabic structures into "2+2" patterns. However, empirical investigations have not consistently supported these theoretical proposals concerning syntactic-prosodic mapping [8]. For example, in noun phrases shorter than four syllables, durational patterns reveal that final syllables are the longest, initial syllables are moderately shorter, and medial syllables are the shortest. Furthermore, the morphosyntactic structure of four-syllable sequences systematically modulates their durational profiles: in sequences consisting of a monosyllabic modifier followed by a trisyllabic head noun, syllable duration follows the ranking 1st > 4th > 2nd > 3rd; in sequences composed of two disyllabic words, the durational hierarchy shifts to 4th > 1st > 3rd > 2nd.

References

[6] Duanmu, S. (1999). “Zhongyin lilun he Hanyu de ci chang xuanze [Metrical theory and word length choices in Chinese],” Zhongguo Yuwen, 4(271), 246–254.

[7] Feng, S. (1998). Lun Hayu De Ziran Yinbu [On the natural foot in Chinese],” Zhongguo Yuwen [Chinese Linguistics], 1, 40–47.

[8] Xu Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns In Mandarin,” Journal of Phonetics, 37(4), 502–520.

Divergence Point Analysis: does it *really* establish the precise timepoint of divergence?

John C.B. Gamboa, Shaiban Alshaibani, Christopher Allison, Leigh B. Fernandez and Shanley E.M. Allen
University of Kaiserslautern-Landau
gamboa@rptu.de

Background: In a Visual World Paradigm (VWP) study, participants hear a sentence while looking at a number of objects on the screen. We can use this paradigm, among other things, to test prediction. Evidence of prediction can be seen when looks to the target occur before it is explicitly spoken. Recently, the Divergence Point Analysis (DPA) method [1] has become a popular way to determine the time in which a prediction occurs. The DPA works by dividing time into bins of (for example) 20ms, and, for each bin, each participant's proportion of looks to the target in relation to a distractor is calculated. Before a prediction, these proportions should be around 50% (i.e., at chance level); and it should increase over time as participants make the prediction. Therefore, at each bin, all participants' proportions are input into a one-sample t-test that compares them with 0.5. That is, the t-test will be "significant" if these proportions are "significantly different" from 50%. The Divergence Point (DP) is defined as the first bin from which a run of (for example) 10 bins is significant. In order to produce confidence intervals, researchers typically use bootstrapping, i.e., they resample the participants' individual trials with replacement and recalculate the DP for a large number (e.g., 2000) of resampled datasets. Intuitively, if we ran a VWP experiment with N participants, each of which saw T trials, and calculated the DP of this data, we would expect it to approximate that of the overall population. That is, increasing N and T should lead it to become closer and closer to the real DP of the population. But is it really the case?

Method: We reanalyzed data from [2], which contained two groups of participants (Younger adults vs. Older adults) and two sets of trials (S1 with 48 trials, and S2 with 32 trials). In their experiment, participants heard sentences such as *The tailor trims the suit* and saw objects such as a *suit* (target) and a *pot* (distractor). To examine the impact of T (the number of trials), we combined the two trial sets (S1+S2) and ran a DPA with a 2000 iteration bootstrap. To examine the impact of N (the number of participants, $N \in \{30, 45, 60, 75, 90, 120, 180, 240, 300\}$), we modified the bootstrapping method slightly (with 2000 iterations): before each iteration, we resampled participants' data so that the dataset contained N participants. Then we performed the DPA procedure normally.

Results: Figure 1 shows the results for varying T. In both groups, the resulting divergence point was outside the confidence interval of S1. In addition, for Younger adults, the confidence interval for S1+S2 moved *earlier* than each trial set individually. Figure 2 shows the results for varying the population size N. For both groups, the DP consistently decreased as population sizes increased.

Discussion: The results suggest that, as the amount of data increases, the calculated DP decreases. This seems to happen because, as N (and, to some extent, T) increases, the t-tests (of the DPA) become more and more likely to produce a "significant" result. The implications of these results are twofold. First, while the DPA does allow "to directly compare timing differences between speaker groups and experimental manipulations", it does not really "establish the precise timepoint at which an effect onsets" [1, p. 833]. Second, upon comparing the DPs of different groups, one should be careful to ensure that the values of N and T do not differ excessively. It is not clear whether or how the DPA can be modified so that it approximates the population DP.

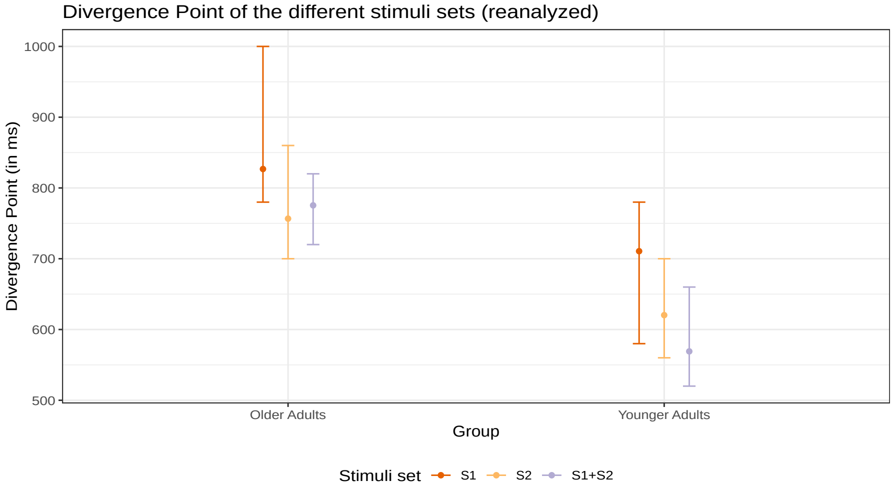


Figure 1: The Divergence Point for the the different stimuli sets. While the results for the Older adults conform better to our expectations, the DP and confidence intervals for Youger adults consistently moved to an earlier point in time.

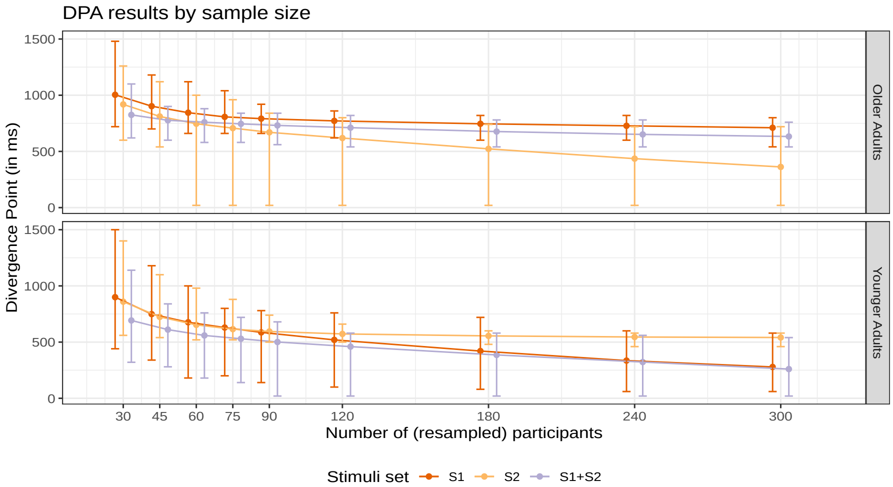


Figure 2: The Divergence Point for the the different population sizes. As population size increases, the DP moves to earlier timepoints. The greater variability is the result of the participant resampling procedure.

References

[1] Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. https://doi.org/10.1017/S1366728920000607

[2] Fernandez, L. B., Shehzad, M., & Hadley, L. V. (2025). Younger adults may be faster at making semantic predictions, but older adults are more efficient. *Psychology and Aging*, *40*(3), 318–325. https://doi.org/10.1037/pag0000872

Does language similarity affect second language prediction in discourse comprehension? Evidence from visual-world Eye-tracking

Binger Lu[1], Julie E. Boland[2], Robert J. Hartsuiker[1]
[1] Ghent University; [2] University of Michigan
Binger.Lu@UGent.be

Previous research has demonstrated that cross-linguistic influence plays a role in second language (L2) predictive processing [1, 2], with L2 prediction partially shaped by learners' first language (L1) experience. Consequently, prediction in L2 comprehension may differ depending on learners' L1 backgrounds and the degree of similarity between their L1 and L2. However, it remains unclear whether such L1–L2 similarity also affects L2 prediction at the discourse level, which is more complex and cognitively demanding than sentence processing. One possibility is that greater similarity between two languages may mitigate the processing difficulty for L2 learners during discourse comprehension.

In our earlier visual-world eye-tracking experiment, we examined discourse prediction in Dutch-English bilinguals (L1: N = 32; L2: N = 32) and observed similar patterns across L1 and L2 in both the magnitude and time course of predictive processing (presented at AMLaP, 2024). Each trial consisted of a four-sentence mini-discourse, with Sentence 1 manipulated to provide either a highly predictive or a neutral context for a target word in Sentence 4 (the critical sentence). Sentences 2 and 3 were identical across conditions. The critical sentence had an SVO structure, in which the verb was semantically compatible with all four visual referents on the screen. Using the same materials and procedure, we now tested 32 Chinese-English bilinguals and 32 English native speakers, allowing us to compare L2 discourse prediction across different language backgrounds and to test for differences between L1 and L2 discourse prediction.

Eye-tracking data were analyzed across two time windows within the critical sentence: the first started from 200 ms after audio onset to 200 ms after verb onset and the second from 200 ms post-verb onset to 200 ms post-noun onset. All groups of participants made rapid predictions in English after the visual objects appeared (Figure 1) in the first time window (pictures were shown from 500 ms before Sentence 4 onset). In the second window, Dutch- and Chinese-English bilinguals diverged: Dutch-English bilinguals maintained their anticipatory looks, whereas Chinese-English bilinguals gradually disengaged from predictive fixations until the target noun was mentioned. Native speakers showed stable and robust prediction throughout, which was stronger than L2.

In sum, native speakers demonstrated more sensitivity to discourse cues in predictive processing than L2 learners. Nevertheless, both bilingual groups showed a clear ability to make rapid predictions in L2 discourse. After such predictive processing, both bilingual groups seemed to decrease their eye gaze on the target and turned their looks to distractors. However, after verb onset, Chinese-English bilinguals appeared to be more strongly affected by the semantics of this neutral verb, leading to a decline in predictive engagement. In contrast, Dutch-English bilinguals confirmed and maintained prediction, similarly to the English native speakers in L1. These findings suggest that language similarity influences the mechanisms underlying L2 discourse prediction. This is in line with evidence that L2 learners show more native-like patterns when their L1 and L2 are more similar [e.g., 3]. Further research needs to investigate why Chinese-English bilinguals seem to process discourse information differently. Additionally, it remains possible that both cross-linguistic influence and L2 proficiency can play a role in L2 discourse prediction simultaneously (L2 proficiency was measured by the LexTALE test of English [4], Chinese-English group: M = 62%, Dutch-English group: M = 77%).

References

[1] Foucart, A. (2021). Language prediction in second language: Does language similarity matter?. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 91–114). John Benjamins.

[2] Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253-267.

[3] Van Bergen, G. & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26-42.

[4] Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*, 325-343.

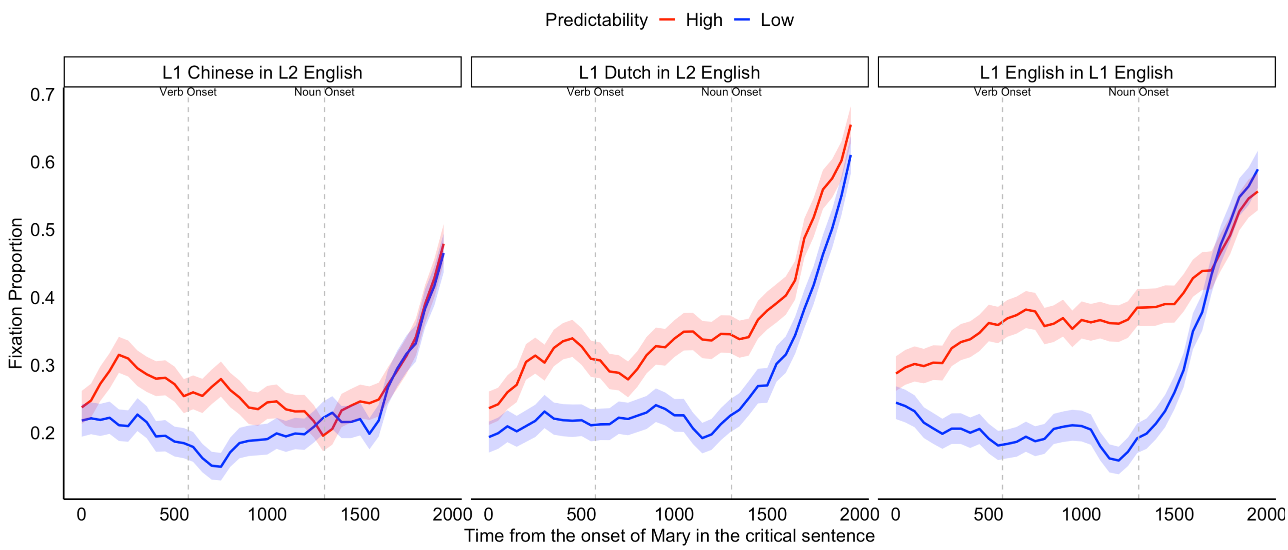
Example Sentences:

1. Highly Constrained: *Mary lives far from school. She always complains about this. She wants to make her life easier. Mary uses the bike.*
2. Neutral: *Mary has a big problem. She always complains about this. She wants to make her life easier. Mary uses the bike.*

Objects: "bike", "pan", "drill", "hammer"

Figure 1

Eye-tracking results of three groups of participants in the English task



Relative clause processing and comprehension in Greek: Effects of academic background

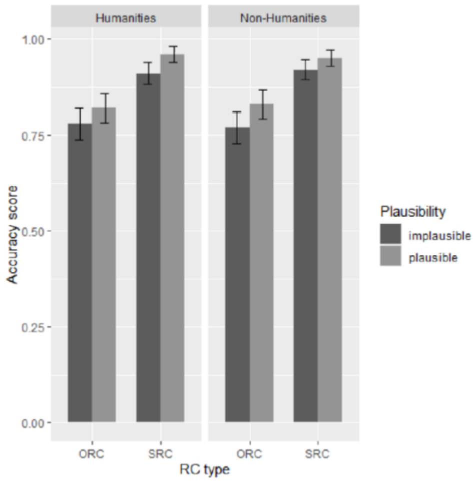
Gerakini Douka & Despina Papadopoulou
Aristotle University of Thessaloniki
dgerakini@lit.auth.gr

Background: Previous research [1] has shown that individuals with lower levels of education comprehend relative clauses (RCs) less accurately than those with higher education, even when the latter are second language speakers. This advantage has been linked to enhanced metalinguistic skills. Building on these findings, our study seeks to refine the understanding of educational effects on RC comprehension by examining variation within highly educated populations—specifically, comparing students from humanities and non-humanities disciplines. We also explore how cognitive and metalinguistic abilities influence RC processing and comprehension.

Method: We explored RC parsing and comprehension in 80 monolingually raised Greek-speaking adults (mean age: 20.60) using an eye-tracking paradigm. Half of the participants were students from the School of Philology at the Aristotle University of Thessaloniki (35 female, mean age=19.30), while the other half were students from the university's Departments of Physics and Biology (18 female, mean age=21.90). We examined center-embedded RCs introduced with the relative pronoun *o opíos*. RCs were manipulated for syntactic structure (subject vs. object RCs) and plausibility (plausible vs. implausible) (Example 1). These manipulations were related to the well-documented SRC-ORC asymmetry pointing towards a penalty for ORCs [2, 3], and to the use of the plausibility heuristic pointing towards a disadvantage for implausible sentences [4]. Additionally, we aimed to explore the role of individual variation in RC processing [5]. To this end, participants completed a battery of assessments targeting general cognitive abilities, such as working memory and inhibitory control, as well as reading rate and metalinguistic skills, namely morphosyntactic awareness, ambiguity awareness and knowledge of metalanguage.

Results: Accuracy results revealed that both humanities and non-humanities participants performed less accurately on object relative clauses (ORCs) compared to subject relative clauses (SRCs), and on implausible RCs compared to plausible ones (see Fig. 1). In terms of online processing, reading times (RTs) indicated that participants from non-humanities backgrounds were generally slower than their humanities counterparts. What is more, an ORC disadvantage was evident across both groups, though on some regions it was more pronounced—or exclusive—to the non-humanities group (see Fig. 2). Additionally, implausible sentences were read more slowly than plausible ones across both groups. Further analyses examining individual differences in total duration of fixations showed that, within the humanities group, higher working memory capacity was associated with a reduced ORC disadvantage at the relativizer (see Fig. 3A). In the non-humanities group, greater reading fluency was linked to shorter RTs at the main verb (see Fig. 3B), whereas increased morphosyntactic awareness was associated with longer RTs at the main verb (see Fig. 3C).

Discussion: Overall, the present study aligns with previous research regarding the ORC disadvantage and the use of the plausibility heuristic [2, 3, 4]. What is more, it demonstrates that (a) academic background influences relative clause processing, and (b) different constructs of individual variation affect RC processing in the two groups of participants.



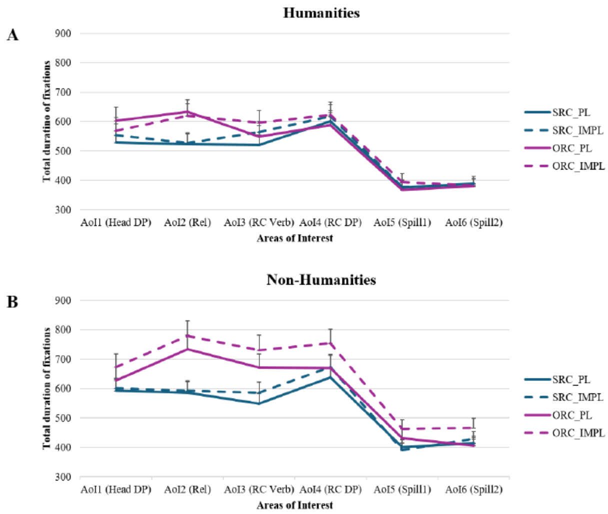
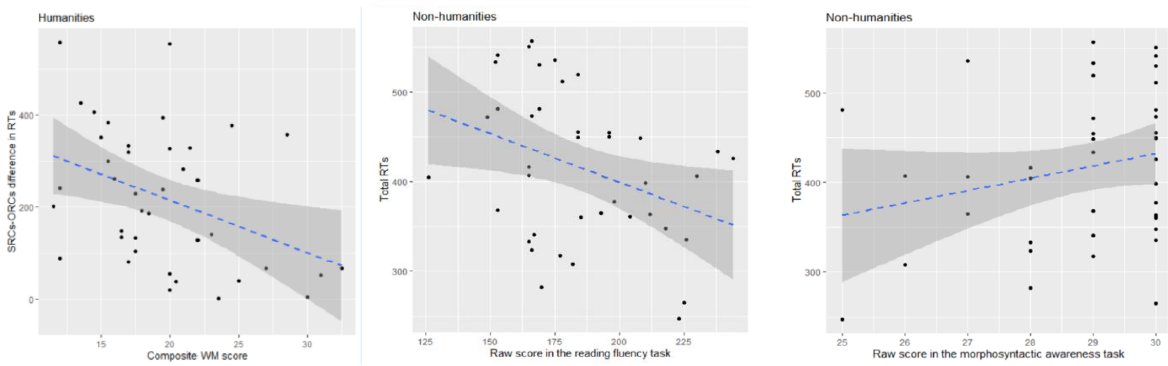


Fig. 1 Accuracy data

Fig. 2 Total duration of fixations



A **B** **C**

Fig.3 Effects of individual variation

References

[1] Street, J. A. (2017). This is the native speaker that the non-native speaker outperformed: Individual, education-related differences in the processing and interpretation of Object Relative Clauses by native and non-native speakers of English. *Language Sciences*, *59*, 192-203.

[2] Papadopoulou, D., Amvrazis, N., Douka, G., & Tantos, A. (2024). Triangulating learner corpus and online experimental data: Evidence from gender agreement and relative clauses in L2 Greek. *The Modern Language Journal*, *108*(4), 932-953.

[3] Papadopoulou, D., Douka, G., & Paspali, A. (2025). Similarity effects in the online and offline comprehension of relative clauses: Evidence from L1 and L2 Greek. *Glossa: a journal of general linguistics*, *10*(1).

[4] Zhou, P., & Christianson, K. (2016). I “hear” what you're “saying”: Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, *69*(5), 972-995.

[5] James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions?. *Journal of Memory and Language, 102,* 155-181.

Example 1

a. SRC, plausible

Ο	*σκύλος*	*ο*	*οποίος*	*δάγκωσε*
o	scílos	o	opíos	ðágose
the$_{\text{NOM.SG.M}}$	dog$_{\text{NOM.SG.M}}$	the$_{\text{NOM.SG.M}}$	who$_{\text{NOM.SG.M}}$	bite$_{\text{PST.PFV.3SG}}$
τον	*άντρα*	*είχε*	*καφέ*	*μάτια.*
ton	ádra	íçe	kafé	mátça
the$_{\text{ACC.SG.M}}$	man$_{\text{ACC.SG.M}}$	have$_{\text{PST.IPFV.3SG}}$	brown$_{\text{ACC.PL.N}}$	eye$_{\text{ACC.PL.N}}$

‘The dog that bit the man had brown eyes.’

b. SRC, implausible

Ο	*άντρας*	*ο*	*οποίος*	*δάγκωσε*
o	ádras	o	opíos	ðágose
the$_{\text{NOM.SG.M}}$	man$_{\text{NOM.SG.M}}$	the$_{\text{NOM.SG.M}}$	who$_{\text{NOM.SG.M}}$	bite$_{\text{PST.PFV.3SG}}$
τον	*σκύλο*	*είχε*	*καφέ*	*μάτια.*
ton	scílo	íçe	kafé	mátça
the$_{\text{ACC.SG.M}}$	dog$_{\text{ACC.SG.M}}$	have$_{\text{PST.IPFV.3SG}}$	brown$_{\text{ACC.PL.N}}$	eye$_{\text{ACC.PL.N}}$

‘The man that bit the dog had brown eyes.’

c. ORC, plausible

Ο	*άντρας*	*τον*	*οποίο*	*δάγκωσε*
o	ádras	ton	opío	ðágose
the$_{\text{NOM.SG.M}}$	man$_{\text{NOM.SG.M}}$	the$_{\text{ACC.SG.M}}$	who$_{\text{ACC.SG.M}}$	bite$_{\text{PST.PFV.3SG}}$
ο	*σκύλος*	*είχε*	*καφέ*	*μάτια.*
o	scílos	íçe	kafé	mátça
the$_{\text{NOM.SG.M}}$	dog$_{\text{NOM.SG.M}}$	have$_{\text{PST.IPFV.3SG}}$	brown$_{\text{ACC.PL.N}}$	eye$_{\text{ACC.PL.N}}$

‘The man that the dog bit had brown eyes.’

d. ORC, implausible

Ο	*σκύλος*	*τον*	*οποίο*	*δάγκωσε*
o	scílos	ton	opío	ðágose
the$_{\text{NOM.SG.M}}$	dog$_{\text{NOM.SG.M}}$	the$_{\text{ACC.SG.M}}$	who$_{\text{ACC.SG.M}}$	bite$_{\text{PST.PFV.3SG}}$
ο	*άντρας*	*είχε*	*καφέ*	*μάτια.*
o	ádras	íçe	kafé	mátça
the$_{\text{NOM.SG.M}}$	man$_{\text{NOM.SG.M}}$	have$_{\text{PST.IPFV.3SG}}$	brown$_{\text{ACC.PL.N}}$	eye$_{\text{ACC.PL.N}}$

‘The dog that the man bit had brown eyes.’

"Write here, write now": Spelling difficulty disrupts parallel planning in sentence production

Jens Roeser[1], Pablo Aros Muñoz[1], Mark Torrance[1]
[1] Department of Psychology, Nottingham Trent University
jens.roeser@ntu.ac.uk

Background: Expressing ideas in language involves a cascade of processes starting with an intended message, moving on to retrieving words, their spelling or phonology, and finally outputting them in speech, onto paper or screen. Our understanding of the timecourse of linguistic planning and, in particular, spelling retrieval in writing is limited. Word spelling may [1] or may not [2] be fully retrieved at word onset and upcoming linguistic information may be planned at linguistic edges or in parallel to production [3]. We tested to what extent difficulty with orthography delays planning of upcoming information. Importantly, only the parallel view predicts that disruptions delay planning of upcoming lexical information.

Method: In three experiments participants (Ns$_{\text{ppts}}$ = 56, 80, 96) were asked to describe spatial arrangements of images (N_{items} = 36) with typed utterances of the form *The N1 and the N2 are above the N3*; see Fig. 1. To assess the extent to which (1) N1 orthography and (2) retrieval of N2 unfolded in parallel to writing, we manipulated N1 spelling: the name of the image of N1 was – according to independent naming data (Ns = 100) – difficult (e.g. *caterpillar*) or easy to spell (e.g. *watermelon*) and 2-4 syllables long. To reduce time for parallel planning we added a one-syllable and easy-to-spell N1 condition (e.g. *fly*). Prior to response, participants saw similar displays with auditory utterances starting with the same noun as the target to control for lexical (but not orthographic) retrieval. Prime sentences had either the same syntactic form as the target, or started with a simple noun phrase (*The N1 is above the N2 and the N3*). To constrain parallel planning, a third prime condition was added to Exp. 2 that elicited sentences without N2-initial determiner (*The N1 and N2 are …*). N2 retrieval was tested in Exp. 3 by providing an N2 preview as prime. Items were counterbalanced across six (Exp. 1), nine (Exp. 2), and twelve (Exp. 3) Latin square lists, presented in randomised order intermingled with 24 filler arrays. Every keystroke and their timing were recorded to detect disruptions of the production timecourse.

Results: Hesitations between adjacent keystrokes were extracted using Bayesian mixed-effects mixture models [3]. Posterior cell-means are shown in Fig. 2. Difficult-long N1 types were associated with more hesitations (compared to the easy-long condition; in all experiments) for mid N1-determiner keystroke-intervals, before N1, for mid N1 keystrokes, and before the following conjunction (all $P(\hat{\theta} < 0)$s < .01). Word length only affected mid-N1 key-intervals ($P(\hat{\theta} < 0)$ < .01). In Exp. 1, no hesitations were observed immediately before N2 ($P(\hat{\theta} < 0)$s = .35, .49). When the N2 determiner was omitted (Exp. 2), hesitations before N2 increased in the short and the difficult-to-spell N1 conditions (both $P(\hat{\theta} < 0)$s < .05) but less so for the long N1 condition ($P(\hat{\theta} < 0)$ = .19). N2 preview (Exp. 3) reduced hesitations immediately before N2 for the short and the difficult-to-spell N1 conditions (both $P(\hat{\theta} < 0)$s > .99) but less so for the long N1 condition ($P(\hat{\theta} < 0)$ = .8).

Discussion: We found that difficulty with orthographic retrieval leads to disruptions of the writing flow before and during the production of the difficult-to-spell word. Hesitations related to N1 spelling affected N2 retrieval but only when there was not enough time for parallel planning. These results demonstrate that written sentence planning unfolds in parallel to production; difficulty with low-level properties (orthography) create a planning bottleneck has a knock-on effect on higher level planning (lexical retrieval of N2).

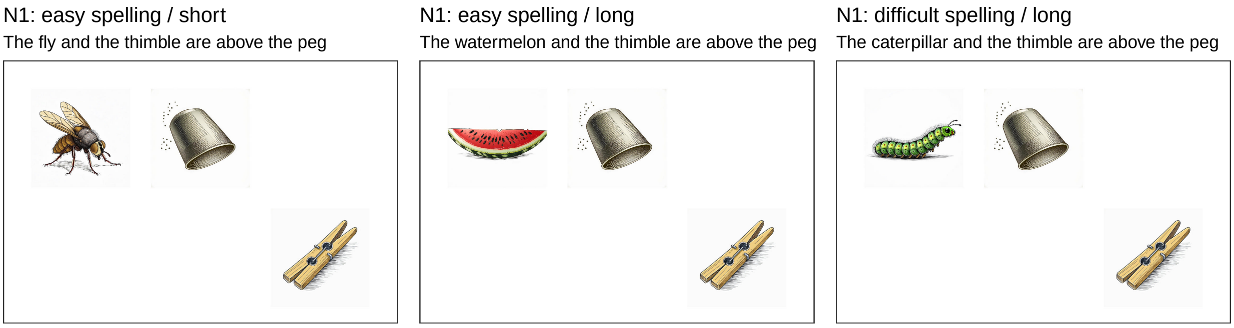


Figure 1: Example array for each N1 type. Target sentences are shown above.

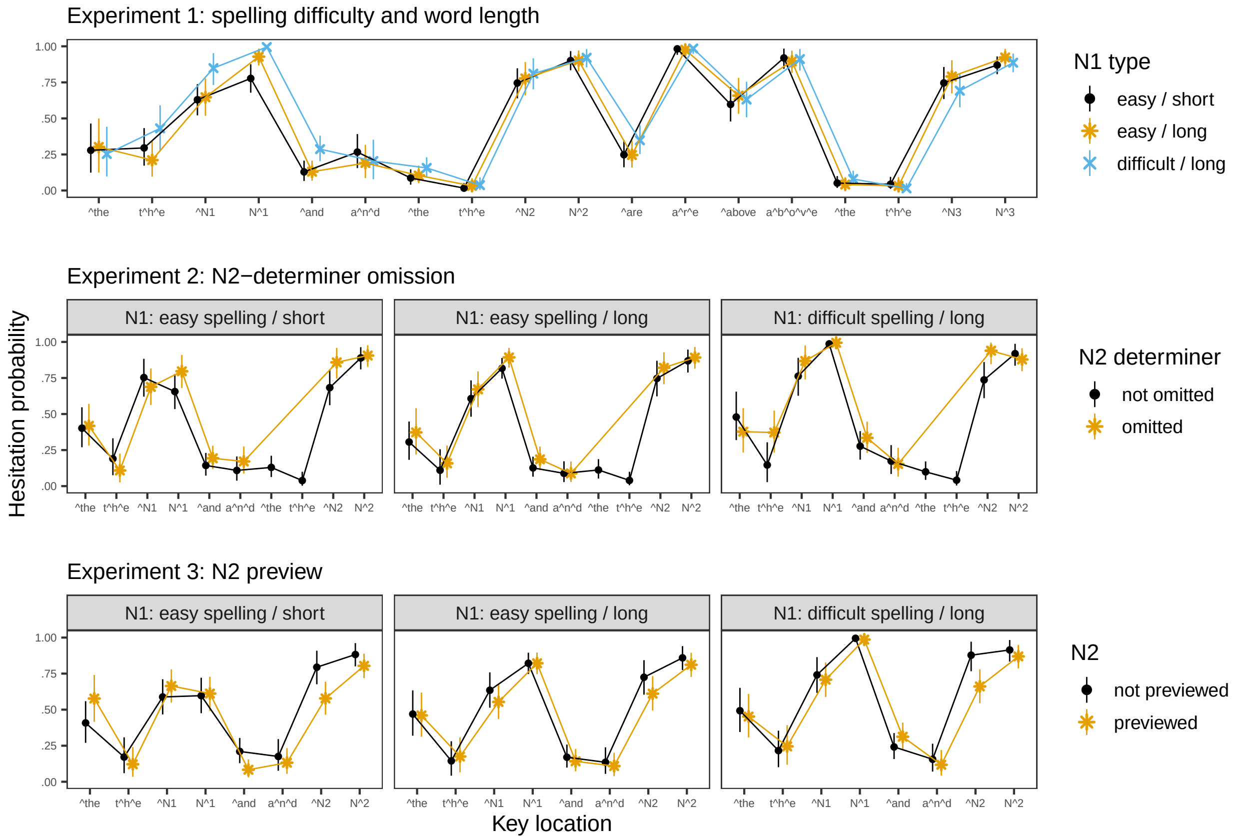


Figure 2: Hesitation probability. "^" indicate key location; "^the" indicates stimulus onset or, if not sentence initial, the transition from space to <t> keypress; "t^h^e" are mid-word transitions. Errorbars are 95% PIs.

References

[1] Caramazza, A. (1991). Orthographic structure, the graphemic buffer and the spelling process. In *Issues in Reading, Writing and Speaking. Neuropsychology and Cognition* (pp. 297–311, Vol. 3). Springer.

[2] Kandel, S. (2023). Written production: The APOMI model of word writing: Anticipatory processing of orthographic and motor information. In *Language Production* (pp. 209–232). Routledge.

[3] Roeser, J., Conijn, R., Chukharev, E., Ofstad, G. H., & Torrance, M. (2025). Typing in tandem: Language planning in multi-sentence text production is fundamentally parallel. *Journal of Experimental Psychology: General*, *154*(7), 1824–1854. https://doi.org/10.1037/xge0001759

Semantic Update as a Predictor of Reading Time: Moving Beyond Word-Level Surprisal

Siddharth Gupta, Alessandro Lopopolo, Milena Rabovsky
Department of Psychology, University of Potsdam
siddharth.gupta@uni-potsdam.de

Reading is a complex cognitive process that involves various partly interrelated stages of prediction, word recognition, lexical access, syntactic parsing, and semantic integration. Understanding these processes could provide valuable insights into the cognitive mechanisms underlying language comprehension. Recent advances in language modeling have fueled interest in using computational measures to predict reading behavior observed in eye-tracking studies. One widely used measure derived from language models, surprisal, reflects the unexpectedness of a word given its preceding context. Surprisal has been ubiquitous in testing psycholinguistic hypotheses about lexical predictability and its effects on reading measures [1]. While surprisal captures predictability at the lexical-level, to predict per-word total fixation duration (gaze duration) we here test a complementary measure derived from an artificial neural network model of sentence comprehension that works at the level of sentence meaning: the Sentence Gestalt (SG) model [2]. This measure, called semantic update (SU), calculates how the introduction of each new word in a sentence updates the SG model's internal, predictive representation of sentence meaning [3, 4]. Mathematically SU, as reflected by the arrival of a new word w_n, is defined as: $SU_n = \sum_{i=1}^{k} |a_i(w_n) - a_i(w_{n-1})|$, where $a_i(w_n)$ is the activation at the i-th unit in the gestalt layer (with k units) as the network (see Fig 1 for the visualisation of encoder-decoder architecture in SG model) encounters the n-th word in the sentence. Hence, SU is the absolute error between the layer's activation before and after encountering a word. SU has previously been successfully applied to predict the N400 event-related potential (ERP) component [3, 4], the most widely used ERP component in research on language comprehension [5].

In the current study, we examine the effect of SU as predictor of gaze duration using an eye-tracking dataset of 46 native English-speaking participants who read 12 English texts [6]. To predict gaze duration, we fitted a linear mixed-effects model that included SU, GPT-2-based surprisal, word length, word position, and word frequency, with random intercepts and slopes for SU and GPT-2-based surprisal, grouped by both participant and word. We included GPT-2-based surprisal [7] to assess whether the SU provides explanatory power beyond this information-theoretic predictor.

The results (Table 1) indicate that SU significantly predicts gaze duration ($\beta = 8.67, t = 2.07, p = 0.04$) beyond the influence of GPT-2-based surprisal, which as expected also has a significant impact ($\beta = 18.83, t = 4.15, p < 0.001$).

The analysis reveals a significant relationship between gaze duration and the size of semantic updates. These findings suggest that larger gaze duration reflect the cognitive effort of updating semantic representations, emphasizing the role of semantic-not just lexical-prediction in reading behavior.

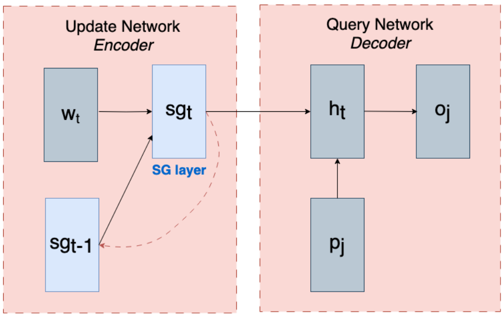


Figure 1: A visualisation of encoder-decoder architecture of Sentence Gestalt (SG) Model

	Estimate	Std. Error	CI	df	t value	Pr(>|t|)
(Intercept)	343.19	9.01	325.52 – 360.85	59.69	38.08	**<0.001**
SU	8.67	4.20	0.44 – 16.89	226.65	2.07	**0.04**
gpt2_surprisal	18.83	4.54	9.93 – 27.73	186.24	4.15	**<0.001**
word_length	59.85	4.09	51.84 – 67.86	398.77	14.65	**<0.001**
word_pos	-5.74	2.44	-10.52 – -0.96	3355.76	-2.35	**0.019**
word_freq	-8.95	4.84	-18.43 – 0.53	383.09	-1.85	0.065

Table 1: Results of the linear mixed-effect model predicting gaze duration as a function of semantic update (SU), GPT-2-based surprisal, and word-level features.

References

[1] Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

[2] McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*(3-4), SI287–SI335. https://doi.org/10.1080/01690968908406371

[3] Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

[4] Lopopolo, A., & Rabovsky, M. (2024). Tracking Lexical and Semantic Prediction Error Underlying the N400 Using Artificial Neural Network Models of Sentence Processing. *Neurobiology of Language*, *5*(1), 136–166. https://doi.org/10.1162/nol_a_00134

[5] Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

[6] Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Kwon, N., Lõo, K., … Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, *54*(6), 2843–2863. https://doi.org/10.3758/s13428-021-01772-6

[7] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners* (tech. rep. No. 1(8)). OpenAI. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Transferring islands across languages

Yimin ZHU[1], Caterina DONATI[1]
[1] Laboratoire de Linguistique Formelle, Université Paris Cité, Paris, France ;
yimin.zhu@etu.u-paris.fr

Background. Chinese and French display different strategies for wh-questions: the former has an in-situ strategy, while the latter exhibits both in-situ and ex-situ strategies. Crucially, the two languages also differ at a more abstract level, concerning locality. In Chinese, only wh-questions on adjuncts are sensitive to strong islands constraints [1], while in French all questions, on arguments and on adjuncts, are equally sensitive [2]. The issue we address is whether bilinguals exhibit crosslinguistic influence (CLI) effects on this abstract dimension, which is related to the C-domain and the syntax-pragmatics interface (as in [3]'s model) but can hardly rely on positive evidence.
Research Question. Following [4,5,6]'s methodology, we investigated with two experiments whether Chinese argument questions (Exp1) and French argument questions (Exp2) display super-additive effects corresponding to islands in two populations: Chinese native L2 French learners (FL2) and Heritage Chinese speakers (CHS) living in Paris. The former were also tested for their French lexical competence, while the latter were tested for their Chinese lexical competence. We were guided by the following research questions: **Q1.** Can CLI concern abstract constraints such as islands? **Q2.** Is there a directionality in CLI? If the directionality is determined by the nature of the grammars involved and it goes form the more restrictive to the less restrictive, we expect to observe super-additive effects in both Chinese and French experiments. **Q3.** Is directionality or strength of CLI affected by factors such as language dominance and competence? If CLI is affected by language dominance, we expect to see super-additive effects in Chinese in CHS but not in FL2. If dominance is *not* a factor, we expect to see the effect in Chinese in both populations, possibly modulated in FL2 by their competence in French.
Methods. Based on preliminary experiments with monolinguals in Mandarin Chinese [7] and French [8], we administered 2 experiments (one in French and the other in Chinese), each of them targeting two populations: CHS (N = 40) and FL2 (N = 50). Both assessed acceptability of argument extraction with a 2 × 2 factorial design, with the following two factors: Dependency Length (short vs. long), Structure (non-island vs. CNP island. There were 24 target items in both experiments, for a total of 96 items and 24 fillers. We used a Latin Square design to distribute the sentences across 4 lists and introduced 4 practice items, for a total of 52 sentences. Each experiment was preceded by a lexical proficiency test: in French (LexTALE_FR. [9]) for FL2, and in Chinese (LexTALE_CH [10]) for CHS.
Results.
<u>Exp 1. Chinese</u>. A significant interaction between Length and Structure was observed in both FL2 [see fig. 1-2], (β = -1.8836, SE = 0.2257, t = -8.345, p < 0.001***) and the CHS [see fig. 3-4] (β = -2.0292, SE = 0.3175, t = -6.392, p < 0.001***). Both populations displayed a pattern significantly different from Chinese monolingual's result as reported in [7] (β = −0.25, SE = 0.16, t = −1.62, p = 0.11). Fig. 5 shows the relation between the super-additive effect and the number of French errors among FL2: participants with lower proficiency (i.e., more errors) exhibited lower island sensitivity. As for the CHS, the more Chinese errors, the weaker the super-additive effect in Chinese (fig. 6).
<u>Exp2. French.</u> A marginally significant interaction between Length and Structure was observed in FL2 [see fig. 7-8] (β = -0.4845, SE = 0.2392, t = -2.026, p = 0.0492*). As for CHS, there was a significant interaction between Length and Structure (β = -1.1865, SE = 0.2503, t = -4.740, p < 0.001***), see fig. 9-10. Only FL2 displayed a pattern different from that observed in French monolinguals as reported in [8] (β = 1.6550, SE = 0.1747, t = 9.475, p < 0.001***). Fig. 11 shows a negative relationship between the French super-additive effect and the number of errors in French in FL2: learners with higher error rates (i.e., lower proficiency) tend to show weaker island effects in French. As for the interaction between the French super-additive effect and the number of errors in Chinese in CHS, the analysis revealed the opposite trend: the more errors in the Chinese test, the weaker the super-additive effect in French (see fig. 12).
Discussion. As for questions in Chinese, we found a CLI effect for island constraints from French to Chinese in both FL2 and CHS, suggesting that grammatical property is a crucial factor. As for questions in French, we only found a CLI effect in FL2(although this directionality of CLI is more difficult). This suggests that CLI is also influenced by dominance factor. Besides, the super additive effect was stronger depending on the French competence in FL2, suggesting that competence in the source language facilitates CLI. The super additive effect depending on the competence of Chinese in CHS suggests that competence in the recipient language does not protect from CLI.

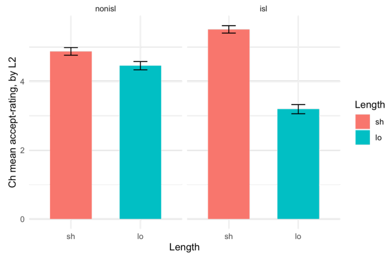


Figure 1. Mean acceptability in Chinese, by FL2

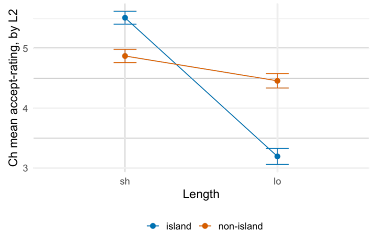


Figure 2. The interaction plots in Chinese in FL2

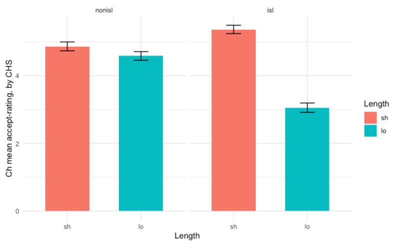


Figure 3. Mean acceptability in Chinese, by CHS

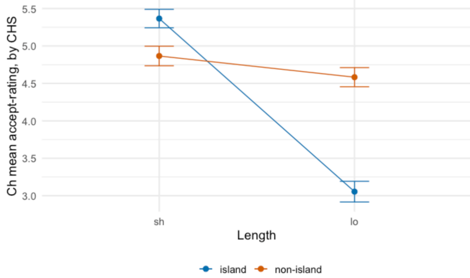


Figure 4. The interaction plots in Chinese in CHS

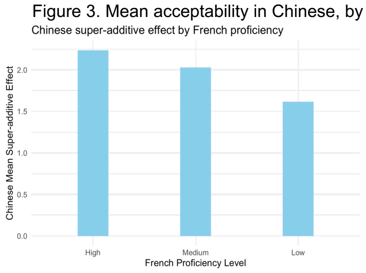


Figure 5. The relationship between French competence and Chinese super-additive effect, by FL2

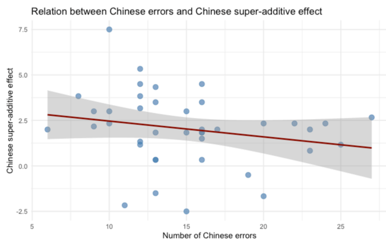


Figure 6. The relationship between Chinese test errors and Chinese super-additive effect, by CHS

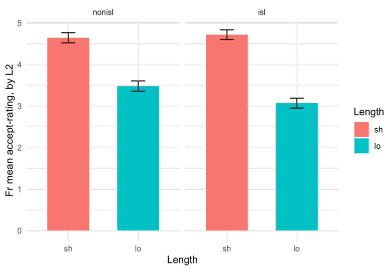


Figure 7. Mean acceptability in French, by FL2

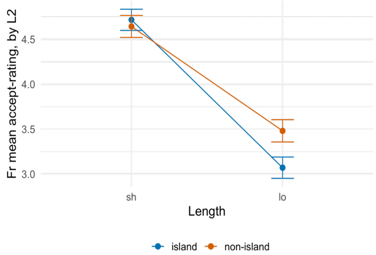


Figure 8. The interaction plots in French FL2

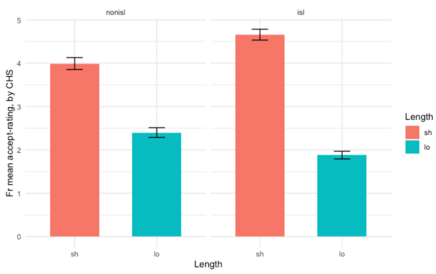


Figure 9. Mean acceptability in French, by CHS

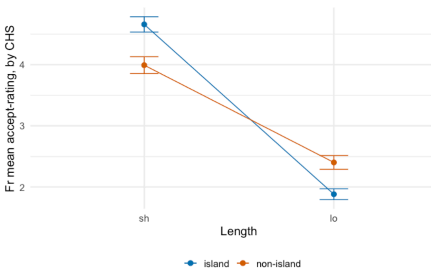


Figure 10. The interaction plots in French in CHS

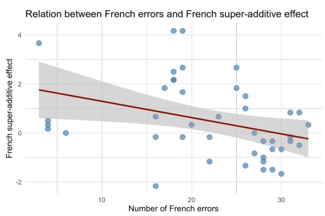


Figure 11. The relationship between French test errors and French super-additive effect, by FL2

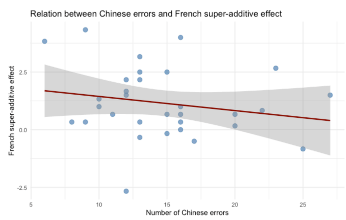


Figure 12. The relationship between Chinese test errors and French super-additive effect, by CHS

Annex: Two sample items in experiment 1 (1-4) and experiment 2 (5-8)

(1) Zhōuyǒng xiǎngzhīdào shéi juéde lièrén huì
Zhouyong wonder who think hunter will
bǔshā jīngyú. (sh + nonisl + arg)
kill whale
'Zhouyong wonders who thinks that the hunter will kill whales.'

(2) Zhōuyǒng xiǎngzhīdào zhèngfǔ juéde
Zhouyong wonder government think
lièrén huì bǔshā
hunter will kill
shénme. (lo + nonisl + arg)
what
'Zhouyong wonders what the government thinks that the hunter will kill.'

(3) Zhōuyǒng xiǎngzhīdào shéi huì chéngfá
Zhouyong wonder who will punish
bǔshā jīngyú de lièrén. (sh + isl + arg)
kill whale Rel hunter
'Zhouyong wonders who will punish the hunter that kills whales.'

(4) Zhōuyǒng xiǎngzhīdào zhèngfǔ huì
Zhouyong wonder government will
chéngfá bǔshā shénme de
punish kill what Rel
lièrén. (lo + isl + arg)
hunter
'Zhouyong wonders what is the thing x such that the government will punish the hunter who kills x.'

(5) Julie se demande qui saura que la fille
Julie REFL wonder.3SG who know.FUT that the girl
payera bientôt la dette. (sh+nonisl+arg)
pay.FUT soon the debt
'Julie wonders who will know that the girl will soon pay the debt.'

(6) Julie se demande ce que tu sauras bientôt
Julie REFL wonder.3SG what you know.FUT soon
que la fille payera. (lo+nonisl+arg)
that the girl pay.FUT
'Julie wonders what is the thing x such that you will soon know that the girl will pay x'

(7) Julie se demande qui invitera la fille
Julie REFL wonder.3SG who invite.FUT the girl
qui payera bientôt la dette. (sh+isl+arg)
who pay.FUT soon the debt
'Julie wonders who will invite the girl who will soon pay the debt.

(8) Julie se demande ce que tu inviteras bientôt
Julie REFL wonder.3SG what you invite.FUT soon
la fille qui payera. (lo+isl+arg)
the girl who pay.FUT
'Julie wonders what is the thing x such that you will soon invite the girl who pays x.'

References

[1]. Huang 1982. [2]. Rizzi 1990. [3] Hulk & Müller 2000. [4] Sprouse 2007. [5] Sprouse 2012. [6] Sprouse 2013. [7] Tian et al 2022. [8] Zhu et al 2023. [9] Brysbaert, M. 2013. [10] Chan, I. L., & Chang, C. B. 2018.

Predicting scalar diversity with crowdsourcing QUD in naturalistic discourse

Xuetong Yuan[1], Minjae Joh[2], Ming Xiang[3]
University of Chicago[1,2,3]
xuetongyuan@uchicago.edu[1], johminjae@uchicago.edu[2], mxiang@uchicago.edu/mxming@gmail.com[3]

Background: Scalar diversity refers to the well-known phenomenon that interlocutors do not draw scalar implicatures (SIs) at the same rate across different lexical scales [5]. To explain how likely a scale is to lead to SI calculation, recent literature has shown that inter-scale variation is reduced when the *explicit* Question Under Discussion (QUD) [2] is manipulated to make the relevant alternatives highly salient to comprehenders [3]. In this study, we examine the hypothesis with *implicit* QUDs using a QUD elicitation task and two inference task experiments in naturalistic discourse contexts. If QUD plays a role in calculating SIs, then the rates of SI calculation across all scales should correlate with the proportion of elicited SI-relevant QUDs.

Experiment: 37 <weak, strong> (e.g., <some, all>) lexical scales were selected from two existing datasets [3,4]. 37 short naturalistic conversations were then selected from [1], with each conversation containing a target "weak" scalar term in the last utterance of the conversation. In **Experiment 1**, participants (N=38) were presented with the 37 conversations and 8 filler trials. For each trial, participants read the entire conversation and were probed for the SI inference using a slider bar on a 0-100 scale. An example trial is shown in Figure 1. To assess the baseline SI rates *without* context, in **Experiment 2** (N = 40), participants answered the SI questions on the same set of target utterances, but without prior conversational context. Figure 2 shows the z-transformed inference ratings from the two experiments. Both experiments demonstrated clear SI diversity across lexical items. In **Experiment 3** the QUD elicitation task (N=120), we adapted the web-based data collection paradigm for crowdsourcing QUDs [6]: conversations were presented to participants with a pause before the target sentences, and a probing point was placed both before and after the target sentences. At each probing point, participants were prompted to enter a question that was evoked by the text up to that point. We tested the same set of conversations (37 items), with each participant annotated 10 items. Here we only report the results from responses elicited at the 2nd probing point (i.e., after the target sentence appears). For each item, we manually clustered elicited questions into bins that represent only the unique questions based on semantic similarities. Next, we manually identified which bin(s) of the unique questions was the most relevant for the SI calculation. The item-specific proportion of the SI-relevant questions was then calculated.

Results: To quantify the relation between the relevant QUD and SI rates, we constructed a mixed effects model predicting the SI rate from Experiment 1, with the proportion of SI-relevant-QUDs as the only predictor. The QUD predictor had a significant effect ($b=0.43$, $t.=3.23$, $p=0.002$). Figure 3 plots the correlation between the proportion of the SI-relevant QUD from context and the z-transformed SI rates. But when the baseline SI rate from Experiment 2 is added into the mixed-effects model, the effect of QUD disappeared ($b=0.17$, $t.=1.6$, $p>.1$), and only the baseline SI rate had a significant effect ($b=0.99$, $t.=5.9$, $p<.0001$).

Discussion: The preliminary results found a potential effect of contextually relevant QUDs on SI calculation in naturalistic discourse contexts, but the effect also appeared not robust when the baseline lexical effect was taken into account. The strong baseline lexical effect could potentially be attributed to the fact that the target sentences were extracted from naturalistic contexts, with some already cueing for rich contextual information. To establish a solid empirical generalization, we plan to increase both the participant sample size and the number of lexical scalar items/naturalistic conversations.

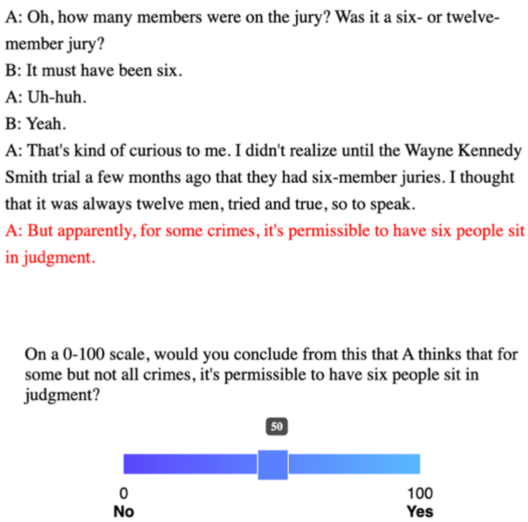


Figure 1. Example trial for Experiment 1.

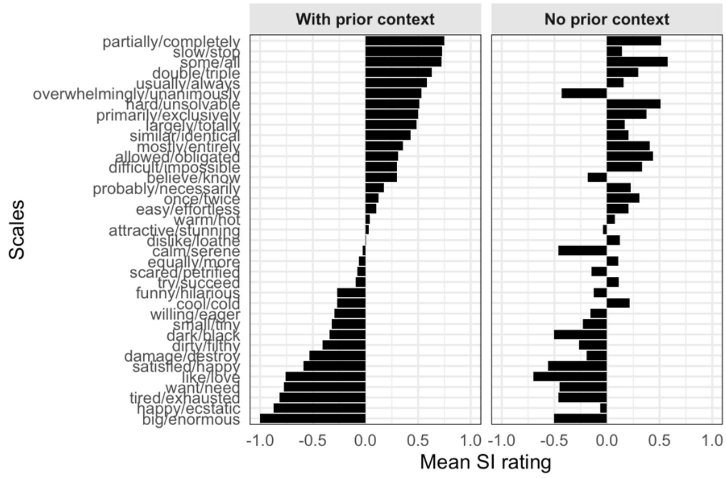


Figure 2. SI rates for 37 different scales, with (Experiment 1) or without contexts (Experiment 2).

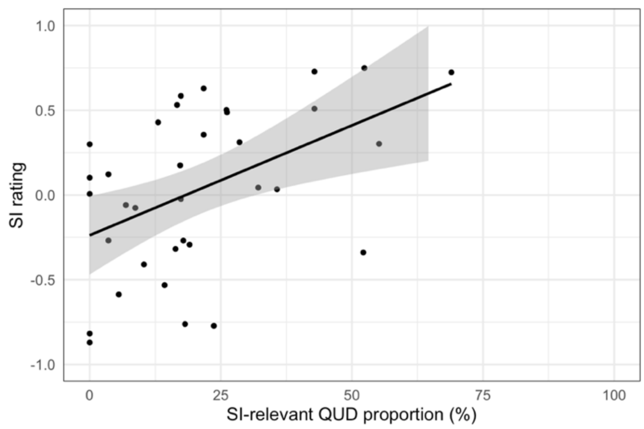


Figure 3. Correlation between SI rate (Experiment 1) and elicited SI-relevant QUD.

References

[1] Godfrey, J., & Holliman, E. (1993). Switchboard-1 Release 2 LDC97S62. Linguistic Data Consortium, 34.

[2] Roberts, Craige (1996/2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. Semantics and Pragmatics, 5(6), 1–69.

[3] Ronai, E., & Xiang, M. (2024). What could have been said? Alternatives and variability in pragmatic inferences. Journal of Memory and Language, 136, 104507.

[4] Shivade, C., De Marneffe, M. C., Fosler-Lussier, E., & Lai, A. M. (2015). Corpus-based discovery of semantic intensity scales. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 483-493.

[5] van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar Diversity. Journal of Semantics, 33(1), 137–175.

[6] Westera, M., Mayol, L., & Rohde, H. (2020). TED-Q: TED talks and the questions they evoke. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 1118-1127.

Action imagination and verb semantics encoded in morphology influence neural responses beyond somatotopic mapping: An fMRI study

Ting-Wu Lee[1], Shiao-Hui Chan[1]
[1.] Department of English, National Taiwan Normal University
olim4101@gmail.com

Background: The embodied cognition framework hypothesizes that action concepts are grounded in the sensory-motor regions corresponding to somatotopic representations [1,2,3], regardless of input modality. However, it remains unclear whether the hypothesis of grounded concepts is cross-linguistically valid and whether action representation is modulated by other task-related factors. We conducted an event-related fMRI experiment to explore action concept representation in Mandarin Chinese with the following questions: (1) Is the embodied cognition framework cross-linguistically valid? (2) Does brain activity elicited by action verb stimuli reflect somatotopic representations? (3) How does the brain respond to verbs of the same action with sematic differences embedded in morphology?

Method: The verbs were carefully selected to associate with three body parts—hands (*hit, wipe*), legs (*kick, step*), and mouths (*speak, eat*)—and were formed in two morphologically distinct verb compounds: attributives (ATTRs) and resultatives (RESs). In ATTR compounds, the modifier precedes the head verb and specifies the manner or intensity of the action (e.g. *qiao*-***da*** [Lit. knock-**hit**], *meng*-***da*** [Lit. intense-**hit**]). In RES compounds, the modifier follows the head verb and indicates the end state or direction of the action (e.g. ***da***-*shang* [Lit. **hit**-hurt], ***da***-*dao* [Lit. **hit**-fall]). The baseline (BASE) was designed with two numbers in Chinese characters (e.g. 二五 [two five]) which served as semantic controls since that numbers do not convey concrete concepts but still carry lexical meaning. During the experiment, verbs were randomly presented on the screen for 2 seconds. Participants were instructed to silently read the verbs and imagine perfoming the described actions. A general linear model was constructed with 5*2 parameters (HAND/LEG/MOUTH x ATTR/RES) along with BASE and FILL (fillers). One sample t-test was conducted for multiple t-contrasts of interest.

Results: The anterior IFG and the posterior MTG were consistently found active across all conditions relative to baseline. In contrast, the motor cortex (PreCG) was not strongly engaged by most of the body-part conditions. While the BOLD signals were detected, only the R-PreCG in the LEG-BASE condition survived FDR-correction. All body-part stimuli significantly activated the supplementary motor area. For the ATTR-RES contrast, a cluster peak extending from the PostCG to the PreCG was observed.

Discussion: Despite weak signals, the PreCG activity observed across conditions relative to baseline partially replicated prior studies supporting the embodied cognition framework, suggesting that action concepts embodied in the corresponding regions are cross-linguistically valid. The reduced motor cortex activity may be attributed to the task and modality differences [4,5,6], as the linguistic stimuli likely relied more on language-related areas such as the anterior IFG and the posterior MTG for lexical-semantic processing. [7, 8, 9]. The SMA activation, in contrast, may reflect implicit motor simulation after participants silently read the words [10]. Greater neural activity in the postcentral and precentral gyrus in the ATTR-RES contrast not only supports the grounded action concept hypothesis, but also suggests that the postcentral gyrus, together with the precentral gyrus, contributes to processing of action details (motion, specificity and intensity), corresponding to richer verb semantics and more detailed mental imagery.

References

[1] Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

[2] Kemmerer, D., Gonzalez Castillo, J., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, *107*(1), 16-43.

[3] Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

[4] Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*, *43*(3), 634-644.

[5] Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes but within the context of a flexible, multilevel architecture for conceptual knowledge. *Pyschonomic Bulletin & Review*, *22*, 1068-1075.

[6] Wu, H., Mai, X., Tang, H., Ge, Y., Luo, Y.-J., & Liu, C. (2013). Dissociable somatotopic representations of Chinese action verbs in the motor and premotor cortex. *Scientific Reports*, *3*(1), 2049.

[7] Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, 24(2), 427-432.

[8] Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262-268.

[9] Wu, D. H., Waller, S., & Chatterjee, A. (2007). The functional neuroanatomy of thematic role and locative relation knowledge. *Journal of Cognitive Neuroscience*, 19(9), 1542-1555.

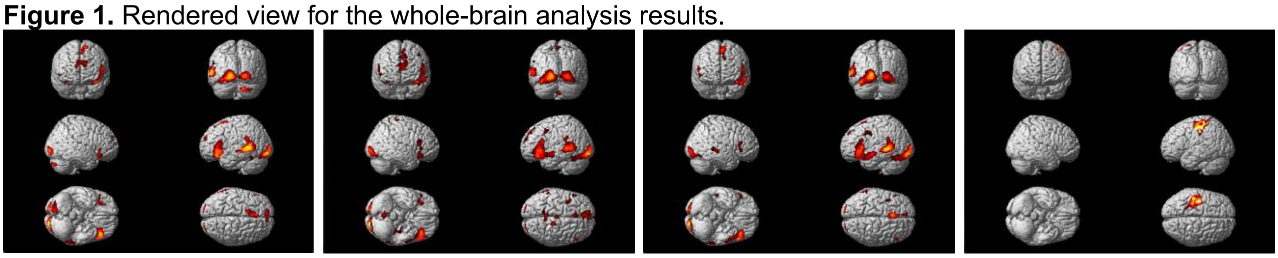
[10] Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, *22*(10), 2387-2400.

Table 1. Result of Whole-brain analysis (k >100)

contrast	region/cluster	side	cluster size	q	x	y	z
HAND-BASE	IFG	L	1117	0.0001*	-36	30	-20
		R	202	0.099	32	30	-16
	MTG	L	1548	0*	-62	-42	-2
	SMA	L	376	0.024*	-14	10	66
	MedSFG	L	314	0.033*	-4	58	28
LEG-BASE	IFG	L	2035	0*	-36	32	-16
		R	196	0.08	56	26	22
	MTG, STG	L	1410	0*	-60	-40	2
	ParaCL, PreCG	L	512	0.003*	-6	-46	62
	MedSFG	L	632	0.001*	-6	46	44
	SMA	L	123	0.184	-4	24	54
		R	106	0.212	2	-20	56
MOUTH-BASE	IFG, aMTG	L	1658	0*	-52	20	4
	pMTG	L	1340	0*	-58	-40	0
		R	254	0.045*	50	-30	-4
	ATL	L	1658	0*	-58	2	-14
	SMA	L	555	0.003*	-4	18	60
	PCG	L	205	0.067	-40	-4	36
ATTR-RES	PostCG, PreCG	L	834	0.008*	-48	-28	56
RES-ATTR				na			

Note. The visual and subcortical areas are intentionally removed from the table due to the limit of space. ATTR: attributive compound; RES: resultative compound; BASE: baseline; (a/p)MTG: (anterior/posterior) middle temporal gyrus; IFG: inferior frontal gyrus; SMA: supplementary motor area; ATL: anterior temporal lobe. medSFG: medial superior frontal gyrus; PreCG: precentral gyrus; PostCG: postcentral gyrus; ParaCL : paracentral lobule; MedSFG: medial superior frontal gyrus; STG: superior temporal gyrus; SMG: supramarginal gyrus; q= FDR-corrected p value; * = $pFDR_{corr.}$ < 0.05 (significant).

Figure 1. Rendered view for the whole-brain analysis results.



Note. From left to right: HAND-BASE, LEG-BASE, MOUTH-BASE, ATTR-RES (p = .005 uncorr., k = 100)

Robust mutual exclusivity in multiparty conversations: Contextual adaptation without speaker-specific effects

Background: Mutual Exclusivity Inference (MEI) drives listeners to map novel terms onto unnamed referents by excluding familiar ones [1,2]. While other pragmatic inferences, such as scalar and contrastive interpretations, adapt to speaker-specific reliability in dyadic settings [3,4], it remains unclear whether MEI exhibits similar flexibility in multiparty conversations. This study examines whether listeners modulate MEI based on the referential consistency of individual speakers within a complex interactional context. **Method:** We conducted two visual-world eye-tracking experiments. Experiment 1 (N = 32, native English speakers) employed a between-subjects design: participants followed pre-recorded instructions from two "speakers," one referentially consistent (always reusing labels) and one inconsistent (alternating labels), versus a control group with two consistent speakers. On each trial, participants chose between a previously named object and an unnamed object upon hearing a novel description. Experiment 2 (N = 44, native Spanish speakers) used a fully within-subjects, face-to-face setup. After a baseline phase with two consistent speakers, participants entered an experimental phase featuring one consistent and one inconsistent speaker, selecting among four images (named, unnamed, seen distractor, unseen distractor). Behavioral choices were analyzed via mixed-effects logistic regression; eye movements were examined using cluster-based permutation analyses to identify the timing of MEI and any adaptation effects. **Results:** In both experiments, participants robustly preferred the unnamed referent (Experiment 1: 71.5%; Experiment 2: 63.5%), indicating MEI emergence. Mixed-effects models revealed no speaker-specific reduction in MEI in either experiment ($ps > .50$). However, Experiment 1 showed a significant context-general decrease in MEI when an inconsistent speaker was present ($p < .001$). Eye-tracking data confirmed that this context-level modulation emerged around 650 ms after description onset, whereas no reliable clusters indexed speaker-specific adaptation at any time window. **Discussion:** These findings indicate that MEI functions as a powerful default heuristic, resistant to rapid, speaker-specific calibration in multiparty settings but sensitive to overall conversational reliability. The early onset of MEI aligns with dual-process models of pragmatic comprehension—an initial automatic inference followed by later contextual integration—while the absence of speaker-specific effects suggests cognitive limits on maintaining granular speaker models under high load. Our results advance understanding of pragmatic flexibility, highlighting the primacy of global referential consistency for effective communication in multiparty interactions.

This page serves for Tables, Figures, and References.

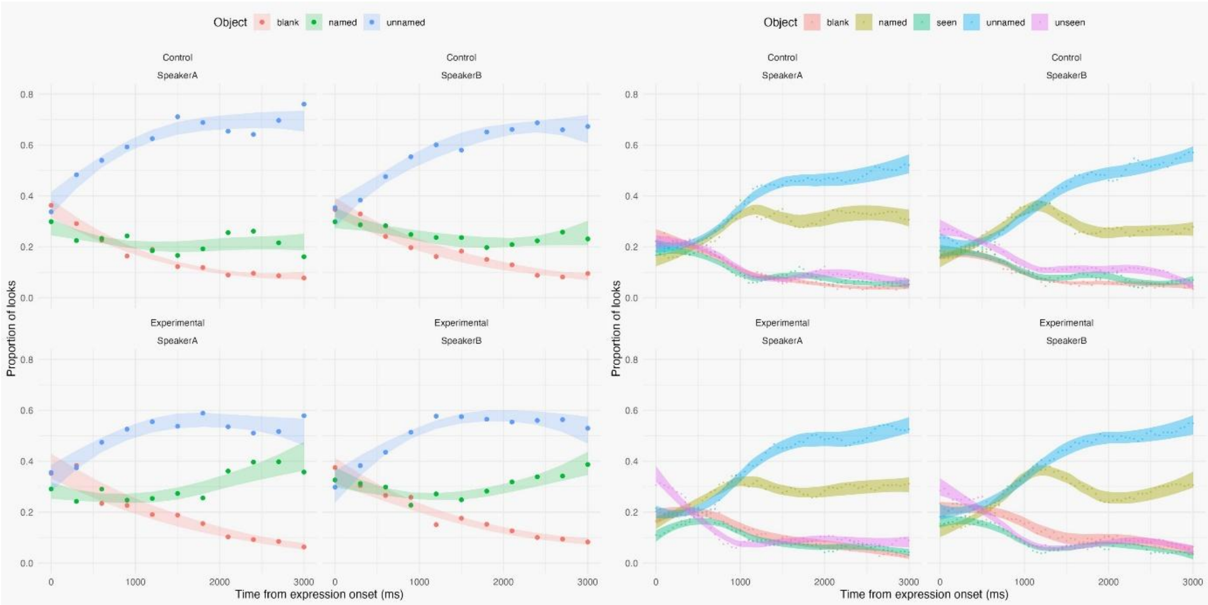


Figure 1. Mean proportion of looks to each object across conditions from expression onset for Experiment 1 (on the left) and Experiment 2 (on the right). The shaded error bands represent the standard error of the mean, computed by bootstrapping across participants and applying a LOESS smoothing function.

References

[1] Markman, E. M., & Wachtel, G. F. (1988). Children’s use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

[2] Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin & Review*, 19(2), 317–324. https://doi.org/10.3758/s13423-011-0194-6

[3] Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). MIT Press. https://doi.org/10.7551/mitpress/9780262015127.003.0010

[4] Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, 43(8), e12769. https://doi.org/10.1111/cogs.12769

This page may be used only for additional information about the generally less-known language you are targeting in the abstract.

Surprisal is Influenced by Syntax and Semantics, but not Equally across Language Models

Alessandro Lopopolo[1], Milena Rabovsky[1]
[1] Department of Psychology, University of Potsdam
lopopolo@uni-potsdam.de

Lexical surprisal, widely used to explain neural responses like the N400 and BOLD signal [1, 2], is often viewed as a measure of lexical prediction. However, it remains unclear how much it reflects syntactic and semantic structure beyond surface-level cues [3]. This study quantifies how surprisal is shaped by features such as semantic distance, and constituency and dependency structure (based on Figure 1A). We estimate word-wise surprisal from nine language models (GPT-2 [4], Falcon [5], BERT [6], RoBERTa [7], BART [8], T5 [9], and N-gram models) widely used in pyscholinguistics studies. Analyses were conducted on 60,000 English sentences.
Analyses: We fitted independent linear regression models predicting each of word-wise surprisal estimates from each of the nine LMs, using either (i) word position and lexical frequency alone (which yielded an average R^2 0.446 $\pm$ 0.082 across models) or (ii) structural (constituency, dependency) and semantic predictors alongside the baseline. Figure 1B shows the gains in explained variance (ΔR^2) from adding structural and semantic predictors. Constituency predictors included syntactic depth and the number of closed phrases per word; dependency predictors captured the number of dependencies per word and the distance to each word's dependency head. Dependency features yielded overall larger ΔR^2, constituency features produced moderate gains, and semantic distance – measured via FastText-based contextual dissimilarity –showed variable effects. GPT-2 and Falcon integrated both dependency and semantic information most strongly. BART showed weaker effects, particularly for constituency. BERT and RoBERTa benefited from syntax, though RoBERTa was less sensitive to semantics. N-gram models relied inconsistently on structure, with modest constituency and semantic effects but relatively strong sensitivity to dependency features, suggesting a shallow structural encoding. Additionally, we conducted SHAP (SHapley Additive exPlanations) analyses to assess each predictor's contribution to the nine surprisal estimates. SHAP quantifies feature impact by computing marginal contributions across feature subsets. Results (Figure 1C) show that semantic distance was the most influential predictor for transformer LMs, particularly GPT-2, Falcon, and BART. The number of left dependencies was key for T5 and N-gram LMs, while constituent depth notably impacted RoBERTa. Overall, transformers relied more heavily on semantic predictors, whereas N-gram models showed a more balanced or weaker feature profile. Hierarchical clustering confirmed distinct grouping patterns between transformer and N-gram architectures based on predictor sensitivities.
Conclusions: Overall, despite the conspicuous explanatory effects of the baseline predictors, structural predictors enhance surprisal estimates, confirming that surprisal reflects not just surface-level properties. However, the nature and extent of this sensitivity is not universal but varies sharply across models, reflecting their architectural and training distinctions. These findings highlight the importance of evaluating surprisal's informational content in model-specific terms when used to interpret cognitive or neural data, and they encourage an empirically driven definition of surprisal—one that reflects how different models relate to both sequential and structural information. In ongoing work, we test whether model-specific sensitivities to structure and semantics shape how surprisal maps onto neural correlates of predictive processing.

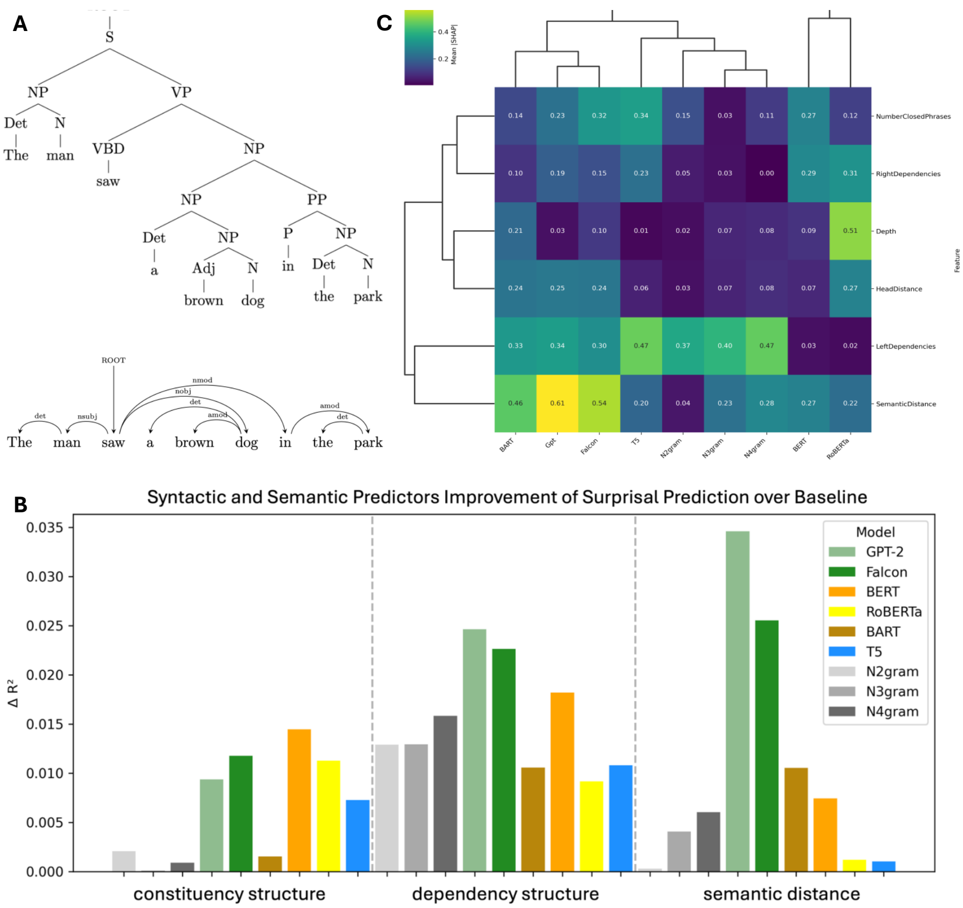


Figure 1: (A) Constituency and Dependency structure representations of a sentence; (B) ΔR^2 from linear models predicting surprisal across 9 language models. Each bar shows the increase in R^2 over a baseline using position and lexical frequency, after adding constituency, dependency, or semantic predictors; (C) SHAP scores and clustering per predictors for each of the 9 surprisal estimates.

References

[1] Willems, R. M., Frank, S., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction During Natural Language Comprehension. *Cerebral cortex*, *26 6*, 2506–2516.

[2] Michaelov, J. A., Bardolph, M. D., Petten, C. K. V., Bergen, B. K., & Coulson, S. (2024). Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. *Neurobiology of Language*, *5*(1), 107–135.

[3] Slaats, S., & Martin, A. E. (2025). What's Surprising About Surprisal. *Computational Brain & Behavior*.

[4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners [Technical Report].

[5] Penedo, G., & et al. (2023). The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and the Importance of Benchmarking.

[6] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*.

[7] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*.

[8] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv*.

[9] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, *21*(140), 1–67.

Nominalised adjectives in Basque: experimental evidence from a self-paced reading experiment

Jon Lapresa Serrano[1], Marianne Hundt[1], Fernando Zúñiga[2]
[1] University of Zurich; [2] University of Bern
jon.lapresaserrano@uzh.ch

Adjectives are traditionally defined as a part of speech distinct from nouns and verbs. However, category boundaries may be fuzzy, allowing for both sub-sective and inter-sective gradience [1, 8]. Grounded in Croft's *Radical Construction Grammar* [2, 3, 4, 5] and Dixon's [6, 7] notion of property concepts, our study argues that the morphosyntactic features of adjectives and nouns may be more context-dependent than previously assumed. This project examines an underdescribed construction in a language isolate (Basque) that nominalises adjectives through genitive expressions: NP-GEN + NomA, e.g., *akatsaren handia* 'the magnitude of the mistake'. To analyse the way speakers process sentences that include this construction, and how they compare to those that use the derivational nominaliser *-tasun*, we conducted a word-by-word self-paced reading experiment with 31 Basque native speakers. All 28 experimental sentences were implemented using a 2x2 within-subjects factorial design. The two independent variables were (i) the nominalisation type, i.e., the suffix *-tasun* and the bare nominalised adjective (NomA), and (ii) the genitive type, i.e., suffixes *-(r)en* and *-ko*. In order to avoid training effects, our design was based on a Latin Square rotation, meaning that each participant was only presented with one version of each sentence per adjective. Dixon's seven main semantic types (VALUE, DIMENSION, PHYSICAL PROPERTY, SPEED, HUMAN PROPENSITY, AGE, and COLOUR) were also added as fixed effects for a consequent statistical model, but they were still taken into account for the selection and distribution of the adjectives. We predicted that two of the conditions would show faster processing times: (i) the *-(r)en* genitive type and (ii) the *-tasun* nominaliser. Unexpectedly, neither the genitive nor the semantic type showed any significant effects, while the *-tasun* nominaliser was found to have an effect by actually slowing down the participants' processing of the construction. These results, as well as the project overall, aim to contribute to the cross-linguistic understanding of porous boundaries of the categorisation of parts of speech, and the interface between semantics, morphology, and syntax. Moreover, we aim to provide new empirical evidence for the interaction between constructional variation and cognitive processing in Basque.

References

[1] Aarts, B. (2004). Modelling linguistic gradience*. *Studies in Language 28*(1), 1–49.

[2] Croft, W. (2000). Parts of speech as language universals and as language-particular categories. In P. Vogel & B. Comrie (Eds.), *Approaches to the Typology of Word Classes* (pp. 65-102). Berlin, New York: De Gruyter Mouton.

[3] Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

[4] Croft, W., & Crude, D. Alan (2004). *Cognitive Linguistics*. Cambridge: Cambridge Univeristy Press.

[5] Croft W. (2022). *Morphology: Constructions of the world's languages.* Cambridge: Cambridge University Press.

[6] Dixon, R. M., & Aikhenvald, A. Y. (Eds.) (2004). *Adjective classes: A cross-linguistic typology*. Oxford: Oxford University Press.

[7] Dixon, R. M. (1982). *Where have all the adjectives gone?: and other essays in semantics and syntax*. Berlin: Mouton De Gruyter.

[8] Keizer, M. E. (2023). Word classes and gradience. In E. van Lier (Ed.), *The Oxford Handbook of Word Classes* (pp. 178-195). Oxford University Press.

Table. Latin Square rotation design for stimuli sentences presented in four different lists (one per participant).

Adjective	**List 1**	**List 2**	**List 3**	**List 4**
A1	*-(r)en* + *-tasun*	*-ko* + NomA	*-ko* + *-tasun*	*-(r)en* + NomA
A2	*-ko* + NomA	*-ko* + *-tasun*	*-(r)en* + NomA	*-(r)en* + *-tasun*
A3…	*-ko* + *-tasun*	*-(r)en* + NomA	*-(r)en* + *-tasun*	*-ko* + NomA
A28	*-(r)en* + NomA	*-(r)en* + *-tasun*	*-ko* + NomA	*-ko* + *-tasun*

Table. Results from pairwise comparisons (Estimated Marginal Means) of genitive type (*-(r)en* vs *-ko*) and nominalisation type (*-tasun* vs NomA) of log-transformed RTs at the critical regions.

Contrast	**Estimate**	**SE**	**t.ratio**	**p.value (adjust.)**	**Sign.**
KO NomA - REN NomA	-0.018	0.022	-0.805	0.8517	
KO NomA - KO TASUN	**-0.113**	**0.022**	**-5.045**	**<0.0001**	***
KO NomA - REN TASUN	**-0.096**	**0.022**	**-4.299**	**0.0002**	***
REN NomA - KO TASUN	**-0.095**	**0.022**	**-4.234**	**0.0003**	***
REN NomA - REN TASUN	**-0.078**	**0.022**	**-3.492**	**0.0038**	***
KO TASUN - REN TASUN	0.016	0.022	0.735	0.8829	

Basque is a non-Indo-European language isolate spoken in the Basque Country, known for features such as ergativity, agglutinative morphology, and the absence of grammatical gender, among others. Adjectives in Basque follow the head noun, except for specific cases like demonyms, and will often take the morphological markers of the noun phrase when they are in final position.

Adjectives are morphologically productive and may be nominalised with, for example, the suffix *-tasun*, producing abstract nouns (e.g. *larri* 'urgent' → *larritasun* 'urgency'). Regarding the genitive case in Basque, it is marked by either the suffix *-(r)en* (general genitive) *emakumea<u>ren</u> etxea* 'the house of the woman', or by the suffix *-ko* (locative or relational genitive), e.g., *plaza<u>ko</u> etxea* 'the house of the square'. The latter, *-ko*, is restricted to inanimate nouns only.

This paper examines an underdescribed nominalisation construction in which bare adjectives function as nouns without any derivational morphology (nominalised adjectives or NomAs). In this construction, the NomAs appear following genitive-marked NPs: NP-GEN + NomA. For example, *akatsaren handia* 'the magnitude of the mistake', lit. 'the "big" of the mistake'. These NomAs are syntactically and semantically similar to their *-tasun* counterparts, *akatsaren handitasuna* 'the magnitude of the mistake', which raises questions of categorial boundaries and usage preferences. Whether these NomAs and *-tasun* nouns are fully interchangeable or carry subtle differences remains an open question.

As an example, compare the following sentences, taken from the experimental sentences used in the SPR task experiment described above:

(1)	*Ez*	*nuen*	***ni-re***	***akats-a-ren***
	not	have.1sgERG.PST[3sgABS]	1sg-GEN	mistake-DET-GEN
	handi-a	*ikusi.*		
	big-DET	see		
(2)	*Ez*	*nuen*	***ni-re***	***akats-a-ren***
	not	have.1sgERG.PST[3sgABS]	1sg-GEN	mistake-DET-GEN
	handi-tasun-a	*ikusi.*		
	big-NMLZ-DET	see		

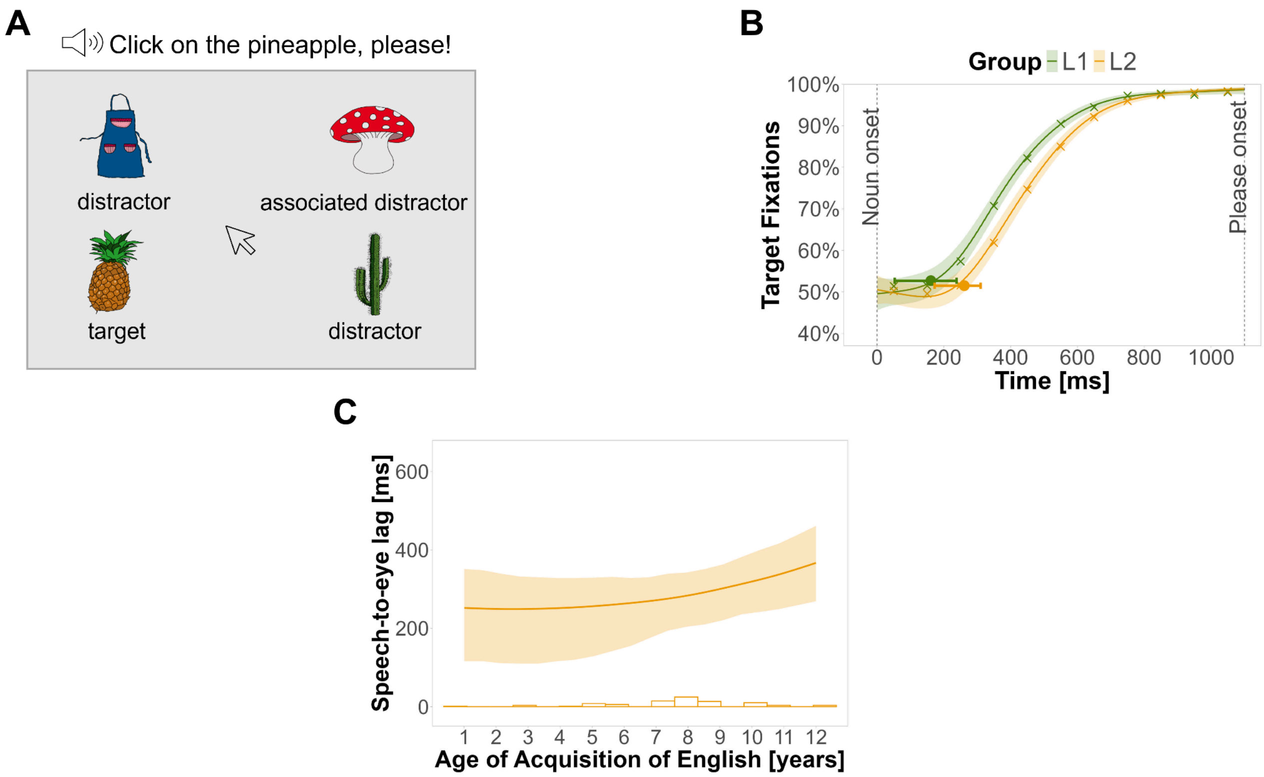
'I did not see the magnitude of my mistake.'

Eyes on delay: Revisiting the timecourse of spoken word recognition in L1 and L2 speakers

Jéssica Gomes[1], João Veríssimo[1], Dan Parker[2], Sol Lago[3]
[1]University of Lisbon; [2]Ohio State University; [2]Goethe University Frankfurt
jgomes4@edu.ulisboa.pt

Background: Visual world studies on language comprehension typically assume that, in the absence of coarticulation, it takes approximately 200 ms for speech input to affect eye movements. This lag has been crucial for estimating real-time psycholinguistic effects such as prediction, and it has been used across different types of populations, including children and adult native speakers [1], second language learners [2], older adults [4], and individuals with aphasia [5]. However, the original estimate of a 200 ms lag was experimentally validated only with young native speakers of English and was obtained through a saccade-based method with some methodological shortcomings, including the removal of a large numbers of trials (more than 50% in [6]) and the use of statistical models that did not account for between-participant variability. We revisited the speech-to-eye lag through a conceptual replication of [6], testing native English speakers (L1) as well as a more diverse group: second language (L2) learners of English. We measured the speech-to-eye lag using both the saccade-based method [6] and a novel method based on generalized additive models [7], which allowed estimating the speech-to-eye lag for each participant and the role of individual predictors, namely age of L2 acquisition and vocabulary size.
Method: Seventy English native speakers (mean score on LexTALE = 91%, SD = 9.2) and seventy-nine German learners of English (mean age of English acquisition = 7.7 years, SD = 2.1; mean score on LexTALE = 84%, SD = 11.1) performed a visual world eye-tracking experiment. In each trial, participants heard instructions like "Click on the pineapple, please!" while looking at displays with four objects: the target object, an associated distractor and two other objects (Figure 1.A). The speech-to-eye lag was defined as the earliest temporal point at which the modelled fixation curve was significantly greater than at onset of the time-window [7]. Participants also completed a demographic questionnaire and the English proficiency test LexTALE. We used these measures to assess the effects of age of acquisition and vocabulary size.
Results and discussion: At the group level, the speech-to-eye lag for L1 speakers was estimated at 164 ms 95% CI [45, 232] ms, and for L2 speakers at 260 95% CI [171, 317] ms (Figure 1.B). The mean difference between the groups was 100 ms 95% CI [-8, 234] ms. Although the difference was not statistically significant, as the confidence interval included zero (data collection is still ongoing), there was a clear tendency for L2 speakers to exhibit a longer speech-to-eye lag. This pattern supports previous evidence that lexical retrieval—a cognitive process contributing to the lag—is slower in second- vs. first-language speakers. For both groups, the saccade-based method of [6] yielded a similar estimate of the group-level lag (L1: 170 ms; L2: 260 ms), providing an internal validation of the novel method. There was evidence for an effect of age of acquisition in the L2 group: Participants who began acquiring English earlier showed shorter lags (Figure 1.C). There was no evidence for an effect of vocabulary size as measured by the LexTALE, likely due to the high scores and low variability observed in both groups. Overall, these results have useful implications for visual world research, suggesting that the originally established 200 ms lag should not be treated as an universal estimate, but instead should be experimentally evaluated in the population of interest, taking into account potential individual differences [8]**.**

Figure 1. (A) Sample visual display in the experiment. Following [6], one of the three distractors was named “associated distractor” and used to compare with the target object in the saccade-detection method. The determiner and noun were cross-spliced to eliminate co-articulatory cues at the determiner. (B) Fitted GAMM curve overlaid onto empirical means of fixation proportions across participants. Points with error bars indicate the median speech-to-eye lag with the 95% highest density interval obtained from the GAMM-based onset detection method. (C) Age of acquisition effect for the L2 group. The ribbons show 95% confidence intervals. The number of participants per age of acquisition is shown by the histogram on the x-axis.



References

[1] Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2013). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience, 29*(7), 877–892.

[2] Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition, 21*(2), 251-264.

[3] Noiret, N., Vigneron, B., Diogo, M., Vandel, P., & Laurent, É. (2017). Saccadic eye movements: what do they tell us about aging cognition?. *Aging, Neuropsychology, and Cognition*, *24*(5), 575-599.

[4] Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging, 27*(1), 80.

[5] Mack, J. E., Ji, W., & Thompson, C. K. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics, 26*(6), 619-636.

[6] Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145-163.

[7] Veríssimo, J. & Lago, S. (2025). A novel method for detecting the onset of experimental effects in visual world eye-tracking [Preprint]. https://doi.org/10.17605/OSF.IO/9ETHQ.

[8] Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154-169.

Introducing the Estonian Auditory Lexical Decision database

Kaidi Lõo [1], Anton Malmi [1], Benjamin V. Tucker [2]

[1] University of Tartu, [2] Northern Arizona University

kaidi.loo@ut.ee

Large-scale lexical decision databases are important for understanding word recognition, yet the few current auditory lexical decision datasets (Dutch: Ernestus & Cutler, 2015; French: Ferrand et al., 2018, and English: Tucker et al., 2019) focus on rather morphological simple languages. To address this gap, we introduce the Estonian Auditory Lexical Decision Database (EALD), a dataset with responses from 412 native Estonian speakers (aged 18-72 years; 321 females, 85 males, 10 non-binary) to 8800 Estonian words and 8800 pseudowords, collected online using the OpenSesame software and the JATOS server Mindprobe (Lange et al., 2015).

The stimuli entail monomorphemic, inflected, derived, and compound Estonian words. Compared to the languages mentioned above, Estonian has a more complex agglutinative and fusional case-marking system, resulting in large morphological paradigms that significantly influence word recognition processes (Lõo et al., 2018). We also provide auditory and lexical information, enabling researchers also to analyze the effects of syllable and morphological structure, phonotactics, word frequency, and both orthographic and phonological neighborhood density on auditory word recognition.

Preliminary regression analyses indicate that reaction times decrease as written Estonian word frequency and inflectional paradigm size increase, while reaction times increase with higher phonological neighborhood density. The data also reveal a nonlinear relationship between age and reaction time, with participants between 30 and 50 being the fastest. Furthermore, performance is affected by the time of day, with slower response times observed during nighttime.

By making this dataset freely available, we aim to facilitate research on morphological processing, spoken word recognition, and computational modeling in Estonian and in other morphologically rich languages. This resource complements existing auditory lexical decision datasets and contributes to a more comprehensive understanding of morphological effects in spoken language processing.

References

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *Quarterly Journal of Experimental Psychology*, *68*(8), 1469-1488.

Ferrand, L., Méot, ... & Grainger, J. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, *50*, 1285-1307.

Lange, K., Kühn, S., & Filevich, E. (2015). Just another tool for online studies (JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS one*, *10*(6), e0130834.

Lõo, K., Järvikivi, J., and Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size
facilitate Estonian case-inflected noun processing. Cognition, 175:20–25.

Tucker, B. V., Brenner, D…, & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior Research Methods*, *51*, 1187-1204.

Aligning to what I don't say: structural alignment and pragmatic inferencing

Alice Rees
Department of Psychology, Lancaster University
A.rees3@lancaster.ac.uk

Background: A key element of successful communication involves correctly retrieving a speaker's intended meaning. Often however, speakers do not explicitly say what they intend. Consider this exchange:

A: "Have you met Lucy's new partner? They're rich and intelligent." B: "They're rich"

Here, in addition to communicating that Lucy's partner is rich B may also communicate that Lucy's partner is *not intelligent*. This is an example of a pragmatic inference [1]. Much of what we know about pragmatic inferencing is rooted in the perspective of a comprehender [2,3] but order to infer, there must be a speaker who has *implied* something. However, successful communication also involves becoming linguistically aligned with your interlocutor. During conversation interlocutors converging on similar syntactic, phonological, and lexical choices through the activation of linguistic representations shared between the comprehension or production systems [4, 5]. This raises an interesting question for pragmatic inferencing, which relies on listeners recovering unsaid meaning [1]: can interlocutors become aligned in their use of pragmatic inferencing? Some work suggests yes [6].

In a referential communication game [6] participants took turns with a confederate to describe cards. When the confederate used implicit constructions requiring an inference for the correct target to be identified, participants reciprocated in their use of implicit constructions. This was found within the global context of the task and at a local, trial-by-trial level. These findings were argued in part to be evidence of "pragmatic priming" and specific pragmatic representations akin to syntactic representations.

The current study aims to firstly replicate these findings [6] in an online study where the confederate is not physically co-present, and secondly to determine the nature of alignment.

Method: Participants (N=21 English speakers) took part in a virtual (Gorilla.sc) scripted reference game where they alternated between comprehension (prime) and production (target) trials. On each trial participants were presented with 4 panels containing 1 or 2 objects (see Fig 1 below).

Experimental target trials referred to either [A] or [AB] panel which share an item thus, creating an ambiguity if referring to the [A] panel with an unmodified noun phrase ("The raspberry"). To resolve this ambiguity a listener would have to infer that, the speaker means the panel with "only the [A]" and not the [B] item. For [AB] targets there would be no ambiguity if speakers referred to the [B] item ("The lemon").

Global priming was manipulated between participants; one group of participants only heard cards described with single NPs, whereas for the other group half of the [A] prime descriptions included a modifying adverb (e.g. "only"). Local priming was manipulated within participants through the prime-target ordering. The dependent measure was the proportion of unmodified single NP utterances participants produced.

Preliminary Results: Data collection is still ongoing. Therefore, a formal analysis has not yet been conducted, and only descriptive statistics are reported below (Figure 2). Looking at the rates of unmodified single NP utterances participants in the Implicit condition are overall more likely to produce single NP utterances than those in the explicit condition (M = .81 vs .46). The picture for local priming also seems to be patterning with prior work where the target utterances are contingent on the prime utterance's construction.

References

[1] Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, *3*, 43-58.

[2] Khorsheed, A., & Gotzner, N. (2023). A closer look at the sources of variability in scalar implicature derivation: a review. *Frontiers in Communication*, *8*, 1187970.

[3] Khorsheed, A., Price, J., & van Tiel, B. (2022). Sources of cognitive cost in scalar implicature processing: A review. *Frontiers in Communication*, *7*, 990044.

[4] Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*, e282.

[5] Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329-347.

[6] Rees, A., & Bott, L. (2020). Overlapping mechanisms in implying and inferring. *Cognitive Science*, *44*(1), e12808.

Figures

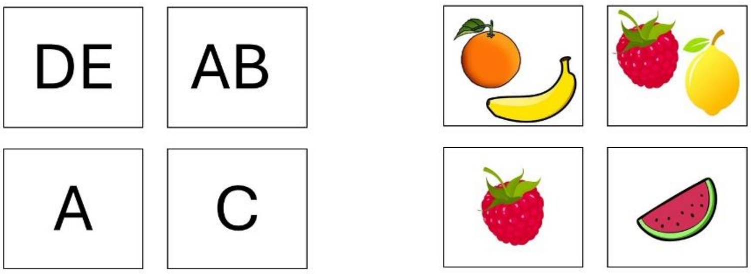


Figure 1. Left panel shows picture configurations. Right panel shows an example trial.

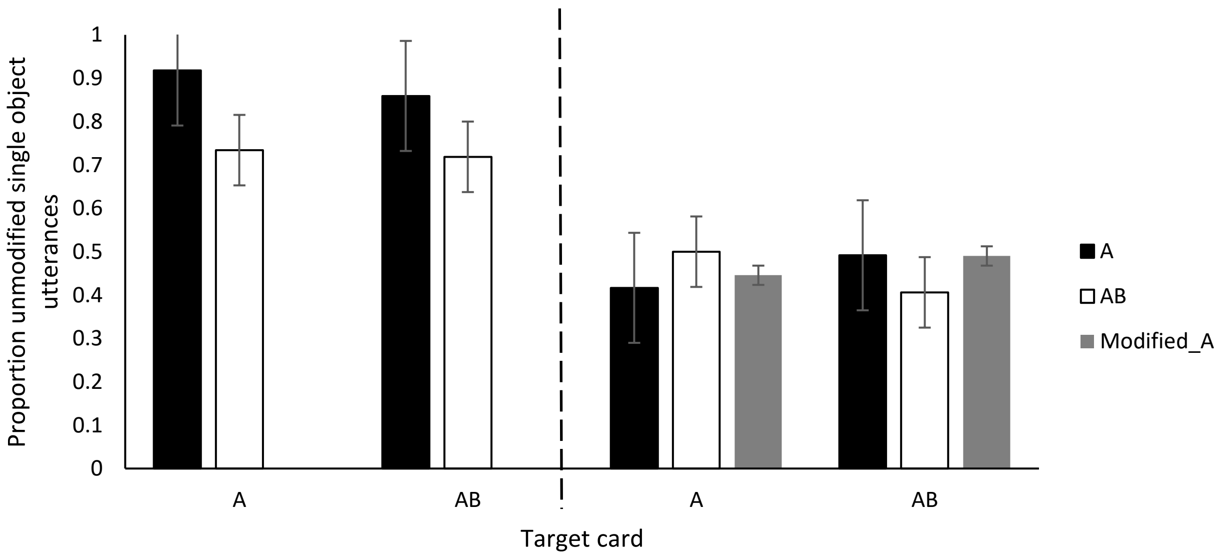


Figure 2. Proportion of implicit responses to targets. Left panel shows results for the implicit condition, right panel shows results for the explicit condition. Black bars correspond to A prime trials, white bars correspond to AB prime trials, and grey bars correspond to modified A prime trials in the explicit condition.

Binding Principle C in Online Processing of Mandarin Cataphoric Pronoun Resolution

Yi-ching Su[1]; Antonella Sorace[2]; Ming-Lei Chen[1]

[1] National Tsing Hua University, Taiwan; [2] University of Edinburgh, UK

ycsu@mx.nthu.edu.tw

Background: Principle C has been claimed to be a universal constraint and children show the knowledge of Principle C as early as they could be tested (e.g., English: Crain & McKee 1985; Italian: Guasti & Chierchia 1999/2000; Russian: Kazanina 2005; Mandarin: Su 2020). Moreover, online processing studies using self-paced reading tasks found evidence for an early effect of Principle C (e.g., English: Kazanina et al. 2007; Yoshida et al. 2014; Russian: Kazanina & Phillips 2010; German: Experiment 2 in Patterson & Felser 2019;). However, studies using eye-tracking reading tasks on German (i.e., Drummer & Felser 2018; Experiment 1 of Patterson & Felser 2019) found gender effects for both constraint and no-constraint conditions, hence a delayed filter effect for Principle C. The goals of the current study are to examine (1) whether Principle C acts as an early filter or a delayed filter in Mandarin online sentence processing, and (2) whether there is a task effect between self-paced reading and eye-tracking reading tasks. Unlike the languages used in previous studies, Mandarin third-person pronouns only show gender distinction in written forms but not in spoken forms, and hence whether this language-specific property in Mandarin may influence the processing pattern will also be discussed.

Methods: Two experiments (one self-paced reading and one eye-tracking reading) were conducted (N=40 in each). Each experiment had 24 sets of experimental trials with 4 conditions as shown in examples. The pronoun in conditions (a)/(b) is the matrix subject NP c-commanding the embedded subject in region 4 and coreference is prohibited by Principle C, whereas the pronoun in conditions (c)/(d) is the possessive of the matrix subject NP and does not c-command the embedded subject, making coreference acceptable.

Results: Data was analyzed with linear mixed-effects models on the pre-critical region (i.e., region 3), the critical region (i.e., region 4), and the three spillover regions. For self-paced reading, only the reading time at the third spillover region (i.e., region 7) in the no-constraint condition showed significant difference (p=0.04). For eye-tracking, the differences at the second and third spillover regions (i.e., regions 6 and 7) in the no-constraint condition was significant for re-reading time (p=0.03) and total reading time (p=0.01), respectively. (All had longer reading times for gender-mismatch than gender-match conditions as shown in Table 1.) None of the comparisons for the constraint condition at the relevant regions reached statistical significance.

Discussion: Unlike previous self-paced reading studies, the gender effects in the current study occurred at the third spillover region rather than at the critical noun region for the no-constraint condition. Moreover, unlike precious eye-tracking reading studies on German, the gender effects were mainly observed for the no-constraint condition at later stages of processing. The results of the two experiments demonstrate that (1) although gender distinction on pronouns appears only on written forms in Mandarin, Principle C still serves as an active filter in Mandarin online cataphoric resolution, and (2) the orthographic but not phonological distinction on gender renders the effect to be observable mainly at the post-critical regions.

Examples for experimental conditions:

(a) Constraint (gender match)
他 星期一 在會議中 聽說 李榮浩 擔任 評審委員 這使得 周杰倫 整天 悶悶不樂
he Monday in meeting hear (male name) serve reviewer this make (male name) all day unhappy

(b) Constraint (gender mismatch)
她 星期一 在會議中 聽說 李榮浩 擔任 評審委員 這使得 楊丞琳 整天 悶悶不樂
she Monday in meeting hear (male name) serve reviewer this make (female name) all day unhappy
"He/She heard that (male name) served as the reviewer in the meeting on Monday, and this made (male/female name) unhappy for the whole day."

(c) No constraint (gender match)
他 的朋友 在會議中 聽說 李榮浩 擔任 評審委員 這使得 周杰倫 整天 悶悶不樂
his friend in meeting hear (male name) serve reviewer this make (male name) all day unhappy

(d) No constraint (gender mismatch)
她 的朋友 在會議中 聽說 李榮浩 擔任 評審委員 這使得 楊丞琳 整天 悶悶不樂
her friend in meeting hear (male name) serve reviewer this make (female name) all day unhappy
"His/Her friend heard that (male name) served as the reviewer in the meeting, and this made (male/female name) unhappy for the whole day."
1 2 3 4 5 6 7 8 9 10

Table 1. Mean reading times with significant difference

	Self-paced reading Region 7	Eye-tracking reading Region 6 Re-reading time	Eye-tracking reading Region 7 Total reading time
No-constraint (gender-mismatch)	444	404	445
No-constraint (gender-match)	416	331	359

References

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. *Proceedings of NELS 15* (pp.94-110). Amherst, MA: GLSA.

Drummer, J. D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non- native sentence comprehension. *Journal of Memory and Language, 101*, 97–113.

Guasti, M. T. & Chierchia, G. (1999/2000). Backward and forward anaphora: Reconstruction in child language. *Language acquisition*, *8*(*2*), 129-170.

Kazanina, N. (2005). *The acquisition and processing of backwards anaphora*. Ph.D. dissertation, University of Maryland, College Park.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*(3), 384–409.

Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology, 63*(2), 371–400.

Patterson, C., & Felser, C. (2019). Delayed application of binding condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research, 48*(2), 453–475.

Su, Y.- C. (2020). Backward/forward anaphora in child and adult Mandarin Chinese. *Language Acquisition*, *27*(*2*), 187-216.

Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience, 29*(7), 761–770.

Animacy and null objects in English

Raffaella Folli[1], Juliana Gerard[1], Heidi Harley[2], Balthazar Lauzon[1], Morgan Macleod[1]
[1] Ulster University; [2] University of Arizona
m.macleod@ulster.ac.uk

Background: Object omission is acceptable in English in specific contexts, as in (1). The phenomenon of object omission in English has long received considerable attention [1, 2, 3, 4], and recent accounts of object omission focus on different aspects of these contexts. For example, Rappaport Hovav & Levin [5] and Folli & Harley [6] note that only animate entities can perform the action denoted by the verb in the absence of the object, as observed in (1) with the animate subject *John*. This predicts the marginal acceptability of (2), with an inanimate subject. However, RH&L propose a **polysemy-style account** based on their analysis of *sweep*, in which the lexicon contains both a 'basic *sweep'*, which always requires an object, and a 'conventionalized *sweep'*, which does not require an object but which requires that the subject be an Agent. In contrast, F&H suggest that such contrasts may relate to a **more pervasive distinction between animate and inanimate subjects**; more specifically, they propose that the cognitive primitive responsible for the observed distributions is that of goal-oriented action, a criterion which aligns with animacy to a considerable (if not complete) extent. Little quantitative research has been performed to establish the extent to which such factors play a role in object omission, or the extent of variation among verbs in their association with object omission. Moreover, the two accounts make different predictions: the approach of RH&L predicts a greater degree of lexical idiosyncrasy, while the syntactic account of F&H predicts greater transparency and uniformity. This project tests these predictions to investigate **the role played by animacy in object omission for English speakers**.

Method: The first phase of this study was an online experiment conducted with adults (N=80), who were native English speakers recruited on Prolific. Participants heard recordings of sentences such as (1) and (2) and rated their acceptability rating in a 7-point Likert format; individual sentences were presented in randomized order, with animate or inanimate subjects (within-subjects) and overt or omitted objects (between-subjects). 12 test verbs were used, with 4 intransitives and 4 obligatory transitives as controls.

Results: We conducted a linear regression analysis with fixed effects of animate/inanimate and overt/omitted object and random effects for participant and verb, and z-scored ratings as the dependent measure. Both main effects were significant (ps < .001); importantly, so was the interaction (p<.001): **with inanimate subjects, there were higher ratings with an overt object than with the object omitted; with animate subjects, no significant differences were found** (Figure 1). Thus, animacy and object presence are significant factors in predicting the acceptability of test sentences. However, variation also existed among different verbs. For some verbs, object omission was significantly less acceptable with inanimate subjects (e.g. *wash*); for others, object omission was acceptable with animate and inanimate subjects (e.g. *cook*).

Discussion: The influence of animacy upon object drop with multiple verbs suggests that animacy-related effects may be more than isolated lexical idiosyncrasies. However, the existence of variation among different verbs suggests that these animacy effects may be best interpreted not as a strict syntactic constraint, but in terms of the grammatical embodiment of speakers' 'making sense' of the world, along the lines of Borer [7]. The subsequent phases of this research will assess the broader patterns represented by the correlations identified in this study.

Examples

(1) *John is cleaning the windows. He isn't cleaning very well.*

(2) *The rain is cleaning the windows. ??It isn't cleaning very well.*

Figures

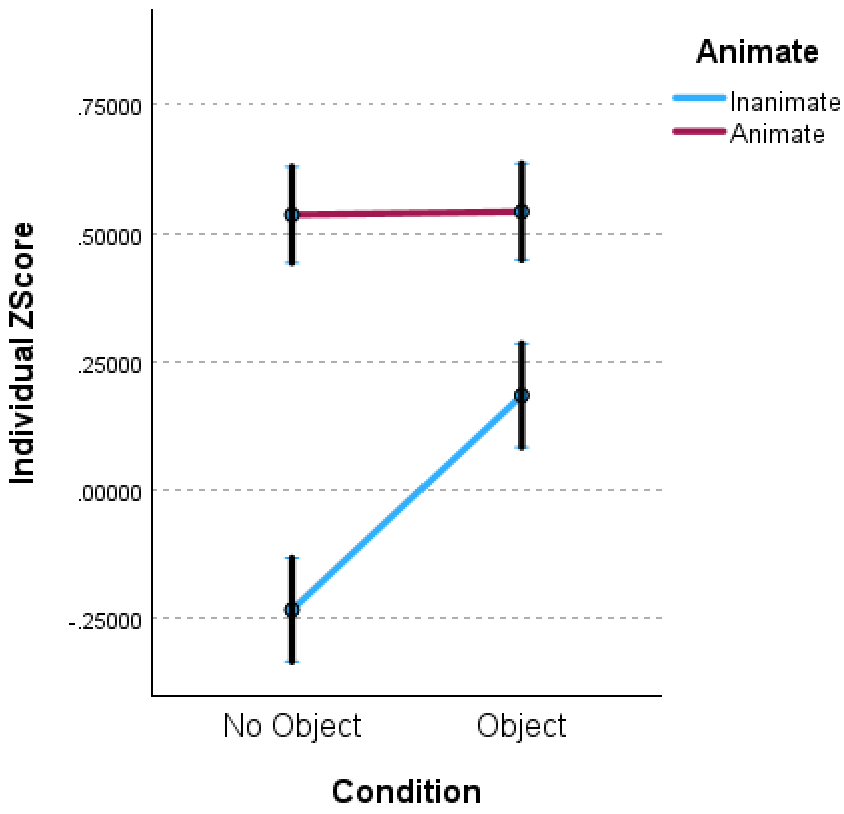


Figure 1: Interaction of animacy and object drop

References

[1] Fillmore, C. J. (1986). Pragmatically controlled zero anaphora in English. Berkeley Linguistics Society 12, 95–107.

[2] Martí, L. (2015). Grammar versus pragmatics: Carving nature at the joints. Mind & Language 30(4), 437–473.

[3] Yankes, A. (2021). Object drop in English: A statistical and Optimality Theoretical analysis. Doctoral dissertation, Carnegie Mellon University.

[4] Glass, L. (2022). English verbs can omit their objects when they describe routines. English Language and Linguistics 26(1), 49–73.

[5] Rappaport Hovav, M. & Levin, B. (2024). Variable agentivity: Polysemy or underspecification? Glossa 9(1), 1–39.

[6] Folli, R. & Harley, H. (Forthcoming). Teleological capability in English object "drop" constructions.

[7] Borer, H. (2005–2013). *Structuring Sense*. Oxford: Oxford University Press.

The impact of multi-accent and L1-accented input on preschoolers' perceptual adaptation to L2 vowels

Monika Kučerová[1,2], Kateřina Chládková[1,2]

[1] Institute of Psychology, Czech Academy of Sciences; [2] Faculty of Arts, Charles University

kucerova@psu.cas.cz

Background: In a multilingual society, being able to adapt to novel accents is indispensable. Children successfully recognize words in an unfamiliar L1 accent at 19 months [1]. In 8-year-old children, exposure to an unfamiliar L2 accent boosts word recognition for words heard previously in that accent [2]. The question is whether children's adaptation to a novel L2 accent is facilitated by prior experience with multi-accent L2 input. We test this by examining L2 vowel adaptation by 5-6 year old children exposed to English non-immersively. We ask whether exposure to varying L2 accents enhances children's ability to adjust to /ɛ,æ/ in an unfamiliar L2 accent. This contrast is difficult to acquire, Czech learners typically map both categories onto L1 [ɛ].

Method: Data was collected from 51 children with L2 experience (mean length of exposure = 2.5 years). Over two consecutive days, children watch three short animated videos in one of three conditions (randomly assigned between subjects). They feature two talkers with the same native accent (General British English, GBE); the same non-native accent (Czech-accented English, CE); or one native and one non-native talker (one CE and one GBE talker, condition Multi). After exposure, a test video followed by a word recognition task is administered using stimuli from two talkers of unfamiliar native accents (White South African English, WSAE; and Singapore English, SE). The task features three types of trials: the participant hears two dissimilar words, members of a minimal pair (MP) produced contrastively (one with [ɛ], the other with [æ]), or as homophones (both with [ɛ]). The contrast we focus on is present in GBE and WSAE, but not in CE and SE. We model word identification accuracy using GLMM with condition, trial type, and progress (responses after training and test).

Results: Multi-accent input was not confirmed to enhance novel accent adaptation. Differences were found in how often the groups replayed sounds on the task: when compared to the Multi group, CE and GBE groups replayed MPs more often than dissimilar words. All groups identified dissimilar words consistently above chance. MPs were not identified with above-chance accuracy at training by the GBE and Multi groups. Surprisingly, the CE group recognized MP words with above chance accuracy at training, even though Czech talkers typically produce them as homophones [3]. Acoustic analyses reveal no consistent cue differences between the stimuli, suggesting possible use of multiple cues to identify the words, or reliance on exemplar representation. This L1 benefit did not carry over onto test stimuli produced by the SE talker. No group showed above chance performance on MP trials at test, suggesting that at this stage of L2 acquisition, children use variable categories that encompass both [ɛ] and [æ] to process L2 speech.

Discussion: Word identification performance of all groups suggests equivalence classification of /ɛ, æ/ tokens [4], despite long-term exposure to native English. This view is compatible with the L1 adaptation account proposed by Kleinschmidt and Jaeger [5]: participants cannot fully exploit distributional learning because they hold strong L1-based beliefs about category characteristics. However, chance-level identification performance does not preclude adaptation: children may be using the equivalence classification mechanism to adapt to novel speech [6], which involves category widening as an adaptive response to novel input.

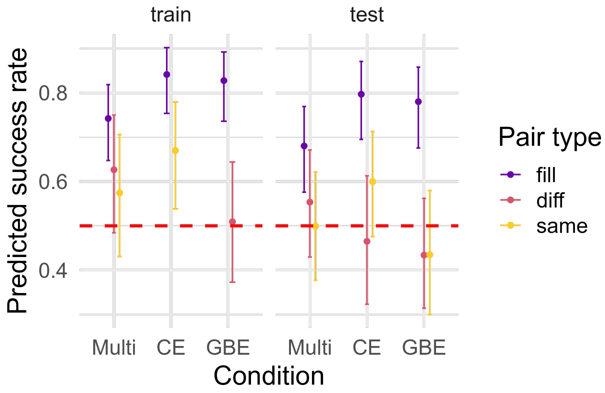


Figure 1: Predicted word identification accuracy scores at training (left) and test (right) for each condition on three trial types (*fill* for two dissimilar words, *diff* for minimal pairs where one member is produced with [ɛ] and the other with [æ], *same* for minimal pairs produced as homophones with [ɛ].)

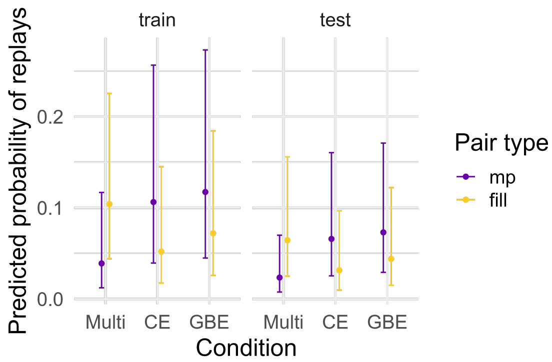


Figure 2: Predicted probability of replaying sounds on the word identification task, separately for dissimilar word trials and minimal pair trials.

References

[1] Johnson, E. K., van Heugten, M., & Buckler, H. (2022). Navigating accent variation: A developmental perspective. *Annual Review of Linguistics*, *8*(1), 365–387.

[2] Hu, C.-F. (2021). Adaptation to an unfamiliar accent by child L2 listeners. *Language and Speech*, *64*(3), 491–514.

[3] Šimáčková, Š. (2003). Engela's fleshes": Cross-linguistic perception and production of english [æ] and [ɛ] by czech efl learners trained in phonetics. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003), Barcelona*, 3–9.

[4] Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, *10*(9781108886901.002).

[5] Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148.

[6] Schmale, R., Seidl, A., & Cristia, A. (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental science*, *18*(4), 664–670.

Exploring ERP Components in Emotional Word Processing: Participant Subjectivity and Embodiment

Hoekeon Choi[1, 2], Haeun Ko[1], Ha-a-yan Jang[1], Jonghyun Lee[3], Sung-Eun Lee[1, 4]
[1]Artificial Intelligence Institute of Seoul National University; [2]Department of English Language and Literature, Seoul National University; [3]Department of English Language and Literature, Sejong University; [4]Department of German Language and Literature, Seoul National University
cristlo5@snu.ac.kr

Background: Emotional meaning is subjective; the same word can evoke different feelings across readers. Because words are symbolic rather than perceptual, this subjectivity plays a larger role in language than in direct visual stimuli such as faces or images, yet most emotion-language studies still treats affective values (valence; postive-negative, arousal; high-low) [1] as fixed stimulus properties. Addressing this gap, the present study investigates how the emotional dimensions of valence and arousal modulate emotional word processing, taking participant subjectivity into account, by examining two ERP components—the Early Posterior Negativity (EPN) and Late Positive Potential (LPP).

Method: Thirty-five Korean adults viewed Korean emotional words that was “originally” classified by pre-fixed ratings into valence (positive, negative) and arousal (high, low) categories, with object words as a neutral category. While each word displayed, participants completed two counter-balanced 9-point rating tasks, one for valence (1 = negative, 9 = positive) and the other for arousal (1 = low, 9 = high), under continuous 64-channel EEG recording. Afterward, each word’s valence and arousal scores were averaged across all participants, and the 30 most extreme items in each category were retained to create a new “reclassified” stimulus set. In both original and reclassified sets, ERP amplitudes for EPN (250–350 ms, temporo-occipital) and LPP (400–800 ms, centro-parietal) [2, 3] were then analysed in separate 3 (Valence: positive, negative, neutral) × 2 (Task: valence-rating, arousal-rating) and 3 (Arousal: high, low, neutral) × 2 (Task) repeated-measures ANOVAs.

Results: Repeated-measures ANOVAs revealed robust main effects of Valence and Arousal for both the EPN and LPP in the original and reclassified word sets (Table 1). The LPP patterns, however, diverged between sets: valence shifted from *Negative* > *Positive* = *Neutral* in the original set to *Negative* = *Positive* > *Neutral* in the reclassified set, indicating a reduced negativity bias, while the arousal effect sharpened from *High* = *Low* > *Neutral* to a graded *High* > *Low* > *Neutral* trend (Table 1, Figure 1 & 2). This trend shift replicated in both rating tasks, showing that the pattern is stable across task focus.

Discussion: Altered LPP pattern after subjective reclassification underscores the role of subjective arousal impressions in later processing of emotional words. This aligns with the relatively weak correspondence between original(pre-fixed) and reclassified(subjective) arousal ratings ($r \approx 0.5$), compared to valence ratings ($r \approx 0.9$). Participant subjectivity must therefore be incorporated to fully capture how emotional dimensions shape neural responses to language [4]. Moreover, because the LPP trend shift persisted across tasks, the subjective reclassification effect appears task-independent.

Tables

Table 1. Summary of Significant Main Effects of Valence/Arousal in RM-ANOVAs for EPN and LPP

ERP Component	Main effect	Post-hoc	
		Original(Pre-fixed) set	Reclassified(Subjective) set
EPN	Valence	Negative < Positive < Neutral**	Negative < Positive < Neutral**
	Arousal	High = Low > Neutral**	High = Low > Neutral*
LPP	Valence	**Negative > Positive = Neutral***	**Negative = Positive > Neutral****
	Arousal	**High = Low > Neutral****	**High > Low > Neutral****

Note: ** $p < 0.001$; * $p < 0.01$

Figures

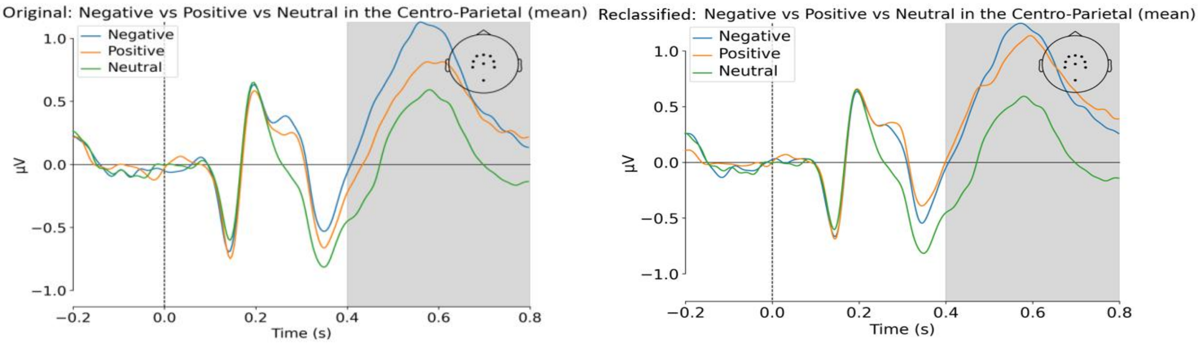


Figure 1. LPP: Differences by Emotional Valence (Left: Original, Right: Reclassified)

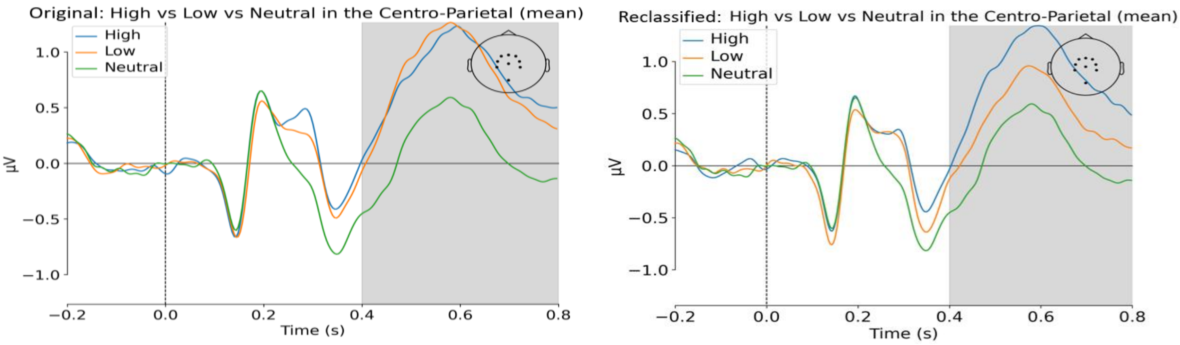


Figure 2. LPP: Differences by Emotional Arousal (Left: Original, Right: Reclassified)

References

[1] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

[2] Schindler, S., Vormbrock, R., Helming, H., & Straube, T. (2023). Dissociating different temporal stages of emotional word processing by feature-based attention. *Scientific Reports*, *13*(1), 16860.

[3] Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive, affective, & behavioral neuroscience*, *16*, 415-432.

[4] Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological bulletin*, *139*(4), 917.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(0670-20240045).

Investigating Cerebellar-Language Network Alterations in Parkinson's Disease Using Open Data

Alaa M. Salem[1] and Shinri Ohta[2]

[1]Kyushu Univ., Grad. Sch. Lit., Dept. Ling., [2]Kyushu Univ., Fac. Hum., Dept. Ling.

Contact: abouzeid.alaa.495@s.kyushu-u.ac.jp

Recent research demonstrates that the cerebellum plays a key role not only in motor control but also in cognitive functions, including language. In healthy adults, cerebro-cerebellar pathways—particularly those involving Crus I/II and cortical language areas—are implicated in syntactic processing, verbal fluency, and semantic integration [1]. Investigating a disorder that affects both language and cerebellar circuits could help unravel underexplored mechanisms. Parkinson's disease (PD) is classically identified by motor deficits; however, many patients develop subtle yet significant language impairments. These may result from disrupted cerebro-cerebellar connectivity [2], which remains poorly understood in PD research.

Our study aims to explore the structure of white matter tracts connecting the cerebellum to cortical language regions such as the left inferior frontal gyrus (LIFG) and left superior temporal gyrus (LSTG). Using advanced diffusion tensor imaging (DTI) and tractography analyses, we will correlate imaging findings with objective language performance measures, including syntax comprehension, semantic processing, and verbal fluency. We propose a multi-modal approach using open-access imaging data from the Parkinson's Progression Markers Initiative (PPMI). Specifically, we will employ DTI and high-resolution T1-weighted MRI datasets to trace cerebro-cerebellar pathways, such as connections between Crus I/II, thalamus, and LIFG/LSTG.

Region of interest (ROI)-based tractography will be used to outline specific tracts and quantify diffusion metrics such as fractional anisotropy (FA) and mean diffusivity (MD). These metrics will be correlated with language scores using regression analyses. We hypothesize that PD patients will show reduced white matter integrity within critical cerebro-cerebellar tracts compared to controls, and that lower FA values will be associated with poorer language performance. To consider the impact of external variables, we plan to include variables such as education level, age of onset, bilingualism, and musicianship in our models.

Our longitudinal DTI analysis will follow patients aged 40–75 over a span of 3–5 years to determine the slope of degeneration and its temporal relationship with language decline. This may help identify early imaging biomarkers for progressive cognitive-linguistic impairment and guide neuromodulation strategies to reduce or delay language deficits. Additionally, we will explore whether changes in white matter integrity precede measurable speech or comprehension problems in early-stage PD.

Future work will investigate how stimulation techniques such as transcranial magnetic stimulation (TMS) and transcranial alternating current stimulation (tACS) affect cerebro-cerebellar pathways in both control and PD patients. By integrating statistical analysis with cutting-edge tractography and open neuroimaging data, our study might provide a novel perspective on the structural basis of language dysfunction in PD. We expect our findings to shed light on potential compensatory mechanisms—such as increased recruitment of cerebellar regions—and contribute to broader research on non-motor symptoms in neurodegenerative disorders.

References:

[1] Massara, M., Delogu, C., Cardinale, L. et al. The lateralized cerebellum: insights into motor, cognitive, and affective functioning across ages: a scoping review. J. Neurol. 272, 122 (2025).

[2] Pietracupa, S., Ojha, A., Belvisi, D. et al. Understanding the role of cerebellum in early Parkinson's disease: a structural and functional MRI study. npj Parkinsons Dis. 10, 119 (2024).

Understanding intertextual relations and numerical processing in L2 multiple-text reading: An eye-tracking study

Yuko Hijikata[1], Masumi Ono[2], Haruka Shimizu[3], Yuko Hoshino[4], Yuji Ushiro[5]
[1]University of Tsukuba, [2]Keio University, [3]Tohoku Gakuin University,
[4]Chiba University, [5]Japan International University
hijikata.yuko.fe@u.tsukuba.ac.jp

Background: Previous studies have shown that source information affects multiple-text reading processes. Specifically, science-based information is reported to have higher credibility [1], although scientific texts are more difficult to understand and recall than texts based on personal cases [2]. Considering that we often read science-based texts and that intertext integration skills are required in our daily and academic lives, several English proficiency tests (e.g., the Test of English for International Communication) include questions that require test-takers to integrate information from multiple texts. Moreover, authentic texts frequently require the interpretation of arithmetic and statistical data (numerical processing), and language and mathematics share domains [3]. However, intertext questions that require numerical processing in intertextual integration have not been extensively researched, specifically in L2. Furthermore, few L2 studies have examined the cognitive processes involved in answering reading comprehension questions using eye-tracking techniques [4]. Therefore, we conducted an eye-tracking study to examine these response processes and compared L2 students' processing of intertext questions that require numerical processing with those that can be answered through language only.

Method: Japanese undergraduate and graduate students who are learning English as a foreign language read 16 pairs of English texts that included numbers. Each material set contains a lead sentence describing the context, two four-sentence-long texts, a comprehension question, and four answer options. Half of the comprehension questions are intratext questions based on the content of one text. The other half consist of intertext questions that require the integration of information across texts. Furthermore, half of the intertext questions require the participants to process numerical-based information across texts, while the other half ask for the integration of linguistic information only. Therefore, this study adopted a 2 (inter vs. intra) x 2 (numerical vs. linguistic) within-subject design. The participants' eye movements are recorded using the EyeLink Portable Duo, and the following areas of interest (AOIs) have been set: (a) Text A, Text B, and the question and answer (Q&A) section, (b) phrases and sentences necessary to answer the questions. After the reading session, stimulated recall interviews were conducted. We established separate models for eye-tracking indices (i.e., fixation duration, fixation counts, and dwell time) as well as comprehension accuracy. Processing time data are analyzed through linear mixed-effects models, and we have developed generalized linear mixed-effects models for accuracy and frequency data.

Results and Discussion: The preliminary results are as follows. First, regarding the dwell time on the Q&A section, the difference between the intertext and intratext questions was greater for numerical questions than for linguistic questions. Second, fixation duration on the question was longer with the linguistic questions than with the numerical questions, while the number of fixation counts was higher with the numerical questions than with the linguistic questions. In summary, our findings support Savvidou et al. (2025) in that argument type influences multiple-text reading processes, and we extend their results to L2 readers.

Sample of the Material Set

[Lead sentence] People use water for different purposes and without careful consideration.
[Text A] Total daily water consumption is 300 liters per person. For example, 50 liters are used in toilets, 40 liters for cooking, 35 liters for laundry, and 80 liters for taking a bath. Given this large amount, we must save and store water for emergencies. Reducing water usage can also help lower costs, as the price of water increases every day.
[Text B] Water is used outside the home, in places like restaurants and parks. The average daily water usage for such places is about 70 liters per person. Although supplying free water to public places is common practice, maintaining water quality is challenging. In fact, some water pipes in Japan are aging, thus being at risk from earthquakes.
[Intertext questions] (Numerical) How much more water is used at home compared to water used outside the home? (a) About 2.5 times (b) About 3 times (c) About 3.5 times (d) About 4.2 times
(Linguistic) What important action is indicated as in need of being taken? (a) Introducing a water supply system (b) Reducing the water rates (c) Preparing for disastrous situations (d) Improving the quality of water
[Intratext questions] (Numerical) On average, how much water is used at home per person weekly? (a) About 230 liters (b) About 490 liters (c) About 2,100 liters (d) About 2,600 liters
(Linguistic) What is indicated in the text? (a) Water rates are high everywhere (b) Water rates are increasing (c) Water consumption is decreasing (d) People avoid paying water rates

Figure 1. Comparison of Question Types

Dwell Time (ms) on the Q&A / Fixation Duration (ms) on the Question / Fixation Counts on the Question

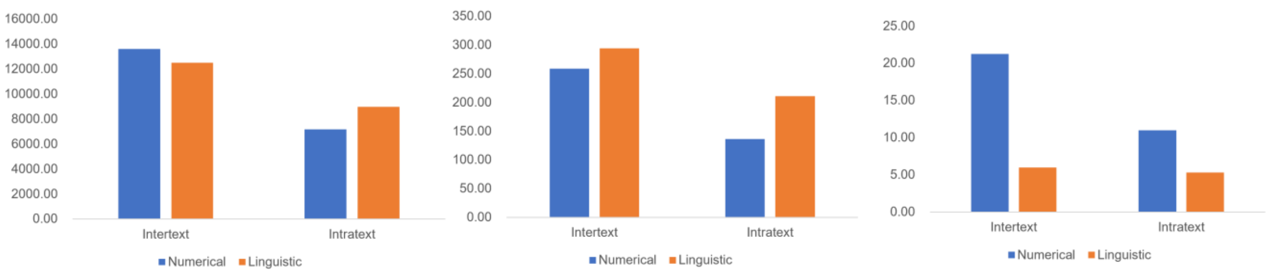
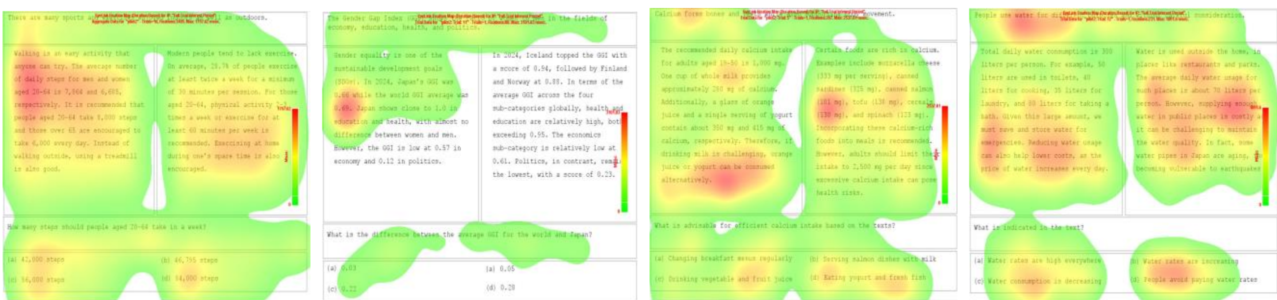


Figure 2. Examples of Heatmaps

Intertext Numerical Intratext Numerical Intertext Linguistic Intratext Linguistic



References

[1] Kiili, C., Bråten, I., Strømsø, H., Hagerman, M. S., Räikkönen, E., & Jyrkiäinen, A. (2022). Adolescents' credibility justifications when evaluating online texts. Education and Information Technologies, 27, 7421–7450.
[2] Savvidou, S. M., Diakidoy, I.-A., & Mason, L. (2025). Multiple-text comprehension and evaluation: The influence of reading goal, belief consistency, and argument type. Reading Research Quarterly, 60, e568.
[3] Scheepers, C., Galkina, A., Shtyrov, Y., & Myachykov, A. (2019). Hierarchical structure priming from mathematics to two- and three-site relative clause attachment. Cognition, 189, 155–166.
[4] Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. Language Testing, 30(4), 441–465.

Cognitive Load and Language Dominance: Bilingual Performance in a Dual-Task Paradigm

Nilanjana Chowdhury[1], Bidisha Som[2], Sukumar Nandi[3]

1. Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati
2. Department of Humanities and Social Science, Indian Institute of Technology Guwahati
3. Department of Computer Science, Indian Institute of Technology Guwahati

Corresponding email: n.chowdhury@iitg.ac.in

In bilingual societies, it is common to find one of the languages more dominant than the other (Oppenheim et al., 2020a), however the dominant language may or may not be the first language of the speakers. Given the nature of modern life, where people often leave their native place for higher education, job etc, dominant language can change depending on a person's life-stage. The current study is focused on young adult bilinguals (Hindi-English bilinguals from India, having a formal education in Hindi and English), who are about to complete their degrees in a national institute. During their stay in campus (between 4-6 years), English has been the primary language and hence was expected to be the domain language of the group. This was investigated using two experiments: (a) a picture naming task and (b) Dual-task paradigm.

At first, 250 images were selected from The International Picture Naming Project CRL-UCSD and their names in both first (L1) and second language (L2) were added to them. These two lists then, was anonymously rated for frequency, familiarity, and correctness of match by 10 participants each for Hindi and English. The rating yielded 150 image-name pairs that scored 80% and above on the three parameters. Same was done for both L1 and L2 lists and the final list was used as a stimulus.

A total of 90 participants in the age group of 18-28 years took part in the experiment. Statistical analysis using Python reveals a statistically significant difference in the L1 Vs. L2 in both accuracy (Hindi: Mean= 91.17647, SD= 4.765899; English: Mean= 92.89273, SD= 3.8714) and Response Time (RT) (Hindi: Mean= 835.0725, SD= 161.838; English: Mean= 632.059, SD= 152.7944). Mann Whitney U test results indicate p-value= 0.004078 for Accuracy and p-value=1.78E-13 for RT. The results indicate that L2, (English), shows better performance in both tasks, indicating dominance. To further understand the nature of the dominance, picture naming task as followed by, Dual Task Paradigm in order to probe, whether with increased cognitive load has any impact on the performance of the two languages. The primary task of this study was the Picture-Word Matching with the same stimulus as the previous one. For the secondary task Digit span recall was selected and used as a 3 back task where after every 3 trials of the primary task the digit sequence appeared and the participants had to respond if the digit sequence is same as before. The Digit sequence was of single digits with a length of not more than 5 digits. Statistical analysis using Python reveals a statistical significance in the primary task for L1 Vs. L2 in both accuracy (Hindi: Mean= 91.00403, SD= 2.881165; English: Mean= 94.10539, SD= 4.999848) and Response Time(RT) (Hindi: Mean= 1208.846, SD= 259.4981; English: Mean= 875.1606, SD= 167.1497). Mann Whitney U test results indicate p-value= 3.00E-06 for accuracy showing high statistical significance and t-test results indicates p-value= 7.10E-08 for RT with high statistical significance. For the secondary task, Mann Whitney U test indicates p-value= 0.316355 for accuracy and p-value= 0.064858 for RT. The results show no statistical significance for the secondary task but for the primary task even with increased cognitive load better performance in the second language was seen. The bilinguals performed both the tasks simultaneously and even though the digit span recall was seen to have a better performance in L2 than in L1 but with minimal or no statistical significance. Overall, the two experiments point towards the second language as the dominant language of the group. We propose the impact of education in a multilingual elite setting as a probable reason for the finding.

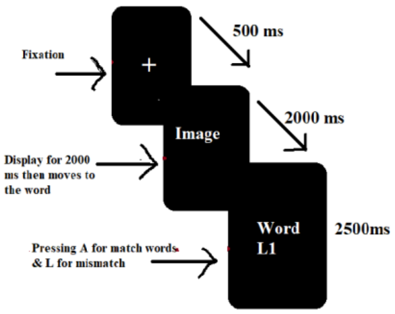


Fig.1: Schematic Representation of Experiment 1

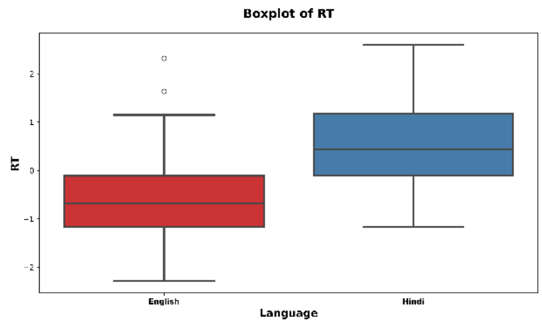


Fig.2: Boxplot of Response Time of Experiment 1

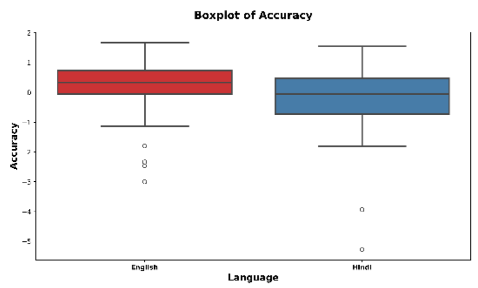


Fig.3: Boxplot of Accuracy of Experiment 1

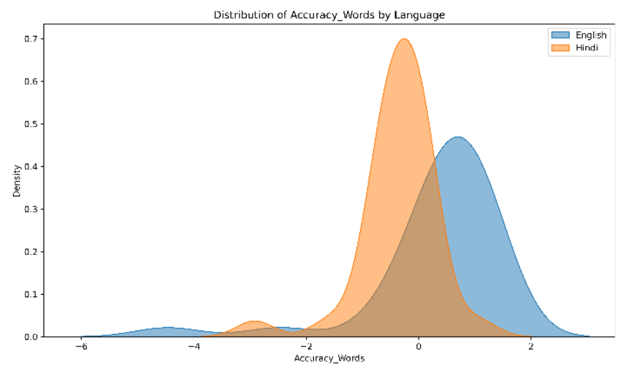


Fig.4: Distribution of Accuracy for L1 & L2 in Exp. 2

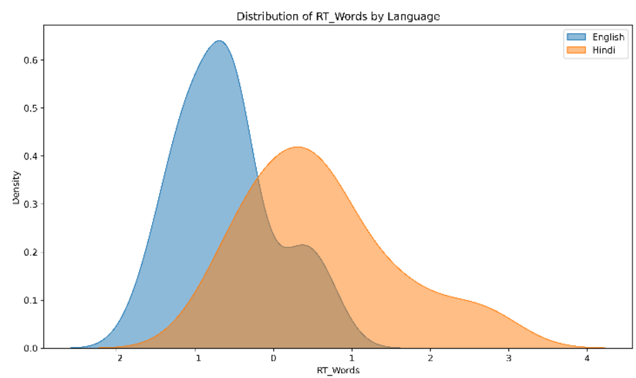


Fig.5: Distribution of RT for L1 & L2 in Exp.2

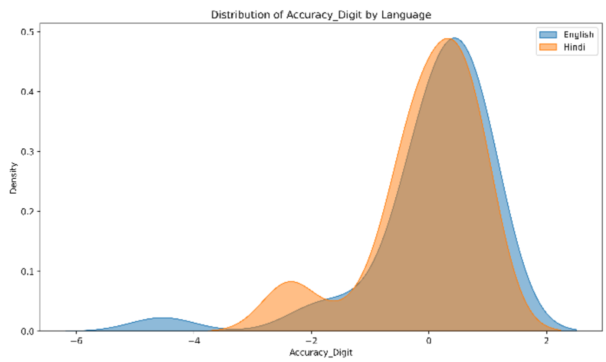


Fig.6: Distribution of Accuracy for Digit span recall in Exp.2

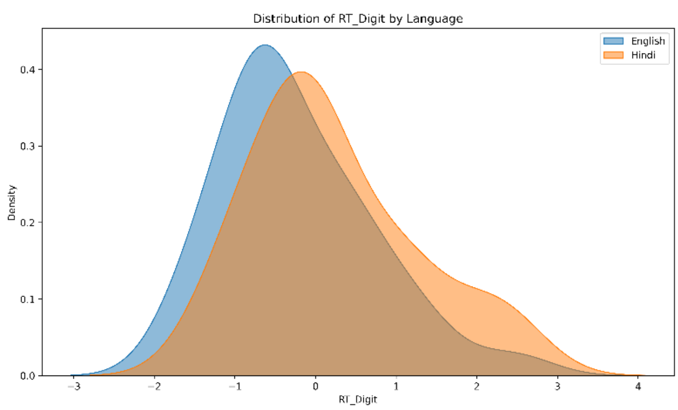


Fig.7: Distribution of RT for Digit span recall in Exp. 2

Cross-Modal Activation in Hearing-Impaired Preschoolers: An fNIRS Study of Speech and Sign Processing

Michaela Svoboda[1,2], Natálie Kikoťová[2,3], Monika Kučerová[1,2], Kateřina Chládková[1,2,4]
[1]Institute of Psychology, Czech Academy of Sciences, Prague, Czechia

[2]Faculty of Arts, Charles University, Prague, Czechia, [3]Graduate School of Neural and Behavioral Sciences, University of Tübingen, Tübingen, Germany, [4]Faculty of Medicine in Hradec Králové, Charles University, Hradec Králové, Czechia

michaela.svoboda@ff.cuni.cz

Background: In our contribution, we introduce a neuroimaging experiment that assesses the degree of language-related cross-modal activation in hearing-impaired (HI) children fitted with hearing prosthesis. Cross-modal activation is a state in which brain structures primarily associated with the processing of stimuli in one modality are involved in the processing of stimuli in different modalities, often as a result of sensory deprivation. Cross-modal activation and neural reorganisation have been studied for decades in people with HI, but its influence on the acquisition of spoken language is not yet understood [1,2].

Method: We used functional near-infrared spectroscopy (fNIRS) to measure hemodynamic responses during exposure to audiovisual speech, silent speech, and Czech Sign Language (CSL). Activity was recorded bilaterally in the superior temporal gyrus and temporal visual association areas (see Fig. 1 for optode placement and Fig. 2 for exp. setup). After artifact rejection, the sample included 11 children with cochlear implants (CIs; 8F, mean age: 66.5 months), 8 with hearing aids (HAs; 6F, 85 months), and 23 typically hearing peers (14F, 63.4 months). We expected stronger visual-only responses in HI children, typical auditory activation in hearing peers, and correlations between HbO levels, CSL exposure, and vocabulary.

Results: Fig. 3 shows grand averages across groups and conditions. Hearing children showed canonical auditory responses [3], while HI children exhibited a double-peaked response. Contrary to expectations, silent speech evoked negative HbO responses. CI users showed cross-modal responses to CSL. The effects of our factors of interest od HbO concentrations weree modelled using linear mixed effect models. In HA users, auditory responses increased with hearing age and negatively correlated with degree of hearing loss. CSL exposure interacted with vocabulary scores in predicting auditory HbO responses, suggesting sign input may influence the neural basis of spoken language.

Discussion: Our findings offer new insights into cross-modal processing in hearing-impaired children. The most surprising result was the negative hemodynamic response during visual speech - a pattern also reported in previous research but not yet fully understood [4]. These and other findings will be explored in more detail at the conference.

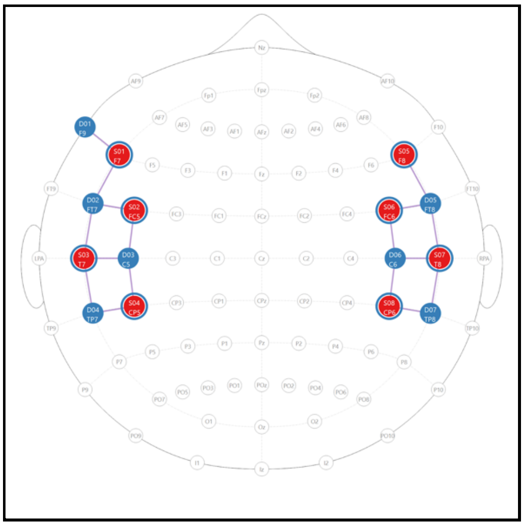


Figure 1: Optode placement.

Figure 2: Experimental setup

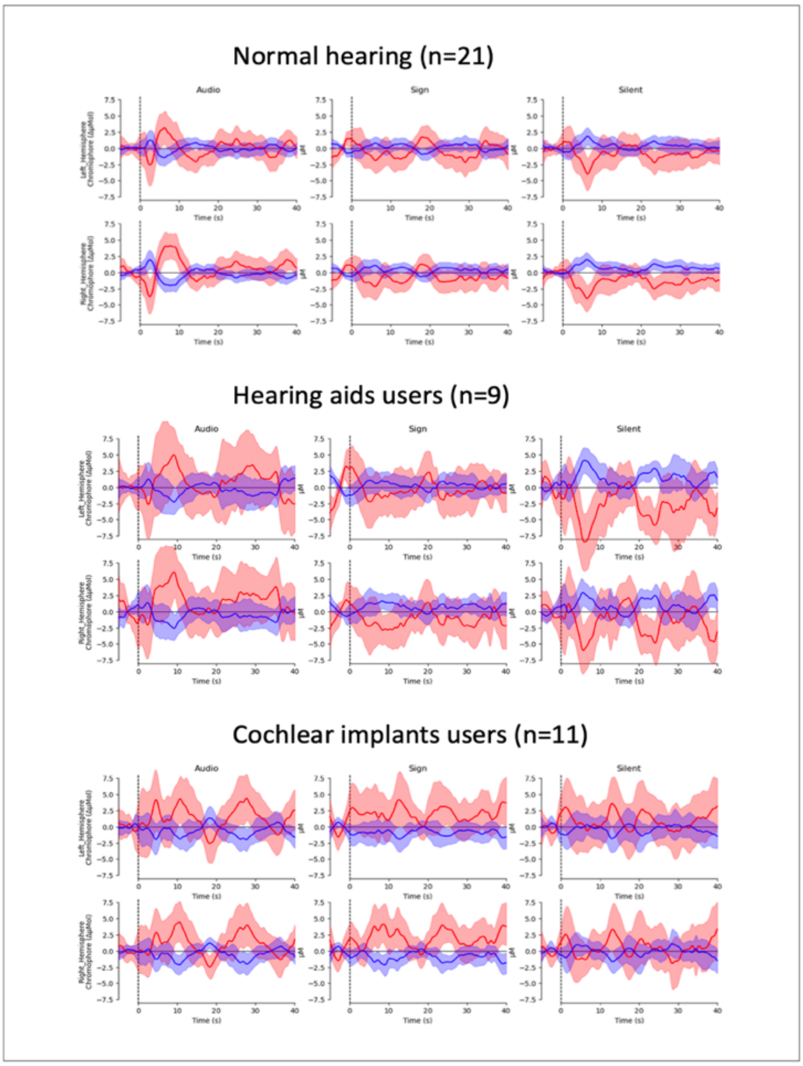


Figure 3: Grand averages across all channels for each group and condition

References

[1] Giraud, A. L., & Lee, H. J. (2007). Predicting cochlear implant outcome from brain organisation in the deaf. Restorative Neurology and Neuroscience, 25(3–4), 381–3

[2] Zhou X-Q, Zhang Q-L, Xi X, Leng M-R, Liu H, Liu S, Zhang T and Yuan W (2023) Cortical responses correlate with speech performance in pre-lingually deaf cochlear implant children. Front. Neurosci. 17:1126813. doi: 10.3389/fnins.2023.1126813

[3] Issard, C., & Gervain, J. (2017). Adult-like processing of time-compressed speech by newborns: A NIRS study. Developmental Cognitive Neuroscience, 25, 176–184. https://doi.org/10.1016/j.dcn.2016.10.006

[4] Steinmetzger, K., Shen, Z., Riedel, H., & Rupp, A. (2020). Auditory cortex activity measured using functional near-infrared spectroscopy (fNIRS) appears to be susceptible to masking by cortical blood stealing. Hearing research, 396, 108069. https://doi.org/10.1016/j.heares.2020.108069

Universal parsing biases: Small Clauses drive RC attachment in Vietnamese

Anh Ngo, Nino Grillo
Department of Language and Linguistic Science, University of York
anhngo.240201@gmail.com, nino.grillo@york.ac.uk

Background: The role of Small Clause (SC) availability in Relative Clause (RC) attachment has been demonstrated across Indo-European languages [1, 2, 3], but its generalization to typologically distinct languages remains unexplored. Investigating Vietnamese—a classifier language with divergent syntactic and morphological properties—can shed light on the universality of these parsing mechanisms. Vietnamese reduced RCs (1-a) under perceptual verbs can also be interpreted as eventive SCs (e.g., perceiving a running event). As in English, this ambiguity arises only with verbs that can take *events* as complement and disappears with verbs that only select *entities* (1-b) and with full RCs (1-c), introduced by a classifier and the relativizer *mà*.

(1) a. Mary thấy người-đàn-ông đang chạy.
Mary see class-man prog run.
Mary sees [the man [$_{RC}$ that is running]] / [$_{SC}$ [the man] [running]].

b. Mary kết-hôn với người-đàn-ông đang chạy.
Mary marry with class-man prog run.
*Mary is married to [the man [$_{RC}$ that is running]]. / *[$_{SC}$ [the man] [running]].*

c. Mary thấy người-đàn-ông người mà đang chạy.
Mary see class-man class that prog run.
*Mary sees [the man [$_{RC}$ that is running]] / *[$_{SC}$ [the man] [running]].*'

Evidence from Indo-European languages show that eventive SC availability robustly modulates RC attachment within and across languages [1, 3, 4, 5, 6] (a.o.), in line with Grillo's proposal that perceptual verbs prefer to embed events over entities and the interaction of this preference with the loss of attachment ambiguity under eventive SC readings (2). In two experiments, we tested whether this account also holds in Vietnamese.

(2) a. Mary thấy người-con-trai của người-đàn-ông đang chạy.
Mary sees the son$_i$ of the man$_j$ [$_{RC}$that is running$_{i/j}$].

b. Mary thấy [[người-con-trai [của [người-đàn-ông]]] [đang chạy]].
Mary sees [$_{SC}$[$_{DP1}$the son$_i$ [of [$_{DP2}$the man$_j$]]] [running$_{i/*j}$]].

Exp.1: We manipulated VERB-TYPE (Event taking/taking)*RC-TYPE (Long/Short) in an attachment questionnaire (N=60), Table 1. If the SCs preference extends to Vietnamese, we predict a higher proportion of HA with Event taking verbs, in particular with short RCs. **Results:** A significant VERB-TYPE*RC-TYPE interaction (β=-1.92, $p<$.001) shows that Event-taking verbs increased Non-local Attachment in all RC types, but the effect was stronger for SC-compatible short RC (β=2.59, $p<$.001) than for long RC (β=0.75, $p<$.001).

Exp.2: We tested online processing of RC-attachment via self-paced reading and semantic disambiguation in a 2VERB-TYPE (Event taking/Entity taking)*2ATTACHMENT (Local/Non-Local) design (N=76), Table 2. **Results:** Unambiguous RCs under entity taking verbs longer RTS for non-local disambiguation (β=69.19, *t-value*=2.9). The opposite trend, albeit not significant (β=-33.28, *t-value*=-1.4), appeared with event-taking verbs.

Discussion: Vietnamese replicates the SC-driven non-local preference observed in Indo-European languages, supporting universality of parsing principles. Offline, SC availability strongly modulated attachment; online, numerical trends aligned with this bias. Our results underscore the need to test psycholinguistic theories in typologically diverse languages to disentangle universal mechanisms from language-specific constraints.

Verb-Type	**RC-Type**	**Stimuli**	**SC-availability**
Event-taking	Long	Mary thấy người-con-trai của người-đàn-ông người mà đang chạy. *Mary sees the son of the man the one that is running.*	RC-only
Entity-taking	Long	Mary kết-hôn với người-con-trai của người-đàn-ông người mà đang chạy. *Mary is married to the son of the man the one that is running* .	RC-only
Event-taking	Short	Mary thấy người-con-trai của người-đàn-ông đang chạy. *Mary sees the son of the man running/that is running.*	SC/RC
Entity-taking	Short	Mary kết-hôn với người-con-trai của người-đàn-ông đang chạy. *Mary marries the son of the man that is running.*	RC-only

Table 1: Experimental design and example stimuli Experiment 1

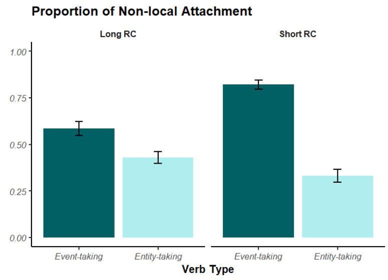


Figure 1: By-subject mean proportion of non-local attachment

Verb-Type	**Attachment**	**Stimuli (critical region in bold)**	
Event-taking	Non-local	John thấy người-con-gái của người-đàn-ông **đang sinh-con** ở bệnh-viện. *John sees the daughter of the man that is **giving birth** in the hospital.*	SC/RC
Entity-taking	Non-local	John sống với người-con-gái của người-đàn-ông **đang sinh-con** ở bệnh-viện. *John lives with the daughterof the man that is **giving birth** in the hospital.*	RC-only
Event-taking	Local	John thấy người-con-trai của người-phụ-nữ **đang sinh-con** ở bệnh-viện. *John sees the son of the woman that is **giving birth** in the hospital.*	RC-only
Entity-taking	Local	John sống với người-con-trai của người-phụ-nữ **đang sinh-con** ở bệnh-viện. *John lives with the son of the woman that is **giving birth** in the hospital.*	RC-only

Table 2: Experimental design and example stimuli Experiment 2

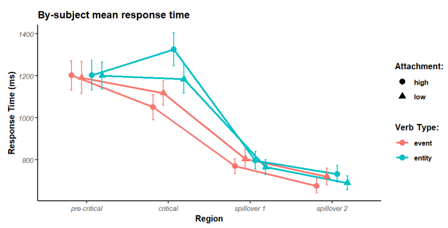


Figure 2: By-subject mean response times by region (Error bars represent standard error)

References

[1] Grillo. (2012). Local and universal. In Bianchi & Chesi (Eds.), *Enjoy Linguistics!* CISCL Press.

[2] Grillo, Costa, Fernandes, & Santi. (2015). Highs and Lows in English attachment. *Cognition*, *144*.

[3] Pozniak, Hemforth, Haendler, Santi, & Grillo. (2019). Seeing events vs. entities. *JML*, *107*.

[4] Grillo & Costa. (2014). A novel argument for the Universality of Parsing principles. *Cognition*.

[5] Aguilar, Ferré, Gavilan, Hinojosa, & Demestre. (2021). The actress was on the balcony. *Cognition*, *211*.

[6] Aguilar & Grillo. (2021). Spanish is not different. *Glossa*, *6*.

Relative Clauses and Eventive Small Clauses in Vietnamese

Vietnamese allows for three different types of post-nominal modifiers to express Relative Clauses:

(3) a. cuốn-sách mà mẹ đang đọc
class-book that mom prog read
the book that mom is reading

b. cuốn-sách mẹ đang đọc
class-book mom prog read
the book mom is reading

c. **cuốn**-sách **cuốn** mà mẹ đang đọc
class-book class that mom prog read
the book the one that mom is reading

The first type is signalled by the relativiser *mà* (3-a).[1] Similarly to English, a reduced version can be formed by omitting *mà* (3-b). In some cases, to emphasize/refer back to the NP that the RC intends to modify, Vietnamese speakers construct an extended RC in which the NP's classifier is repeated and precedes the RC (3-c).

We have argued that–in the context of event taking predicates, e.g. perceptual reports– the reduced RC variant (3-b) also allows a Small Clause (SC) parse. This is supported by several diagnostics.

i. **Availability with Proper Names:** Eventive SCs (4-a), but not RCs (4-b), freely take proper names as subjects (lack of comma intonation indicates that (4-a) is not an appositive).

(4) a. Quỳnh thấy Anh đang nhảy.
Quynh saw Anh dancing.

b. *Quỳnh là chị-gái của Anh đang nhảy.
**Quynh is the sister of Anh that danced.*

ii. **Availability with Pronouns:** Pronouns cannot head RCs, but can be SC subjects.

(5) a. Quỳnh thấy cô-ấy đang nhảy.
'*Quynh saw her dancing.*'

b. *Quỳnh là chị-gái của cô-ấy đang nhảy.
**Quynh is the sister of her that danced.*

iii. **Animacy:** Antecedent of animate arguments of Pseudoclefts have to be animate, (6-a). Since SC complements of perceptual verbs denote (inanimate) events, they form pseudoclefts with inanimate antecedents (6-b).

(6) a. Người/*Điều tôi biết là Anh.
*Who/*What I know is Anh.*'

b. Điều tôi thấy là Anh đang nhảy.
'*What I saw was Anh dancing.*

This last test allows us to further support our claim that the attachment ambiguity disappears with SCs. One way to show this is to force a SC reading using pseudoclefts with an inanimate antecedent (7): since the attachment ambiguity is gone, forcing Local Attachment (7-b) results in an implausible interpretation:

(7) a. Việc tôi thấy là người con gái của người đàn ông đang sinh con.
What I saw was the daughter of the man (that was) giving birth.

b. *Việc tôi thấy là người cha của cô gái đang sinh con.
**What I saw was the father of the girl (that was) giving birth.*

[1] Interestingly, the first type Relative Clauses (RCs) in (3-a) shows no attachment ambiguity, i.e. only Local Attachment is licensed. This suggests that, despite appearances, their syntax is very different from that of run-of-the-mill RCs. For this reason the present work focuses on the RC-types in (3-b) and (3-c), both of which allow for both Non-local and Local attachment.

Task adaptation in web-based self-paced reading

Pia Schoknecht[1], Shravan Vasishth[1]

[1] University of Potsdam

pia.schoknecht@uni-potsdam.de

Background: The replication crisis has demonstrated how problematic low-powered studies are for science [1, 2]. In recent years, researchers have become more aware that it is important to run sufficiently powered studies. One way to increase statistical power for a given experiment design is to increase the sample size. Recruiting larger numbers of participants is relatively easy if one uses online platforms. With the increased acceptance of web-based studies, it has become common to collect relatively large amounts of data. However, increasing the number of participants is not the only way to increase power. The amount of data that is collected per participant, i.e., the number of trials, can also be increased. Increasing the number of trials per participant is also important for investigating individual differences, a topic of growing importance in psycholinguistics. Many trials per participant are needed in order to obtain more precise estimates of individual-level effects [3].

Collecting large amounts of data per participant can help to address one's research question, but what are the downsides? It is common to be concerned about waning attention and fatigue during long experiments. The present data suggest that task adaptation, i.e., practice effects, deserve just as much attention.

Method: We used data from 160 native speakers reading German sentences in a standard self-paced reading moving-window format [4]. In this experiment, participants completed two experimental sessions with 100 trials each. We analyzed log-transformed reading times (RTs) of nouns, e.g., *Lehrer* 'teacher', which occurred as the second word of the sentence and were always preceded by a definite article. This word was chosen to keep word position within the sentence constant and to not introduce confounds due to different sentence structures used in [4]. We estimated the effects of session, trial, word length and their interactions on RTs using Bayesian hierarchical mixed effects models. Session was a categorical predictor with dummy contrast coding (1st session 0, 2nd session 1). Trial was a continuous predictor. The trial id was rescaled to span from 0.01 to 1. These modeling choices reflect the belief that the behavior in the early trials of 1st session is unaffected by task adaptation. Word length was a centered and scaled continuous predictor.

Results: An increase in word length led to the well-known increase in RTs (see Figure 1A). Increases in trial (from 1 to 100) as well as session (from 1 to 2) led to remarkable decreases in RTs. Participants started the 2nd session with RTs that were almost as fast as they were at the end of the 1st session. Overall, RTs from the 1st trial in the 1st session to the last trial in the 2nd session decreased by almost 200 ms (see Figure 1C). The effect of word length decreased over the course of an experimental session: A long word at the end of the session was read as fast as a short word at the beginning of the session (see Figure 1B). This was the same in both experimental sessions.

Discussion: RTs decreased dramatically as a function of trial and experimental session. Because decreases in the mean entail decreases in variance [5], a decrease of RTs due to task adaptation might lead to decreases in effect size (see Figure 1B showing that in the present data the difference between long and short words decreased as a function of trial). Task adaptation might be widespread and should not be ignored: adaptation could bias the average estimates of the effect of interest [6, 7].

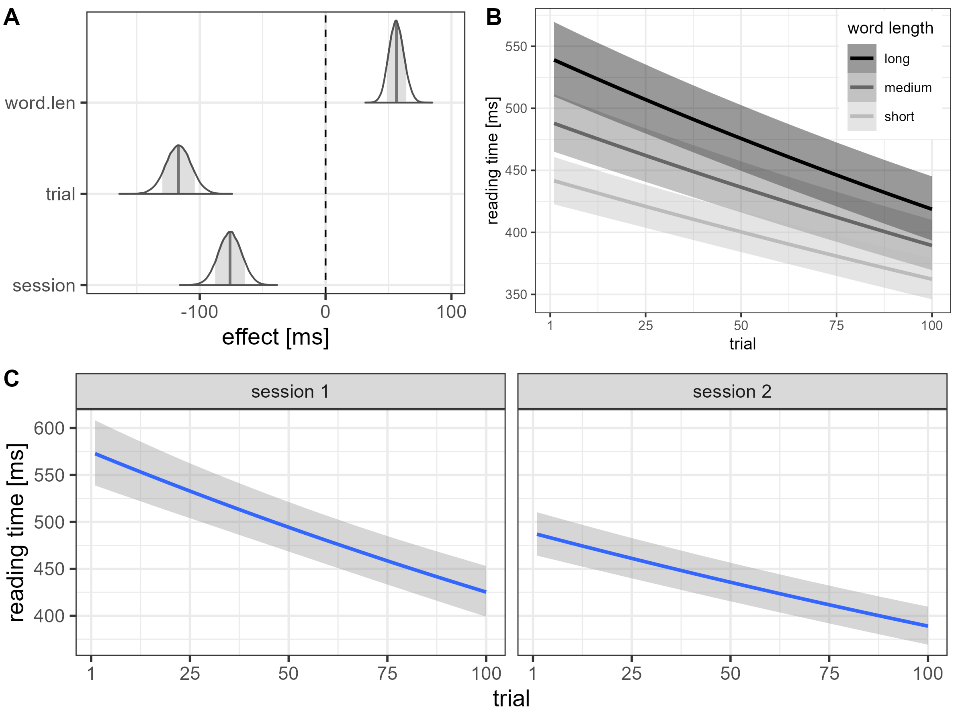


Figure 1: Posteriors of the estimates of the main effects of trial, session and word length (A) and conditional effects (as provided by brms [8]) of the interactions of trial $\times$ word length (B) and trial $\times$ session (C).

References

[1] Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

[2] Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651.

[3] Yadav, H., Paape, D., Smith, G., Dillon, B. W., & Vasishth, S. (2022). Individual differences in cue weighting in sentence comprehension: An evaluation using approximate bayesian computation. *Open Mind*, *6*, 1–24.

[4] Schoknecht, P., Yadav, H., & Vasishth, S. (2025). Do syntactic and semantic similarity lead to interference effects? evidence from self-paced reading and event-related potentials using German. *Journal of Memory and Language*, *141*, 104599.

[5] Wagenmakers, E.-J., Grasman, R. P., & Molenaar, P. C. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*, *49*(3), 195–204.

[6] Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1156–1172.

[7] Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), e77661.

[8] Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Vocabulary Size as Prediction Error: A New Method for Lexical Assessment in Adults

Bálint József Ugrin[1], Péter Rácz[1], Ágnes Lukács[1]
[1] Department of Cognitive Science, Faculty of Natural Sciences,
Budapest University of Technology and Economics
balintugrin@edu.bme.hu

Standardised vocabulary size estimation methods often assume a homogeneous distribution of word exposure across speakers, treating variation primarily as a function of word frequency [1]. However, the rarer a word is, the more likely its distribution is uneven—for instance, highly frequent in specific subpopulations (e.g., technical terms among professionals) but virtually absent elsewhere [2]. This challenges unidimensionality - a critical assumption of ability performance tests [3], since we cannot reliably predict whether knowing a rare word implies the knowledge of a more common word.

To address this problem, we developed a framework to test word knowledge using an error prediction approach. In our framework, we define a baseline performance model representing the expected response (i.e. knowing a word) of the 'average' speaker. The model is constructed using corpus linguistic tools and participant responses allowing us to incorporate the frequencies of words and to capture local interdependence effects: domain-specific words often co-occur, and knowing one increases the likelihood of knowing others within the same semantic field [2]. Individual performance is then compared to this model, and deviation patterns—interpreted as model prediction errors—are calculated. Crucially, knowledge of a low-probability (rare) word yields a larger positive deviation than knowledge of a high-probability (common) word. We assume that higher model prediction errors signal larger vocabulary. We hypothesise that this approach offers an ecologically more valid model of vocabulary, accommodating both probabilistic word knowledge and semantic clustering effects.

To empirically test our model, we developed a novel paradigm to assess word knowledge of Hungarian native speakers. Participants were shown panels of 16 words and asked to select those they recognized. Each panel included nonce words and existing words varying in their frequency, and a very frequent word as a control for attentiveness. The rate of frequent-infrequent words was randomised in each panel, simulating natural lexical variation. The panels and their compositions were also randomised for each participant. Furthermore, we aimed to explore the effects of vocabulary divergence, stimuli were therefore set up to include 500 existing Hungarian words (verbs, nouns, and adjectives): 50 very common words, 150 less common words, and 300 rare and very rare words to cover many areas of the semantic space (e.g., occupational terms, obsolete expressions). Nonce words were generated using phoneme bigrams and selected to match Hungarian phonotactics.

Reanalysis of data collected earlier (507 participants) confirms our assumption that the homogeneity of word distribution decreases as word frequency declines. With our ongoing data collection, we aim to demonstrate the accuracy of our vocabulary estimation approach and explore the relationship between vocabulary size and its lexical composition, offering insights into how domain-specific knowledge influences overall lexical breadth.

References

[1] Hashimoto, B. J. (2021). Is Frequency Enough?: The Frequency Model in Vocabulary Size Testing. *Language Assessment Quarterly, 18(2), 171–187.* doi:10.1080/15434303.2020.1860058

[2] Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond Word Frequency: Bursts, Lulls, and Scaling in the Temporal Distributions of Words. *PLoS ONE*, 4(11), e7678. doi:10.1371/journal.pone.0007678

[3] Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning. *Topics in Cognitive Science*, 6(1), 5–42. doi:10.1111/tops.12078

Effects of Surprisal and Contextual Entropy on L2 and Heritage Language Processing

Hailin Hao, Zuzanna Fuchs
University of Southern California
hailinha@usc.edu

Computational theories of language processing like Surprisal Theory [1] provide testable predictions for broad-coverage evaluations using reading time (RT) corpora [2-3]. However, most prior research has focused on monolingual adults (but see [4]), limiting the generalizability across diverse linguistic populations. This study evaluates how surprisal [1] and contextual entropy [5-6] impact RTs in L2 and heritage language (HL) speakers, to shed light on integrative and anticipatory processing mechanisms in non-dominant languages. Importanly, L2 speakers differ from L1 speakers in early exposure, while HL speakers typically receive less written input.

Predictions Surprisal Theory [1] suggests that processing difficulty correlates with a word's surprisal—its contextual probability given prior context (Eq. 1). Others have proposed that contextual entropy, which measures the degree of uncertainty about how the context will continue (Eq. 2), may also influence RTs, if a speaker estimates processing time for a word before encountering it [5-6]. Unlike surprisal, which might reflect integration difficulty, entropy effects provide evidence for anticipatory processing [6-7]. If L2 and HL speakers use incremental contextual cues for prediction, both surprisal and entropy should affect L2 RTs. However, prior research shows L2 speakers are less effective at using linguistic context for prediction than L1 speakers [8-9]. For HL speakers, despite target-like performance in non-reading tasks [10], limited written input [11] may reduce sensitivity to surprisal and entropy during reading.

Analysis 1: Using the MECO-L2 corpus [12], we analyzed eye movements of L1 and L2 English speakers with diverse L1 background while reading English. Linear mixed-effects models were fitted on three log-transformed eye-tracking measures—first fixation (FF), gaze duration (GD), and total reading time (TT), representing early, medium, and late processing stages, respectively. Predictors included surprisal and entropy (estimated from GPT-2 small [13]), word-level controls (e.g., length, frequency), and spillover (formula in Eq. 3). **Results:** Fig 1 plots coefficients for surprisal and entropy (with 95% confidence intervals) for each group. Surprisal significantly predicted FF in 8 groups and GD/TT across all groups. Entropy was significant for 2 groups in FF and 9 groups in GD/TT. Interestingly, entropy effects were not significant in L1 English in FF/GD and marginally significant in TT. A combined model including speaker group (L1 vs. L2) as a predictor and its interactions revealed weaker surprisal effects and stronger frequency effects in L2 speakers compared to L1.

Analysis 2: Using a Russian eye movement corpus [14], we analyzed L2 and low-proficiency HL speakers of Russian whose dominant language is English. The analysis is identical to Analysis 1, with surprisal and entropy estimated from mGPT [15]. **Results (Fig. 2):** For HL speakers, higher surprisal and entropy generally leads to longer procesisng time, with surprisal marginally significant in TT and entropy significant in GD and TT. However, neither surprisal nor entropy had an effect in the L2 group. Models with speaker group as a predictor showed that L2 speakers relied less on surprisal and more on frequency compared to HL speakers.

Discussion Overall, we found strong evidence of surprisal and entropy effects on L2 English processing across diverse L1 backgrounds, demonstrating their generalizability. However, our analyses also revealed that L2 speakers differ from L1 speakers in that they rely more on word-level than contextual cues. Notably, we are also the first to show surprisal and entropy effects in HL processing. Despite limited written input, HL readers are influenced by these factors and utilize contextual cues more effectively than L2 Russian speakers.

Eq. 1 $\text{surprisal}(w) = -\log P(w|c)$

Eq. 2 $\text{entropy}(w_t) = -\sum_{w\in V} P(w|c)\cdot\log P(w|c)$

Eq. 3 log(FF/GD/TT) ~ surprisal + entropy + logfreq + prev.surprisal + prev.logfreq + prev.len.scaled + prev.entropy + (1|Subject) + (1|Word)

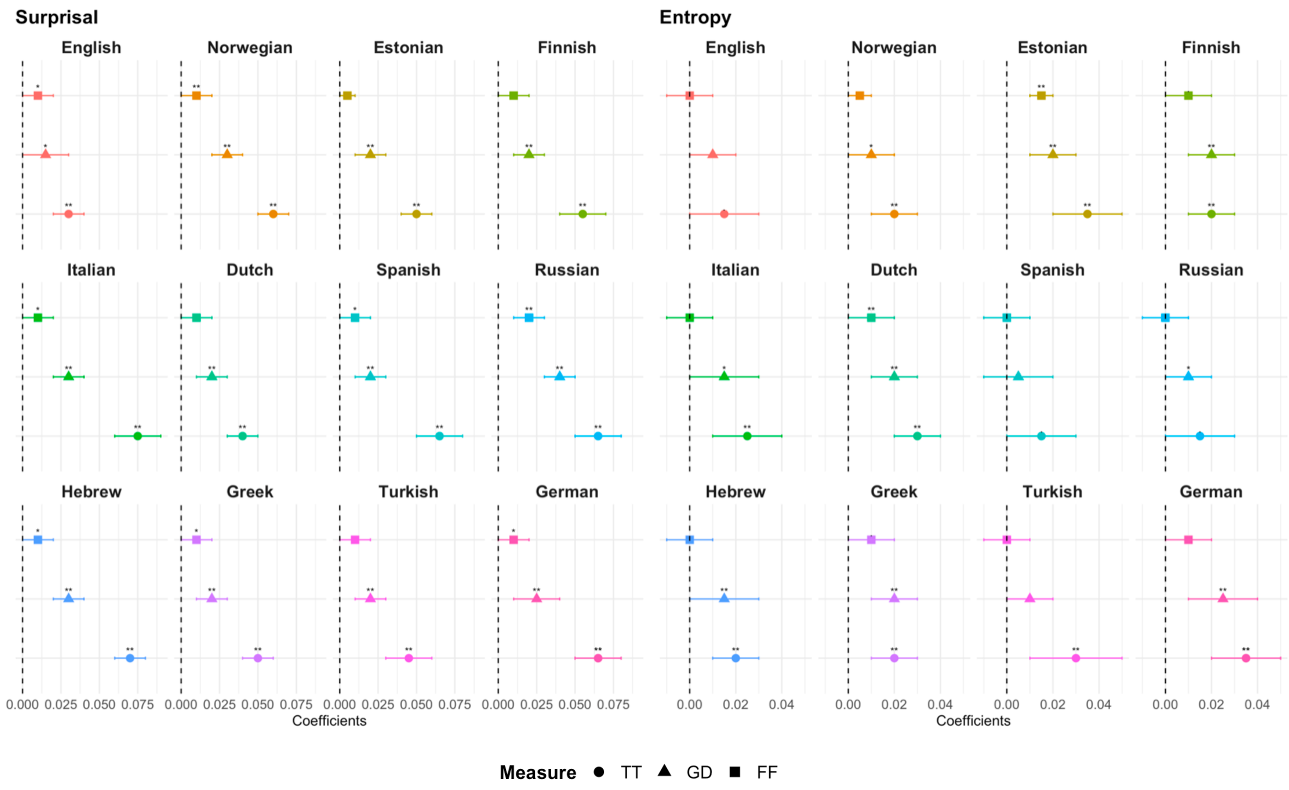


Fig. 1. Effects of surprisal and contextual entropy across speaker groups in MECO

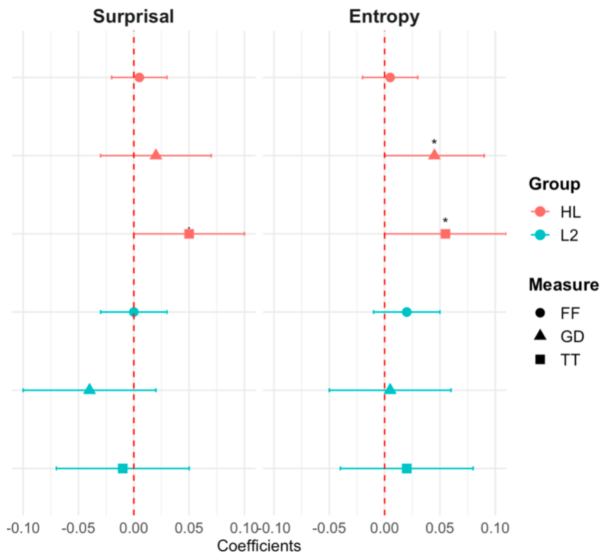


Fig. 2. Effects of surprisal and contextual entropy in L2 and HL Russian

References: [1] Levy (2008) Cognition. [2] Smith & Levy (2013) Cognition. [3] Shain et al. (2024) PNAS. [4] Varda & Marelli (2022) ACL. [5] Hale (2006) Cogn. Sci. [6] Pimentel et al. (2023) TACL. [7] Pickering & Gambi (2018) Psychol. Bull. [8] Grüter, Lew-Williams, & Fernald (2012) Second Lang. Res. [9] Grüter & Rohde (2021) Appl. Psycholinguist. [10] Fuchs (2021) Linguist. Approaches Biling. [11] Rothman (2007) Int. J. Biling. [12] Kuperman et al. (2022) Stud. Second Lang. Acquis. [13] Radford et al. (2019) OpenAI. [14] Parshina, Laurinavichyute, & Sekerina (2021) Biling.: Lang. Cogn. [15] Shliazhko et al. (2022) *arXiv.*

Sentence imitation and its relation to working memory and language skills

Klára Matiasovitsová, Filip Smolík

Institute of Psychology, The Czech Academy of Sciences

matiasovitsova@psu.cas.cz

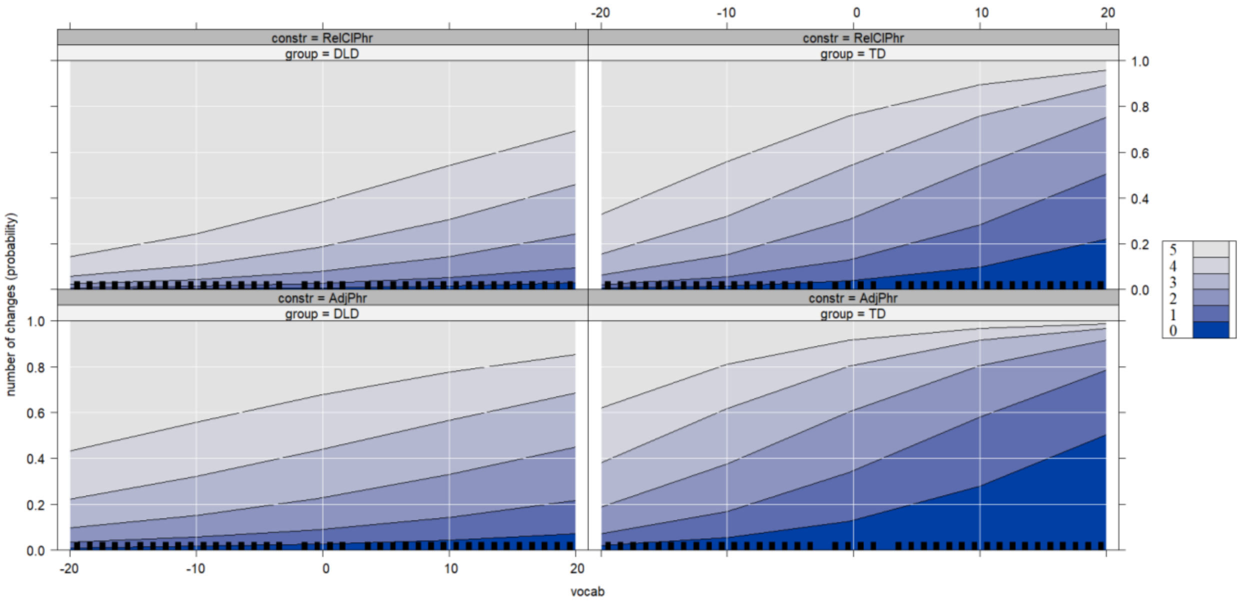
Sentence imitation (SI) is widely used in both research and the assessment of developmental language disorder (DLD; [1, 2]). As the task requires the accurate repetition of a heard sentence, it is thought to reflect the functioning of working memory – the phonological loop [1, 2, 3] and the central executive [2, 4] - as well as language proficiency [3, 5]. However, the exact contributions of language skills and working memory to SI performance remain not completely understood. Previous studies reported increased demands on central executive functioning in morphosyntactically complex sentences [1, 6]. The roles of memory and language knowledge may also differ between children with and without DLD [1, 2]. This study examines the effects of language proficiency and memory variables on the imitation of sentences of varying complexity in Czech-speaking children with DLD and typically developing (TD) peers. There is limited research on SI performance in morphologically rich languages.

Sixty-three children with DLD (6;5-9;6) were matched on gender and vocabulary with 63 TD children (3;7-6;7). In addition to SI (scored by number of errors) and receptive vocabulary, they completed tasks measuring phonological memory and central executive function (nonword repetition, listening span). The target structures were relative clauses (RCs) and simple sentences with adjectival nominal phrases. These structures share some morphological properties but differ in syntactic complexity.

The effects of sentence type, DLD status, and vocabulary/memory measures, and their interactions, on the number of errors in SI were analyzed using cumulative link mixed models. A significant interaction between vocabulary and group ($z = 2.03$, $p = .04$), and a marginally significant interaction between vocabulary, group and construction ($z = -1.82$, $p = .07$), indicated that the relation between vocabulary and SI performance was weaker in children with DLD than in TD children, but may be stronger for RCs than for simple sentences in children with DLD (see Figure). Models with independent variables of group, construction and nonword repetition/listening span showed that children with higher scores in nonword repetition and listening span made fewer errors in SI ($z = -3.07$, $p = .002$ and $z = -5.98$, $p < .001$, respectively), although the effect of listening span was weaker in children with DLD ($z = 1.88$, $p = 0.06$).

The results indicate that SI is a valid marker of language skills. However, the effects of language knowledge may differ in children with DLD. SI also reflects memory functioning, but no differences were found in memory effects across sentence types. A plausible explanation is that the morphological richness of inflected languages increases overall demands on the central executive, even in syntactically simpler sentences. Furthermore, the central executive may function differently in children with DLD, possibly due to difficulties in efficiently engaging it during sentence repetition or insufficient entrenchment of syntactic representations in long-term memory.

Figure. Relations between sentence imitation of relative clauses and simple sentences, vocabulary and group. The darkest and lightest band show the probability of zero-error or five-plus error responses, respectively. Increases in the dark blue bands show the increase of probability for imitations with lower number of errors.



References

[1] Frizelle, P., & Fletcher, P. (2015). The Role of Memory in Processing Relative Clauses in Children With Specific Language Impairment. American Journal of Speech-Language Pathology, 24(1), 47–59.

[2] Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. International Journal of Language and Communication Disorders, 47(5), 499–510.

[3] Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). Journal of Child Psychology and Psychiatry and Allied Disciplines, 42(6), 741–748.

[4] Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children’s working memory capacity and processing speed on their sentence imitation performance. International Journal of Language and Communication Disorders, 48(3), 329–342.

[5] Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of cantonese. Journal of Speech, Language, and Hearing Research, 49(2), 219–236.

[6] Delage, H., & Frauenfelder, U. H. (2020). Relationship between working memory and complex syntax in children with Developmental Language Disorder. Journal of Child Language, 47(3), 600–632.

Slower reading on interlingual homographs can be a surprisal effect

Iza Škrjanec[1], Irene Winther[2], Merit Huisman[2], Vera Demberg[1], Sybrine Bultena[2], Stefan L. Frank[2]
[1]Saarland University; [2]Radboud University
skrjanec@lst.uni-saarland.de

Background: Words from different languages can have identical orthography. *Cognates* share both form and meaning between languages (e.g., *winter* in English and Dutch) whereas *interlingual homographs* (IHs) share form but not meaning (e.g., *stem*, meaning ‘voice’ in Dutch). Behavioral studies find that cognates facilitate processing in bilingual speakers [1, 2] compared to non-cognate control words. In contrast, IHs in sentence context result in inhibition that surfaces as longer reading times [3, 4].

One account of the observed cognate facilitation effect is cumulative frequency: bilinguals encounter cognate words more often than non-cognates (and more than monolinguals do) due to the identical form, leading to a facilitated lexical access. Using computational language models trained on Dutch and English, [5] show that cumulative frequency indeed drives the cognate facilitation effect: cognates have lower surprisal than non-cognates in bilingual models but not in monolingual models. Following this, a similar effect of facilitation (and not inhibition) should be found for IHs, as they occur in both languages as opposed to non-overlapping control words. We investigate to what extent IH effects can nevertheless be reduced to surprisal.

Method: We estimate surprisal from a monolingual Dutch as well as bilingual Dutch-English LSTM language model trained on Wikipedia to predict the next word. We use the models by [5] which account for cognate facilitation effects. We estimate surprisal of IHs (e.g., *stem*) and their control words (e.g., *kerk*). A total of 32 word pairs are embedded in Dutch sentences with low semantic constraint (e.g. *Anna rilde toen ze een **stem/kerk** opmerkte*; ‘Anna trembled when she noticed a **voice/church**’). We use linear mixed-effects regression to test the effects of IH word status, log frequency and model type (mono- or bilingual) on word surprisal (and thus estimated processing effort).

Results: Our results reveal an interaction of IH status and model type (Fig. 1 and Table 1). In bilingual models, IH words have higher surprisal than control words. This mirrors the inhibition effect found in RTs of interlingual homographs in bilinguals. In monolingual models, it is the other way around: control words have a higher surprisal than IHs. In the monolingual training data, control words are less frequent, so their surprisal is proportionally higher.

Discussion: Our results are not in line with the cumulative frequency account, i.e., more frequent encounters do not necessarily lead to lower surprisal. Surprisals of IHs are not similar to those of cognates in [5]. Rather, our results match those from human readers [3] in that bilingual models assign higher surprisal to IHs than to controls.

We speculate that the linguistic contexts of IH words are more diverse than those of control words as they appear in two languages, which causes the bilingual language model’s predictions of IHs to be less accurate (i.e., higher surprisal). The same does not need to hold for cognates because they share their part-of-speech and meaning between languages. We are currently working on the discrepancy in the frequency estimates as well as extending this analysis to the French-English items of [3].

Predictor	β	*SE*	t	p
(Intercept)	4.78	0.97	4.93	<.001
IH status	-0.07	0.08	-0.82	>.05
Model type	0.1	0.08	1.2	>.05
Frequency	-1.81	0.15	-11.83	<.001
IH status : Model type	0.44	0.09	5.23	<.001

Table 1: Regression coefficients and test statistics. Regression model: $surprisal \sim ih\_status * model\_type + frequency + (1|item)$. Contrast coding: IH status (-1 control, 1 IH) and model type (-1 mono-, 1 bilingual).

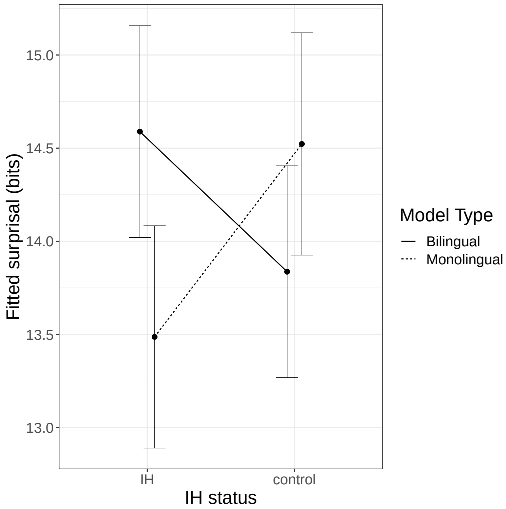


Figure 1: Predicted surprisal estimates of interlingual homographs (IH) and their control words in mono- and bilingual language models. The error bars indicate the standard error.

References

[1] Dijkstra, T., Grainger, J., & van Heuven, W. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518.

[2] Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*, 1214–1241.

[3] Libben, M., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–90.

[4] Pivneva, I., Mercier, J., & Titone, D. A. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 787–796.

[5] Winther, I. E., Matusevych, Y., & Pickering, M. J. (2021). Cumulative frequency can explain cognate facilitation in language models. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.

Local context in quantifier scope ambiguity resolution in Swedish

David Pagmar, Asad Sayeed
Department of Philosophy, Linguistics, and the Theory of Science, University of Gothenburg
david.pagmar@gu.se

Background: This work aims at clarifying an aspect of the syntax-semantics interface in an interactive context: the resolution of quantifier scope ambiguities (QSA). Despite a long history of research, there are important aspects of this phenomenon that have not been well-explored experimentally, especially as the work has largely been performed in English (e.g. [1], although exceptions exist [2].) In particular, there is still considerable room to explore competitive effects between the syntactic and morphological factors that signal such ambiguities and the larger pragmatic context: when and how does the wider context of an utterance override the sentence-internal syntax and morphology that promotes the ambiguity? The sentence ”every road leads to a town” is ambiguous. It can be interpreted in at least two different ways: either the roads lead to one and the same town (singular), or they lead to different towns (plural). Potential ambiguities are relatively common in every day language use [3]. Language users can easily recognise QSA and even though there are more than one way of understanding a given QSA sentence, there will often be a preferred interpretation. Can individual preferences for specific QSA sentences be controlled by different contexts? If so, to what degree? In Swedish, the indefinite article (”en/ett”) has two genders, utrum and neutrum. In one of these genders, neutrum, the indefinite article (”ett”) is also the default word for counting (”ett, två, tre, …”). Our hypothesis is that the neutrum article will establish a bias towards singular readings, due to ”ett” having a stronger assosiation to the numeral one.
Method: We ran a forced choice judgment task via the online platform Pavlovia. Participants were presented with written prompts and questions, with directions for how to reply through their keyboard. Each experimental run consisted of 30 trials (20 critical and 10 distractors). The content of each trial revolved around a group of tourists. Each trial involved a context sentence followed by a critical sentence with QSA over either an utrum or neutrum indefinite article. After each QSA sentence, the participant was asked make a choice, revealing a singular or plural interpretation of the QSA. A total of 28 participants took part in the experiment. All had Swedish as their first language and were above the age of 18. Examples of the task can be seen in figure 1.
Results: We conducted a multilevel logistic regression analysis, with random intercepts for participants, using the *glmer* function of the *lme4* package (version 1.1-35.3) in R. The analysis follows the equation: $\text{logit}(P(y_{ij} = 1)) = \gamma_{00} + \beta_1 \text{SIN}_{ij} + \beta_2 \text{UTR}_{ij} + u_{0j}$. In the analysis, every trial is analyzed as an individual observation (N=560). The binary dependent variable is the QSA reading (plural = 0, singular = 1) and the predictors are the two conditions contextual sentences and grammatical gender. The multilevel logistic regression analysis showed an effect of context condition, but not of grammatical gender (table 1) with odds ratios in figure 2.
Discussion: Contextual alterations influenced the interpretations of the QSA sentences to a very high degree. However, no context shifted 100% of interpretation to either plural or singular, showing that there is still room for different interpretations of each sentence. The grammatical gender conditions did not show any reliable effect. Our hypothesis stated that neutrum gender would prompt a bias towards singular interpretations. The results go in the *opposite* direction (figure 3). While other readings that go in the opposite direction of the contextual information end up at around 20%, singular readings under the plural-neutrum conditions are at 10%. A potential explanation connects to [4], where demonstrative determiners were used as an experimental control (e.g., *every child climbed* **that** *tree*). Reading time slowed down for singular demonstratives rather than the indefinite article. A similar effect could be at work here, in that counting interpretation (”one single tree”) of ”ett” introduces a discourse referent and additional cognitive load. This could be further investigated with a larger sample size.

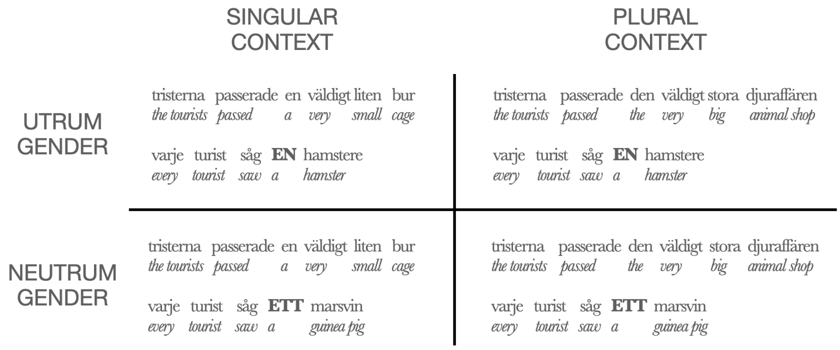


Figure 1: An example of how contextual prompts and grammatical genders were combined. If a participant were exposed to utrum gender in the singular context, they would be exposed to the neutrum gender in the plural context. In this way, no participant were exposed to the same context or grammatical gender twice.

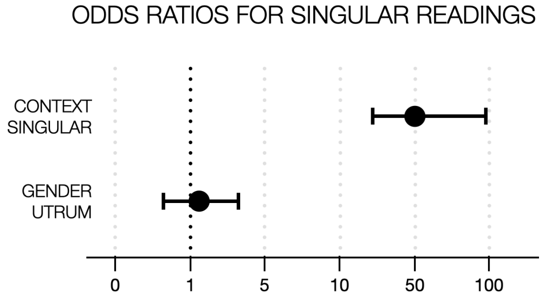


Figure 2: The Odds Ratios result for singular readings of QSA sentences show an sizable effect of contextual sentences, OR = 50.20, 95%CI[24.32, 103.65]. The results show no reliable effect for grammatical gender, OR = 1.26, 95%CI[.69, 2.31].

	Est.	St.Er	z	*p*
Intercept	-2.19	.36	-6.05	<.001
contextSIN	3.91	.36	10.58	<.001
genderUTR	.23	.30	.76	.43

Table 1: Logistic regression analysis.

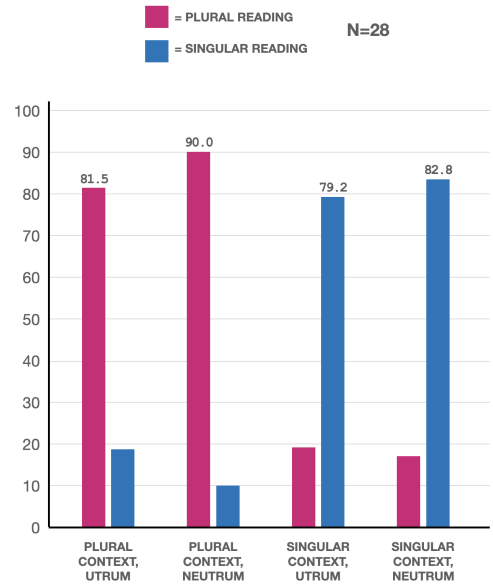


Figure 3: Percentages of QSA interpretations for each combination of conditions. Contextual condition provides substantial effects on QSA interpretations.

References

[1] Paterson, K. B., Filik, R., & Liversedge, S. P. (2008). Competition during the processing of quantifier scope ambiguities: Evidence from eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *61*(3), 459–473.

[2] Sayeed, A., Lindemann, M., & Demberg, V. (2019). Verb-second effect on quantifier scope interpretation. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 134–139.

[3] Koller, A., Thater, S., & Pinkal, M. (2010). Scope underspecification with tree descriptions: Theory and practice. *Resource-Adaptive Cognitive Processes*, 337–364.

[4] Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PloS one*, *8*(11), e81461.

Untangling musical and linguistic processing using low-resolution EEG

Lily Arrom, Samantha Wray
Dartmouth College
samantha.c.wray@dartmouth.edu

Background: Music and language are universal, with no known human communities lacking either. Cognition of both stems from a shared foundation of pattern recognition processes and utilizes processing of many of the same acoustic features, such as pitch, prosody, timing, but research also shows resources for processing them are ultimately separable, with distinct neural responses for unexpected or syntactically invalid items [1]. However, like so much of psycholinguistic research, existing research can be limited in its scope to WEIRD societies [2], missing structurally more complex musical systems [i.e 3] or rare practices such musical speech surrogacy, which encodes language into the acoustic features of a musical instrument [4]. As a first step to closing this gap, the current project aims to validate usage of a high-portability, low-resolution, commercial-grade EEG system well-suited to field sites that allow the study of both language and music outside of lab settings to expand the scope of current research.

Method: N=30 right-handed participants (age M=20.7 SD=1.9) read sentences presented in rapid serial visual presentation while a tonal sequence played simultaneously. The sentences varied across three conditions: grammatical, semantically implausible, and syntactically ungrammatical. Tonal sequences varied across four: regular tonic sequences and music-syntactically irregular final notes (Neapolitan), with both having the possibility of an instrument change partway through (stimuli structure largely following [1]). Two tasks ensured participant attention: an occasional grammaticality judgement, and monitoring for the instrument change. The experiment was implemented using PsychoPy [5] while EEG data were recorded on a 14 sensor EEG system (Emotiv Epoc X). EEG data were processed using EEGlab [6], with data bandpass filtered from 0.1 to 30 Hz, and re-referenced to the common average. Eye-blink and other motion artefacts were removed using visual inspection. Data were epoched from −100 ms prior to the onset of the last tone/word and to 700 ms post onset.

Results: An ERP analysis revealed: (a) N400 for semantically implausible items (b) left posterior negativity (LPN) ~400ms for syntactically ungrammatical items (c) ELAN and N500 for musically irregular sequences (d) attenuated LPN for simultaneous musically irregular sequences with syntactically ungrammatical items and (e) attenuated N500 for simultaneous musically irregular sequences with semantically implausible items. These effects are largely consistent with but do diverge from attested effects in [1] in the following ways: [1] showed a left anterior negativity (LAN) was elicited by syntactically ungrammatical items and an ERAN was elicited by musically irregular sequences.

Discussion: While previous work has established feasibility of high-portability EEG systems for coarse-grained sensory processing [7], the current work expands this into finer-grain domains and demonstrates their feasibility as well, replicating work showing shared but separable neural resources for the processing of music and language.

Figures

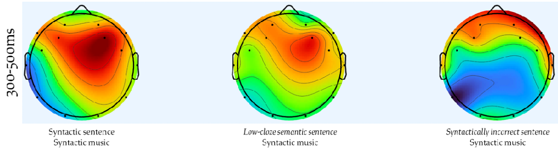


Figure 1: Topographic maps showing activity associated with grammatical, semantically incongruent/ low-cloze (exhibiting N400), and syntactically incongruent (exhibiting left posterior negativity) sentences.

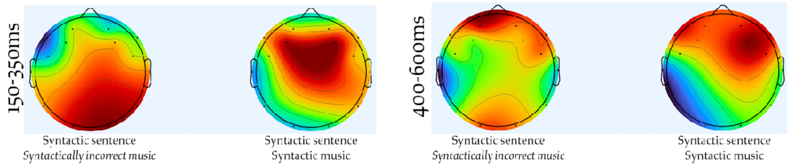


Figure 2: Topographic maps showing ELAN (left) and N500 (right) elicited by irregular music sequences

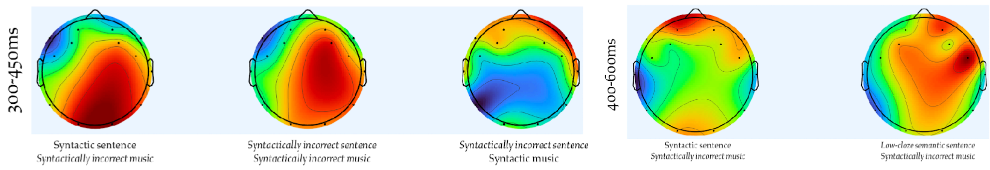


Figure 3: Topographic maps showing shared neural resources for music and language, with attenuated LPN (left) and N500 (right) elicited by simultaneous syntactic and semantic incongruencies, respectively

References

[1] Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169-1178.

[2] Bylund, Emanuel. "10 WEIRD Psycholinguistics". *Struggles for Multilingualism and Linguistic Citizenship*, Bristol, Blue Ridge Summit: Multilingual Matters, 2022, pp. 183-200.

[3] Sanyal, S., Banerjee, A., Basu, M., Nag, S., Ghosh, D., & Karmakar, S. (2020, December). Do musical notes correlate with emotions? A neuro-acoustical study with Indian classical music. In Proceedings of meetings on acoustics (Vol. 42, No. 1). AIP Publishing.

[4] McPherson, L., & Winter, Y. (2022). Surrogate Languages and the Grammar of Language-Based Music. *Frontiers in Communication*, *7*, 838286.

[5] Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162(1-2), 8-13.

[6] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

[7] Williams, N. S., McArthur, G. M., de Wit, B., Ibrahim, G., & Badcock, N. A. (2020). A validation of Emotiv EPOC Flex saline for EEG and ERP research. *PeerJ*, *8*, e9713.

Reading Tiramisu in Czech and English:
Robust Processing Speed Differences in Translation Equivalent Stimuli

Jan Chromý[1], Markéta Ceháková[1], Michael Ramscar[2]
[1] Faculty of Arts, Charles University; [2] University of Tübingen
jan.chromy@ff.cuni.cz

Background: In contemporary sentence processing research, there is increasing interest in crosslinguistic differences [1]. However, studies that compare sentence comprehension across languages using matched translation-equivalent stimuli remain relatively rare. A recent investigation [2] uncovered notable differences in agreement attraction effects between Czech and English. Interestingly, a closer analysis revealed another pattern: native English speakers processed individual words in their language significantly faster than Czech speakers did in theirs. The present study seeks to explore this phenomenon in greater depth.

Method: We created 70 pairs of translation-equivalent sentences (Table 1), matched in length in words between Czech and English. These sentences included words that have identical graphical forms in both languages (e.g., tiramisu or Robert). While some of these words were declinable in Czech (e.g., judo, router, Adam), meaning they exhibited distinctive inflectional paradigms with varying endings for different cases (e.g., genitive forms *juda*, *routeru*, *Adama*), others were indeclinable (e.g., *fantasy*, *origami*, *blues*), retaining the same form regardless of case. Native speakers of Czech (N=176) and English (N=176) read these sentences, along with 48 fillers, in a self-paced reading task using a moving-window presentation. After each sentence, participants answered a yes-no comprehension question. Reaction times (RTs) for individual words served as the dependent variable, while the independent variables were language, word length (in syllables), and declinability in Czech.

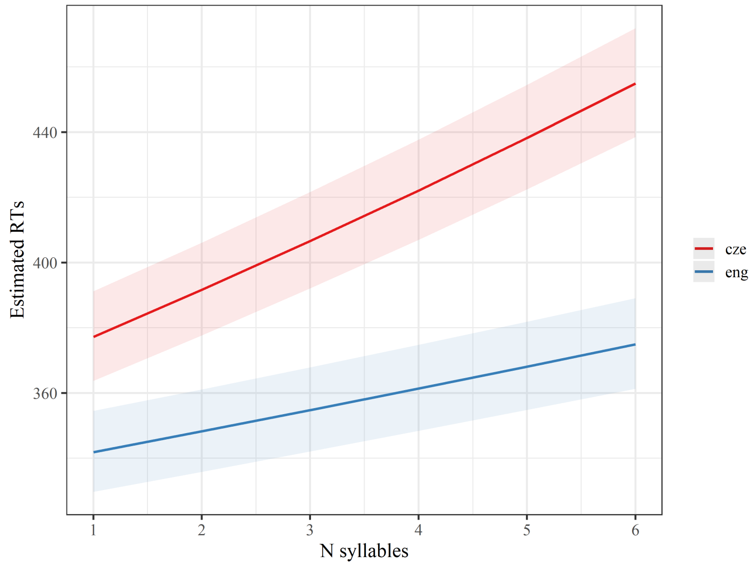
Results: To assess the comparability of the two samples, we examined response accuracy on the comprehension questions. The Czech sample showed an average accuracy of 96.74% (SD = 3.18), while the English sample averaged 95.66% (SD = 3.82). This difference was not significant in the logit mixed-effects model. We then ran a linear mixed-effects model with reaction times as a dependent variable, language and word length as fixed effects, and participant and item as random effects, revealed significant effects for both factors and their interaction (Figure 1). Specifically, (i) English words were processed faster than Czech words, (ii) longer words took more time to process, and (iii) the effect of word length was more pronounced for Czech.

Discussion: The results show robust differences in RTs between English and Czech. We argue that these differences are driven by the greater morphological complexity (i.e. efficiency) of Czech compared to English [3], which increases the cognitive load (i.e., prior) on Czech speakers even when processing identical words.

Table 1: Item example in Czech and English.

Moje	teta	obvykle	pije	Bordeaux	a	další	francouzská	vína.
My	aunt	usually	drinks	Bordeaux	and	other	French	wines.

Figure 1: Estimated RTs from the linear mixed-effects model targeting the effects of language and word length.



References

[1] Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., et al. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, 1–21. https://doi.org/10.3758/s13428-021-01772-6

[2] Chromý, J., Brand, J., Laurinavichyute, A., & Lacina, R. (2023). Number agreement attraction in Czech and English comprehension: A direct experimental comparison. *Glossa Psycholinguistics*, *2*. https://doi.org/10.5070/G6011235

[3] Sadeniemi, M., Kettunen, K., Lindh-Knuutila, T., & Honkela, T. (2008). Complexity of European Union Languages: A comparative approach□. *Journal of Quantitative Linguistics*, *15*(2), 185–211. https://doi.org/10.1080/09296170801961843

Friday Afternoon Posters

Sleep Patterns and Language Acquisition in Cantonese-Speaking Preschoolers: Preliminary Evidence for the Role of Sleep Regularity

Lu Li[1], Jiayi Lu[1], Jueyao Lin[1], Changsheng Li[1], Zhengqin Liu[1], Cehao Yu[1], Caicai Zhang[1]
[1] The Hong Kong Polytechnic University
lu-lucia.li@connect.polyu.hk

Background: According to the Active Systems Consolidation Theory and the Vigilance Hypothesis [1, 2], sleep contributes to both offline memory consolidation and the maintenance of cognitive alertness for online learning—factors essential for language acquisition in early childhood. During the preschool years—a period marked by rapid language development and shifting sleep patterns—most children are still transitioning from naps, making both daytime and nighttime sleep crucial components of their overall sleep structure [3]. While previous studies have largely linked preschoolers' vocabulary to sleep duration and quality, findings remain inconsistent [4, 5]. Beyond these traditional aspects, recent studies have identified positive associations between younger children's vocabulary and two additional sleep dimensions: regularity (the consistency of sleep across timepoints) and maturation (the developmental stages of day/night sleep) [6, 7]. By integrating both traditional sleep aspects (duration, quality) and underexplored dimensions (regularity, maturation), while also assessing broader language skills beyond vocabulary, this study aims to identify key sleep predictors and provide a more comprehensive understanding of how sleep is related to early language development.

Method: Twenty-two typically developing Cantonese-speaking children (4.49 ± 0.25 years; 10 males) in Hong Kong participated by wearing a GENEActiv actigraph for seven days, while parents completed daily sleep diaries. Sleep metrics were calculated using data from the diaries and the actigraph via the van Hees algorithm (GGIR v3.2-0 in R), and categorized into four sleep constructs. Language abilities were assessed using the Hong Kong Test of Preschool Oral Language (TOPOL), yielding an overall standard language score and composite scaled scores. Pearson correlations were computed between sleep metrics and language scores. To further evaluate the relative contribution of significant sleep metrics (Table 1) to total language scores, multiple linear regression was performed, adjusting scaled predictors for best model fit and controlling for gender, age, and scaled Pictorial Test of Nonverbal Intelligence (PTONI) scores (Table 2).

Results & Discussion: Table 1 reveals that four sleep metrics are significantly correlated with total language scores. Total sleep duration (TSD; $r = 0.48$, $p < .05$) and sleep regularity index (SRI; $r = 0.66$, $p < .01$) show positive correlations, suggesting that longer and more consistent all-day sleep patterns are associated with better overall language abilities. In contrast, sleep onset latency (SOL; $r = -0.46$, $p < .05$) and social jet lag (SJL; $r = -0.58$, $p < .01$) exhibit negative correlations, indicating that difficulties in falling asleep at night and weekday–weekend nighttime mid-sleep timing discrepancies are linked to lower overall language abilities. Significant correlations are also found in language subdomains, with SRI showing the widest associations across all subdomains. Moreover, Table 2 shows that SRI has the largest and statistically significant association with overall language outcomes ($\beta = 10.85$, $p < .05$), suggesting that its sensitivity to both daytime and nighttime sleep-wake patterns, as well as daily variability, potentially serves as a more nuanced and developmentally meaningful measure of sleep compared to other metrics.

Conclusion: Preliminary findings indicate that sleep regularity, represented by SRI, may strongly predict preschoolers' language outcomes, highlighting consistent full-day sleep patterns favorable for language development, aligning with theories on sleep's role in both offline consolidation and online language learning.

Table 1. Pearson Correlations Between Sleep Metrics and TOPOL Language Composites.

Sleep Constructs	Sleep Metrics	TOPOL Total Standard	Vocabulary (1)[a]	Grammar (2+3)[a]	Narrative (4)[a]	Production (1+3)[a]	Comprehension (2+4)[a]
Sleep Duration	Total Sleep Duration (N=21)	**0.48***	0.3	**0.57****	**0.50***	0.42	**0.50***
	Nighttime Sleep Duration (N=22)	0.39	0.27	0.41	0.42	0.29	**0.47***
Sleep Quality	Sleep Onset Latency (N=21)	**-0.46***	-0.27	**-0.46***	**-0.55****	-0.37	**-0.50***
Sleep Regularity	Intraindividual Variability in Wake Timing (N=22)	-0.39	-0.17	-0.41	**-0.56****	-0.29	**-0.47***
	Social Jet Lag (N=19)	**-0.58****	**-0.48***	-0.45	**-0.58****	**-0.47***	**-0.63****
	Sleep Regularity Index (N=19)	**0.66****	**0.50***	**0.59****	**0.64****	**0.62****	**0.62****
Maturation of Sleep Patterns	Day/Night Sleep Ratio (N=21)	-0.04	-0.01	-0.03	-0.07	0.05	-0.12
	Napping Frequency (N=21)	0.03	-0.03	0.05	0.1	0.06	0.01

Note. Values represent Pearson correlation coefficients. *p < .05. **p<.01. Sample sizes vary across sleep metrics due to partially missing data required for certain calculations.

[a]1: Expressive Vocabulary; 2: Receptive Grammar; 3: Expressive Grammar; 4: Story Comprehension

Table 2. Multiple Linear Regression Analysis of Sleep Metrics Predicting TOPOL Total Standard Scores.

Variable	Coefficient (β)	Std. Error	t-value	p-value	95% CI (Lower)	95% CI (Upper)	Adjusted R²	F	Prob > F
Intercept	84.71	61.52	1.38	0.1986	-35.55	205.99			
GenderM	-3.61	5.84	-0.62	0.5505	-15.06	7.84			
Age in Months	0.67	1.13	0.59	0.5671	-1.54	2.87			
PTONI	-0.37	3.12	-0.12	0.9074	-6.49	5.75	0.48	3.44 (df = 6; 10)	0.0414*
Total Sleep Duration	3.88	3.07	1.27	0.2342	-2.12	9.88			
Sleep Onset Latency	2.94	4.12	0.71	0.4923	-5.13	10.99			
Sleep Regularity Index	10.85	3.77	2.88	0.0164*	3.46	18.23			

References

[1] Hoedlmoser, K., Peigneux, P., & Rauchs, G. (2022). Recent advances in memory consolidation and information processing during sleep. Journal of sleep research, 31(4), e13607.

[2] Vriend, J., Davidson, F., Rusak, B., & Corkum, P. (2015). Emotional and cognitive impact of sleep restriction in children. Sleep medicine clinics, 10(2), 107-115.

[3] Spencer, R. M., & Riggins, T. (2022). Contributions of memory and brain development to the bioregulation of naps and nap transitions in early childhood. Proceedings of the National Academy of Sciences, 119(44), e2123415119.

[4] Lam, J. C., Mahone, E. M., Mason, T., & Scharf, S. M. (2011). The effects of napping on cognitive function in preschoolers. Journal of Developmental & Behavioral Pediatrics, 32(2), 90-97.

[5] Vaughn, B. E., Elmore-Staton, L., Shin, N., & El-Sheikh, M. (2015). Sleep as a support for social competence, peer relations, and cognitive functioning in preschool children. Behavioral Sleep Medicine, 13(2), 92-106.

[6] Schlieber, M., & Han, J. (2018). The sleeping patterns of Head Start children and the influence on developmental outcomes. Child Care, Health and Development, 44(3), 462–469.

[7] Gliga, T., Hendry, A., Kong, S. P., Ewing, B., Davies, C., McGillion, M., & Gonzalez‐Gomez, N. (2023). More frequent naps are associated with lower cognitive development in a cohort of 8–38‐month‐old children, during the Covid‐19 pandemic. JCPP advances, 3(4), e12190.

The (Non-)Effect of Grammatical Gender on Early Perception: ERP Study in Simultaneous Bilinguals

Authors: Oleksandra Osypenko[1], Silke Brandt [1], and Aina Casaponsa [1]
1 – Lancaster University

Grammatical gender has been shown to influence cognitive processes, such as categorisation (Osypenko et al., 2025) and memory recall (Osypenko et al., in press), but the nature of these effects remains a subject of debate. This pre-registered study aims to investigate whether previously found behavioural effects were due to more conscious manifestation of language or unconscious activation of grammatical gender. We examine whether grammatical gender modulates early perceptual processing in Ukrainian-Russian simultaneous bilinguals tested in their genderless L2 (English). Participants completed a non-verbal object-face association task using conceptually neutral stimuli with either matching or mismatching grammatical genders across their two L1s. Both event-related potentials (ERPs) and behavioural data (reaction times and response types) were examined.

Each depicted item was followed by a gendered face, resulting in four conditions depending on the grammatical gender of the nouns and biological sex of faces: (A) congruent across Ukrainian and Russian (e.g., a fork – feminine in both, paired with a female face), (B) incongruent across Ukrainian and Russian (a fork, paired with a male face), (C) congruent in Ukrainian/incongruent in Russian (an umbrella – feminine in Ukrainian and male in Russian, paired with a female face), (D) incongruent in Ukrainian/congruent in Russian (an umbrella, paired with a male face).

We predicted that if grammatical gender unconsciously affected conceptual representations, we would find faster reaction times and biased responses for conditions congruent across both L1s (A vs B) or in participants more dominant/proficient L1 (C vs D). For ERP markers we predicted to find modulations of N1 (marker of early, automatic attentional shift), P2/VPP (marker for categorical face perception) and N300 (linked to violated semantic expectations) components as a result of grammatical gender effects. All data and pre-registration is available on OSF (https://osf.io/2vr7k).

Contrary to our predictions, no significant modulations were found in the ERP components of interest, regardless of condition (figure 1). Behaviourally, no effects of grammatical gender were observed, although participants exhibited a general response bias. Furthermore, when grammatical genders of a noun differed across L1s, language dominance or proficiency did not affect behavioural or neural responses.

These results contrast with prior ERP studies (Boutonnet et al., 2012; Sato et al., 2020) showing early grammatical gender effects in sequential bilinguals and suggest that grammatical gender, in isolation and absent conceptual cues, may not automatically influence early perceptual encoding. Also, these results contrast sharply with our prior behavioural experiments (Osypenko et al., 2025) using the same participant group, in which we observed significant effects of grammatical gender on cognitive processes. We discuss whether the previous findings reflect true Whorfian effects on perception or rather a more strategic, task-driven processes (e.g., “thinking-for-speaking”, Wolff & Holmes, 2011).

References

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79. https://doi.org/10.1016/j.brainres.2012.08.044

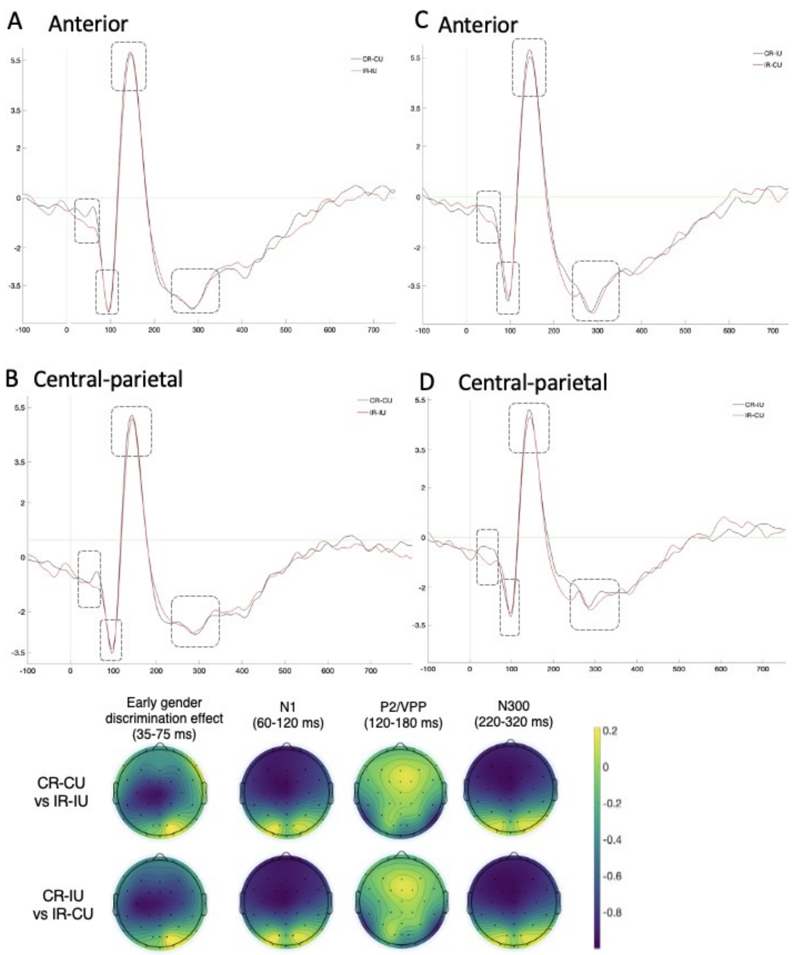
Osypenko, O., Brandt, S., & Athanasopoulos, P. (2025). The influence of three-gendered grammatical systems on simultaneous bilingual cognition: The case of Ukrainian-Russian bilinguals. *Language and Cognition*, *17*, e25. https://doi.org/10.1017/langcog.2024.73

Sato, S., Casaponsa, A., & Athanasopoulos, P. (2020). Flexing Gender Perception: Brain Potentials Reveal the Cognitive Permeability of Grammatical Information. *Cognitive Science*, *44*(9), e12884. https://doi.org/10.1111/cogs.12884

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Figure 1.

Grand average ERP waveforms of the N1, P2/VPP, and N300 components (circled with a grey line) showing the effects of grammatical gender congruency for Ukrainian-Russian bilinguals across conditions



Interacting with someone shapes prediction in spoken-language comprehension

Angèle Brunellière[1,2], Laurent Ott[1], Solène Kalénine[1], Martin Pickering[3]

[1] Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France; [2] Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 - CRIStAL, F-59000, Lille, France, [3] University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, Scotland, United Kingdom

angele.brunelliere@univ-lille.fr

Background: While listening to a spoken message, predicting upcoming information from a previous sentential context ensures a successful comprehension [1,2]. However, prediction in spoken-language comprehension is not always found [3,4]. In order to explain this flexibility, both prediction-by-association and prediction-by-production have been proposed [2]. While prediction-by-association would be an automatic process related to the spreading activation of conceptual features from the sentence representation, prediction-by-production, leading to the preactivation of words predicted from the sentence representation and their properties thanks to the production system, would be optional. In this study, we investigated the flexibility of prediction-by-association and prediction-by-production through a social interaction. This investigation is motivated by two key components of social interactions: mutual comprehension at the conceptual level and interplay between comprehension and production systems.

Method: In the Visual World Paradigm, thirty-two native French speakers listened to forty-eight highly and weakly constraining sentences which were associated with a visual scene containing three distractor objects and one of four critical objects (Target, Semantic Competitor, Phonological Competitor, Unrelated). Each sentence was presented only once before or after a social interaction, which was not related to the content of the sentences (see Figure 1A). The Target object referred to the word predicted from a sentence whose effect could reflect the two types of predictions. The Semantic and Phonological Competitor objects referred respectively to a word sharing either semantic category or phonological onset overlap with the predicted word to explore prediction-by-production. Used as a control condition, the Unrelated object referred to a word with neither phonological onset overlap nor semantic relationships with the predicted word. Fixations on the different objects were recorded during sentence comprehension. After hearing the sentence, participants had to determine whether one of the objects was mentioned in the sentence. During the social interaction, participants had to find the correct position for five Tangram pictures into a grid of ten pictures under time pressure by actively collaborating with their partner and describing shapes via an audioconference device.

Results: Cluster-based permutation analyses were performed on fixation proportion differences between related and unrelated objects. For highly constraining sentences, fixation proportions between Target and Unrelated objects differed in a time window between -320 and 980 ms after word onset before the interaction (p=.001, see Figure 1B). This difference emerged 160 ms earlier after the interaction in a time window between -480 ms and 980 after word onset. Fixation proportions between Semantic Competitor and Unrelated objects differed in a time window between -380 and -80 ms after word onset before the interaction (p=.004, see Figure 1B). No other significant effects were found for highly or weakly constraining sentences.

Discussion: Consistent with optional prediction-by-production [2], a predictive effect due to Semantic Competitor objects only occurred before the social interaction. In contrast, the predictive effect of Target objects was speeded-up after the social interaction. This pilot study suggests that social interaction may shift prediction toward finer based-concept processing. Findings will be discussed in line with models of social interaction and language comprehension and further studies are needed to confirm the role of interactions.

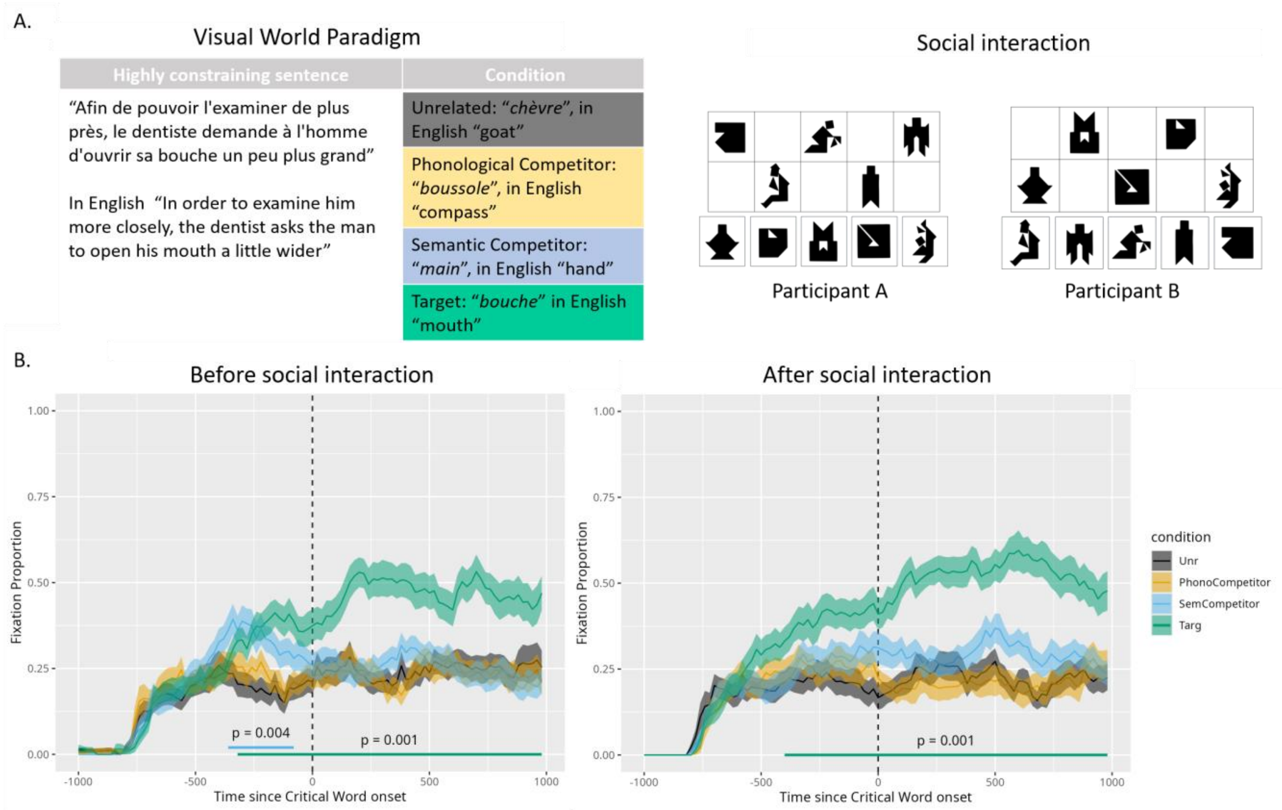


Figure 1. A. Example of stimuli in the Visual World Paradigm and social interaction task; B. Gaze-related changes during the listening of highly constraining sentences before the social interaction and after the social interaction. Graphs showing fixation proportions on Target (Targ, the predicted word), Semantic Competitor (SemCompetitor, a semantic competitor with the predicted word), Phonological Competitor (PhonoCompetitor, a phonological competitor with the predicted word), and Unrelated (Unr) objects as a function of time since display onset. Each type of objects defined an experimental condition. Semantic and Phonological competitors and Unrelated objects did not evoke plausible words in the sentences. Time 0 shows the time in milliseconds at which the target word was mentioned in the sentence. The green line along the below indicates significant differences between Target and Unrelated objects ($p < .05$). The blue line along the below shows significant differences between Semantic Competitor and Unrelated objects ($p < .05$).

References

1. Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension. *Language, Cognition and Neuroscience*, *31*, 32–59. https://doi.org/10.1080/23273798.2015.1102299
2. Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*, 1002–1044. https://doi.org/10.1037/bul0000158
3. Brunellière, A., Vincent, M., & Delrue, L. (2022). Oscillatory correlates of linguistic prediction and modality effects during listening to auditory-only and audiovisual sentences. *International Journal of Psychophysiology*, *178*, 9–21. https://doi.org/10.1016/j.ijpsycho.2022.06.003
4. Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

The effect of visual cuing during simultaneous interpreting

Dongpeng Pan[1], Kilian Seeber[1]
[1] Faculty of Translation and Interpreting, University of Geneva
Dongpeng.Pan@unige.ch, Kilian.Seeber@unige.ch

Background: Visual context critically shapes real-time spoken language comprehension. Eye-tracking reveals that listeners exploit scenes incrementally: e.g., a second apple prevents garden-path errors [1], and verbs like "eat" trigger anticipatory gazes to edibles [2]. Deictic signals refine this: reflexive gaze-cueing shifts attention [3,4], accelerating reference resolution in dialogue [5] and sentences [6]. Pointing gestures similarly direct focus [7]. Pupillometry indexes load, scaling with ambiguity [8,9] or noise [10]; congruent gaze eases noun processing [11], while incongruent gestures enlarge pupils [12] and mismatched cues slow responses [13–15]. Simultaneous interpreting (SI) tests multimodal cueing under extreme load: congruent gestures boost accuracy even in noise [16–18], but mismatches inflate effort [19]. Pupillometry shows dilation shrinks in easier SI conditions [20,21]. This study extends visual-world-paradigm (VWP) findings to SI, probing deictic pointing's effects on attention and load.

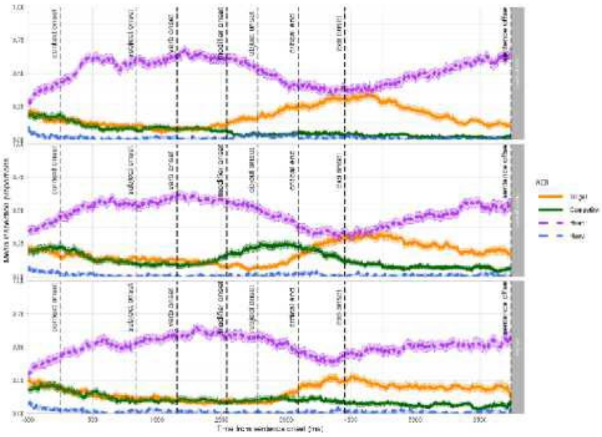
Objectives: We investigated how congruent, incongruent, and neutral deictic gestures influence interpreters' real-time processing, using VWP with pupillometry. Unlike prior studies on general comprehension or passive listening, this targets professional SI's multitasking demands, highlighting originality in combining behavioural/physiological measures to dissect gesture congruence's role in extreme bilingualism. We hypothesised congruent gestures facilitate anticipation/reduce load, incongruent induce conflict/heighten load, and neutral impose greatest burden via auditory reliance.

Method: Twenty-four Geneva-based conference interpreters (22 women, 2 men; Age: M=41.7 yrs, SD=10.8; Experience: M=11.6 yrs, SD=9.0; 12 L1-French, 12 L1-Spanish; English as B/C) interpreted 30 English clips into French/Spanish while monocular gaze was sampled at 1 kHz (EyeLink Portable Duo). Motivation for interpreters: their expertise in concurrent comprehension/production under load amplifies multimodal effects, bridging lab findings to real-world bilingualism. Videos showed one male speaker (with visible mouth movements, no eye-bar in experiment; bar in figures for anonymity) on left, two pictures (target/competitor, e.g., camel/cactus) on right (Fig. 1). Trials: critical sentence ("In the desert, the explorer sees a large camel.") + numeral sentence for buffer. At qualifier ("large") onset, speaker pointed to target (congruent), distractor (incongruent), or stayed still (neutral; 10 each). Targets counterbalanced; sentences corpus-normed via LimeSurvey. Binary fixations to AOIs (face, hand, target, competitor) and baseline-corrected pupil diameters indexed attention/load. After excluding trials with propositional distortions (e.g., lexical substitutions/omissions altering referents, ensuring valid comprehension) or >50% track-loss, 606/720 retained. Analysed in windows (qualifier, object, spill-over) via GAMMs [22] with lag-1 for autocorrelation [23,24].

Results: Fixations (Fig. 2) showed Time×AOI smooths (target vs. competitor) across conditions (Fig. 3). Congruent: early/sustained target advantage (qualifier: edf=2.98, χ^2=27.85; object: edf=3.10, χ^2=38.72; spill-over: edf=2.00, χ^2=24.19; all p<.001). Incongruent: initial competitor bias flipping to target mid-object (qualifier: edf=2.00, χ^2=19.85; object: edf=2.46, χ^2=28.50; spill-over: edf=2.00, χ^2=56.68; all p<.001). Neutral no qualifier bias (edf=2.00, χ^2=4.47, p=.11; brief competitor edge), target rise in object (edf=2.00, χ^2=7.12, p=.028)/spill-over (edf=3.74, χ^2=23.32, p<.001). Pupillometry (Fig. 4, blue=congruent, orange=incongruent, green=neutral; colors checked for accessibility): no qualifier effects; neutral > incongruent in object/spill-over (predicted via GAMM differences; congruent vs. incongruent non-significant, as focus was on mismatches' costs).

Discussion: Interpreters prioritised pointing: congruent enabled pre-noun target shifts; incongruent misdirected then corrected; neutral delayed until noun, amplifying load. Pupil pattern (unpredicted but post-hoc interpretable): any spatial anchor (even misleading) narrows search space vs. auditory-only, with incongruent's correction cost lower than neutral's sustained effort—images alone insufficient without gesture to fully pre-activate/resolve under SI load, unlike simpler tasks. Extends VWP [11] to SI, showing automatic gesture integration survives production demands. Future: naturalistic discourse, expertise effects on cue weighting.

Figure 1. Example frames from the three cueing conditions for the critical sentence “In the desert, the explorer sees a large camel.” In (A) congruent the speaker begins pointing to the camel at the onset of the qualifier “large,” in (B) incongruent the speaker instead points to the cactus, and in (C) neutral the speaker maintains forward gaze with no gesture; the two pictures on the right display the potential referents.



Qualifier window Object window

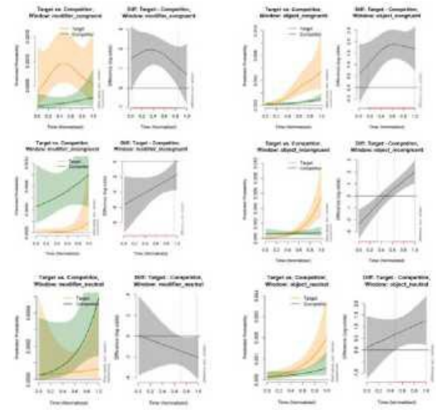


Figure 2. Fixation proportions to target (orange), competitor (green), speaker’s head (purple), and hand (blue) across congruent, incongruent, and neutral conditions.
Solid lines show mean fixation probability; ribbons indicate ±1 SE. Vertical dashed lines mark key linguistic events: context onset, subject onset, verb onset, modifier onset (gesture onset), object onset, critical sentence end, extro onset, and sentence offset.

Figure 3. Rows show congruent, incongruent, and neutral conditions; columns show the modifier (left) and object (right) windows. In each group, the left plot gives GAMM-predicted fixation probability for target (orange) versus competitor (green) with 95 % CI shading, and the right plot shows the log-odds difference (target – competitor) with 95 % CIs and

red ticks marking timepoints of significant divergence.

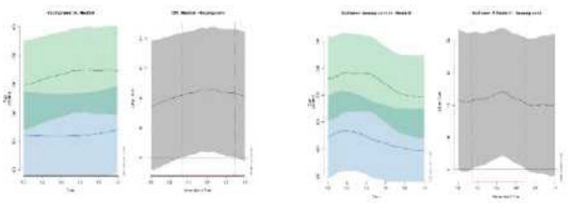


Figure 4. Left: baseline-corrected pupil diameter with shaded 95 % CIs for incongruent (teal) versus neutral (green) trials in the object window, and the pointwise difference (neutral – incongruent) with its 95 % CI; vertical red lines bracket the interval where neutral dilation exceeds incongruent. Right: the same format for the spillover window.

References

[1] Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*. *268*(5217), 1632–1634.

[2] Altmann, G.T. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*. *73*(3), 247–264.

[3] Friesen, C.K. & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*. *5*(3), 490–495.

[4] Frischen, A., Bayliss, A.P., & Tipper, S.P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*. *133*(4), 694–724.

[5] Hanna, J.E. & Brennan, S.E. (2007). Speakers’ eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*. *57*(4), 596–615.

[6] Knoeferle, P. & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*. *3*.

[7] Langton, S.R., Watt, R.J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*. *4*(2), 50–59.

[8] Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*. *91*(2), 276–292.

[9] Engelhardt, P.E., Ferreira, F., & Patsenko, E.G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly journal of experimental psychology (2006)*. *63*(4), 639–645.

[10] van Engen, K.J. & McLaughlin, D.J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*. *369*, 56–66.

[11] Sekicki, M. & Staudte, M. (2018). Eye'll Help You Out! How the Gaze Cue Reduces the Cognitive Load Required for Reference Processing. *Cognitive Science*. *42*(8), 2418–2458.

[12] Morett, L.M., Roche, J.M., Fraundorf, S.H., & McPartland, J.C. (2020). Contrast Is in the Eye of the Beholder: Infelicitous Beat Gesture Increases Cognitive Load During Online Spoken Discourse Comprehension. *Cognitive Science*. *44*(10), e12912.

[13] Kelly, S.D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*. *21*(2), 260–267.

[14] Staudte, M., Crocker, M.W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*. *133*(1), 317–328.

[15] Abashidze,D. & Knoeferle, P. (2021).Influence of Actor's Congruent and Incongruent Gaze on Language Processing. *Frontiers in Psychology*. *12*.

[16] Arbona,E.,Seeber, K.G., & Gullberg, M. (2023). Semantically related gestures facilitate language comprehension during simultaneous interpreting. *Bilingualism: Language and Cognition*. *26*(2), 425–439.

[17] Arbona,E.,Seeber, K.G., & Gullberg, M. (2023). The role of manual gestures in second language comprehension: a simultaneous interpreting experiment. *Frontiers in Psychology*. *14*, 1188628.

[18] Arbona,E.,Seeber, K., & Gullberg, M. (2024). The role of semantically related gestures in the language comprehension of simultaneous interpreters in noise. *Language, Cognition and Neuroscience*. *39*(5), 584–608.

[19] Chmiel, A., Janikowski, P., & Lijewska, A. (2020). Multimodal processing in simultaneous interpreting with text Interpreters focus more on the visual than the auditory modality. *Target*. *32*(1), 37–58.

[20] Seeber, K.G. (2011). Cognitive load insimultaneous interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*. *13*(2), 176–204.

[21] Seeber, K.G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target*. *25*(1), 18–32.

[22] Wood SN. (2017). *Generalized additive models: An introduction with R*. CRC Press/Taylor & Francis Group.

[23] Cho, S.-J., Brown-Schmidt, S., Boeck, P. de, & Naveiras, M. (2022). Space-time modeling of intensive binary time series eye-tracking data using a generalized additive logistic regression model. *Psychological methods*. *27*(3), 307–346.

[24] Brown-Schmidt, S., Cho, S.-J., Fenn, K.M., & Trude, A.M. (2025). Modeling spatio-temporal patterns in intensive binary time series eye-tracking data using Generalized Additive Mixed Models. *Brain Research*. *1854*.

The effect of speech rate on two prediction stages in older adults with and without hearing loss

Naomi Nota[1]*, Muzna Shehzad[2]*, Ruth E. Corps[3], Martin J. Pickering[1], Graham Naylor[2], & Lauren V. Hadley[2]
[1]University of Edinburgh, [2]University of Nottingham, [3]University of Sheffield
Naomi.Nota@ed.ac.uk

Language comprehension is facilitated when listeners predict what they are likely to hear (e.g., Altmann & Kamide, 1999), which could be particularly beneficial for people with hearing loss who often experience communication challenges. Previous research suggests that language prediction occurs in two stages (Corps et al., 2022; Pickering & Gambi, 2018). The first associative stage is automatic, allowing listeners to quickly predict a noun that is related to a verb after hearing a sentence like *Tonight, it is likely James will wear a…* (with more fixations to target *tie* than to distractor *drill*). In contrast, the second strategic stage of prediction is non-automatic and slower, allowing listeners to hone predictions by taking context such as the agent's perspective into account (i.e., more fixations to the gender-stereotypical *tie* than *dress*).

It has been theorised that people with hearing loss are delayed and less able to use prediction as listening consumes more resources (Fernandez et al., 2024). In addition, speakers often accommodate to listeners with hearing loss by speaking more slowly, as it improves intelligibility and reduces effort (Winn & Teece, 2021). Furthermore, there is evidence for speech rate affecting prediction (Fernandez et al., 2020). In this preregistered study, we therefore investigated the effect of speech rate on associative and strategic prediction stages in listeners with and without hearing loss.

We tested English speakers from Glasgow with hearing loss (HL, n = 38, mean age = 70.74, better ear pure tone audiogram mean = 44.37) and with no hearing loss (NH, n = 37, mean age = 64.62, better ear pure tone audiogram mean = 17.52) using a visual-world eye-tracking paradigm. Participants listened to sentences with speech rates of 4 and 5.5 syllables per second in an interleaved design. They listened to sentences from female and male speakers, viewed object images, and answered comprehension questions, following Corps et al. (2022).

Divergence point bootstrapping analyses revealed that at both rates, the two groups showed associative prediction (HL, slower rate 407 ms, speeded rate 567 ms; NH, slower rate 303 ms, speeded rate 440 ms) and strategic prediction (HL, slower rate 947 ms, speeded rate 711 ms; NH, slower rate 736 ms, speeded rate 656 ms). At the slower rate, both groups showed significantly earlier associative than strategic predictions. However, at the speeded rate, this difference in timing between the two prediction stages was significant only for NH. Associative prediction was marginally delayed for HL vs. NH at the speeded rate, but there were no significant differences between groups for strategic prediction.

Our findings show that for HL the differentiation between the two prediction stages is disrupted at the speeded rate but not at the slower rate. Additionally, when speech is speeded, HL make associative predictions more slowly than NH. This suggests that faster speech negatively impacts prediction in older adults with hearing loss but not in those without it. Investigating these effects may enhance our understanding of how hearing loss impacts communication, and provide potential routes to intervention in the population with hearing loss.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, *122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Fernandez, L. B., Pickering, M. J., Naylor, G., & Hadley, L. V. (2024). Uses of Linguistic Context in Speech Listening: Does Acquired Hearing Loss Lead to Reduced Engagement of Prediction? *Ear and Hearing*, *45*(5), 1107–1114. https://doi.org/10.1097/AUD.0000000000001515

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Winn, M. B., & Teece, K. H. (2021). Slower Speaking Rate Reduces Listening Effort Among Listeners With Cochlear Implants. *Ear and Hearing*, *42*(3), 584–595. https://doi.org/10.1097/AUD.0000000000000958

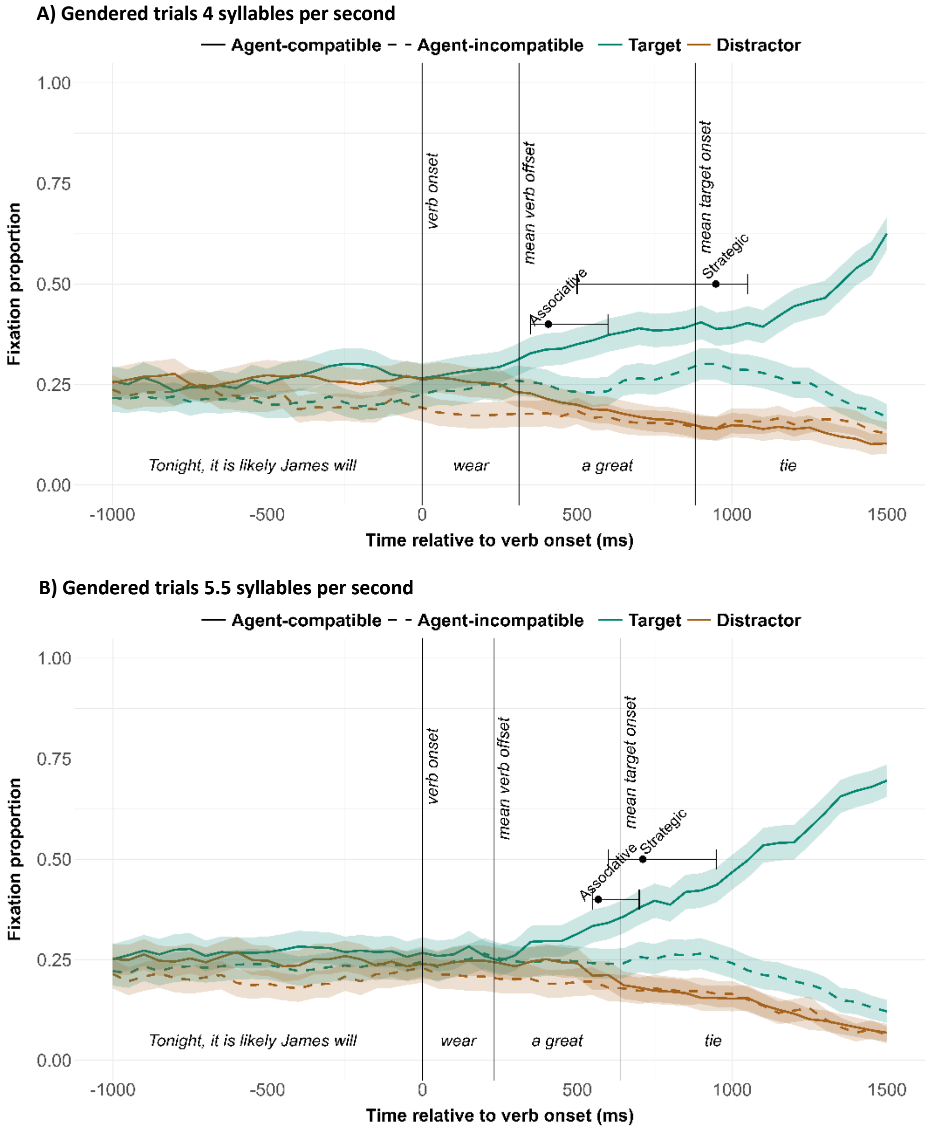


Figure 1. Mean fixation proportions on all objects for participants with hearing loss on gendered trials with speech rates of 4 (A) and 5.5 (B) syllables per second. Shaded areas around the lines represent CIs. The lower horizontal line with a circle indicates the bootstrapped CI and mean onset of divergence for associative prediction, comparing target to distractor objects (*tie* vs. *drill*). The upper horizontal line with a circle shows the bootstrapped CI and mean onset of divergence for strategic prediction, comparing stereotypically gender-compatible target objects to gender-incompatible target objects (*tie* vs. *dress*).

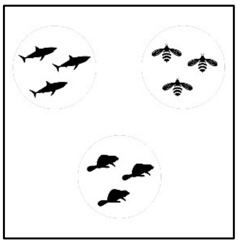
How information structure, prosodic prominence, and speech act affect reference resolution: Evidence from eye-tracking

Sophie Repp, Heiko Seeliger, Judith Schlenter, Petra B. Schumacher
University of Cologne
sophie.repp@uni-koeln.de

Prosodic prominence (PProm) is known to affect reference resolution because it signals information structure (InfS): if a nominal expression is unaccented, listeners quickly infer that the denoted referent is discourse-given while accentuation may indicate a reference shift because it may indicate contrast. These effects have been observed for the processing of statements (i.e., declaratives) in the visual world paradigm, where we see rapid use of prosodic information [1,2,5]. Recent prosodic production studies [3,4] show that in speech acts (SpA) other than statements, like exclamations, InfS may not be reliably tied to PProm, which has been argued to result from SpA and InfS imposing conflicting requirements on prosodic realization: exclamations require high PProm while the InfS status *given* requires low PProm. In German, speakers resolve this conflict by not marking givenness by low PProm in exclamations. **Experiment:** Using the visual-world paradigm, we explored if processing of PProm is affected by the SpA. Thirty-two German speakers looked at a display and listened to a target clause with a plural object (e.g., *Biber* 'beavers') that, as manipulated via the context, presented either given or contrastive information; Fig. 1. Next to InfS (given/contrastive), we manipulated whether the object's first syllable was accented (L+H*) or unaccented (cross-spliced). In one experimental session, target sentences were statements, and in another, exclamations (sessions a week apart, order balanced; each 28 items, 32 fillers). The display showed three images, depicting the target (e.g., beavers), a phonological competitor with an identical first syllable (e.g., *Bienen* 'bees'), and a distractor (e.g., *Haie* 'sharks'). Participants clicked on the image matching the target and responded to a subsequent verification statement. **Analysis** (GLMMs of fixations to target (vs. other), aggregated in 100ms bins; three target windows aligned to object onset; Fig. 2) revealed for the pre-disambiguation window (0-300ms) main effects of InfS and accent due to a higher likelihood of target looks for given vs. contrastive objects and for unaccented vs. accented objects across SpA. In the disambiguation window (second syllable of the object, 300-600ms), these effects were qualified by interactions: there were more looks to the target for given unaccented vs. given accented objects for exclamations but not for statements, and no effect for the (non-)accentuation of contrastive objects. In the post-disambiguation window (600-900ms), the interaction InfS×accentuation remained significant. **Discussion**: Our results align with previous research showing an impact of PProm and InfS on reference resolution. Further, exclamations show a greater sensitivity to PProm than statements. We propose that deaccentuation is taken as a strong signal for givenness in exclamations because it is rarely used in production in this SpA. The pre-disambiguation effects may be due to the 'medium' PProm of the adjective before the (cross-spliced) noun, which may have been taken as a predictor for low PProm of the object.

Sample item auditory presentation *(Context English translation only)*: Last week Ben's task in his drawing class was to draw various animals. Initially, he was not so sure and he rather hesitantly experimented with a few ***{beavers$_{GIVEN}$ / bees$_{CONTRAST}$}***. But then Ben really got going.

STATEMENT	*Dann hat er ganz viele* ***{BIBER$_{L+H*}$ / Biber$_{unaccented}$}*** *gezeichnet.*
(verb-second)	then has he very many beavers beavers drawn
EXCLAMATION	*Hat der viele* ***{BIBER$_{L+H*}$ / Biber$_{unaccented}$}*** *gezeichnet!*
(verb-first)	has he many beavers beavers drawn



STATEMENT: 'Then he drew lots of beaves.'

EXCLAMATION: 'Did he draw a lot of beavers!'

Figure 1. Sample item set: Context and target sentence with corresponding visual display.

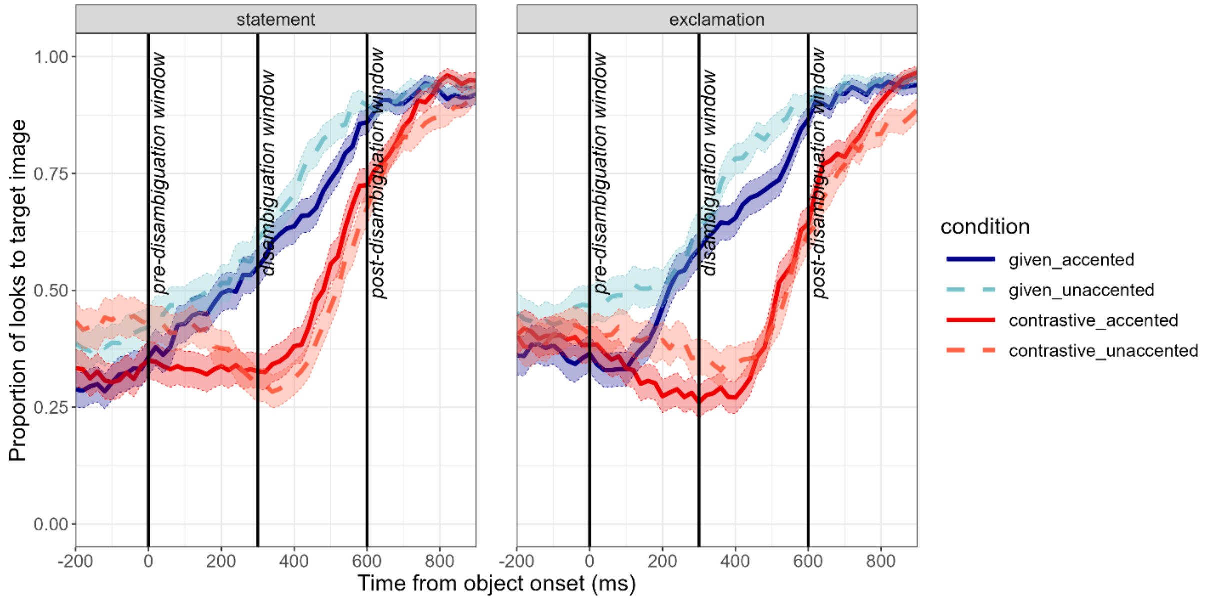


Figure 2. Proportion of looks to target out of looks to all three interest areas from object onset. The vertical lines indicate the onset of the analysis windows.

[1] Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *J Memory and Language* 47(2), 292–314.

[2] Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *J Memory and Language* 58(2), 541-573.

[3] Repp, S. (2020). The prosody of *wh*-exclamatives and *wh*-questions in German: Speech act differences, information structure and sex of speaker. *Language & Speech* 63(2) 306 –361.

[4] Seeliger, H., & Repp, S. (2023). Contrast, givenness, and (the lack of) accent shift and deaccentuation in non-assertive speech acts. *Laboratory Phonology* 14(1), 1–46.

[5] Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49(3), 367–392.

Visual Cues Not Only Facilitate Online Sentence Comprehension But Also *License* Ellipsis Resolution: A Self-Paced Reading Study of English Verb Phrase Ellipsis

Kohei Haneda[1,2,3,4,*], Anja Schüppert[2], and Roel Jonkers[2]

[1]International Doctorate for Experimental Approaches to Language And Brain (IDEALAB), [2]University of Groningen, Groningen, Netherlands, [3]Macquarie University, Sydney, Australia, [4]University of Potsdam, Potsdam, Germany, *k.haneda@rug.nl

Language processing does not exist in a vacuum. Robust evidence has been accumulating to demonstrate that various contextual factors can impact real-time language processing. Among the studies that investigated the roles of nonlinguistic information in online sentence comprehension, those that took advantage of visual cues appear to have been particularly successful.[1, 2, 3, 4, 5, 6, 7, 8] This is not surprising considering vision is the most prominent in all perceptual modalities among primates.[9] The extant literature has been successful in affirmatively answering the yes-no question of whether visual information can immediately modulate language comprehension; however, it has yet to respond to the question of how much and to what extent nonlinguistic visual information can yield such effects. The current study thus aims to delineate the capacity of nonlinguistic visual cues to influence real-time sentence comprehension by examining whether or not the parser can capitalize on visual inputs in lieu of linguistic inputs to comprehend ellipsis. Ellipsis resolution typically requires that a syntactically isomorphic antecedent be stored in working memory and later retrieved, so as to fill the “gap” downstream. What if this syntactic antecedent is replaced with visual information such as an image depicting an action? Can comprehenders still integrate the visual cue with the unfolding discourse representation and resolve the ellipsis? Investigating ellipsis with a visual antecedent thus provides an ideal testing ground for putting the visual system to a “stress test” in probing its true capacity to impact sentence comprehension. Sixty L1 English readers viewed images depicting actions followed by sentences with elliptical gaps. Participants were first shown a sentence fragment such as “Last Friday afternoon, Joanna was”, followed by an image of a woman knitting a scarf. Subsequently, a sentence with an elliptical gap was displayed: “Elise was too because…”. Participants’ reading times at the ellipsis sites (“was too”) and spillover regions were compared to those of the controls to examine whether the elliptical resolution with a visual antecedent was successful. Linear mixed model analyses show that visually-situated processing of ellipsis appears to have some processing advantage over its linguistic counterpart–but surprisingly, not at the ellipsis region itself but rather at the subsequent spillover regions. At the critical verb region, the reading times for the ellipsis sentences with visual antecedents did not significantly differ from those of the ellipsis sentences with linguistic antecedents. But at the two subsequent regions, the reading times were significantly shorter for the visual antecedent condition compared to the linguistic antecedent condition. The results may suggest that the mental representation of the unfolding discourse constructed via the visual modality might be less cognitively taxing in working memory. Thus, the current study presents the first evidence that (1) comprehenders are indeed capable of capitalizing on nonlinguistic visual contexts to execute a complex retrieval operation such as ellipsis resolution in real-time sentence comprehension, and (2) such visually-situated comprehension may impose less cognitive demand on the language parsing system and working memory, compared to linguistically-situated processing, revealing the highly multimodal nature of the linguistic system, and opening up new doors for further cross-linguistic and cross-methodological replications.

References

1. Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

2. Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95-127.

3. Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive science*, *30*(3), 481-529.

4. Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519-543.

5. Guerra, E., & Knoeferle, P. (2014). Spatial distance modulates reading times for sentences about social relations: evidence from eye tracking. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

6. Guerra, E., & Knoeferle, P. (2017). Visually perceived spatial distance affects the interpretation of linguistically mediated social meaning during online language comprehension: an eye tracking reading study. *Journal of Memory and Language*, *92*, 43-56.

7. Guerra, E., & Knoeferle, P. (2018). Semantic interference and facilitation: Understanding the integration of spatial distance and conceptual similarity during sentence reading. *Frontiers in Psychology*, *9*, 718.

8. Burmester, J., Sauermann, A., Spalek, K., & Wartenburger, I. (2018). Sensitivity to salience: Linguistic vs. visual cues affect sentence processing and pronoun resolution. *Language, Cognition and Neuroscience*, *33*(6), 784-801.

9. Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality?. *Frontiers in psychology*, *10*, 481030.

The Role of Task Framing and Context Source in Scalar Implicature Detection

Michael Vrazitulis
University of Potsdam
vrazitulis@uni-potsdam.de

Scalar implicatures (SIs) are commonly studied using sentence verification tasks. In these tasks, participants judge underinformative sentences such as *Some trees are plants*. A standard linking assumption holds that rejecting such sentences as false reflects a pragmatic interpretation (but see [7]). Accepting them as true reflects a literal interpretation [1, 4].

Literal vs. pragmatic response rates for the English scalar *some* vary widely across previous studies (e.g., from 38% literal responses [9] up to 77% literal responses [6]). One underexplored source of variation is task framing—whether judgments are about truth or felicity (as suggested by [7]). Another is the source of context provided, i.e., either world knowledge (WK) or accompanying pictures (sentence–picture verification, [2]). We tested the impact of these two factors on SI response behavior.

Across three web-based experiments (N = 252, 252, and 576), we used a 2×2 between-subjects design. We manipulated framing (truth vs. felicity) and context source (WK vs. picture). Participants judged underinformative targets and control items, responding with either *true/false* or *good/bad* judgments. Figure 1 illustrates an example item. There were 24 items overall.

Experiments 1 and 2 included 24 trials per participant, evenly distributed across three conditions: target, control true, and control false. Experiment 3 also included 24 trials per participant, but featured three additional control types using the stronger scalemate *all*. Further, Experiment 2 differed from the other two in response format: In Experiments 1 and 3, participants chose between explicit polarity pairs (e.g., *Is this a good or a bad statement?*). In Experiment 2, they answered yes/no questions instead (e.g., *Is this a good statement?*).

All three experiments showed the same basic result pattern. Participants gave more pragmatic responses in the felicity condition than in the truth condition (95% credible interval: Exp. 1 [+6.5%, +30.3%], Exp. 2 [+0.6%, +15.5%], Exp. 3 [+25.1%, +47.9%]). We also observed a consistent main effect of context source: Participants responded more pragmatically in the picture-based condition than in the WK-based one (95% credible interval: Exp. 1 [+4.6%, +29.3%], Exp. 2 [−0.9%, +13.5%], Exp. 3 [+18.1%, +44.5%]). Literal response rates in the WK condition were especially high, sometimes even reaching over 80%. Figure 2 displays the average literal response rates across conditions and experiments. These results suggest that SI processing is indeed modulated by task framing and context source.

Taken together, our findings indicate that task design can systematically influence how underinformative sentences are interpreted. Future work should identify which design factors most reliably affect response rates. Crosslinguistic differences may also play a role: For example, all else being equal, English *some* (e.g., [6, 7]) seems to elicit more literal responses than French *certains* (e.g., [1, 3, 8]) or Dutch *sommige* (e.g., [4, 5]). A clearer understanding of how contextual features shape response rates will help clarify which aspects of SI processing are stable and which depend on particularities of the experimental setup.

Experiment scripts for PCIbex [10], the collected response data, and R code for statistical analysis can be accessed here (anonymized link):
https://osf.io/86r4u/files/osfstorage?view_only=61773f9ca6e84e5db09e317d57617947

Condition	Sentence	Question
a	Some trees are plants.	Is this a true or a false statement?
b	Some of the trees are palms.	Is this a true or a false description of the picture?
c	Some trees are plants.	Is this a good or a bad statement?
d	Some of the trees are palms.	Is this a good or a bad description of the picture?

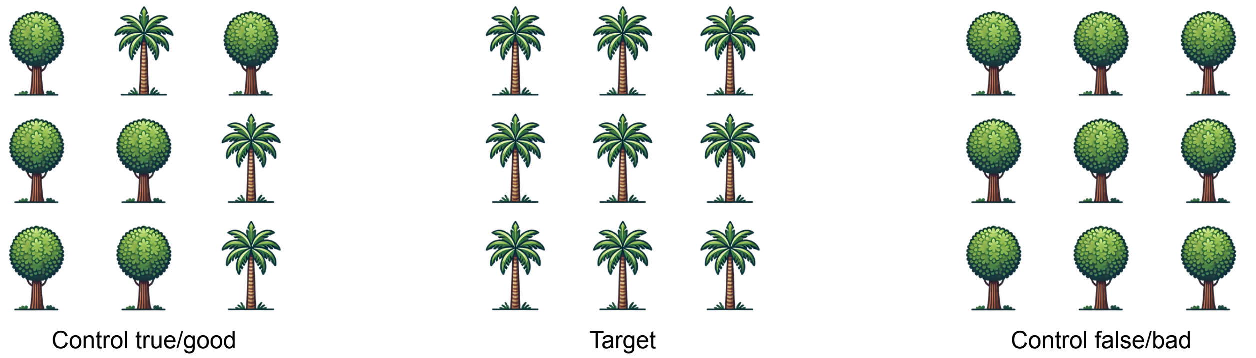


Figure 1: Above: Example item shown in each of the four conditions in the 2×2 design. Conditions a and b use world knowledge as context; c and d use pictures. Conditions a and c use truth-based framing; b and d use felicity-based framing. Below: Pictures for the example item. For the sentence *Some of the trees are palms*, the control picture on the left requires an unambiguous *true/good* verification. The target picture in the middle is pragmatically ambiguous. The control picture on the right requires an unambiguous *false/bad* verification.

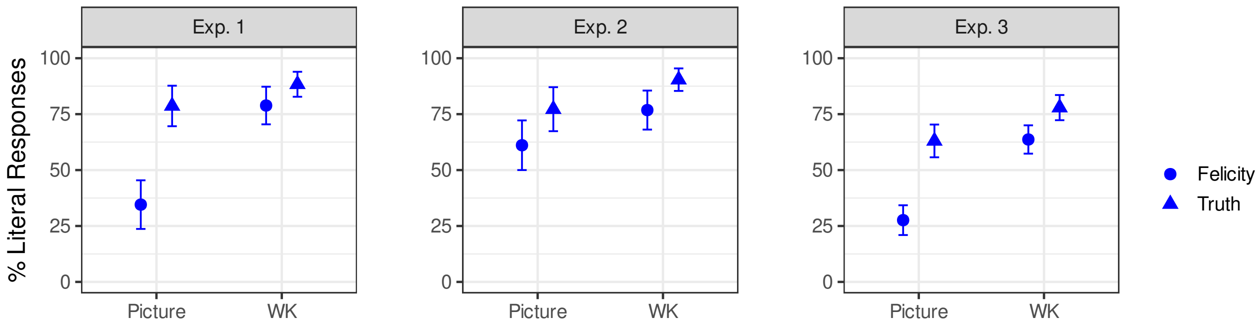


Figure 2: Average literal response rates across the four conditions in Experiments 1–3. Error bars represent 95% confidence intervals computed over subject-level means.

References

[1] L. Bott and I. A. Noveck. “Some utterances are underinformative: The onset and time course of scalar inferences”. In: *Journal of Memory and Language* 51.3 (2004), pp. 437–457. [2] H. H. Clark and W. G. Chase. “On the process of comparing sentences against pictures”. In: *Cognitive Psychology* 3.3 (1972), pp. 472–517. [3] A. Cremers and E. Chemla. “Direct and indirect scalar implicatures share the same processing signature”. In: *Pragmatics, semantics and the case of SIs*. Springer, 2014, pp. 201–227. [4] W. De Neys and W. Schaeken. “When people are more logical under cognitive load: Dual task impact on scalar implicature”. In: *Experimental Psychology* 54.2 (2007), p. 128. [5] K. Dieussaert et al. “Some effort for some: Further evidence that scalar implicatures are effortful”. In: *Quarterly Journal of Experimental Psychology* 64.12 (2011), pp. 2352–2367. [6] A. Feeney et al. “The story of some: Everyday pragmatic inference by children and adults”. In: *Canadian Journal of Experimental Psychology* 58.2 (2004), pp. 121–132. [7] M. Kissine and P. De Brabanter. “Pragmatic responses to under-informative some-statements are not scalar implicatures”. In: *Cognition* 237 (2023), p. 105463. [8] I. A. Noveck and A. Posada. “Characterizing the time course of an implicature: An evoked potentials study”. In: *Brain and Language* 85.2 (2003), pp. 203–210. [9] B. van Tiel, E. Pankratz, and C. Sun. “Scales and scalarity: Processing scalar inferences”. In: *Journal of Memory and Language* 105 (2019), pp. 93–107. [10] J. Zehr and F. Schwarz. *PennController for Internet Based Experiments (IBEX)*. 2018.

Consistency and Frequency Effects in Japanese Kanji Nonword Reading by L1-Chinese Speakers

Nan Kang, Satoru Saito
Graduate School of Education, Kyoto University
kang.nan.66r@st.kyoto-u.ac.jp

The phonological processing of logographic writing systems has attracted sustained scholarly attention for decades. Despite the relatively low regularity and systematicity in the orthography-to-phonology correspondence of Japanese Kanji words, research on native Japanese speakers has consistently demonstrated that words with greater reading consistency or higher lexical frequency confer advantages in both reading speed and accuracy [1, 2, 3, 6, 7]. These effects, known respectively as the consistency effect and the frequency effect, have also been observed in oral reading tasks involving Japanese two-character Kanji nonwords (e.g., 予少), which are composed of two real Kanji characters but do not form an existing lexical item with semantic meaning [1].

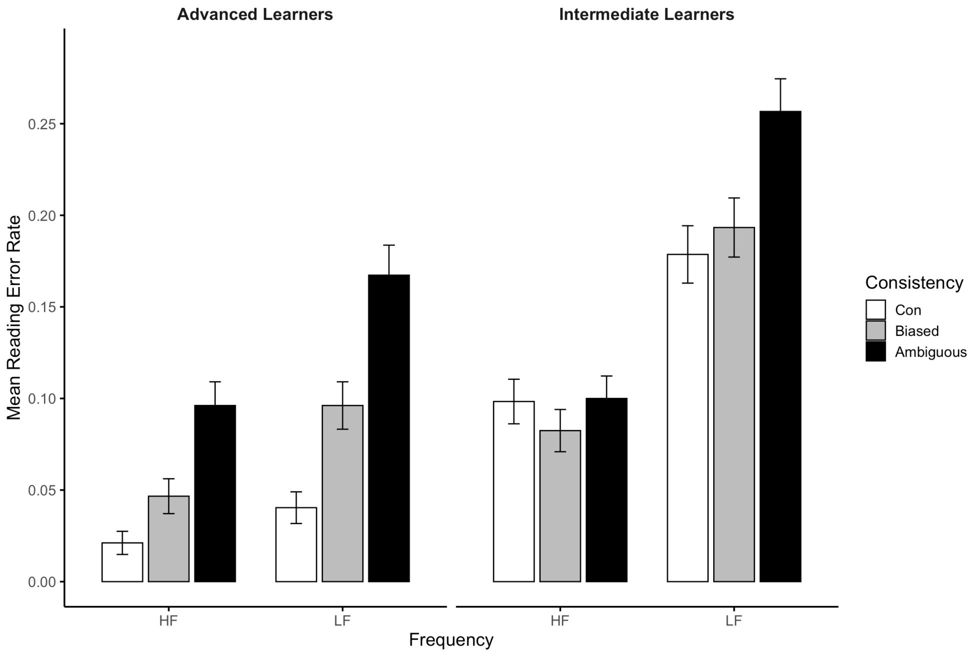
This study re-examined the consistency and frequency effects in the reading of Kanji nonwords by second-language (L2) learners of Japanese whose native language is Chinese. Two experiments were conducted with university students: 28 advanced learners and 31 intermediate learners. A total of 120 nonwords were used [1], classified into three consistency categories based on a corpus analysis of the dominant readings of their component Kanji characters. Consistent nonwords were composed of characters with a single dominant pronunciation. Inconsistent-biased nonwords included at least one character with multiple readings, but with one clearly dominant option (e.g., 決所). Inconsistent-ambiguous nonwords consisted of characters with several equiprobable readings and no clear preference (e.g., 間物). These categories corresponded to high, medium, and low levels of reading consistency, with 40 nonwords in each group. Participants were instructed to read each nonword aloud, presented in random order, within a 5-second time limit. Mixed-effects modeling was used to analyze reading times (RTs) and accuracy [4, 5].

Results from the advanced learner group revealed robust consistency and frequency effects in both RT and reading accuracy. Notably, the consistency effect exhibited a graded pattern: the inconsistent-ambiguous nonwords elicited the longest RTs and highest error rates, while the consistent nonwords resulted in the shortest RTs and lowest error rates. In contrast, the consistency effect was not statistically significant in the intermediate learner group. Further analysis indicated that L2 proficiency level (advanced vs. intermediate) modulated the consistency effect.

These findings suggest that the process of generating a reading response to Kanji nonwords in L2 learners is influenced by statistical properties—reading consistency and character frequency—derived from the real-world Japanese lexicon. Reading consistency, in particular, reflects the regularity of character-sound correspondences across orthographic neighborhoods. The results indicate the functional role of statistical learning in L2 phonological processing of Japanese nonwords, and by extension, in the reading of novel Kanji words. They further highlight the role of statistical learning in L2 Kanji reading and suggest that ability to utilize probabilistic character-sound mappings develops with increasing proficiency.

Key words: Japanese Kanji, L2 learning, reading consistency, statistical learning

Figure: Reading Error Rates across Levels of L2 Proficiency



Note. The figure depicts the mean error rates of Experiments 1 and 2 in each word group. con = consistent, typical = inconsistent-typical, atypical = inconsistent-atypical nonwords; HF = high-character frequency, LF = low-character frequency nonwords. Error bars are 95% confidence intervals for within-subject comparisons.

References

[1] Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese kanji. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 382–407. https://doi.org/10.1037/0096-1523.25.2.382

[2] Fushimi, T., Komori, K., Ikeda, M., Lambon Ralph, M. A., & Patterson, K. (2009). The association between semantic dementia and surface dyslexia in Japanese. *Neuropsychologia*, *47*(4), 1061–1068. https://doi.org/10.1016/j.neuropsychologia.2008.12.030

[3] Kang, N. & Saito, S. (2025). Phonological processing of Japanese Kanji word by L1-Chinese learners of Japanese: verification of consistency and frequency effects. Applied Psycholinguistics. https:// doi.org/10.1017/S0142716425000128

[4] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest* package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

[5] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest (Version 3.1-3)* [Computer software]. https://github.com/runehaubo/lmerTestR

[6] Wydell, T. N., Butterworth, B., & Patterson, K. (1995). The inconsistency of consistency effects in reading: The case of Japanese kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1155–1168. https://doi.org/10.1037/0278-7393.21.5.1155

[7] Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for kanji: Is a row still a rose in Japanese kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 491–514. https://doi.org/10.1037/0278-7393.19.3.491

The effects of perceived cooperativeness of lexical alignment, memory, and social judgments

Nitzan Trainin[1,2] & Einat Shetreet[1,2]
[1] Sagol School of Neuroscience, Tel Aviv University; [2] Department of Linguistics, Tel Aviv University
nitzant@mail.tau.ac.il

Background: In conversation, people often align in how they refer to objects—a phenomenon called lexical entrainment or alignment [1–4]. Debate continues on whether alignment is egocentric or socially motivated [5–7]. Some view it as a form of imitation [8], which, based on motor imitation literature [9], suggests social relevance. Yet [8] found no social gain from linguistic imitation, questioning whether alignment is considered pro-social. Contrary to that, in our earlier work [10], we found evidence suggesting it plays a role in speech planning. This study directly tests whether perceived cooperativeness affects alignment and social judgments.

Methods: Two experiments used a pseudo-interactive picture-selection task (PST). Participants interacted with two simulated confederates whose cooperativeness was manipulated based on whether they reused participants' lexical choices when naming repeated images. <u>Experiment 1</u> (N = 140, between-participants): On each trial, the Director instructed the Matcher on which image to choose. The task followed a fixed block order: In *Block 1*, participants acted as Directors, describing images that could be described by two ~equally acceptable words (the exposure-critical trials; based on [8]). In *Block 2*, participants acted as Matchers and were instructed by their simulated confederates. In this block, the cooperative speaker always reused participants' descriptions from Block 1, and the non-cooperative speaker never reused them. *Block 2* also included images described using a disfavored term (alignment-critical trials; based on [11]). In *Block 3*, participants acted again as Directors, describing the alignment-critical images. Additionally, we measured participants social judgments regarding their partner, using 6 questions – half about 'warmth' and half about 'competence' (similarly to [12]). <u>Experiment 2</u> (N = 140) was similar to Experiment 1, but in a within-participant design and included a memory test, assessing participants' recall of which confederate used each word.

Results: In <u>Experiment 1</u>, cooperativeness significantly affected alignment: disfavored words were more likely to be used with the cooperative speaker than with the non-cooperative one (*OR* = 1.910, *p* = 0.022; Fig. 1A). Further, the cooperative speaker was rated as 'warmer' (Δ Rating = 0.315, *HDI*: 0.008 – 0.609, *Probability of PD* = 95.47%; Fig. 1B). In <u>Experiment 2,</u> there was no effect of cooperativeness on alignment (*OR* = 1.230, *p* = 0.269; Fig 2A). However, the 'warmth' effect was replicated (to a smaller degree; Δ Rating = 0.093, *HDI:* -0.012 – 0.185, *PD* = 92.95%). Moreover, cooperativeness influenced memory retention, such that the cooperative speaker's words were remembered better (*OR* = 1.179, *p* = 0.019; Fig. 2C). In both experiments, 'warmth' correlated with alignment rates (see Fig. 1C and Fig. 2D).

Discussion: Although Experiment 1 showed greater alignment with a cooperative speaker, this effect did not replicate in Experiment 2, where both speakers were directly contrasted. The difference in the partner-specific alignment effect between within-participants and between-participants design merits further exploration. In both experiments, however, social effects did emerge: cooperative speakers were rated as warmer, and their word choices were better remembered. These results indicate that cooperativeness enhances partner-specific encoding and social impressions. Furthermore, higher warmth ratings were associated with higher alignment rates. Therefore, while cooperativeness does not necessarily directly affect alignment (Ex. 2), cooperative behaviour might make a person be perceived as warmer, and could thus bear social gain, including, at least in isolation (Ex. 1), increased likelihood of alignment.

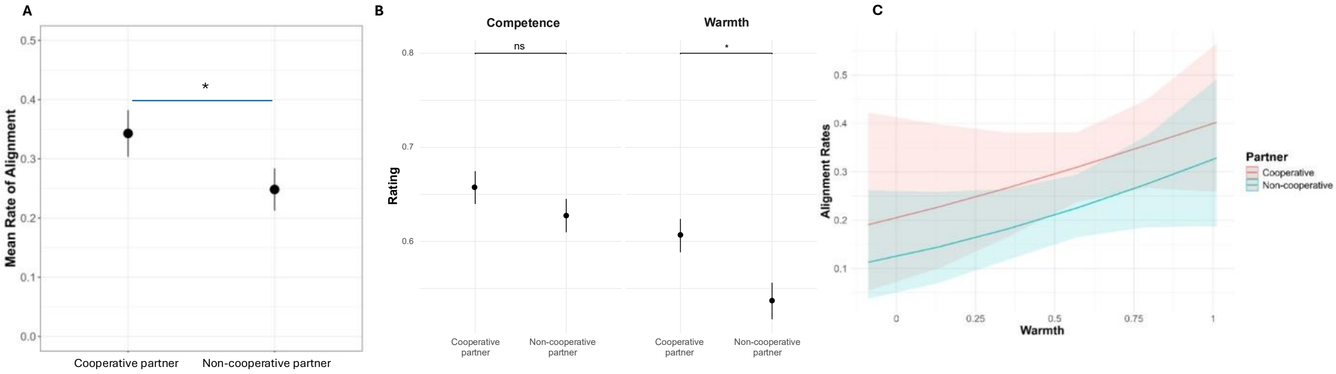


Figure 1. The results from Experiment 1. (A) mean alignment rates by the partner's cooperativeness status (* $p < 0.05$); (B) Competence and Warmth ratings for both partners (*PD > 95%); (C) Alignment Rates by Warmth ratings

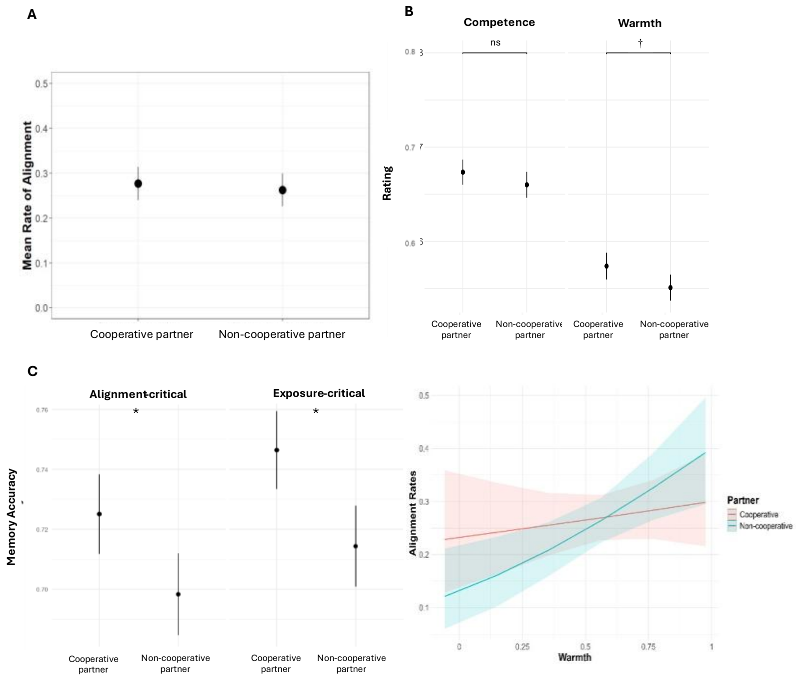


Figure 2. The results from Experiment 2. (A) mean alignment rates by the partner's cooperativeness status; (B) Competence and Warmth ratings for both partners († PD > 90%); (C) Accuracy in the memory test for the words that appeared in the experiment, by partner (* $p < 0.05$); (D) Alignment rates by Warmth ratings.

References

[1] Garrod, S., & Anderson, A. (1987). *Cognition*. **[2]** Brennan, S. E., & Clark, H. H. (1996). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **[3]** Pickering, M. J., & Garrod, S. (2006). *Research on Language and Computation*. **[4]** Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **[5]** Barr, D. J., & Keysar, B. (2002). *Journal of Memory and Language*. **[6]** Horton, W. S., & Gerrig, R. J. (2002). *Journal of Memory and Language*. **[7]** Heller, D., & Brown-Schmidt, S. (2023). *Cognitive Science*. **[8]** Lelonkiewicz, J. R., Pickering, M. J., & Branigan, H. P. (2021). *Royal Society Open Science*. **[9]** Chartrand, T. L., & Bargh, J. A. (1999). *Journal of Personality and Social Psychology.* **[10]** The authors (2024). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **[11]** Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2020). *Journal of Experimental Psychology: Learning, Memory, and Cognition.* **[12]** Beltrama, A., & Papafragou, A. (2023). *Glossa Psycholinguistics*.

The Mid-dot in Gender-Inclusive French : A Reading Study

Sarah Michel, Céline Pozniak , Saveria Colonna
University of Paris 8, Structures Formelles du Langage
sarah.michel12@univ-paris8.fr

In french, the masculine plural nouns as in *"Paul a rencontré les étudiants"* (Paul met the studentsM) can lead to two interpretations : the *specific masculine*, meaning that the group is composed only of men, and the *generic masculine*, meaning that the group is composed of individuals of any gender. However, the use of masculine as generic lead to a male biased interpretation. Inclusive writing (IW) has been proposed as a solution to reduce the male bias induced by masculine plural nouns. Among the various forms of IW, this study focuses on contracted forms using a mid-dot between masculine and feminine endings (*étudiant·e·s*, studentsM·F·PL). One of the main criticism is that such forms impede reading. However, studies have shown that these forms in French only increases reading times at the first occurrence of the contracted forms [1, 2, 3]. Furthermore, [2] suggests that the ease with which contracted forms can be pronounced influences silent reading ; reading times (RTs) would be longer for contracted forms that are difficult to pronounce, such as *chirgurgien·ne* (surgeonM·F).

The main goal of our research is to determine in what extent contracted forms influence RTs, and if so, to identify which factors account for the longer RTs of some forms compared to others. To test whether the contracted forms influence reading, four experiments were conducted. Experiments 1, 2, and 3 used the eye-tracking method, while experiment 4 used the self-paced reading method. Experiment 1 (*N* = 42) aimed to test whether the most common contracted form (*étudiant·e·s*, studentsM·F·PL) affects RTs compared to the feminine form (*étudiantes*, studentsF). Experimental items consisted of 24 plural nouns inserted in sentences with an identical structure (p.3). Experiment 2 (*N* = 41) tested a less pronounceable form (*mécanicien·ne·s*, mechanicsM·F·PL). Experiment 3 (*N* = 38) tested a more pronounceable form (*agriculteur·rice·s*, farmersM·F·PL). Experiment 4 (*N* = 78) compared a more pronounceable form (*chanteur·euse·s*, singersM·F·PL) with a less pronounceable one (*chanteur·se·s*, singersM·F·PL). Analyses were conducted using Bayesian mixed models on the region of interest (the role noun) and the spillover region. For the eye-tracking experiments, three dependent variables were analyzed (first-pass RTs, probability of regressions into the region of interest, and total RTs). For the self-paced reading task, RTs were analyzed.

Experiment 1 showed longer RTs for the *·e* form only during first-pass, not in late measures (Fig.1). Experiment 2 revealed that the *·ne* form had an impact on total RTs (Fig.2). This effect was influenced by self-reported exposure : the more participants reported being exposed to IW, the smaller the RTs. Experiment 3 showed that the *·rice* form was initially difficult to process (Fig.3). However, this reading cost decreased over the course of the experiment. Finally, Experiment 4 showed a difference between the feminine and inclusive forms and that this effect persisted in the spillover region only for the less pronounceable form *chanteur·se·s* (Fig.4) . To conclude, results showed that several factors influenced RTs of contracted inclusive forms. First, familiarity with inclusive writing (IW) influences reading times : frequently used forms had little impact on reading times (Exp. 1), self-reported exposure to IW also affected reading times (Exp. 2), and an adaptation effect was observed as participants became more familiar with the contracted forms throughout the experiment (Exp. 3). Finally, the ease of pronunciation of the contracted forms influenced reading times (Exp. 4).

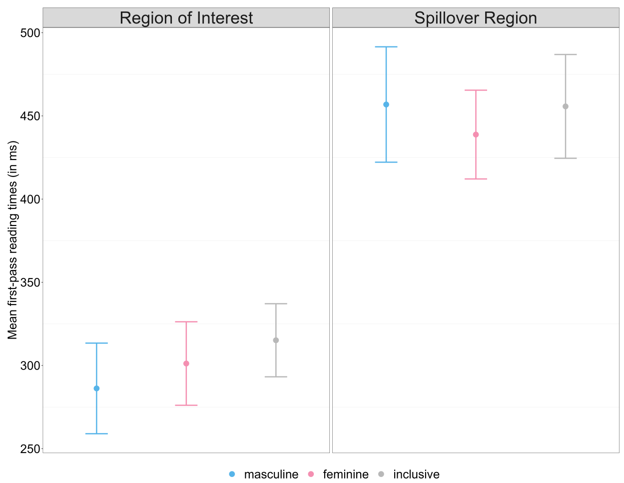


Figure 1 – Mean first-pass RTs for the suffix ·*e*.

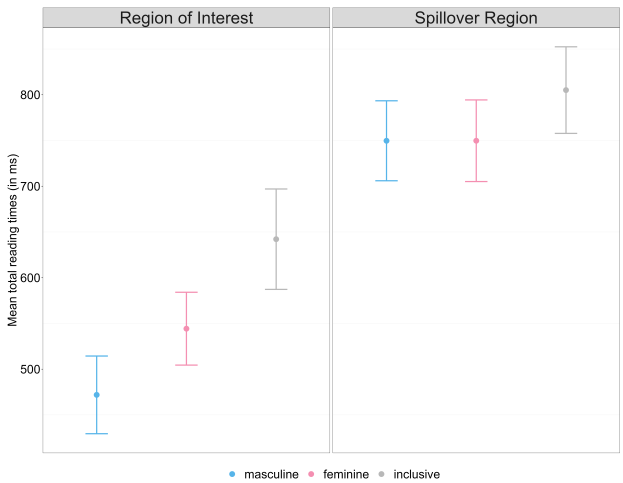


Figure 2 – Mean total RTs for the suffix ·*ne*.

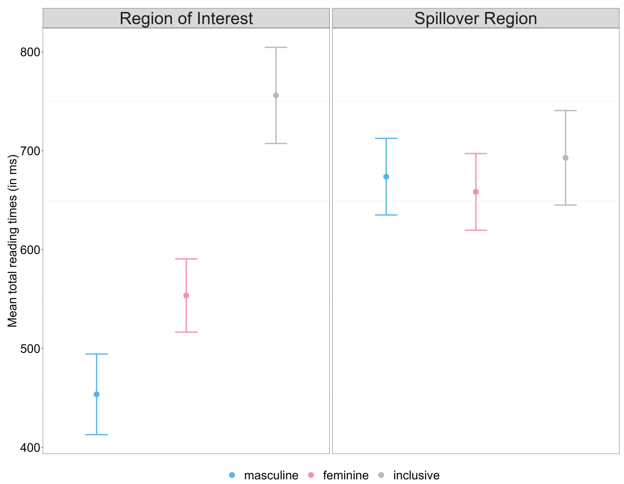


Figure 3 – Mean total RTs for the suffix ·*rice*.

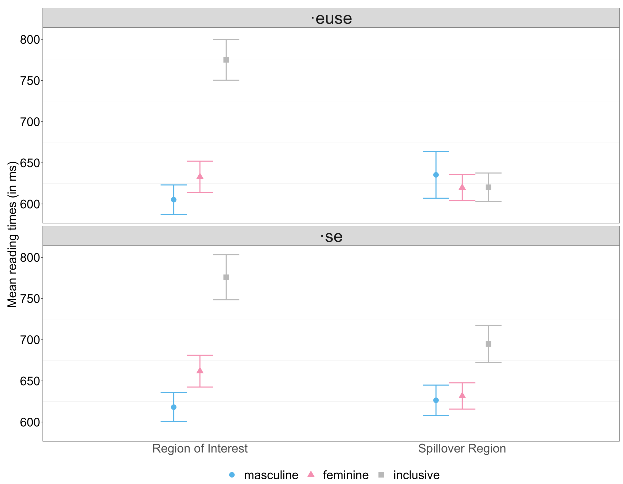


Figure 4 – Mean RTs for both endings, ·*euse* and ·*se*.

Experiment	**Coef.Est**	**P(coef > 0)**	**P(coef < 0)**	**CrI 95%**
Experiment 1	0.035	0.810		[-0.043, 0.114]
Experiment 2	-0.094		0.923	[-0.224, 0.034]
Experiment 3	0.131	0.996		[0.038, 0.224]
Experiment 4	-0.004		0.881	[-0.010, 0.003]

Table 1 – Main results from the mixed-effects models

References

[1] Gygax, P., & Gesto, N. (2007). Féminisation et lourdeur de texte. *L'Année psychologique*, *107*(02), 239. https://doi.org/10.4074/S0003503307002059
Number : 02.

[2] Liénardy, C., Tibblin, J., Gygax, P., & Simon, A.-C. (2023). Écriture inclusive, lisibilité textuelle et représentations mentales. *Discours*. https://doi.org/10.4000/discours.12636

[3] Girard, G., Foucambert, D., & Le Mené, M. (2022). Lisibilité de l'écriture inclusive : apports des techniques d'oculométrie [Type : roceedings of the 2022 annual conference of the Canadian Linguistic Association.].

Experiment 1 : Example of an item presented in masculine (a), feminine (b) and inclusive (c)

(1)	a.	Agathe a sélectionné les	/ étudiants	/ pour le concours	/ d’éloquence.
		Agathe has selected	studentsM	for the contest	of eloquence.
	b.	Agathe a sélectionné les	/ étudiantes	/ pour le concours	/d’éloquence.
		Agathe has selected	studentsF	for the contest	of eloquence.
	c.	Agathe a sélectionné les	/ étudiant·e·s	/ pour le concours	/ d’éloquence.
		Agathe has selected	studentsM·F·PL	for the contest	of eloquence.

Experiment 2 : Example of an item presented in three conditions masculine (a), feminine (b) and inclusive (c)

(2)	a.	Zoé a rémunéré les	/ électriciens	/ après les travaux	/ de rénovation.
		Zoe has paid the	electriciansM	after the work	of renovation.
	b.	Zoé a rémunéré les	/ électriciennes	/ après les travaux	/ de rénovation.
		Zoe has paid the	electriciansF	after the work	of renovation.
	c.	Zoé a rémunéré les	/ électricien·ne·s	/ après les travaux	/ de rénovation.
		Zoe has paid the	electriciansM·F·PL	after the work	of renovation.

Experiment 3 : Example of an item presented in masculine (a), feminine (b) and inclusive (c)

(3)	a.	Lucie a accompagné les	/ agriculteurs	/ au forum	/ annuel.
		Lucie accompanied the	farmersM	to the forum	annual.
	b.	Lucie a accompagné les	/ agricultrices	/ au forum	/ annuel.
		Lucie accompanied the	farmersF	to the forum	annual.
	c.	Lucie a accompagné les	/ agriculteur·rice·s	/ au forum	/ annuel.
		Lucie accompanied the	farmersM·F·PL	to the forum	annual.

Experiment 4 : Example of an item presented in masculine (a), feminine (b) and inclusive with the pronounceable form (c) and the less pronounceable form (d)

(4)	a.	Claire a auditionné	/les chanteurs	/ pour un concert	/ de charité.
		Claire auditioned the	singersM	for a concert	of charity.
	b.	Claire a auditionné	/ les chanteuses	/ pour un concert	/ de charité.
		Claire auditioned the	singersF	for a concert	of charity.
	c.	Claire a auditionné	/ les chanteur·euse·s	/ pour un concert	de charité.
		Claire auditioned the	singersM·F·PL	for a concert	of charity.
	d.	Claire a auditionné	/ les chanteur·se·s	/ pour un concert	/ de charité.
		Claire auditioned the	singersM·F·PL	for a concert	of charity.

The neglected role of sensorimotor information in the processing and representation of polysemous words

Sara Božić[1,2] & Dušica Filipović Đurđević[1,2,3]

[1]Department of Psychology, Faculty of Philosophy, University of Belgrade

[2]Laboratory for Experimental Psychology, Faculty of Philosophy, University of Belgrade

[3]Laboratory for Experimental Psychology, Faculty of Philosophy, University of Novi Sad

sara.andjelic@f.bg.ac.rs

Polysemous words have multiple *senses* that are semantically similar and often share etymology (e.g., *head* of the body, of the nail, of the company, etc.). It is postulated that different polysemous senses, although represented separately, have partially overlapping representations consisting of the shared core and unique domains [1,2]. For that reason, it is expected that the facilitatory priming would occur between two senses. However, experimental evidence from priming studies is equivocal [3–7], thus questioning the theory of a shared core of polysemous senses [4,5].

Therefore, the goal of this research was to shed further light on the representation of polysemous words by focusing on the dynamics of polysemous sense activation in a novel way – by seizing the embodiment cognition framework. Within this framework, the representation of concepts is rooted in the sensorimotor experience [8–10]. Based on that, we predicted that sensorimotor experience shared among polysemous senses would be represented in the core (e.g., visual experience of *a sealed letter* and *an italic letter*), compared to the experience true for one sense only (e.g., tactile experience of *a sealed letter,* but not *an italic letter*). We further hypothesised that the activation of the core can be boosted by focusing on the processing of the shared sensory experience, which in turn would enhance the facilitatory priming. To test the prediction, we selected 30 Serbian polysemous words, and for each word, we selected two sense pairs. Each sense was embedded in a short phrase describing a certain sensory experience (e.g. have seen an italic letter). The phrases were consecutively presented on the screen, and participants (N=102) were instructed to answer by buttonpress if they ever had the experience in question (*Have you ever… [yes/no]*). However, the sequence of phrases was arranged in such a way that in half of the trials, the target sense was preceded by the related sense (primed condition; e.g. *have seen a sealed letter*), and in half of the trials the target sense was preceded by a filler task (unprimed, i.e. control condition; e.g., *A is before B. AB [yes/no]*). In half of the primed trial pairs, the shared sensory modality was focused (true for both word senses; e.g. *have seen a sealed letter – have seen an italic letter*) and in the other half, the sensory modality unique to the prime was focused (e.g. *have touched a sealed letter – have touched an italic letter*). Linear-mixed effects regression to reaction latency showed that targets preceded by the related sense were processed faster than ones preceded by the filler stimulus ($\beta = -.14$, CI [-.17 – -.11], $t = -8.509$, $p < .001$) and the targets referring to the shared modality were processed faster compared to ones referring to the unique modality ($\beta = .08$, CI [.05 – .13], $t = 4.175$, $p < .001$). Crucially, the priming effect was larger in the shared modality condition ($\beta = .05$, CI [.01 – .10], $t = 2.422$, $p < .05$; Figure 1). By targeting specific sensorimotor features of polysemous senses, we managed to dissect the mental representation of polysemous words, by boosting the activation of either the unique (non-shared) or shared (core) part of the mental representation. We managed to evoke a higher activation level of non-intended sense when the shared core of intended sense was activated, compared to when the unique part of intended sense was activated. This differential effect of two modalities corraborates the shared core theory and illustrates the potential of using the embodied cognition framework for better understanding of polysemy representation and processing.

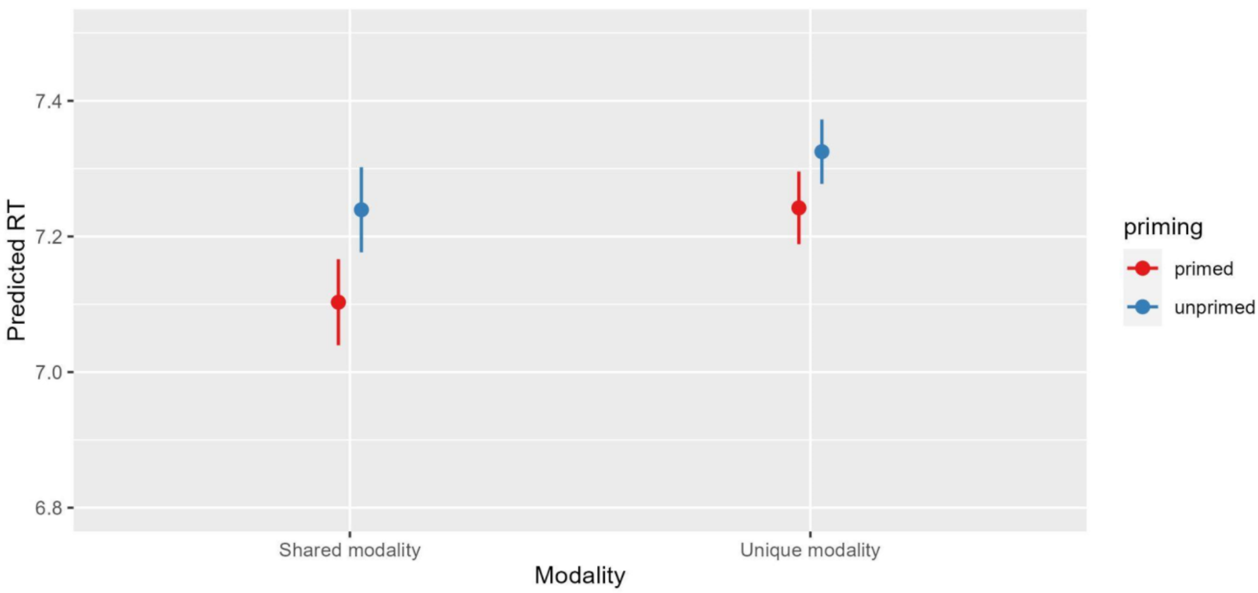


Figure 1. Interaction effect of modality type and priming condition to target reaction latency

References

[1] Rodd, J. M. (2018). Lexical Ambiguity. In Gaskell, M. G. & Rueschemeyer, S.-A. (Eds.), Oxford Handbook of Psycholinguistics. Pre-print: https://osf.io/preprints/psyarxiv/yezc6

[2] Armstrong, B. C. (2012). The temporal dynamics of word comprehension and response selection: Computational and behavioral studies (Unpublished doctoral dissertation). Carnegie Mellon University Psychology Department, Pittsburgh, PA.

[3] Williams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research, 21,* 193-218.

[4] Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45*(2), 259-282.

[5] Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language, 47*(4), 548-570.

[6] Brown, S. W. (2008). Polysemy in the Mental Lexicon. *Colorado Research in Linguistics, 21*.

[7] Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: the comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1534.

[8] Barsalou, L. W. (1999). *Perceptual symbol systems. The Behavioral and brain sciences/Cambridge University Press.*

[9] Barsalou, L. W. (2008). *Grounded cognition. Annu. Rev. Psychol., 59*(1), 617-645.

[10] Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology, 22*(3-4), 455-479.

Inclusive language, then and now: a self-paced reading experiment in Argentina

Ana Zarwanitzer, Santiago Estremero, Gala Esperanza Coronas, Carlos Gelormini-Lezama

Universidad de San Andrés

cgelormini@udesa.edu.ar

Background: In the last decade, inclusive morphemes have emerged in Argentina and other Spanish-Speaking countries—a phenomenon also observed in English and other languages [1, 2, 3, 4]. Experimental data from 2018 in Argentina [5], showed that sentences with inclusive language (using *-e* or *-x* morphemes) yielded longer reading times than sentences using the traditional generic masculine (*-o*). Two interpretations were proposed at the time: (1) the mental grammar of native Spanish speakers rejects inclusive language as unacceptable—a claim often echoed by the Royal Spanish Academy—or (2) the increased reading times reflected processing difficulty due to lower exposure, rather than ungrammaticality.

Methods: We ran a sentence-by-sentence self-paced reading experiment to test processing times of sentences with inclusive language compared to the traditional masculine form to obtain data from 2025. A set of 18 items like the one in Table 1 was constructed. Sentence 1 introduced a context of place/time. Sentence 2 contained an animate noun like *hijos* ("children" in Spanish) in three different conditions: generic masculine indicated by the morpheme *-o*, and two versions of inclusive language, *-e* and *-x*. A set of 18 filler items was also constructed to reduce the predictability of the experimental items. Each item ended with a comprehension question to check for accuracy. Participants were 29 monolingual speakers of River Plate Spanish, 24 women and 5 men, aged 20–40, undergraduate or graduate students at the Universidad de San Andrés, Buenos Aires, Argentina. The dependent variable was the reading time of Sentence 2, where the inclusive or generic masculine morphemes appear.

Results: Table 2 shows the mean reading times of the critical Sentence 2 in all three conditions. A repeated measures ANOVA with Bonferroni adjustment tested the effect of condition (*-o*, *-e*, *-x*) on reading times. Mauchly's test confirmed that the assumption of sphericity was met ($p = .566$). The analysis revealed a significant main effect of condition on reading times, $F(2, 28) = 7.25$, $p = .002$, $\eta^2 = .341$. Pairwise comparisons showed that *-e* forms elicited significantly longer reading times than both masculine (mean difference = 288.45 ms, $p = .001$) and *-x* forms (mean difference = 250.96 ms, $p = .002$). Importantly, no significant difference was found between generic *-o* and *-x* forms (mean difference = -37.49 ms, $p = .679$).

Discussion: Unlike previous findings [5], the *-x* morpheme does not elicit longer reading times, indicating that this type of inclusive form is no longer penalized. Importantly, this study shows that, from the point of view of processing, it is inappropriate to speak of "inclusive language" as a unified category. The inclusive morpheme *-e* may introduce ambiguity, as it can be mistaken with other words in Spanish that are phonologically or orthographically acceptable. In contrast, *-x* clearly breaks these rules, making it more readily identifiable as a marker of inclusive language. Overall, the acceptance of *-x* as the preferred inclusive morpheme shows a shift in its acceptability by native speakers. This, in turn, suggests that the social and language changes underlying the appearance of this inclusive morpheme are a firm ground on which this inclusive morpheme is beginning to consolidate as part of the mental grammar of speakers. Beyond the individual level, it seems as if society is beginning to accept an extremely unusual morphological change, which had not occurred in the Spanish language since the collapse of the Latin neuter gender and the consolidation of the two-gender system in the early Middle Ages [6].

Table 1. Sample experimental item in all three conditions.

Sentence 1	Estuvo lloviendo toda la mañana. *It was raining all morning.*		
	Generic Masculine “o”	**Inclusive Morpheme “e”**	**Inclusive Morpheme “x”**
Sentence 2	Mis **hijos** prefieren quedarse en casa. *My children prefer to stay at home.*	Mis **hijes** prefieren quedarse en casa. *My children prefer to stay at home.*	Mis **hijxs** prefieren quedarse en casa. *My children prefer to stay at home.*

Table 2. Mean reading times in milliseconds for sentence 2 in all three conditions.

	Conditions		
	Generic Masculine “o”	**Inclusive Morpheme “e”**	**Inclusive Morpheme “x”**
Sentence 2	1691.58	1980.03	1729.07

References

[1] Penfield, J. (1987). *Women and language in transition*. State University of New York Press.

[2] MacKay, D. G., & Fulkerson, D. C. (1979). On the comprehension and production of pronouns. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 661–673. https://doi.org/10.1016/S0022-5371(79)90369-4

[3] Vergoossen, H. P. (2019). *Cognitive demands of gender-neutral language: The new genderless pronoun in the Swedish language and its effect on reading speed and memory* [Conference presentation]. 3rd International Convention of Psychological Sciences, Palais des Congrès, Paris, France.

[4] Arnold, J., Mayo, H., & Dong, L. (2020). *Personal pronouns matter: Singular they understood better after explicit introduction* [Conference presentation]. 33rd Annual CUNY Conference on Human Sentence Processing, UMass Amherst, Massachusetts, United States.

[5] Zarwanitzer, A., & Gelormini-Lezama, C. (2023). Tiempos de lectura de oraciones con lenguaje inclusivo en español: un estudio psicolingüístico. *Vertex: Revista Argentina de Psiquiatría, 34*(159), 40–46. https://doi.org/10.53680/vertex.v34i159.366

[6] Penny, R. (2002). A History of the Spanish Language. Cambridge University Press, 2002.

An Emerging Non-Binary Stereotype? An Experimental Assessment of the NB-ness of French Nouns

Emma Kious[1], Gabriel Thiberge[2], Anne Abeillé[1], Céline Pozniak[3], Heather Burnett[1]
[1]LLF Université Paris Cité, [2]CLLE Université Toulouse Jean Jaurès, [3]SFL Université Paris 8

Linguistic choices and judgments are not made in a vacuum: they are subject to interpretation and representational biases, both implicit and explicit. Psycholinguistic measures of these automatic processes show just how permeable we are to extralinguistic factors. Gendered representations are a particularly salient example, given that we encode both grammatical and social gender information during online processing (Irmen 2007). Non-binary (NB) issues are relatively new in the field of inclusive writing studies, and insufficiently documented for French. Brauer (2009), Gygax et al. (2009) and Xiao et al. (2023), among others, found that gender-fair language forms help mitigate male-biased representations and interact with stereotypes for professions.

If inclusive writing "forces reference" beyond androcentrism (Burnett & Pozniak 2021), then such strategies should allow us to access other representations, such as that of NB-related biases. Our goal with this study is twofold: provide a new set of norms to test the inclusiveness of French doublets for role/group nouns, and establish a means of measuring the NB-ness of such nouns.

Building on Misersky et al. (2014), we presented participants (L1 French) with group/role nouns and asked them to estimate proportions of gendered individuals in each group with a set of 3 slider scales (women/men/NBs) from 0 to 100 (Fig.1). Unlike Misersky et al.(2014), who presented a single scale ranging between feminine to masculine plural forms, we presented them above the scales with middot. Overall, we tested 585 nouns, most of which are job-related (N=485), but others are nobility titles (N=9), kinship nouns (N=17), religious roles (N=6), government functions (N=19), and some are not found in Lexique (New et al. 2001) nor tested in Misersky et al. (e.g. *youtubeur 'youtuber'*). Our study was run in two stages. Stage 1 tested 240 nouns, split into 4 lists of 60 items per participant (N=120). Stage 2 tested 345 nouns, split into lists of 69 items per participant (N=155). Our results, reported in Fig.1, are compatible with Misersky et al. (2014), with whom we had 208 nouns in common. We found a strong correlation with our M-scale (ρ=-0.93) and F-scale (ρ=0.93), and a weak correlation with our NB-scale (ρ=0.33). If M and F scales are highly inversely correlated (ρ=-0.96), we found a small correlation between F and NB scales (ρ=0.3). The tendency to estimate a higher proportion of men in a given role/group is also confirmed: 384 nouns received a rating of above 50% on the M-scale, but only 147 on the F-scale. On the NB-scale, 200 nouns are above 10% and only 8 are rated above 20%, a proportion which is likely overestimated (Goodman et al. 2019). What appears clear is that a lack of bias (a score of around 0.5) found by Misersky et al. (2014) is not a predictor of a high estimated proportion of NBs for a given group. In other words, NB-ness is a bias in itself, rather than the absence of a (gendered) bias. Regarding the morphology of nouns, we found that common gender is a significant predictor of higher NB-ness (Est. 2.36, $p<0.001$) and a higher proportion of women (Est. 2.65, $p<0.05$), but predict a lower proportion of men (Est. -2.46, $p<0.05$). When looking at alternating forms in particular, -eur/-euse (e.g. *serveurs/serveuses* 'waiters/waitresses') and -eur/-rice (e.g. *enquêteur/enquêtrices* 'investigators.m/f') endings significantly predict lower NB-ness.

The problematic bipolarity of using a single scale for such measures has been identified by Sato et al. (2025). We show that the representation of gender in groups is more than asymmetrically binary: it is at least ternary. We provided a means to evaluate the NB-ness of role/group nouns in French, and collected data which seems to point to the emergence of an NB-related gender stereotype. We also show that common gender forms mitigate male-bias and open representations of groups to female and NB-ness.

[1] Misersky, J., Gygax, P., & al. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871.

[2] Sato, S., Gygax, P., & al. (2025). Does inclusive language increase the visibility of women, or does it simply decrease the visibility of men? A missing piece of the inclusive language jigsaw. *Collabra:Psych.*, *11*(1).

[3] Elmiger, D. (2015). Masculin, féminin: Et le neutre? *Implications philosophiques*.

[4] Irmen, L. (2007). What's in a (role) name? formal and conceptual aspects of comprehending personal nouns. *Journal of psycholinguistic research*, *36*, 431–456.

[5] Xiao, H., Strickland, B., & al. (2023). How Fair is Gender-Fair Language? Insights from Gender Ratio Estimations in French. *Journal of Language and Social Psychology*, *42*(1), 82–106.

[6] Burnett, H., & Pozniak, C. (2021). Political dimensions of gender inclusive writing in parisian universities. *Journal of Sociolinguistics*, *25*(5), 808–831.

[7] Gygax, P., Gabriel, U., & al. (2009). Some grammatical rules are more difficult than others: The case of the generic interpretation of the masculine. *European journal of psychology of education*, *24*, 235–246.

[8] Brauer, M. (2008). Un ministre peut-il tomber enceinte? l'impact du générique masculin sur les représentations mentales. *L'année psychologique*, *108*(2), 243–272.

[9] Goodman, M., Adams, N., & al. (2019). Size and distribution of transgender and gender nonconforming populations: A narrative review. *Endocrinology and Metabolism Clinics*, *48*(2), 303–321.

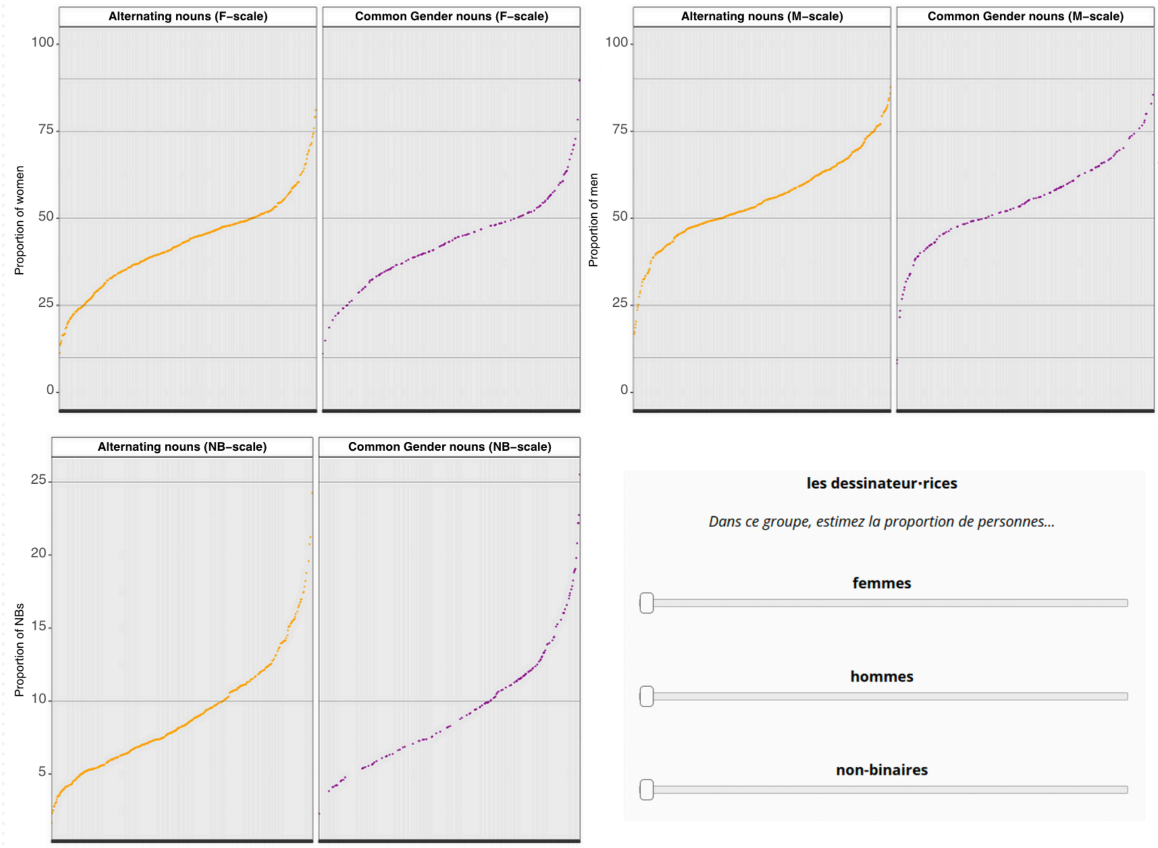


Figure 1: Estimated proportions per scale depending on noun morphology: alternating (N=444, e.g. *serveurs/serveuses ‘waiters/waitresses’*) *vs* common gender nouns (N=139, e.g. *guides ‘guides’*).

The bottom-right figure is an example item for *dessinateur·rices ‘drawers’*.

Does the verb type modulate the ERPs for Thematic Reversal Anomalies? The case of Subject and Object experiencer verbs in Malayalam.

S Shalu[1], R Muralikrishnan[2] & Kamal Kumar Choudhary[1]

[1]Indian Institute of Technology, Ropar, [2]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

Shalu.20hsz0009@iitrpr.ac.in ; R.Muralikrishnan@aesthetics.mpg.de ; kamal@iitrpr.ac.in

Recent studies have shown that the ERP responses associated with Thematic Reversal Anomalies (TRA) vary depending on several factors: the importance of word order for role assignments [1], verb type differences [1, 2, 3], methodological parameters [3], and the typological features of the language under consideration [2]. Among these, verb type has emerged as a key factor in modulating the ERP correlates, and the differences based on verb type remain consistent. TRA research on verb-medial languages with different verb types has shown that verb-based linking properties [1] and the thematic aspects related to verb types [2, 3] play a crucial role in these processes. Several previous studies have compared activity verbs and subject experiencer (SE) verbs. However, object experiencer (OE) verbs, which are different in terms of how grammatical relations (subject, object) map on to thematic roles (agent/experiencer, theme/stimulus), have attracted little attention in TRA studies. Most studies that compared different verb types were on English, in which it is challenging to compare OE with other verbs: the anomaly is only realized at the sentence-final object position for OE verbs, whereas it occurs already at the verb (i.e., sentence-medially) for other verbs. A verb-final language would be an ideal choice for examining verb-type differences, as in these languages, the TRA for all verbs - including OE verbs - would be realized at the sentence-final verb, allowing for a more consistent comparison. To this end, we investigated the processing differences between SE and OE verbs in Malayalam through the present ERP study. We employed a 2x2 design, manipulating the verb type (subject versus object experiencer), and the animacy of the subject, such that the experiencer and theme roles are either correct according to the requirements of the verb, or constituted a reversal anomaly at the verb. EEG from 36 first-language speakers of Malayalam (mean age: 27.08; 17 female; 19 male) was recorded when they read the critical sentences (Table1) and performed acceptability judgement and probe detection tasks. ERP analyses at the verb revealed that reversal anomalies of SE and OE verbs evoked a negativity effect (350-550 ms) followed by a late-positivity effect (750-950 ms) as opposed to their correct counterparts. Our results stand in contrast to findings from Turkish, which is typologically similar to Malayalam [1], but are in line with previous findings from German [4]. In Malayalam, the parser relies on case marking to assign thematic roles. On encountering the verb, its logical structure reveals the required mappings of grammatical function of the arguments (based on case and animacy) to thematic roles. When these requirements are not met, a negativity effect ensues. The subsequent late-positivity can be interpreted as reflecting a domain general, binary categorization of well-formedness [1]. Our results show that both types of experiencers verbs, namely SE and OE verbs, elicit qualitatively similar ERP correlates for TRA, suggesting a uniform processing strategy for thematic reversal anomalies irrespective of the type of experiencer verb involved.

Keywords: Sentence processing, Thematic Reversal Anomaly, subject experiencers, object experiencers, Malayalam

Supplementary material.

Condition	Sample Stimulus	Acceptability (in %)
Non-TRA (SE)	രവി ഇരുട്ടിനെ സ്നേഹിച്ചു. Ravi iruttine-e snehichu Ravi darkness (Inm)-Acc love-PST “Ravi loved darkness.”	77.9 (41.5)
TRA (SE)	ഇരുട്ട് രവിയെ സ്നേഹിച്ചു. Irutt raviye snehichu Darkness ravi(Anm)-Acc love-PST Darkness loved Ravi.	31.6(46.5)
Non-TRA (OE)	ഇരുട്ട് രവിയെ പേടിപ്പിച്ചു. Irutt ravi-e pedippichu Darkness ravi(Anm)-Acc frighten-PST. Darkness frightened Ravi.	78(41.5)
TRA (OE)	രവി ഇരുട്ടിനെ പേടിപ്പിച്ചു. Ravi iruttine pedippichu Ravi darkness frighten-PST. Ravi frightened darkness.	25.6(43.7)

References

[1] Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*(3), 133–152.
[2] Bourguignon, N., Drury, J. E., Valois, D., & Steinhauer, K. (2012). Decomposing animacy reversals between agents and experiencers: An ERP study. *Brain and Language, 122*(3), 179–189.
[3] Kyriaki, L., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Semantic reversal anomalies under the microscope: Task and modality influences on language-associated event-related potentials. *European Journal of Neuroscience, 52*(7), 3803–3827.

[4] Schlesewsky, M., & Bornkessel-Schlesewsky, I. D. (2009). When semantic P600s turn into N400s: On cross-linguistic differences in online verb-argument linking. In *Papers from Brain Talk*. Lund University, Media-Tryck, Lund.
[5] Krishnan, G. G. (2020). *Malayalam morphosyntax: Inflectional features and their acquisition*.
[6] Nizar, M. (2010). *Dative subject constructions in South Dravidian languages* (Unpublished master's thesis). University of California, Berkeley.
[7] De Swart, P. J. F. (2007). *Cross-linguistic variation in object marking*. Utrecht: LOT.
[8] Jayan, J. P., & Kumar, J. S. (2008). *Levin's classification on Malayalam verbs*.
[9] Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2019). *Ethnologue: Languages of the World* (22nd ed.). SIL International. Retrieved from https://www.ethnologue.com
[10] Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge University Press.

Malayalam Language

Malayalam is an agglutinative South Dravidian language with a canonical SOV order, spoken mainly in the state of Kerala and union territories Lakshadweep and Puducherry, belonging to the Indian Union, spoken by about 34.8 million speakers as their first language [9]. Malayalam emerged as a separate language from the west coast dialect of Tamil around the 9th century [10], and is characterized by its rich inflectional morphology, including a three-way tense marking system (verbs are inflected for present, past and future), three genders (masculine/feminine/neuter) and two numbers (singular/plural). Both the subject and object arguments can be omitted from a sentence if they can be understood from the context [5]. In Malayalam, like in other Dravidian languages, grammatical relations and semantic roles are indicated through a series of case suffixes. Thus, changing the word order in a sentence typically does not change its meaning, as the roles and relations are primarily conveyed through these suffixes. The case system of Malayalam includes seven cases: nominative, accusative, dative, sociative, locative, genitive and instrumental. Malayalam exhibits non-nominative subject constructions [6] and differential object marking, whereby only animate objects are marked accusative, with inanimate objects marked accusative only as an exception [7]. The feature that makes Malayalam distinct from other sister languages in the Dravidian language family is the absence of subject-verb agreement.

Malayalam verbs can be categorized into various types, with one significant type being experiencer verbs or psych verbs [8]. These verbs encompass the semantic notions related to experiencing, wanting, feeling, liking, etc. Malayalam has both subject experiencer verbs and object experiencer verbs, and these differ in terms of the realization of the experiencer argument. A subject experiencer in Malayalam may be expressed as a nominative subject or a dative subject depending upon the verb and the type of experience it expresses, whereas object experiencers are always expressed as accusative arguments. All sentences in our study had nominative subjects and accusative objects.

Speech disfluencies and implicit word learning: Fluency shapes preference, not performance

Emma Libersky[1], Kimberly Crespo[2], Margarita Kaushanskaya[1]
[1]Univeristy of Wisconsin—Madison, [2]Boston University
libersky@wisc.edu

Background: Speech disfluencies are a common feature of spontaneous language, occurring approximately 6 times per 100 words.[1] Studies of language processing have revealed that participants use one type of disfluency—filled pauses—to anticipate low-frequency, discourse-novel, or otherwise "difficult" words.[2] However, the consequences of filled pauses for word learning are less well understood. A large body of literature has indicated that instructor disfluency impacts students' instructor ratings.[3] However, few studies regarding the impact of disfluency on word learning have been conducted, and they have exclusively focused on explicit word learning processes.[4, 5] Here, we focus on the impact of disfluency on adults' word learning, using an implicit cross-situational word learning paradigm.

Method: In Experiment 1, 62 monolingual, English-speaking adults listened to audio recordings introducing two speakers: one who produced an elevated rate of filled pauses and one who did not. Both speakers were introduced as knowledgeable and familiar with the target pseudowords. The speakers then named novel objects with English-like pseudowords in a cross-situational word learning task; both speakers fluently labeled the novel objects. After being tested on their retention of the novel word-object mappings, participants were asked to rate the two speakers and indicate which speaker they preferred to learn from in the future. In Experiment 2 (N = 59), the speaker introduction was removed, and disfluencies were embedded directly into the cross-situational word learning task. As a consequence, participants in Experiment 2 were not explicitly informed that the speakers were familiar with the pseudowords. In addition to the experimental task, participants filled out demographics and language background questionnaires and completed measures of English language ability and sustained attention.

Results: The analytical approach was the same across experiments. Learning data were analyzed at the item level via logistic regression in R. We regressed accuracy on speaker fluency (fluent vs. disfluent), using a model comparison approach to select the best model. Participants learned above chance in both conditions, across experiments, but neither experiment yielded a significant effect of fluency on learning. Speaker ratings were analyzed via t-tests. In both experiments, participants rated the fluent speaker as significantly more effective, knowledgeable, prepared, and organized (see Tables 1 and 2), and indicated that they would prefer to learn a new set of words from the fluent speaker over the disfluent speaker.

Discussion: This study is the first to examine the impact of disfluency on implicit word learning. Across two experiments, we found that adults learned novel words equally well from fluent and disfluent speakers but demonstrated a preference for the fluent speaker. Results suggest that cross situational word learning is robust against speaker disfluencies. Taken together, our findings indicate that disfluency shapes attitudes towards instructors but does not appear to disrupt how learners track co-occurring regularities between objects and labels over time.

References

[1] Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738.

[2] Bosker, H. R., Tjiong, J., Quené, H., Sanders, T. J. M., & De Jong, N. H. (2015). Both native and non-native disfluencies trigger listeners' attention. *Proceedings of Disfluency in Spontaneous Speech 2015*, 1–4.

[3] Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. Journal of Experimental Psychology: Applied, 22(2), 161–172. https://doi.org/10.1037/xap0000077

[4] Libersky, E., Neveu, A., & Kaushanskaya, M. (2023). One fish, uh, two fish: Effects of fluency and bilingualism on adults' novel word learning. *Psychonomic Bulletin & Review*, *30*(2), 763–773. https://doi.org/10.3758/s13423-022-02189-8

[5] White, K. S., Nilsen, E. S., Deglint, T., & Silva, J. (2020). That's thee, uuh blicket! How does disfluency affect children's word learning? *First Language*, *40*(1), 3–20. https://doi.org/10.1177/0142723719873499

Figure 1. Predicted probability of correctly selecting the target object in each experiment as a function of condition. Learning performance did not differ across conditions.

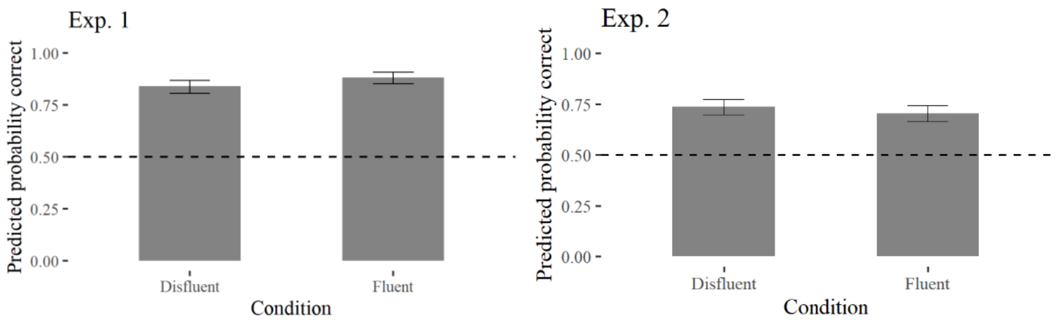


Table 1. t-test results for speaker ratings, Experiment 1. For all questions, the fluent instructor was rated more favorably than the disfluent instructor.

Dependent Variable	Disfluent Mean (SD)	Fluent Mean (SD)	t	df	p	d	CI_lower	CI_upper
Effective	2.95 (0.91)	4 (0.77)	7.66	61	1.70e-10	0.97	0.67	1.27
Knowledgeable	3.27 (0.81)	3.94 (0.77)	5.35	61	1.42e-06	0.68	0.40	0.95
Organized	2.71 (0.93)	4.11 (0.7)	8.77	61	2.14e-12	1.11	0.79	1.43
Prepared	2.5 (1.13)	4.21 (0.77)	9.06	61	6.85e-13	1.15	0.83	1.47

Table 2. t-test results for disfluent vs fluent speaker ratings, Experiment 2. For all questions, the fluent instructor was rated more favorably than the disfluent instructor.

Dependent Variable	Disfluent Mean (SD)	Fluent Mean (SD)	t	df	p	d	CI_lower	CI_upper
Effective	2.97 (1.16)	3.95 (1.02)	-4.95	58	6.74E-06	-0.64	-0.92	-0.36
Knowledgeable	3.08 (1.12)	4.03 (0.76)	-6.42	58	2.76E-08	-0.84	-1.13	-0.54
Organized	2.83 (1.09)	4.03 (0.81)	-6.99	58	3.10E-09	-0.91	-1.21	-0.6
Prepared	2.69 (1.25)	4.24 (0.8)	-8.82	58	2.63E-12	-1.15	-1.47	-0.82

Velocity is key:
Morphological structure affects planning and execution stages in typing

Julia Muschalik
English Language and Linguistics, Heinrich Heine University Düsseldorf
julia.muschalik@hhu.de

Background: A growing body of research indicates that typing as a form of language production is sensitive to linguistic properties and structure similar to other production modalities [1]. With its unique combination of language, visual processing and motor control, typing appears to be particularly suitable for researching core psycholinguistic topics [2], such as the processing of morphologically complex words [3]. While it has been demonstrated that morphological structure affects error performance [4] and typing timing [5, 6], evidence is often inconclusive about magnitude, directionality, and locus of the found effects [7]. A recent study [8] finds no effect of morphology on execution timing (inter-key intervals) for English pseudowords. Their results suggest that either processing units other than morphemes dominate in typing or that morphological processing occurs during earlier production stages, before motor execution is initiated [9]. The present study probes these results for English real words, investigating whether morphological structure exerts an influence during initial stages (i.e. on response latencies) or execution stages (i.e. on execution timing) or both.

Method: We tested 301 native speakers of English in an online single word typing task, using a two-tier decision and recall paradigm. Participants typed words that contained one of four identical target letter sequences (s-e, m-e, t-e, and k-e) across disyllabic morphologically complex words and monosyllabic simplex words, spanning syllabic, morphemic or no linguistic boundary. To pinpoint potential effects, we measured both writing onset times (WOT) as well as inter-key intervals (IKI). The retained data (n = 41173 key presses across 11384 single words) were analyzed using linear mixed-effects regression in R following standard procedures [10] with logWOT and logIKI as dependent variables.

Results: We find an effect of morphological complexity for both WOT and IKI (see Figures 1 and 2). Our results suggest that typing complex words involves additional processing, as reflected in delayed WOT for inflected words compared to both derived and simplex words, and in inflated IKI at morpheme boundaries relative to syllable boundaries and transitions without underlying linguistic boundaries. In addition, WOT in complex words appears to be less delayed in faster typers, while their execution timing is more sensitive to morphological structure, suggesting that increased motor automation allows more resources for linguistic processing.

Discussion: Morphological structure influences typing timing, both during planning and execution stages. Different typers employ different strategies and there seems to be a trade-off between planning and execution. Additionally, increased motor processing seems to overshadow effects of linguistic processing. Our results suggests that morphological effects might not be categorical but rather gradient and highly individual. Among other things, our results highlight the importance of accounting for individual differences in language production and reiterate the need for more cross-modal comparison. Typing data seems to be a particularly well-suited testbed for both. With its unique features, research on typing can offer new insights into the nature of linguistic representation in production to shed further light on more general patterns in language processing and the intricate interplay between language modalities.

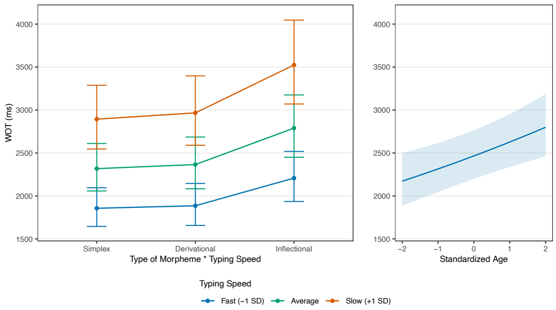


Figure 1 Partial effects plot of the main effects for the WOT model.

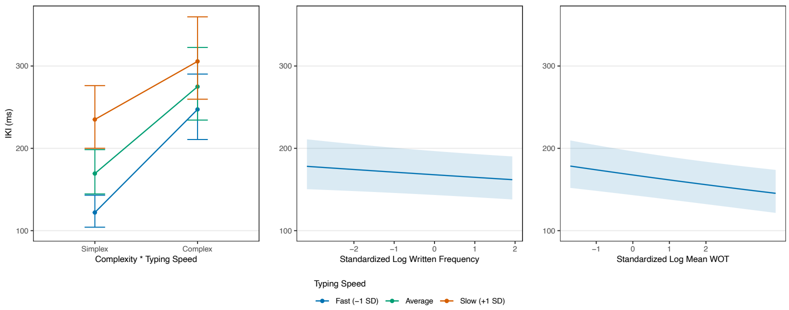


Figure 2 Partial effects plot of the main effects for the IKI model.

References

[1] Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). typing is writing: Linguistic properties modulate typing execution. Psychonomic Bulletin & Review, 23(6), 1898–1906.

[2] Nozari, N., Pinet, S., & Muylle, M. (2025, March 4). What can typing tell us about language production?. https://doi.org/10.31234/osf.io/s8gnd_v1

[3] Libben, G. (2020). What can we learn from novel compounds? In S. S. Walde & E. Smolka (Eds.), The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective (pp. 107–127). Berlin: Language Science Press.

[4] Gahl, S., & Plag, I. (2019). Spelling errors in english derivational suffixes reflect morphological boundary strength. The Mental Lexicon, 14(1), 1–36.

[5] Will, U., Nottbusch, G., & Weingarten, R. (2006). Linguistic units in word typing. Written Language & Literacy, 9(1), 153–176.

[6] Gagné, C. L., & Spalding, T. L. (2016). Written production of English compounds: Effects of morphology and semantic transparency. Morphology, 26(2), 133–155.

[7] Gagné, C., Spalding, T. L., & Taikh, A. (2023). Impact of morphology on written word production: An overview of empirical evidence and theoretical implications. In D. Crepaldi (Ed.), Linguistic Morphology in the Mind and Brain (pp. 96–107). London: Routledge.

[8] Muschalik, J., Schmitz, D., Kakolu Ramarao, A., & Baer-Henney, D. (2024). Typing /s/—morphology between the keys? Reading and Writing. https://doi.org/10.1007/s11145-024-10586-9

[9] Lambert, E., & Quémart, P. (2019). Introduction to the special issue on the dynamics of written word production: methods, models and processing units. Reading and Writing, 32(1), 1–12.

[10] Torrance, M., & Conijn, R. (2024). Methods for studying the writing time-course. Reading and Writing, 37(2), 239–251.

The role of expectedness in L1 and L2 ditransitive prediction in Mandarin-English late bilinguals and heritage bilinguals

Vera Yunxiao Xia, Johanne Paradis, and Juhani Järvikivi
University of Alberta
yunxiao4@ualberta.ca

Background: Structural priming studies, mainly with monolinguals, have shown that language users make syntactic predictions based on incoming input; greater predictive changes result when input is less expected, relative to prior input experiences.[1, 2] Though late bilinguals (LBs) and heritage bilinguals (HBs) can share the same L1 and L2, they have had different experiences in each language: LBs became bilingual as adults and received more L1 input, whereas HBs became bilingual as children and received more L2 input. Thus, the same new input may have different levels of expectedness to LBs vs. HBs. Therefore, we asked whether LBs and HBs show different patterns of prediction after priming, in both of their languages.

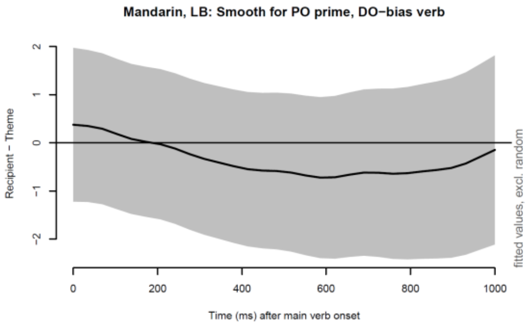
Method: Our ongoing Visual World eye-tracking study investigates how L1 Mandarin L2 English LBs and HBs predict ditransitives in Mandarin and English (blocked design, counterbalanced). Participants (currently 33 LBs, 20 HBs) first listened to a prime sentence with either double-object (DO) or prepositional-object (PO) structure, containing a verb biased toward the same structure (expected prime) or the opposite (unexpected prime). Then, they listened to a target ditransitive while viewing images of the agent, theme, and recipient. Each language had 16 prime–target pairs. We analyzed the empirical logit-transformed recipient advantage (proportion of looks to recipient minus proportion of looks to theme) within 1000 ms after onset of each target verb using a generalized additive mixed model. Significant changes over time indicated prediction; higher recipient advantage indicated DO prediction, and lower indicated PO. Independent variables were Language (Mandarin, English), Group (LB, HB), Prime (DO, PO), Prime Verb Bias (DO, PO) and Time (continuous). As well, separate proficiency tests confirmed that the LBs were stronger in Mandarin, and the HBs in English.

Results: For each group's stronger language (Mandarin for LBs and English for HBs), only unexpected PO primes (PO primes with DO-biased verbs) induced changes in recipient advantage. However, the LBs were primed by structure (PO prediction, Fig. 1a), whereas the HBs were primed by verb bias (DO prediction, Fig. 2b). Both groups showed less predictable patterns for their weaker languages (English for LBs and Mandarin for HBs). The LBs demonstrated PO prediction after expected DO primes (Fig. 2a), as did the HBs (Fig. 1b), the opposite of a priming effect; the HBs also demonstrated DO prediction after unexpected DO primes (Fig. 1c). Interestingly, neither group demonstrated changes over time after expected PO primes in either Mandarin or English; we note that PO is the most common type of verb bias in both languages.

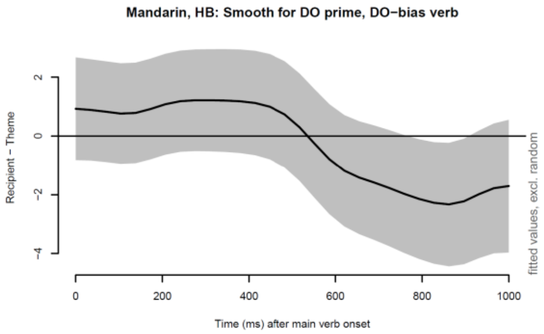
Discussion: Results so far suggest that both LBs and HBs are sensitive to expectedness in primed prediction, with the most expected input being unable to induce prediction. Where priming effects are present, LBs' and HBs' predictions can demonstrate an overall inverse pattern, likely due to inverse proficiencies and input experiences in each language. This explains why only unexpected primes induce prediction in the LBs' Mandarin and the HBs' English (each group's stronger language). Each group's weaker language may have less explicable patterns due to inconsistent exposure to the relevant structures or verbs from participant to participant. When prediction does occur in any language, however, whether the prediction is based on prime structure or prime verb bias is variable. We speculate that this may reflect individual differences in how participants' predictive strategies integrate lexical compared to syntactic information.

Figures

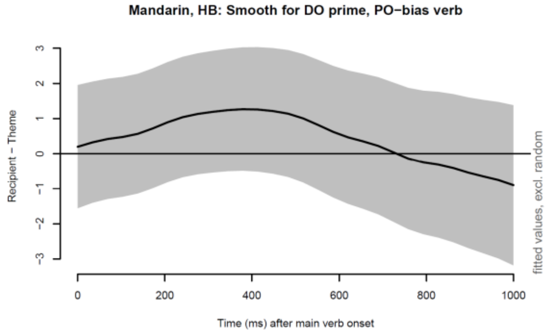
Figure 1. L1 Mandarin smooths showing significant changes in recipient advantage over time: (a) LBs, PO-prime, DO-bias verb (p = .04); (b) HBs, DO prime, DO-bias verb (p < .001); (c) HBs, DO prime, PO-bias verb (p = .03). Shading represents 95% confidence intervals.



a)

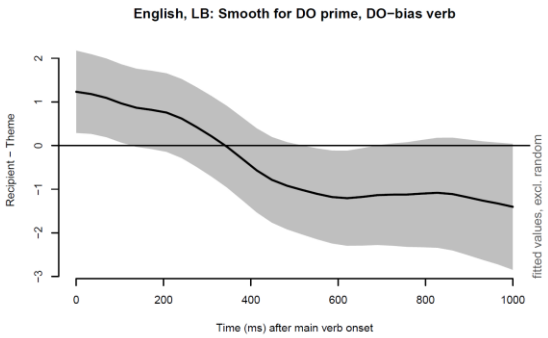


b)

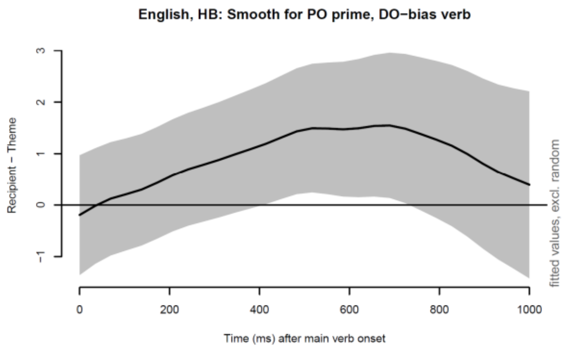


c)

Figure 2. L2 English smooths showing significant changes in recipient advantage over time: (a) LBs, DO prime, DO-bias verb (p = .02); (b) HBs, PO prime, DO-bias verb (p = .01). Shading represents 95% confidence intervals.



a)



b)

References

[1] Chen, X., Wang, S., & Hartsuiker, R. J. (2022). Error-based structure prediction in language comprehension: Evidence from verb bias effects in a visual-world structural priming paradigm for Mandarin Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 60–71. https://doi.org/10.1037/xlm0001048

[2] Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*(1), 57–83. https://doi.org/10.1016/j.cognition.2012.10.013

Compositional Parsing in Adjective-Noun Phrases: The Role of Adjective Semantics

Oliwia Iwan, Eva Wittenberg
Department of Cognitive Science, Central European University, Vienna, Austria
iwan_oliwia@phd.ceu.edu

A key challenge in language comprehension is how individual word meanings integrate during composition. Minimal adjective-noun phrases (e.g., "red square", "big boat") provide a valuable window into this process. Some theoretical accounts propose that adjective-noun composition follows a uniform compositional mechanism [1-4], while others emphasize variability depending on adjective type [5-7]. We tested whether the semantic type of adjectives, specifically intersective (e.g., "red") versus subsective (e.g., "big"), influences integration efficiency when matching adjective-noun phrases to visual objects.

We conducted three pre-registered visual verification experiments (n = 128 each, 4 blocked trial types) where participants judged whether linguistic descriptions matched on-screen images. **Experiment 1** (N = 48, in-lab) investigated intersective adjectives, replicating Bocanegra et al. [8], with word order manipulations (e.g., “red square”, “square red”, “red”, “square”). **Experiment 2** (N = 48, web-based) used the same task structure with subsective adjectives (e.g., “big car”, “car big”, “big”, “car”). **Experiment 3** (N = 98, online) removed word order manipulations, using intersective-only (e.g., “orange”), subsective-only (e.g., “big”), and canonical adjectivenoun phrases (e.g., “orange boat”, “big boat”). This within-subject design enabled a direct comparison of intersective and subsective adjectives combined with identical noun contexts. Participants responded via keypress to indicate image-description matches, with randomized trial order and counterbalanced blocks. **Experiments 1 and 2** replicated a compositional advantage (i.e., faster responses to two-word phrases) for both intersective and subsective adjectives (Exp 1: $F(1, 47) = 15.07$, $p < .001$, $\eta_g^2 = .036$; Exp 2: $F(1, 47) = 8.66$, $p = .005$, $\eta_g^2 = .015$; see Figure 1A and 1B, respectively.). Crucially, **Experiment 3** revealed a dissociation: only intersective adjectives showed a compositional advantage. The repeated-measures ANOVA showed a significant main effect of feature condition ($F(1, 97) = 4.75$, $p = .032$, $\eta_g^2 = .047$), a robust main effect of adjective type ($F(1, 97) = 38.58$, $p < .001$, $\eta_g^2 = .285$), and a marginal interaction between the two factors ($F(1, 97) = 3.47$, $p = .066$, $\eta_g^2 = .035$). To clarify the interaction pattern, exploratory follow-up contrasts using a linear mixedeffects model showed that the compositional advantage was significant for intersective adjectives (estimate = –0.033, SE = 0.0058, $p < .001$), but absent for subsective adjectives (estimate = 0.0007, SE = 0.0060, $p = .90$; see Fig. 2).

These findings inform whether a unified formal account can predict how adjective-noun composition works in the mind. **Experiments 1** and **2** showed that both intersective and subsective adjectives can elicit a compositional advantage, with broadly similar patterns of results. This suggests that, under the right conditions, subsective adjectives can support efficient integration, approximating the behavior of intersective adjectives. However, **Experiment 3** revealed a clear contrast: only intersective adjectives maintained the advantage when both types were tested within participants. While subsective adjectives can elicit compositional facilitation in isolation, their advantage disappears in mixed contexts, indicating greater contextual dependence. The absence of word order effects across experiments highlights a dominant role for semantic over syntactic cues. Together, the results underscore how interpretive constraints (rooted in semantic type) influence real-time language comprehension.

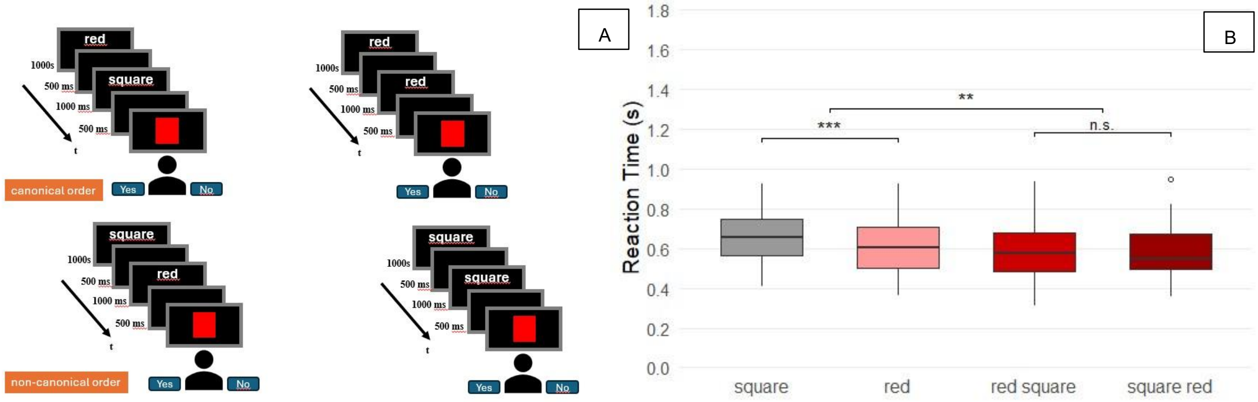


Figure 1
(*A*) Trial structure for **Experiment 1**. Participants read a linguistic description (e.g., “red square”) and judged whether it matched the visual object shown on screen. Each trial consisted of a fixation cross, followed by the linguistic stimulus and then a picture of a colored shape. Participants responded via keypress. This task structure was consistent across all three experiments, with variations in adjective type and stimulus content. (*B*) Mean reaction times per condition in **Experiment 1** with intersective adjectives (in seconds, N = 48). Reaction times were trimmed to exclude values below 0.1 s and above 2 s. Asterisks indicate statistically significant pairwise differences based on repeated-measures ANOVA and post hoc tests

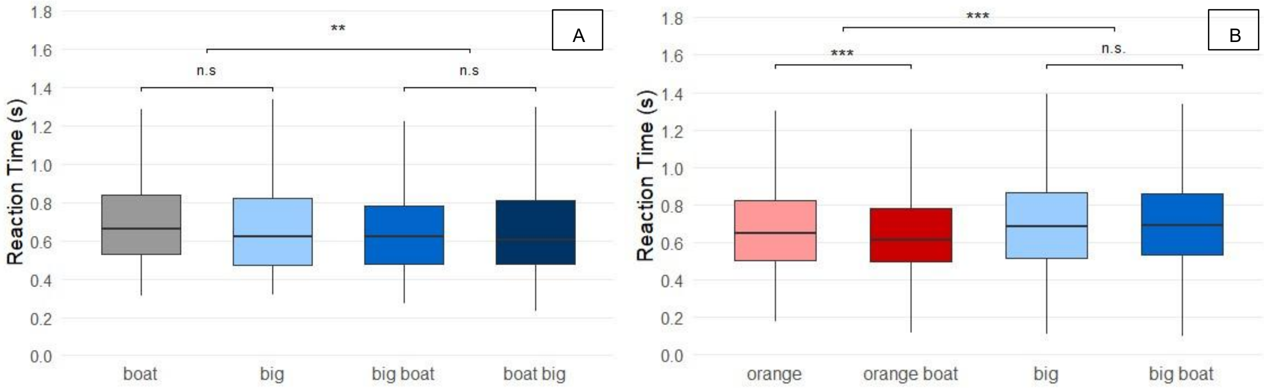


Figure 2
(A) Mean reaction times per condition in **Experiment 2,** with subsective adjectives (in seconds, N = 48). Reaction times were trimmed to exclude values below 0.1 s and above 2 s. Asterisks indicate statistically significant pairwise differences based on repeated-measures ANOVA and post hoc tests.
(B) Mean reaction times (in seconds) per condition for **Experiment 3** (N = 98) with intersective and subsective adjectives. Reaction times were trimmed between 0.1 s and 2 s. Asterisks indicate significant pairwise differences from repeated-measures ANOVA and post hoc tests.

References

[1] Fritz & Baggio, *Language, Cognition and Neuroscience*, 2022; [2] Abdullah & Frost, *Advances in Artificial Intelligence (Canadian AI)*, 2005; [3] Vecchi, Marelli, Zamparelli & Baroni, *Cognitive Science*, 2017; [4] Kamp, *Meaning and the Dynamics of Interpretation*, 2013; [5] Drašković, Pustejovsky & Schreuder, *Advances in Generative Lexicon Theory*, 2012; [6] Ziegler & Pylkkänen, *Neuropsychologia*, 2016; [7] Cinque, *Studies in Chinese Linguistics*, 2014; [8] Bocanegra, Poletiek & Zwaan, *Journal of Memory and Language*, 2022.

The Impact of Semantic Distance on Multiple Phonological Activation in Chinese Speech Production: Evidence from a Picture-Word Interference Study

Panpan Bi[2], Cheng Wang[2], Chen Feng[1]

[1]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.

[2]School of Psychology, Zhejiang Normal University, Jinhua 321004, China

Email: fengc@psych.ac.cn

Background: Word production involves a transition from semantic to phonological representations. A central debate in this process concerns whether non-target phonological information is activated. While **serial models** suggest a strict sequence from lexical selection to phonological encoding[1],, **cascaded models** propose overlapping activation[2,3]. Studies in Indo-European languages largely support cascaded activation[4], but evidence from Chinese remains limited and mixed[5,6,7], raising questions about cross-linguistic generalizability. One key factor influencing non-target activation is semantic distance: semantically closer words exert stronger interference, suggesting higher activation levels[8,9]. This, in turn, may impact whether phonological information of non-target words is detectable. Based on this, the present study examined how semantic distance modulates multiple phonological activation in Chinese speech production.

Method: We used a picture–word interference paradigm in which participants named pictures while ignoring superimposed distractor words. The distractors were either semantically related (same category) or phonologically related to a semantic neighbor (mediated related). Semantic related pairs were further divided into close and distant groups based on conceptual proximity. For each of these, a corresponding mediated condition was created by pairing the target with the phonological relative of the semantic distractor. Each related condition had a corresponding unrelated condition created by re-pairing targets and distractors to remove associations. Two experiments were conducted: Experiment 1a (N=48) manipulated semantic distance between subjects, while Experiment 1b (N=24) used a within-subjects design to control for individual variability on semantic distance.

Results: Results of the two experiments consistently demonstrated that semantic distance modulated semantic interference: in the close-distance condition, semantically related distractors elicited greater naming delays than unrelated ones (23.4 ms vs. 17.8 ms), indicating stronger competition between semantically proximal concepts. Crucially, multiple phonological activation—indexed by slower responses to mediated distractors—emerged only in the far-0distance condition, lending support to the cascaded model. The absence of such activation in the close-distance condition suggests that semantic competition may suppress phonological co-activation under high semantic overlap.

Discussion: The findings of the present study have important theoretical implications for models of speech production. First, they demonstrate that semantic distance modulates the magnitude of semantic interference. Second, they suggest that semantic distance also influences the activation level of non-target words in Chinese: only phonological information associated with semantically distant non-target words receives sufficient activation to be detected.

Figure 1. Mean response latencies (ms) by distractor word type, semantic distance, and relatedness condition in the Experiment 1a

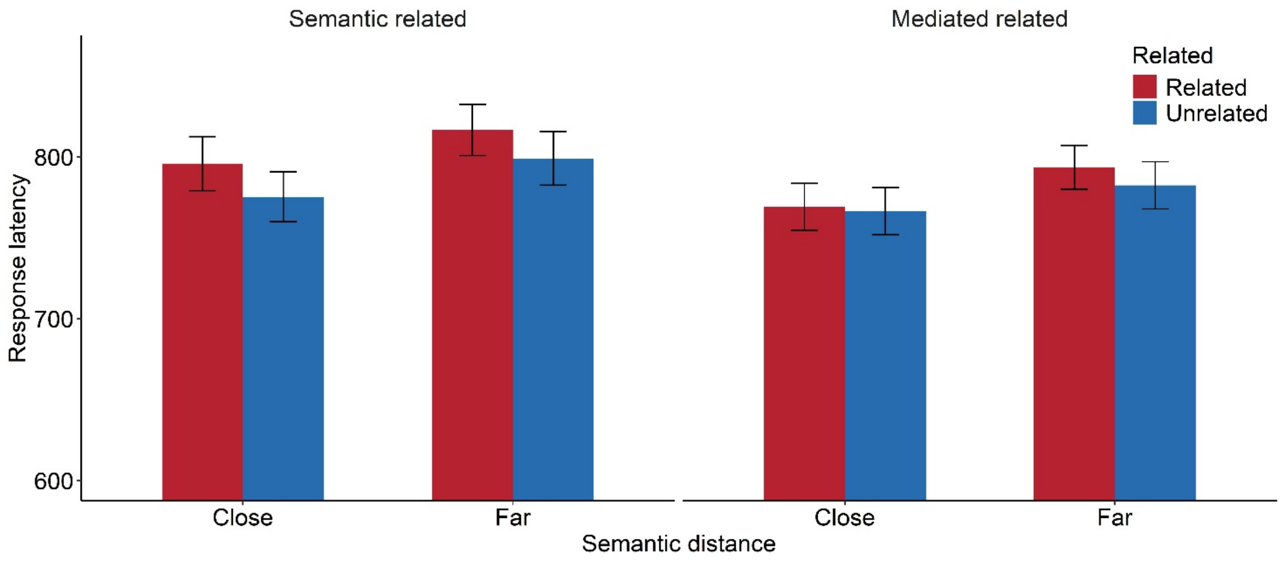
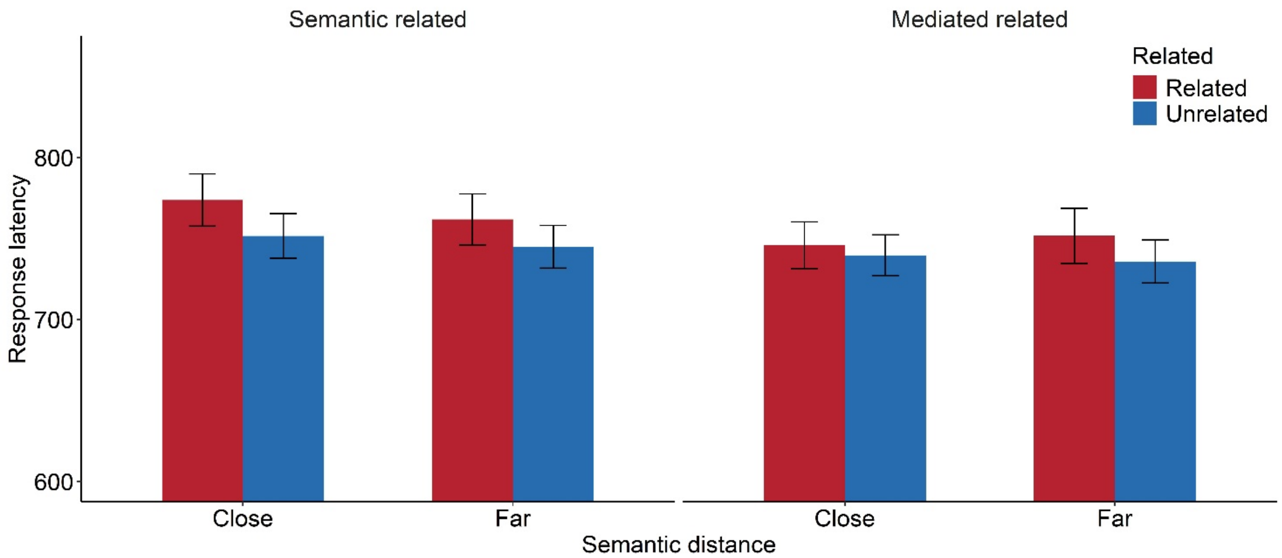


Figure 2. Mean response latencies (ms) by distractor word type, semantic distance, and relatedness condition in the Experiment 1b



References

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, 5(1), 67-104.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359-377.

Zhang, Q., Zhu, X., & Damian, M. F. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in Mandarin. *Applied Psycholinguistics*, *39*(5), 835-860.

Zhuang, J., & Zhou, X. (2003). The interaction between semantics and phonology in the speech production of Chinese. *Acta Psychologica Sinica*, *35*(03), 300.

Bao, L., Qian, Z., & Zhang, Q. (2023). The multiple phonological activation in Chinese spoken word production: An ERP study supporting cascaded model. *Behavioural Brain Research*, *451*, 114523.

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 753.

Vieth, H. E., McMahon, K. L., & De Zubicaray, G. I. (2014). Feature overlap slows lexical selection: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology*, *67*(12), 2325-2339.

Evidentiality and Speaker Commitment: An fEMG Study

Harriet Yates, Corien Bary, Peter de Swart, Bob van Tiel
Radboud University, Nijmegen
harriet.yates@ru.nl

Background This study explores how different syntactic forms of evidential constructions affect perceived speaker commitment, using facial electromyography (fEMG) to detect subtle affective responses from interlocutors.

When reporting speech, speakers not only convey information but also signal their degree of accountability for that information. Evidential constructions serve important pragmatic functions by modulating a speaker's perceived commitment to the truth of their propositional content. This study aims to gather reliable empirical data on whether different syntactic forms of evidentials impact the degree of speaker commitment: specifically, embedding ("I hear that Tom is a bachelor") versus parenthetical ("Tom is a bachelor, I hear") constructions.

Theoretical accounts on this diverge. Some argue that both embedding and parenthetical evidentials express a core proposition and source attribution, with commitment strength depending (probabilistically) on the source, rather than syntactic form [4]. In contrast, others claim that structure matters: parentheticals foreground the propositional content, aligning the speaker more closely with its truth [3, 1]. In this view, contradicting a parenthetical (e.g., "Tom is a bachelor, I hear. But he is married.") would be less acceptable than doing so after an embedding form.

This study uses facial electromyography (fEMG), the recording of facial muscle activity in Muscle Unit Action Potentials, to measure interlocutor's affective, evaluative responses. Earlier research shows that activation in the *corrugator supercilii* (the inner-eyebrow "frowning muscle") is a reliable indicator of negative affect triggered by social norm violation [2] and for commitment violations in the case of assertions [5]. We extend this method to evidential structures.

Methodology A 3x1 design contrasts three forms of propositional content (plain assertion vs. embedding vs. parenthetical) followed by incongruent continuations. We include a control condition of assertion and congruent follow-up, for qualitative comparison. Fifty-seven native Dutch speakers (43 female, 12 male, 2 other; age range = 18–71 years; *M* = 26) read 96 Dutch vignettes (see Table 1): 16 per condition, 32 fillers, plus 12 comprehension questions.

Results Corrugator activation in the critical period was expressed as a proportion of baseline muscle activity per trial (see Figure 1). We found a significant difference in mean activation measured over the whole critical period, between the assertion-incongruent and both kinds of evidentials ($\hat{\beta}$ = 6.10, p = .036), but not within the evidentials ($\hat{\beta}$ = 0.22, p = .95).

This supports the view that evidentials reduce perceived commitment, without providing evidence for differential effects [4]. The results offer novel empirical evidence on syntactic variation in evidentials and speaker commitment, advancing our understanding of how linguistic structure shapes pragmatic interpretation.

Experimental Conditions		
	Context:	Jan and Natasha are chatting about their colleague Tom. Jan says:
assertion		“Tom is a bachelor. But he is married.”
embedding evidential		“I hear that Tom is a bachelor. But he is married.”
parenthetical evidential		“Tom is a bachelor, I hear. But he is married.”
Control Condition		
assertion (congruent)		“Tom is a bachelor. He is single.”

Table 1: (Translated from Dutch) example items for all conditions

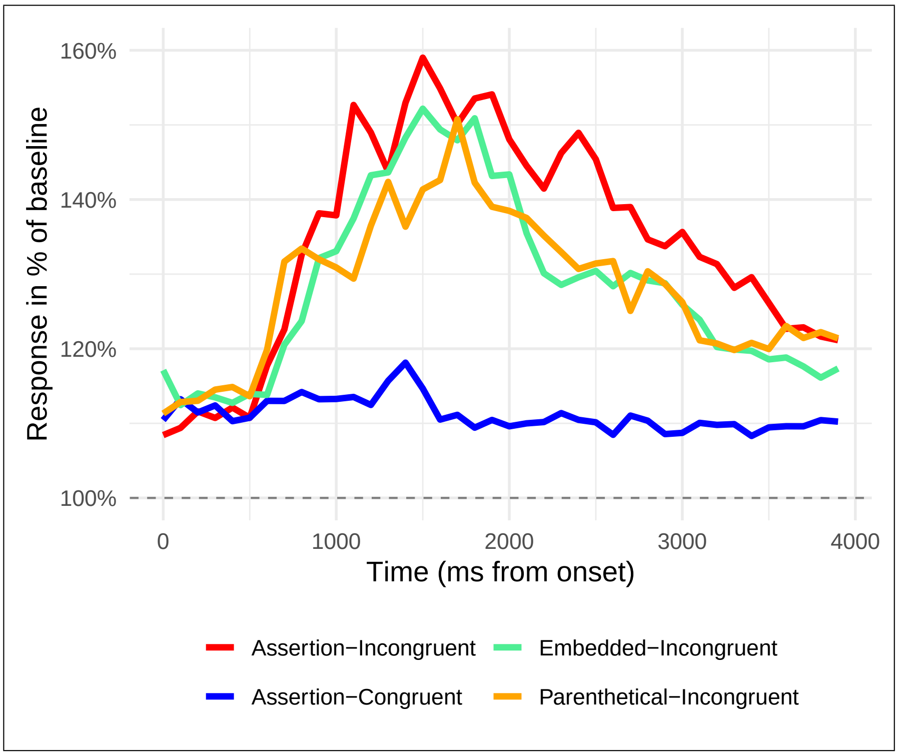


Figure 1: Timeplot of critical (follow-up) period

References

[1] Nicholas Asher. Truth conditional discourse semantics for parentheticals. *Journal of Semantics*, 17(1):31–50, 2000.

[2] Bruce D. Bartholow et al. A psychophysiological examination of cognitive processing of and affective responses to social expectancy violation. *Psychological Science*, 12:197–205, 2001.

[3] Ray Jackendoff. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA, 1972.

[4] Todor Koev. Parentheticality, assertion strength, and polarity. *Journal of Semantics*, 38(1):1–41, 2021.

[5] Harriet Yates, Corien Bary, Peter de Swart, and Bob van Tiel. femg as a window into conversational commitments. In *Proceedings of Sinn und Bedeutung 29*, 2025. Available at https://lingbuzz.net/lingbuzz/008885.

Sign language processing in deaf early signing children – an ERP study

Janika Stille[1], Anne Wienholz[2], Annika Herrmann[2], Ivo Weber[2] & Barbara Hänel-Faulhaber[1]

[1] Faculty of Education, University of Hamburg; [2] Faculty of Humanities, University of Hamburg

janika.stille@uni-hamburg.de

Background: To mark syntactic information, sign languages use the three-dimensional signing space. For example, once referents are located in the signing space, agreement verbs may move between these locations and thus mark the subject and object of an utterance solely by the starting and end point of the movement. Event-related brain potentials (ERP) have not yet been used to investigate how deaf children process syntactic violations. For deaf adults, it has been shown that similar neural correlates occur during sign language processing of semantically or syntactically anomalous sentences when compared with hearing adults for spoken languages [e.g., 1, 2]. Regarding deaf children, it is currently unknown if similar neural correlates are present when they perceive sign language. Hearing children listening to a spoken language have been described to demonstrate an N400 for semantic violations and a P600 for morphosyntactic violations compared to correct sentences, just like adults. However, children typically exhibit a later and larger N400 and a later and smaller P600 than adults [3, 4, 5]. We expected that deaf children would show comparable neural correlates with an N400 for semantic violations and a P600 for morphosyntactic violations.

Method: To investigate the neural processes engaged during sign language processing of different verb types in deaf children, we conducted an ERP study. Twenty-six children with German Sign Language (DGS) as their first language (mean AoA = 10 months [0-30 months]), mean age = 9;11 years [8;7-11;11 years]) watched videos of signed DGS sentences. Here, we focus on a subset of our experiment, i.e., sentences containing agreement verbs (N=45), which were grammatically correct or contained either semantic (implausible object) or morphosyntactic violations (incorrect direction of movement). To maintain the children's attention, they received a probe verification task. ERPs were recorded from 32 scalp electrodes.

Results: We analyzed mean amplitudes from the point in time when the handshape of the verb sign was fully recognizable (target handshape) until 1000ms in 100ms time windows using generalized additive mixed effects models. For each time window, the data was modeled as a function of electrode position and experimental condition. Our data demonstrates a higher negativity at central-posterior locations that was particularly pronounced between 500-600ms for semantic violations compared to correct sentences. For morphosyntactic violations compared to correct sentences, we detected a central posterior distributed negativity between 600-700ms comparable to the effect for the semantic violations.

Discussion: Deaf early signing children demonstrated an N400 effect for sentences containing semantic violations in a time window later than in adults, as previously shown for hearing children. However, the effects for morphosyntactic violations were rather unexpected suggesting that deaf children process this violation more like a semantic violation. Comparable results for specific agreement violations in deaf adults have been reported in some other sign language studies [1, 6]. We will discuss other potential explanations including task effects or handling thematic role assignment [7, 8].

References

[1] Capek, C. M., Grossi, G., Newman, A. J., McBurney, S. L., Corina, D. P., Röder, B., & Neville, H. J. (2009). Brain systems mediating semantic and syntactic processing in deaf native signers: Biological invariance and modality specificity. Proceedings of the National Academy of Sciences, 106(21), 8784–8789. https://doi.org/10.1073/pnas.0809609106

[2] Hänel-Faulhaber, B., Skotara, N., Kügow, M., Salden, U., Bottari, D., & Röder, B. (2014). ERP correlates of German Sign Language processing in deaf native signers. BMC Neuroscience, 15(1), 62. https://doi.org/10.1186/1471-2202-15-62

[3] Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. Journal of Cognitive Neuroscience, 16(7), 1302–1318. https://www.doi.org/10.1162/0898929041920504

[4] Benau, E. M., Morris, J., & Couperus, J. W. (2011). Semantic processing in children and adults: Incongruity and the N400. Journal of Psycholinguistic Research, 40(3), 225–239. https://www.doi.org/10.1007/s10936-011-9167-1

[5] Schneider, J. M., & Maguire, M. J. (2019). Developmental differences in the neural correlates supporting semantics and syntax during sentence processing. Developmental Science, 22(4), e12782. https://www.doi.org/10.1111/desc.12782

[6] Hosemann, J., Herrmann, A., Sennhenn-Reulen, H., Schlesewsky, M., & Steinbach, M. (2018). Agreement or no agreement. ERP correlates of verb agreement violation in German Sign Language. Language, Cognition and Neuroscience, 1–21.

[7] Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. Frontiers in Psychology, 8, 1327. https://ww.doi.org/10.3389/fpsyg.2017.01327

[8] Kos, M., Vosse, T., Van Den Brink, D., & Hagoort, P. (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. Frontiers in Psychology, 1, 222. https://doi.org/10.3389/fpsyg.2010.00222.

Focus Shifts in Contextual and Lexical Cue Interactions in GPT Models

Wonil Chung, Keonwoo Koo, Myung-Kwan Park

Dongguk University

wonil@dongguk.edu

Transformer-based large language models (LLMs), such as GPT (Generative Pre-trained Transformer), look very similar to human language in many ways by employing expectation-driven processing during their pre-training based on human language data. This approach enables them to predict and generate text by anticipating the next word or phrase based on prior context, closely mirroring human-like language understanding. To evaluate whether GPT models process language similarly to humans, researchers are increasingly examining not only grammatical structures but also the subtle ways in which meaning is constructed and conveyed. For instance, language understanding involves managing dependencies like *wh*-dependencies (e.g., *who, what*) that establish relationships between question words and their referents. These dependencies are crucial for processing questions, especially when the referent extend across multiple clauses or sentences, requiring models to maintain a coherent discourse-level understanding. In addition to these dependencies, lexical focus cues like *only* play a key role by signaling emphasis and exclusivity, influencing how different parts of a sentence are interpreted. These semantic phenomena involve subtle shifts in meaning that present a significant test of a model's language processing abilities, as they require attention to both syntactic structure and contextual cues simultaneously.

To detect this processing contextual and lexical cues to focus in transformer-based language models, this study investigates how GPT models process double-object constructions containing contextual and lexical focus cues, followed by remnant continuations, as well as the syntactic and semantic biases these models acquire during pre-training. Adopting a psycholinguistic framework, we employed experimental materials from Sauermann, Filik, and Paterson (2013) and collected word-by-word surprisal (negative log probabilities) assigned by GPT-2 (small, medium, large, XL) and GPT-Neo. In current study, we focus on three primary research questions: (1) To what extent do GPT models capture *wh*-dependency in double-object constructions? (2) How do these models represent the interaction between *wh*-dependencies and lexical focus when processing elliptical remnants in double-object constructions, as indicated by changes in surprisal values? (3) How does the placement of lexical focus within double-object constructions affect the processing of elliptical remnants in the context of *wh*-dependency across different GPT models?

To answer these questions, we design three experiments that treat neural network language models as subjects in a psycho/neurolinguistic experiment. The experimental results provide insight into how contextual and lexical focus cues interact the processing of elliptical remnants in GPT models. Across the experiments, the dominant influence of contextual focus was evident through WhFocus effects, with distinct differences between match and mismatch conditions. Notably, Experiment 3 revealed significant LexicalFocus effects when the lexical focus appeared closer to the remnant, suggesting that proximity plays a key role in guiding the model's interpretation. Consistent with Wilcox et al. (2023), who investigated model generalizations through surprisal analysis, our results show that autoregressive language models not only acquired the fundamental *wh*-dependency between interrogative forms and their corresponding answers in double-object constructions, but also encoded sensitivity to lexical focus cues. These patterns closely parallel human processing outcomes reported by Sauermann et al. (2013).

Reference

Hale, John. 2001. “A probabilistic earley parser as a psycholinguistic model.” In *Second meeting of the north american chapter of the association for computational linguistics*.

Levy, Roger. 2008. “Expectation-based syntactic comprehension.” *Cognition* 106.3:1126-1177.

Sauermann, Antje, Ruth Filik, & Kevin B. Paterson. 2013. Processing contextual and lexical cues to focus: Evidence from eye movements in reading. *Journal Language and Cognitive Processes*, 28(6), 875-903.

Wilcox, Ethan, Richard Futrell, and Reoger Levy. 2021. “Using computational models to test syntactic learnability.” *Proceedings of the Annual Meeting of the Cognitive Science Society* 43.

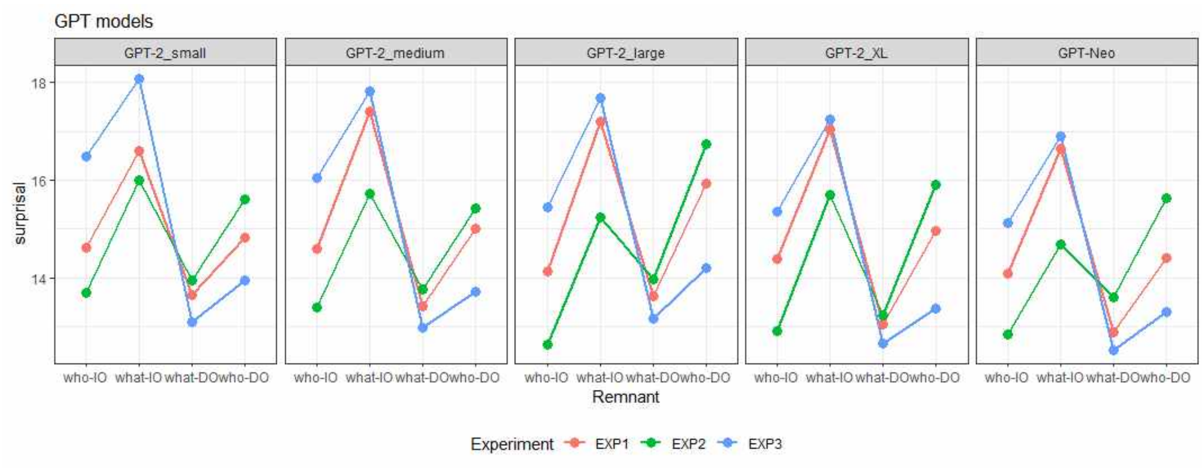


Figure 1 Sensitivity to the focus structure of the double object construction in GPT models

Table 1 An example of 4 conditions. Critical remnant regions are highlighted.

A. WhFocus(match IO-remnant), LexicalFocus(match IO-remnant, EXP 2)/ LexicalFocus(mismatch IO-remnant, EXP 3)

John wondered who Sally would pass the apples. Sally passed [only] the children [only] the apples but not the **grownups**, because they did not want them.

B. WhFocus(mismatch IO-remnant), LexicalFocus(match IO-remnant, EXP 2)/ LexicalFocus (mismatch IO-remnant, EXP 3)

John wondered what Sally would pass the children. Sally passed [only] the children [only] the apples but not the **grownups**, because they did not want them.

C. WhFocus(match DO-remnant), LexicalFocus(mismatch DO-remnant, EXP 2)/ LexicalFocus(match DO-remnant, EXP 3)

John wondered what Sally would pass the children. Sally passed [only] the children [only] the apples but not the **cherries**, because they did not want them.

D. WhFocus(mismatch DO-remnant), LexicalFocus(mismatch DO-remnant, EXP 2)/ LexicalFocus (match DO-remnant, EXP 3)

John wondered who Sally would pass the apples. Sally passed [only] the children [only] the apples but not the **cherries**, because they did not want them.

Online cataphoric pronoun resolution in L1-, L2- and L3-Mandarin: the Maze task

Ryoko Uetomi[1], Leah Roberts[1], and Heather Marsden[1]
[1]University of York
ru537@york.ac.uk

Background: While the Maze is known to capture frequency effects in incremental sentence processing with high locality [1], its application to discourse processing in multilingualism is relatively new. Addressing this gap, we use the task to explore how L1- , L2-, and L3-Mandarin speakers resolve cataphoric dependencies with Mandarin third-person pronouns *ta* in biclausal sentences such as:

(1) Dang ta$_F$ kan jingse de-shihou, Mary zai-he pijiu. ("ta$_F$" = "她 (she)")
During she look scenary when, Mary PRG-drink beer.
'While she looks at scenary, Mary is drinking beer.'

Although cataphoric coreference between pronouns (*ta$_F$* 'she') and sentence-internal referents (*Mary*) is syntactically licensed, some languages like Mandarin, Japanese, and Russian disallow this [2, 3]. This only applies to sentences with temporal subclauses and thus is viewed as a language-specific discourse-related constraint. In online processing (self-paced reading) by L1-Russian speakers, the constraints emerged as a gender 'match' penalty; readers slowed down when main subjects and preceding pronouns unexpectedly 'matched' in gender, at three words after the main subject [3]. This was later than a syntax-triggered gender 'mismatch' penalty observed in languages like English without such constraints. Thus, in online L1-Mandarin processing, a similar penalty is expected after main subjects due to the high locality of the task and the nature of the constraint. The same is expected from L2/L3-Mandarin learners, if target-like.

Method (Appendices A-B): We used coreference judgement (offline) and maze (online) tasks with L1-Mandarin, L1-English L2-Mandarin (L2), and L1-Japanese L2-English L3-Mandarin (L3) speakers. The offline task gives contexts with two gender-matched entities, followed by a target sentence like (1), and probes pronoun referents from sentence-internal, sentence-external, or 'either' (A). In the online Maze task, contexts with two gender-mismatched entities are given, followed by target sentences like (1) but designed in two conditions: gender-matched (encouraging coreference x 8 items) or gender-mismatched (encouraging disjoint x 8 items) mixed with 64 fillers (B). Participants proceeded through the sentences word-by-word, by selecting the word that makes sense from pairs of alternatives. Reaction times at and after main subjects were compared.

Results (Figures 1-2): The L1 group (n=36) accepted sentence-internal coreference 6.6% of the time. Online, they had longer RTs in gender-matched than gender-mismatched conditions in both regions (p=.004 and p=.02 from linear mixed-effect models fit to log-transformed RTs). The L2 group (n=21) accepted sentence-internal coreference 71% of the time and showed no between-condition RT differences. The L3 group (n=28) accepted sentence-internal coreference 9.8% of the time and showed no between-condition RT differences.

Discussions: The gender match penalty observed in the L1 group reflects their offline dispreference for sentence-internal coreference, and is captured earlier by the Maze than in the previous L1-Russian self-paced reading study. This suggests the possibility of active search for gender-mismatched entities before and in addition to the application of the discourse constraint. The L2 group was non-target-like both offline and online, presumably due to influence from the lack of constraint in their L1. The L3 group, though target-like offline, did not show a gender match penalty online as the L1 group did. We discuss this in relation to possible limitations in non-native online processing and the possible counteract impact from the other prior language – L2 English.

References:

[1] Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41(1), 163–171.

[2] Su, Y.C., Sorace, A., Hirose, Y., Mochizuki, K., Hirakawa, M., Nakayama, M., Chen, C.H., & Hsieh, H.Y. (2025). Backward anaphora in L2 acquisition of English, Mandarin and Japanese. *Second Language Research*. https://doi.org/10.1177/02676583251320031

[3] Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, 63(2), 371–400.

Figure 1. By-group Referent Choice Distribution in the Offline Coreference Judgement Task

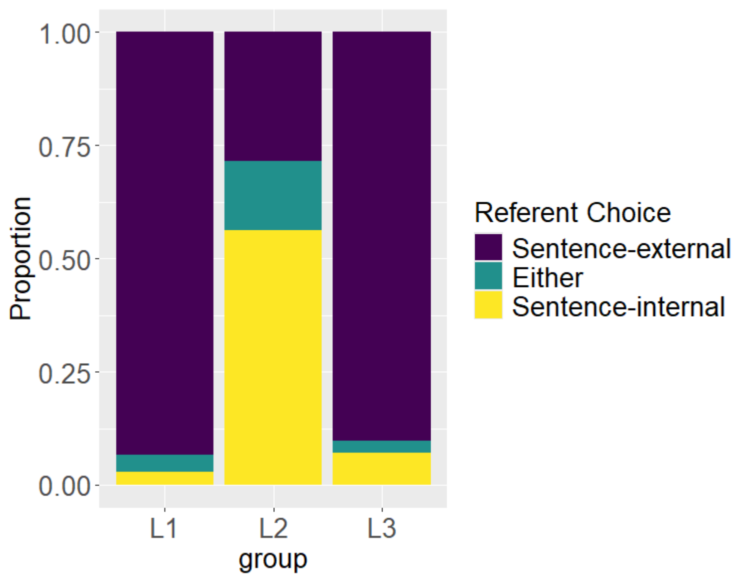
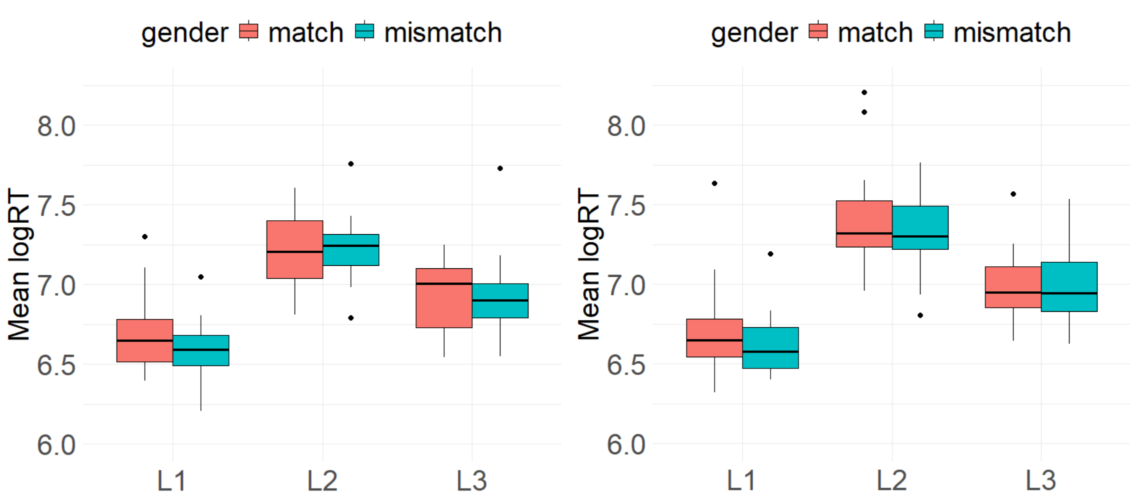


Figure 2. By-group Between-Condition Mean LogRTs per Subject at Each Region in the Maze task

(a) Main subject region

(b) Region after the main subject



Appendices: Example Items (in Bold) with English Gloss and Translation

Appendix A: Example Item for Offline Coreference Judgement Task

Context: **Mary he Lilly zuo huoche qu Beijing.**
Mary and Lilly take train go Beijing.
'Mary and Lilly go to Beijing by train.'

Target: **Dang ta$_F$ kan jingse de shihou, Mary zai-he pijiu.** ("ta$_F$" = "她 (she)")
During she look scenery when, Mary PGR-drink beer.
'While she looks at scenery, Mary is drinking beer.

Question: **Shei zai-kan jingse?**
Who PGR-look scenery?
'Who is looking at scenery?

Answer options (presented in the randomised order):

1. **Mary** (sentence-internal referent)
2. **Lilly** (sentence-external referent)
3. **Mary or Lilly** (either)

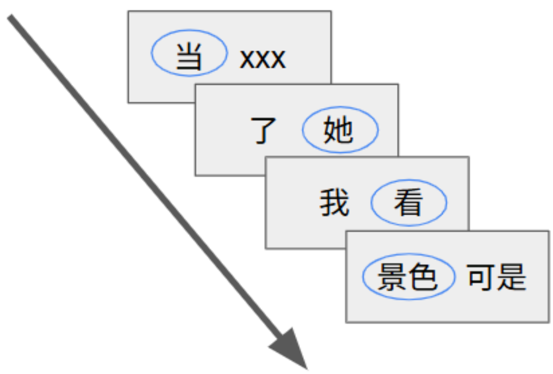
Appendix B: Example Item for the Maze Task and How the Task Proceeds

Context: **Mary he Jack zuo huoche qu Beijing.**
Mary and Jack take train go Beijing.
'Mary and Jack go to Beijing by train.'

Target: (Encouraging coreference reading, gender-matched)
Dang ta$_F$ kan jingse de shihou, Mary zai-he pijiu. ("ta$_F$" = "她 (she)")
During she look scenery when, Mary PGR-drink beer.
'While she looks at scenery, Mary is drinking beer.

(Encouraging disjoint reading, gender-mismatched)
Dang ta$_M$ kan jingse de shihou, Mary zai-he pijiu. ("ta$_M$" = "他 (he)")
During he look scenery when, Mary PGR-drink beer.
'While he looks at scenery, Mary is drinking beer.

*RTs at underlined regions were measured



Not Native, Yet Dominant: The Role of Language Context and Social Value in Multilingual Language Processing

Opangienla Kechu and Bidisha Som
Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati
Corresponding author: aienkechu@iitg.ac.in

This study investigates the interaction between language dominance, linguistic context, and cognitive control in multilinguals from Nagaland, India. Focusing on two non-native languages—Nagamese and English—used in distinct social and linguistic contexts, the research examines how multilingual speakers process audio-visual stimuli under varying contextual conditions. Nagamese, a creole, is widely used in various social settings and it serves as a lingua franca for communication among the Naga tribes (who speak different, mutually unintelligible languages). English, on the other hand is the official language of the State and medium of instruction in schools holding high social and economic prestige. Research on second (L2) and third languages (L3) has gained attention due to their capacity to provide insights into language control, learning, and cognitive effort, as these languages require more executive control than the typically automatic first language (L1) (Green, 1998; Green & Abutalebi, 2013). Additionally, L2/L3 processing exhibits more neural and behavioural variability, which reflects language experience and plasticity (Li et al., 2014). As multilingualism continues to rise, such studies from societies that have been traditionally multilingual (while avoiding attrition of the first language), offer valuable insights into real-world language processing (Grosjean, 2010).

The present study utilized an audio-picture matching task across three blocks: an English block, a Nagamese block, and a Mixed block (both languages). Participants were from different universities (n = 30), aged 20-40 years, who completed a language history questionnaire, the Lextale test for English proficiency (M = 84.90%), and a Nagamese proficiency test (M = 84.52%). Results revealed a significant main effect of language, with reaction times (RTs) being significantly faster for English compared to Nagamese. Slower RTs for Nagamese suggests higher cognitive load associated with its processing. This effect was statistically significant, indicating that English holds the functional dominance in this multilingual context, despite both languages being non-native (languages) and the participants having comparable proficiency (in both). A Block × Language interaction revealed that the block effect was stronger for Nagamese in the Mixed block, indicating a higher switch cost. Additionally, a significant Language × Type interaction indicated that the match advantage (faster RT for match trials) was larger for English compared to Nagamese.

Our finding supports the view that language dominance is not solely determined by proficiency but also by socio-cultural factors, including perceived prestige and social utility (Grosjean, 2010); broadly underscoring the bi/multilingual experience (Green and Abutalebi, 2013). The results contribute to the growing body of research on bilingual cognitive control and the impact of real-world, multimodal language environments on the same. Simultaneously, this also indicates that ‘nativity’ and proficiency may not decide the social hierarchy among languages; but utility sure does.

References

- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. doi:10.1017/S1366728998000133

- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. Journal of Cognitive Psychology, 25(5), 515–530. doi: 10.1080/20445911.2013.796377

- Grosjean, F. (2010). Bilingual: Life and reality. Harvard University Press. ISBN 9780674066137

- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. Cortex; a journal devoted to the study of the nervous system and behavior, 58, 301–324. https://doi.org/10.1016/j.cortex.2014.05.001.

Graph

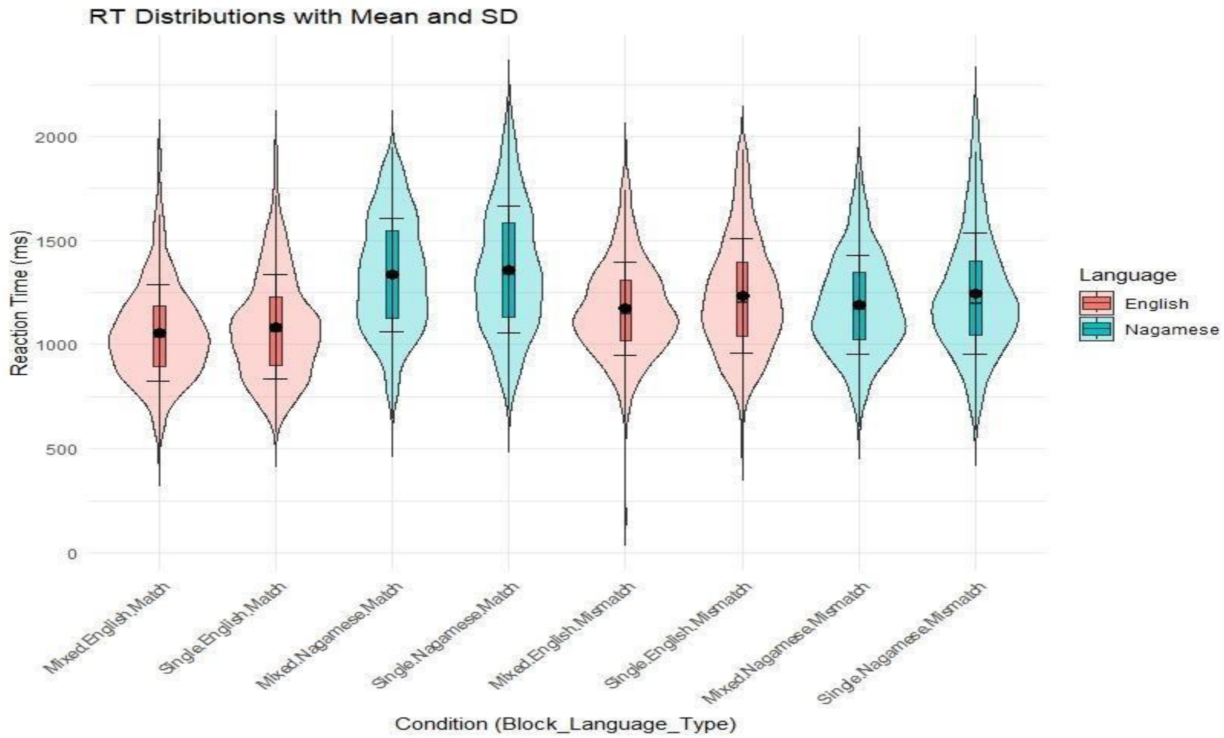


Figure : Violin plots showing the distribution of reaction times (RTs) across conditions defined by Block type (Mixed vs. Single), Language (English vs. Nagamese), and Trial Type (Match vs. Mismatch). Each violin illustrates the RT density distribution for a specific condition, with the black dot indicating the mean RT and the vertical line showing one standard deviation.

Exploring semantic priming effects using piece-wise additive mixed models

Thomas Lieber, Giovanni Cassani, Emmanuel Keuleers, Peter Hendrix
Tilburg School of Humanities and Digital Sciences, Tilburg University
T.Lieber@tilburguniversity.edu

Background: Semantic priming is crucial for understanding word processing, offering insights into the timing of automatic and controlled linguistic processes. Manipulating stimulus onset asynchrony (SoA) suggests that short SoA (100 to 300 ms) leads to fast, automatic responses, while longer SoA (over 300 ms) results in slower, strategic responses [1]. Retrospective accounts propose priming arises after both the prime and target are processed, whereas prospective accounts argue the prime pre-activates the target [2]. These two accounts have been studied using methods that fit ex-gaussian distributions to groups of prime-target pairs with low and high similarity and relate differences in the shape, variance, and skewness of these distributions to differences in semantic priming effects [3, 4, 5]. According to the prospective account, semantic priming shifts the distribution, while the retrospective account proposes that the effect appears gradually and gets more prominent over time. Piecewise Additive Mixed Models (PAMMs), a technique for survival analysis [6], offer more insight into this these theories in two ways: first, they can estimate when and how the effect of prime-target similarity changes over time, while controlling for other relevant variables; second, they allow for modeling prime-target similarity as a continuous variable by computing the similarity between their semantic representations [7]. The aim of this study is thus to explore the semantic priming effect using PAMMs, focusing on when the effect of semantic priming emerges and whether its nature changes over time - while controlling for other relevant variables.

Method: We analyzed the lexical decision task data from the Semantic Priming Project [8] using a PAMM. These data consisted of the correctly answered trials from this task. We defined the prime-target semantic similarity as the cosine similarity between FastText word embeddings of prime and target words [7]. Our analyses further included target length, prime length, target frequency, prime frequency, mean target bigram frequency, mean prime bigram frequency, target orthographic neighborhood density, prime orthographic neighborhood density, target semantic neighborhood density, and target orthographic neighborhood density.

Results: The results are visualized in Figure 1. Effects of semantic priming were present from the start of the response time window. In the 200 ms SoA condition, the semantic priming effect was significant until approximately 650 ms, after which the effect weakened. High prime-target similarity was associated with a higher probability of a response in this time window, while low prime-target similarity was associated with a lower probability of a response. In the 1200 ms SoA condition, the effect, although weaker, showed a similar shape, and was significant until approximately 550 ms. In the 200 ms SoA condition the effect remained significant longer when the prime-target similarity was low, while in the 1200 ms condition there was no such difference.

Discussion: Our statistical approach highlights that the effect of semantic priming on primed-lexical decisions occurs early in the response window and decreases over time. When the SoA is shorter, the effect is stronger and is significant for a longer duration. This result is inline with the prospective and automatic accounts, which predicts a strong, short lasting effect early in the response time window [1,2]. Future studies could investigate whether the observed patterns of semantic priming and the influence of prime-target similarity on response probability hold across different contexts and participant groups.

References

[1] Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & cognition*, *4*(5), 648-654.

[2] Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003.

[3] Hoedemaker, R. S., & Gordon, P. C. (2017). The onset and time course of semantic priming during rapid recognition of visual words. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 881.

[4] Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495-523.

[5] Yap, M. J., Balota, D. A., & Tan, S. E. (2013). Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 140.

[6] Bender, A., Groll, A., & Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, *18*(3-4), 299-321.

[7] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135-146.

[8] Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., ... & Buchanan, E. (2013). The semantic priming project. *Behavior research methods*, *45*, 1099-1114.

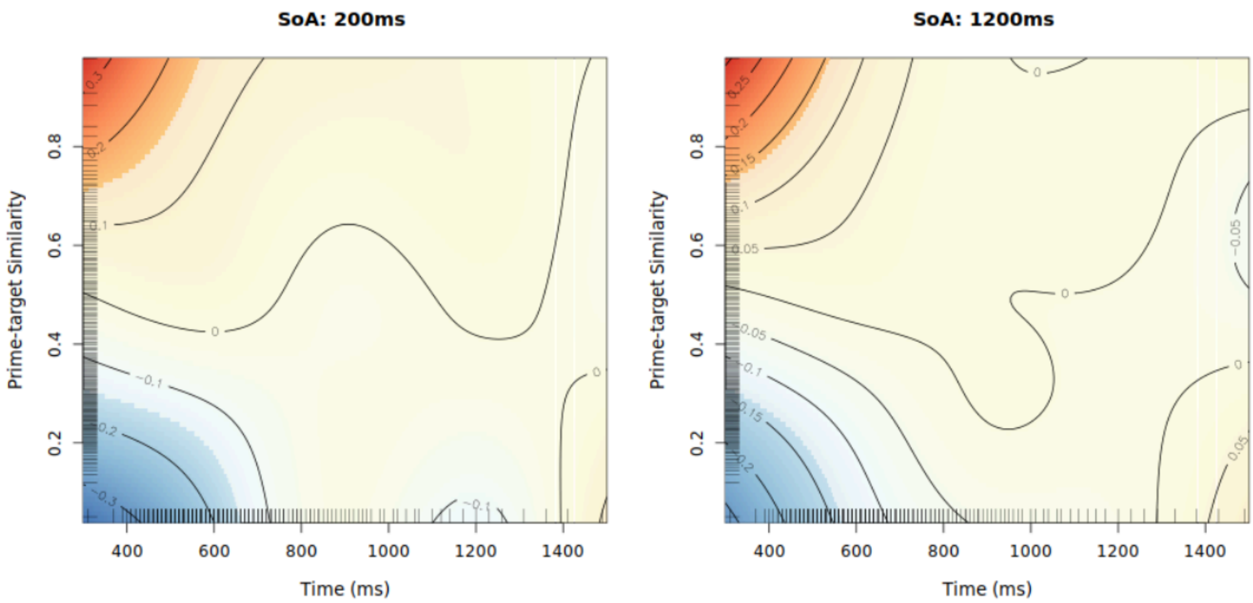


Figure 1. Time-dependent effect of prime-target semantic similarity for words only for short SoA (200ms; left panel) and long SoA (1200ms; right). Warmer colors indicate higher (log) hazard rates, indicating a higher chance for a response to occur. Areas in the plot that show saturated colors indicate a significant effect.

Characterising the affective content of sentences and its role in reading and memory

Juan Haro[1], Daniel Huete-Pérez[1], Miguel Ángel Pérez-Sánchez[2], José Antonio Hinojosa [3,4], Pilar Ferré [1]
[1] Universitat Rovira i Virgili; [2] Universidad de Murcia; [3] Universidad Complutense de Madrid; [4] Universidad Nebrija
juan.haro@urv.cat

Background: Psycholinguistic research on the interplay between language and emotion has primarily relied on single words. This research has greatly benefited from the development of normative datasets on affective variables (e.g., valence, arousal, different emotions) [1, 2]. To elaborate them, researchers collect subjective ratings from many participants for large sets of words. Research conducted in this field has provided ample evidence of the prioritized processing of emotional words (e.g., party, gun) compared to neutral words (e.g., table) in psycholinguistic tasks [3]. However, words are integrated into a context in more natural situations. In search of greater ecological validity, research with supra-lexical linguistic units (e.g., sentences) is needed. Some studies have appeared in recent years in this direction [4, 5], although they are still scarce. As with single-word research, the development of normative datasets can be a great contribution to the field. The aim of this study was twofold. First, we aimed to collect normative affective data on a large set of sentences that were deliberately designed to be emotionally neutral except for a single critical noun, whose emotional valence and arousal varied systematically. Second, we examined the role of these affective variables in sentence reading and memory, focusing on whether emotional valence (positive and negative) influences processing and memory when only one word carries the emotional load.

Method: With respect to the first aim, we collected affective ratings on 415 pairs of Spanish sentences on a series of affective variables: valence (i.e., the quality of the emotional response evoked from the sentence, from unpleasant to pleasant), arousal (i.e., the intensity of that emotional response, from relaxing to activating), and the relationship with the emotions of joy, sadness, disgust, anger, and fear. We also collected plausibility, cloze probability and predictability ratings. Each pair of sentences was exactly the same, except for one word, which gave them their potential emotional or neutral nature (e.g., *Miguel encontró un <u>cuchillo/sombrero</u> en el suelo*, Michael found a <u>knife/hat</u> on the floor). We used varied syntactic structures across the dataset. We analyzed the relationship between the affective variables as well as the relationship between the affective properties of the critical words (i.e., knife/hat) and those of the whole sentences. Regarding the second aim, we conducted a reading experiment with emotional (positive and negative) and neutral sentences. With the goal of testing the sentences in different paradigms, half of the participants performed a comprehension task, in which accuracy and the reading times (RT) of the whole sentences were recorded. The other half performed a self-paced reading task, in which the RT of the critical words were recorded. Furthermore, participants were asked to remember/recognize the sentences at the end of the task.

Results: The affective ratings of sentences were highly reliable and showed a U-shaped relationship between valence and arousal. Furthermore, the affective properties of the whole sentences were mostly determined by the affective properties of the critical word. The reading experiment is currently in progress. We expect to find a facilitation for emotional sentences in both reading and memory.

Discussion: The results will be compared with those obtained in single-word studies and their possible explanation in terms of the mechanisms proposed in that literature, such as attentional capture, will be discussed. In addition, the need for normative datasets of sentences is underscored. These datasets fill a gap in understanding how emotional content embedded in context affects language processing and memory.

References

[1] Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. A. (2017). Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. Behavior Research Methods, 49, 1082-1094.

[2] Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, 45(4), 1191–1207.

[3] Ferré, P., Sánchez-Carmona, A. J., Haro, J., Calvillo-Torres, R., Albert, J., & Hinojosa, J. A. (2024). How does emotional content influence visual word recognition? A meta-analysis of valence effects. Psychonomic Bulletin & Review. Advanced online publication.

[4] Knickerbocker, H., Johnson, R. L., & Altarriba, J. (2015). Emotion effects during reading: Influence of an emotion target word on eye movements and processing. Cognition and Emotion, 29(5), 784–806.

[5] Lüdtke, J., & Jacobs, A. M. (2015). The emotion potential of simple sentences: additive or interactive effects of nouns and adjectives? Frontiers in psychology, 6, 1137.

Funding

This study was supported by Ministerio de Ciencia, Innovación y Universidades of Spain and FEDER, U.E. (PID2023-149606NB-I00/AEI/10.13039/501100011033), Ministerio de Universidades de España (predoctoral contract FPU20/03345), the Horizon Europe Framework Programme (HORIZON-MSCA-2023-SE-01. Ref. 101182959), and the Universitat Rovira i Virgili (2022PFR-URV-47).

Pronoun interpretation in German speech reports

Yvonne Portele[1, 2], Sebastian Walter[1]
[1] Goethe University Frankfurt; [2] Free University Berlin
portele@lingua.uni-frankfurt.de

Background: Speakers can report another person's speech using direct discourse (DD), quoting the speaker verbatim, or indirect discourse (ID), summarizing the content. DD involves a perspective shift to interpret pronouns (e.g., *I* refers to the original speaker), while ID typically does not (e.g., a 3rd person pronoun must be used to refer to the original speaker). Perspective shifts are cognitively demanding [1, 2] and usually avoided unless necessary. Pronouns might be easier to interpret in DD due to positive effects on story comprehension [3] and earlier acquisition of DD before ID in language production [4]. However, research on Dutch has shown that, except 1st person pronouns, pronouns are interpreted correctly more often in ID than in DD [5]. We replicated the overall pattern in *Masked* for German ID reports using an embedded V2 order. 1st person pronouns may render a perspective highly prominent [6], causing comprehenders to wrongly perform a perspective shift, therefore showing a deviating pattern compared to the other pronouns. In this paper, we report a follow-up study targeting ID reports with an embedded verb final order that structurally resembles Dutch ID reports to investigate whether the absence/presence of a complementizer influences ID comprehension (see page 3 for examples illustrating the structural differences).

Method: We used a forced choice referent selection paradigm. Participants (n = 96) read contexts introducing three male referents. Targets (n = 24) were presented as stick figure with a speech bubble representing one of the referents (cf. Fig. 1). We manipulated REPORT TYPE (DD vs. ID), PRONOUN (1st vs. 2nd vs. 3rd person), and in ID conditions also COMPLEMENTIZER presence (with/without *dass* 'that') (between-subjects). After reading contexts and targets, participants indicated referent selection for a given pronoun.

Results: German speakers identified pronouns more accurately in ID than DD, except for 1st person pronouns, where DD outperformed ID [5]. Interestingly, performance in DD was overall worse in German compared to Dutch data. A common error in ID with 1st person pronouns was participants incorrectly attributing pronouns to the reporting speaker, highlighting an unintended perspective shift. Crucially, the presence of the complementizer significantly increased accuracy for 1st person pronouns in ID. Additionally, our data showed unusually high error rates in DD conditions for both 2nd and 3rd person pronouns (Fig. 2).

Discussion: The improved accuracy with a complementizer suggests that structural clarity reduces unintended perspective shifts. German ID without complementizer structurally resembles DD, likely contributing to higher errors compared to Dutch, where ID clearly differs structurally from DD due to the obligatory presence of a complementizer resulting in verb final order. Our findings indicate perspective shifts in pronoun interpretation are sensitive to syntactic cues, with complementizer presence facilitating correct interpretation in ID. However, unexplained high error rates in DD for 2nd and 3rd person pronouns remain. We suggest they are linked to cognitive overload during perspective shifts or ambiguity introduced by the similar structure of DD and ID sentences without complementizer. Future research should investigate how syntactic and cognitive factors interact in perspective taking during pronoun interpretation, exploring why certain pronouns consistently trigger higher error rates and whether additional visual or contextual support could mitigate these challenges.

References

[1] Eerland, A., & Zwaan, R. A. (2018). The influence of direct and indirect speech on source memory. *Collabra: Psychology*, *4*(1), 5.

[2] Harris, J. A. (2021). Extended perspective shift and discourse economy in language processing. *Frontiers in Psychology*, *12*, 613357. https://doi.org/10.3389/fpsyg.2021.613357

[3] Groenewold, R., Bastiaanse, R., Nickels, L., Wieling, M., & Huiskes, M. (2014). The effects of direct and indirect speech on discourse comprehension in Dutch listeners with and without aphasia. *Aphasiology*, *28*(7), 862–884.

[4] Nordqvist, A. (2001). The use of direct and indirect speech by 1 $1/2$-to 4-year-olds. *Psychology of Language and Communication*, *5*(1), 57–66.

[5] Köder, F., Maier, E., & Hendriks, P. (2015). Perspective shift increases processing effort of pronouns: A comparison between direct and indirect speech. *Language, Cognition and Neuroscience*, *30*(8), 940–946. https://doi.org/10.1080/23273798.2015.1047460

[6] Saure, C., Hinterwimmer, S., & Jordan-Bertinelli, A. P. (2023). An experimental investigation of the interaction of narrators' and protagonists' perspectival prominence in narrative texts. *Zeitschrift für Sprachwissenschaft*, *42*(2), 341–372.

Figures

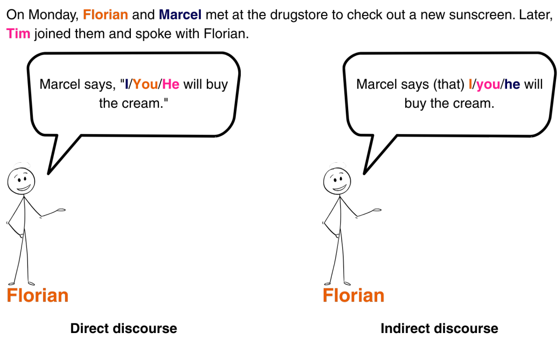


Figure 1: Sample target item for the different conditions, translated from German.

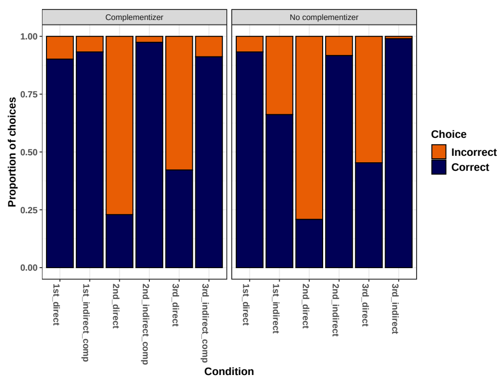


Figure 2: Proportion of correct and incorrect choices in the different conditions.

Structural options of direct discourse (DD) and indirect discourse (ID) in Dutch and German speech reports:

DD including a 2nd person pronoun in Dutch ((1a); taken from [5]) and in German (1b):

(1) a. Olifant zei: “Jij krijgt de voetbal”.
    b. Elefant sagt: “Du bekommst den Fußball”.
    Translation: *Elephant said: “You get the footbal”.*

ID including a 1st person pronoun in Dutch ((2a); taken from [5]). In German, the ID speech report can be expressed with (2b) or without the complementizer *dass* (2c), leading to changes in positioning of the verb. The German option provided in (2c) is not valid in Dutch.

(2) a. Olifant zei **dat** ik de voetbal krijg.
    b. Elefant sagt, **dass** ich den Fußall bekomme.
    c. Elefant sagt, ich bekomme den Fußball.
    Translation: *Elephant said that I get the football.*

Parsing strategies in Hebrew and Italian Relative Clauses: Shall I Avoid Gaps?

Lion Oks[a], Francesca Foppolo[b], Carlo Cecchetto[b,c] and Caterina Donati[a]
[a]LLF/CNRS - Université Paris Cité, [b]University of Milan-Bicocca, [c]SFL - CNRS/Paris 8
lion.oks@etu.u-paris.fr

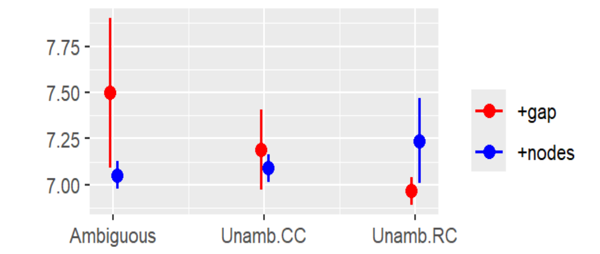
Previous psycholinguistic research provided evidence of structural parsing heuristics during language processing of temporary ambiguity, beyond frequency-based accounts [1]. **Q1.**What happens when two independently motivated heuristics make opposite predictions about the continuation of a sentence? One such case would be the processing of a temporary ambiguity triggered by verbs like 'convince' followed by a that-clause, that could be: an object relative clause (ORC) (1a), involving a Gap, or a ditransitive clause (1b), involving no gap but requiring more syntactic nodes at the moment of disambiguation. Verbs in (1c-d) unambiguously select one of the two structures, thus not posing specific challenges to the parsing mechanism. Verbs in (1a-b) are ambiguous: An **Avoid Gap** heuristic would disfavor (1a), an **Avoid nodes** heuristic (1b).

(1)	a. Mary <u>convinced</u> the students that the principal will have to meet __ soon.	Amb.RC [-Nodes,+Gap]
	b. Mary <u>convinced</u> the students that the principal will meet <u>Tom</u> soon.	Amb.CC [+Nodes,-Gap]
	c. Mary <u>called</u> the students that the principal will appoint __ soon.	Unamb.RC [-Nodes,+Gap]
	d. Mary <u>told</u> the students that the principal will appoint <u>Tom</u> soon.	Unamb.CC [+Nodes,-Gap]

In a crosslinguistic perspective, a second question arises: **Q2**. Does the Avoid Gap heuristics extend to languages where a gap is replaced by a resumptive pronoun (RP), such as Hebrew? Three experiments were conducted in 2 languages with 24 experimental items in 4 conditions, as in (1), tested within participants in a Latin square design (+ fillers). These languages varied w.r.t. *dependency*: Italian RCs involve a Gap, Hebrew may have a RP instead. In Italian the same sentences were used in a G-maze task (N=39) and in a self-paced reading task (SPR, N=61), cf. (2). Hebrew study (N=60) was a SPR with oblique ORCs like (3), in which the RP is obligatory. Experiments were implemented in PCIbex [2] and run online.

Results. In **Italian G-Maze**, a preference to select CC continuations (*p*=.002) at the point of disambiguation (w0) and a cost (in terms of RTs) at w1 emerges when the RC is selected instead (Figure 1A. Continuation*Condition: t=3.3, p<.001). In **Italian SPR,** a cost of Continuation (RC/CC) emerges in the disambiguation region (vs. before) of ambiguous (vs. unambiguous) sentences (Figure 1B: Continuation*Condition*Region: t=2.1, p=.039). **Hebrew SPR** shows no cost for RC over CC in the Ambiguous condition: although the ACC marker, *et*, (RC) is significantly slower than the preposition, *al***,** (CC) in the Unambiguous condition (consistently with the complexity of ORCs), it is not so in the Ambiguous condition (Fig. 3-left, Continuation*Condition: t=2.3, p=.023).

Conclusions. Considering the type of dependency (Gap, as in Italian vs. RP, as in Hebrew) our results provide evidence for an Avoid Gap heuristic in language processing, compatible with Minimal Chain Principle [3]. Data on Hebrew show that resumptives are exempted from Avoid Gap. We plan two follow-up studies in Hebrew including (i) oblique ORCs, where RPs are optional, and might possibly involve silent RPs, and (ii) subject RCs, where RPs are impossible, thus presumably involving a proper gap. If our reasoning is correct, we expect study (i) to reveal the same pattern displayed in Oblique ORCs (no cost associated with RCs over CCs in the ambiguous condition), and study (ii) to display the same pattern as Italian, namely, a cost associated with RCs.



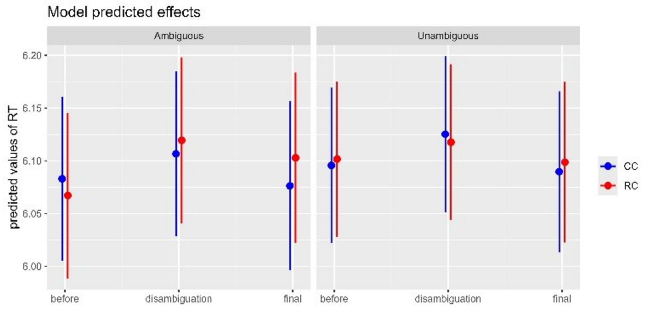
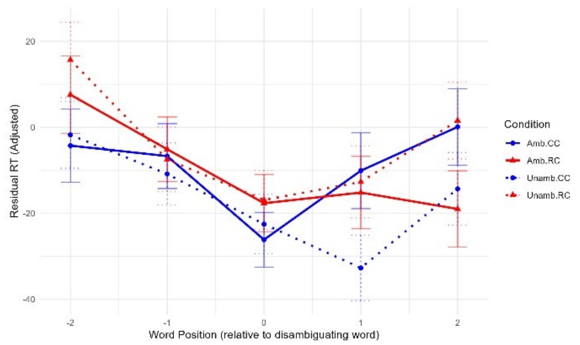


Figure 1 (A-B). Plots of predicted effects of linear mixed models on log(RT) including Condition (Amb vs. Unamb) and Continuation (RC=+gap vs. CC=+nodes), both contrast-coded as +.5,-.5, trial order, frequency and length of the word (in characters) as (centered) predictors, Condition*Continuation interaction, random intercepts and random slopes for Participants and Items (when possible). In Italian SPR, we also included the interaction with region (contrast coded as disambiguation vs. before and disambiguation vs. after).

Figure 1A - Italian G-Maze
Plot at w1, after +gap or +nodes was selected at w0 (i.e., a 'mistake' in Unamb, a choice in Amb).

Figure 1B - **Italian SPR**
Model plot of condition*continuation*region (Before includes w-5:w1; Disambiguation: w0:w3; Final: w4:w5).



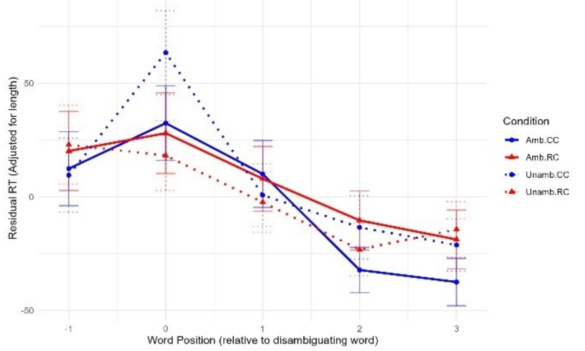


Figure 2A. Residual RTs of Hebrew SPR (with RP). Nested models show a cost of RC over CC in Unambiguous (p=.009), but no difference of CC/RC in Ambiguous (p=.516).

Figure 2B. Residual RTs of Italian SRP. Nested models show a cost of RC vs. CC in Ambiguous condition in Disambiguation region vs. Before (p<.001), but no difference of CC/RC in Unambiguous (p= .525), cf. Figure 1B.

Selected references.

[1] Staub, A., Foppolo, F., Donati, C., & Cecchetto, C. (2018). Relative clause avoidance: Evidence for a structural parsing principle. *Journal of Memory and Language*, *98*, 26-44. [2] Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). [3] De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian: The minimal chain principle* (Vol. 12). Springer Science & Business Media.

Experimental stimuli

(2) Italian
(in G-Maze, the critical choice was between stasera/garanzie in Amb/Unamb.CC/Unamb.RC)

a. Il mediatore convinse gli ostaggi che i rapitori dovevano rilasciare **stasera*** prima di procedere. Amb.RC
b. Il mediatore convinse gli ostaggi che i rapitori dovevano rilasciare **garanzie** prima di procedere. Amb.CC
c. Il mediatore rassicurò gli ostaggi che i rapitori dovevano rilasciare **stasera*** prima di procedere. Unamb.RC
d. Il mediatore assicurò agli ostaggi che i rapitori dovevano rilasciare **garanzie** prima di procedere. Unamb.CC
The mediator convinced/reassured/assured the hostages that the kidnappers will have to release (gap) tonight / (no gap) guarantees before proceeding.

*Note that in the case of RC continuation, at w0 (e.g., stasera, *tonight*) the sentence is never completely disambiguated, given that an adverbial or other prepositional modifiers are always possible in Italian after the verb and before a NP that would force a CC continuation. For example, "…che i rapitori dovevano rilasciare stasera entro la mezzanotte uno dei prigionieri" (…*that the kidnappers had to release tonight before midnight one of the prisoners*.) Conversely, in the case of CC continuation, full disambiguation occurs at w0 (garanzie, *guarantees*), as the presence of a NP/DP rules out the possibility of having a gap, thus an ORC in this case.

(3) Hebrew

a. šixnati et ha-oved še-ha-menahel ca'ak al-av aval lo ***et*** ha-ovedet ha-axeret
convinced.1.SG.ACC the-worker.M that-the-manager.M yelled on-him but not **ACC** the-worker.F the-other.F
'I convinced the (male) worker that the manager yelled at __ but not the other (female) worker.' Amb.RC+RP

b. šixnati et ha-oved še-ha-menahel ca'ak al-av aval lo ***al*** ha-ovedet ha-axeret
convinced.1.SG.ACC the-worker.M that-the-manager.M yelled on-him but not **on** the-worker.F the-other.F
*'I convinced the (male) worker that the manager yelled at **him** but not at the other (female) worker.'* Amb.CC

c. ra'iti et ha-oved še-ha-menahel ca'ak al-av aval lo ***et*** ha-ovedet ha-axeret
saw.1.SG.ACC the-worker.M that-the-manager.M yelled on-him but not **ACC** the-worker.F the-other.F
'I saw the (male) worker that the manager yelled at __ but not the other (female) worker.' Unamb.RC+RP

d. amarti l-a-oved še-ha-menahel ca'ak al-av aval lo ***al*** ha-ovedet ha-axeret
said.1.SG.ACC to-the-worker.M that-the-manager.M yelled on-him but not **on** the-worker.F the-other.F
*'I said(to) the (male) worker that the manager yelled at **him** but not at the other (female) worker.'* Unamb.CC

Grammatical gender, number and case in processing: experimental studies on Russian

Daria Antropova[1], Natalia Slioussar[1], Elizaveta Galperina[2], and Olga Kruchinina[2]

([1]HSE University, Moscow and Saint Petersburg; [2]Sechenov Institute of Evolutionary Physiology and Biochemistry, RAS, Saint Petersburg)

daria.antropos@gmail.com

Background: We present a series of experiments on Russian revealing curious differences between gender, number and case in processing. Previous studies on other languages compared gender and number mismatches using diverse stimuli and tasks. The results were controversial. In some studies, gender mismatches caused longer RTs [1], larger P600 amplitudes [2] and longer P300 latencies [3], associated with the processing cost for mismatches and repair processes. Other studies reported larger and longer P600 effects for number mismatches [4, 5]. In our study, we compared gender and number agreement processing in Russian in predicative and attributive constructions using different methods and added case to the picture.

Method: We conducted four experiments. Participants of all studies were adult native Russian speakers (see Table 1 for details). In Exp. 1, we compared gender (GEN), number (NUM) and case (CASE) mismatches in participle-noun agreement (attributive, ATTR). Participants read stimulus sentences in four conditions (see (1)) word-by-word in the self-paced reading mode. In the second self-paced reading study (Exp. 2), we compared GEN and NUM in subject-verb agreement (predicative, PRED) (see (2)). In Exp. 3, we employed a speeded grammaticality judgment task (SGJT) with both ATTR and PRED stimuli sets from Exp. 1-2. Exp. 4 was an ERP study in which we compared GEN, NUM and CASE in ATTR (1) and GEN and NUM in PRED (2). Participants read the sentences in the rapid serial visual presentation mode and made grammaticality judgments for 25% of them while continuous EEG was recorded.

Results: CASE effects were the least prominent. The P600 amplitude was for CASE was the smallest. RTs in CASE stimuli did not significantly differ from those in correct (CORR) stimuli. The SGJT accuracy for CASE was the lowest. Some differences between GEN and NUM were also observed. In ATTR, NUM effects on ERPs were the most prominent. NUM elicited an N400 and a positive effect on 200-300 ms, which were not observed for other mismatch types. NUM and GEN effects on RTs occurred on the word after the participle and were equally salient. SGJT accuracy for NUM was significantly lower than for CORR, while GEN only marginally differed from CORR. In PRED, both GEN and NUM elicited a LAN/N400 and a P600. For GEN, LAN/N400 effects occurred earlier and had a broader distribution across the ROIs. NUM effects on RTs occurred earlier (on the verb) and persisted longer (up to the 2nd word after the verb) than GEN effects (observed only on the word after the verb). Both NUM and GEN significantly differed from CORR in the SGJT accuracy and did not significantly differ from each other.

Discussion: Observed differences between the features can be explained by their different characteristics. Gender is a property of the lexeme, while number and case are the properties of the word form. Case depends on the syntactic context, which makes it the least salient. Case is not directly associated with semantics, gender is not semantically loaded in most nouns, while number is. In our studies, number and gender mismatches were costlier than case mismatches. In ATTR, number was more salient in ERPs. In PRED, gender was more salient in ERPs, while number was more salient in RTs. We hypothesize that predicative agreement processing is more basic and immediate,[1] so it allows to observe the earliest effect of gender mismatches on ERPs (gender is an inherent property of the noun, so the mismatch is noticed faster). But number mismatches become more salient later in the course of processing because number is semantically loaded.

(1) Conditions in Exp. 1, 3, 4: ATTR: correct, GEN, NUM, CASE):
Učenyj natknulsja na stat'ju, ***soderžavš-uju/*-ij/*-ie/*-ej*** *interesnye fakty.*
scientist stumbled upon paper$_{\text{F.ACC.SG}}$ **containing**$_{\text{F.ACC.SG / }\mathbf{M}\text{.ACC.SG / ACC.}\mathbf{PL}\text{ / F.}\mathbf{LOC}\text{.SG}}$ interesting facts

(2) Conditions in Exp. 2, 3, 4 (PRED: correct, GEN, NUM):
Iva ***stojal-a/*-Ø/*-i*** *u pruda, ukrašavšego tenistyj park.*
willow$_{\text{F.SG}}$ **stood**$_{\text{F.SG / }\mathbf{M}\text{.SG / }\mathbf{PL}}$ near pond adorning shadowy park

Table 1. Participants and materials of Exp. 1-4

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Participants	N = 88 (32 m, 56 f), 18-54 y.o (M = 33)	N = 68 (28 m, 40 f), 18-54 y.o (M = 29)	N = 150 (59 m, 91 f) 18-57 y.o. (M = 37)	N = 49 (17 m, 32 f) 18-36 y.o (M = 23)
Stimuli	64 target (ATTR), 176 fillers	48 target (PRED), 112 fillers	112 target (ATTR + PRED), 48 fillers	320 target (160 ATTR, 160 PRED), 80 fillers

Table 2. Descriptive statistics for Exp. 1-3. For Exp. 1-2: mean RTs (in ms) and SDs across conditions for the segments with significant differences. For Exp. 3: mean SGJT accuracies and SDs across conditions

Agreement type	**Condition**	**Mean reading times (and SDs)**			**Mean response accuracy (and SD)**
		W (target)	**W + 1**	**W + 2**	
Attributive agreement	Correct		473 (159)		.94 (.08)
	Gender		513 (190)		.92 (.13)
	Number		510 (186)		.91 (.15)
	Case		500 (175)		.81 (.25)
Predicative agreement	Correct	493 (173)	419 (109)	453 (144)	.98 (.06)
	Gender	516 (246)	461 (145)	457 (146)	.91 (.14)
	Number	562 (302)	481 (158)	475 (153)	.90 (.14)

References: [1] Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, 11(1), 75-101. **[2]** Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137-153. **[3]** Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, 39(3), 465-482. **[4]** Popov, S., & Bastiaanse, R. (2018). Processes underpinning gender and number disagreement in Dutch: An ERP study. *Journal of Neurolinguistics*, *46*, 109-121. **[5]** Muralikrishnan, R., & Idrissi, A. (2021). Salience-weighted agreement feature hierarchy modulates language comprehension. *Cortex*, 141, 168-189.

[1] After seeing a subject the reader is sure that the predicate is coming, sooner or later, while postpositive attributive constructions that we used are strictly optional. We chose these constructions so that the agreement controller preceded the target in all our stimuli. We are currently running experiments in which this parameter is manipulated.

A brief overview of attributive and predicative agreement in Russian

Attributive agreement

In Russian, adjectives and participles agree with nouns in gender (only in singular), number, and case. The number feature has two values (Sg, Pl), the gender feature has three values (M, F, N) and the case feature has 6 values (Nom, Gen, Dat, Acc, Ins, Loc). In our study, stimulus sentences had target nouns in Sg in the ATTR conditions. Two genders (M, F) and two cases (Acc, Loc) were considered. The relevant fragment of the paradigm for the participle *soderžavšij* ‘containing’ is presented below.

Table 3. A fragment of the paradigm of the participle *soderžavšij* ‘containing’

	Sg		**Pl**	
	Acc	**Loc**	**Acc**	**Loc**
M	*soderžavš-**ij***	*soderžavš-**em***	*soderžavš-**ie***	*soderžavš-**ix***
F	*soderžavš-**uju***	*soderžavš-**ej***		

Predicative agreement

In Russian, verbs agree with subject noun phrases in number and person in non-past tenses or in number and gender (only in singular) in the past tense. Masculine forms have null endings. The relevant fragment of the paradigm for the verb *stojat’* ‘to stand’ in the past tense is presented below.

Table 4. A fragment of the paradigm of the verb *stojat’* ‘to stand’ in the past tense

	Sg	**Pl**
M	*stojal-**Ø***	*stojal-**i***
F	*stojal-**a***	

ERPs reveal differential processing of three types of gender violations in Norwegian

Sarah Cameron[1], Natalia Kartushina[1], Björn Lundquist[2] & Sendy Caffarra[3,4]

[1]University of Oslo; [2]UiT The Arcitc University of Norway; [3]UniMore Università degli studi di Modena e Reggio Emilia; [4]Stanford University

Corresponding author: sarah.cameron@iln.uio.no

Background: EEG studies on grammatical gender processing reveal consistent P600 and less reliable LAN responses when there's mismatch between a noun and an article, determiner, or adjective, and a few studies have also reported N400 responses to the same violations [1]. The precise reasons behind the presence or absence of LAN and N400 remain unclear. Two ERP studies [2,3] have so far been published on gender processing in Norwegian. Both reported P600 responses but analyzed only the P600 time window (500-900 ms post-target), thus overlooking earlier responses. Therefore, no comprehensive findings exist regarding the full response pattern of gender violations in Norwegian. This study seeks to deepen our understanding of gender processing in Norwegian by adopting a data-driven approach to identify responses that occur in earlier time windows, such as LAN or N400, typically observed around 300-500ms. By investigating the emergence of early anterior responses in this new language context, we aim to enrich the existing literature and move closer to uncovering the underlying mechanisms at play. We used three types of gender violations, a mismatch between: a) the noun and the definite noun suffix; b) the determiner and the noun, and c) the noun and the predicative adjective. We anticipate observing P600 responses for all violation types, possibly accompanied with an earlier negativity (LAN or N400).

Method: Data was collected from 36 right-handed, neurologically healthy native speakers of Norwegian, who were shown written sentences displayed word by word on a computer screen while we recorded EEG from 32 electrodes. The stimuli comprised 320 sentences, distributed over the three gender violation types and a set of fillers – each sentence type consisting of 40 grammatically correct sentences and 40 sentences containing a morphosyntactic violation. The ungrammatical fillers contained a definiteness violation. Examples of the stimuli can be seen in table 1.

Results: Cluster-based permutation analyses revealed strong P600 responses for all three sentence types, starting at 424 ms post onset for suffix violations (p=0.002), at 452 ms for determiner violations (p=0.002), and at 548 ms. for adjective violations (p=0.002) and continuing for the full duration of the 1000 ms epochs for all three sentence types. Further, there was an anterior negative response for adjective violations (p=0.04) starting in the left hemisphere at 328 ms and shifting more towards central electrode sites before ending at 548 ms. Finally, determiner violations yielded a later frontal negativity (p=0.026), starting at 660 ms and lasting throughout the duration of the epoch.

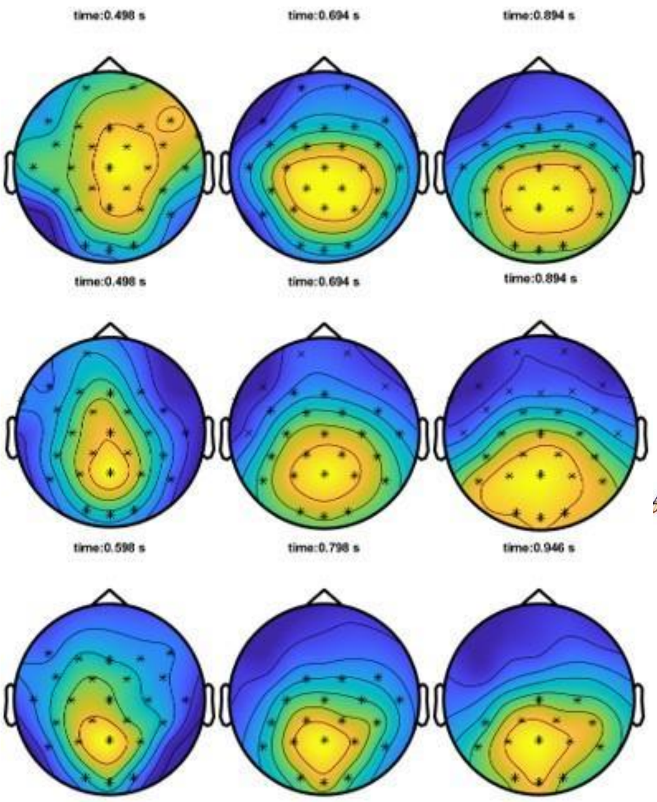
Discussion: In line with previous studies, all three types of gender violations elicited P600 responses in native Norwegian speakers. The P600 was greater for the suffix violations, possibly due to a higher reanalysis cost [4]. Only the noun- adjective violations elicited a LAN response, indicating that the process behind the morphosyntactic reanalysis of these violations differ from the other two. The reason for the occurrence of the late anterior negativity for the determiner-noun violations is less clear but could possibly be a phase reversal of the P600 (i.e., the negative dipoles of the positive response).

References

1. Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019
2. Kubota, M., Alonso, J. G., Anderssen, M., Jensen, I. N., Luque, A., Soares, S. M. P., Prystauka, Y., Vangsnes, Ø. A., Sandstedt, J. J., & Rothman, J. (2024). Bilectal Exposure Modulates Neural Signatures to Conflicting Grammatical Properties: Norway as a Natural Laboratory. *Language Learning*, *n/a*(n/a). https://doi.org/10.1111/lang.12608
3. Sandstedt, J., Kubota, M., Anderssen, M., Darby, J. A., Helset, S., Tavakoli, E., Vangsnes, Ø. A., & Rothman, J. (2025). Bidialectal language representation and processing: Evidence from Norwegian ERPs. *Journal of Memory and Language*, *140*, 104557. https://doi.org/10.1016/j.jml.2024.104557
4. Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153. https://doi.org/10.1162/0898929052880101

Table 1: Example stimuli. Critical items are underlined

	Agreement condition	Disagreement condition
Suffix violation	Jon setter seg på **benken** i hagen *Jon sits down on bench.MASC-the.MASC in the garden*	Jon setter seg på ***benket** i hagen *Jon sits down on bench.MASC-the.NEUT in the garden*
Determiner violation	Kelneren på **den** dyre **kaféen** jobber hardt *The waiter in the.MASC expensive café.MASC works hard*	Kelneren på ***det** dyre **kaféen** jobber hardt *The waiter in the.NEUT expensive café.MASC works hard*
Adjective violation	Det hyggelige **bakeriet** er **travelt** om morgenen *The nice bakery.NEUT is busy.NEUT in the morning*	Det hyggelige **bakeriet** er ***travel** om morgenen *The nice bakery.NEUT is busy.MASC in the morning*
Fillers	Fuglen mister **et frø** på bakken *The bird drops a.INDEF seed.INDEF on the ground*	Fuglen mister ***et frøet** på bakken *The bird drops a.INDEF seed.DEF on the ground*



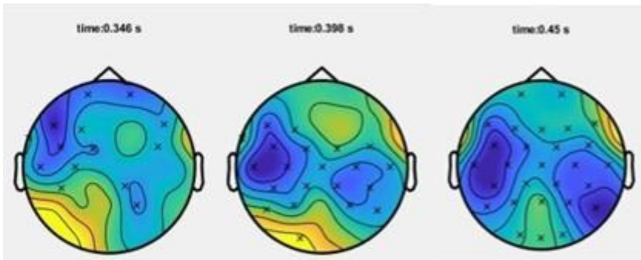


Fig 2: Topographical plots of LAN cluster for adjective violations

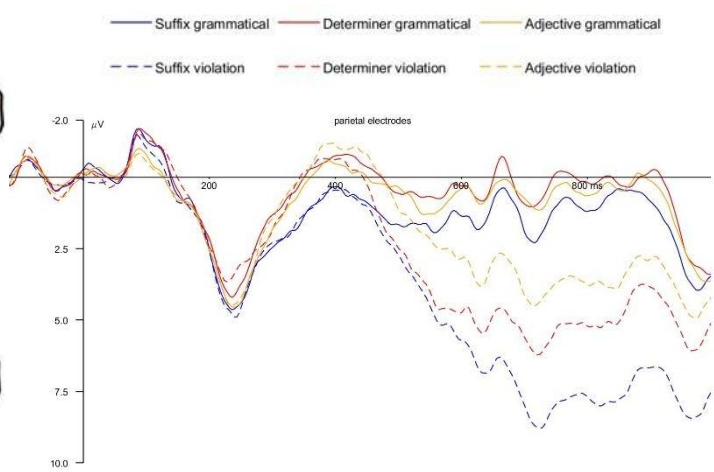


Fig 1: Topographical plots of P600 clusters for suffix violations (top), determiner violations (middle) and adjective violations (bottom). Note negative frontal cluster for determiner violations, marked by x-es.

Fig 3: Waveform of P600 electrode sites (average of P3, Pz and P4) for all three sentence types. Negative plotted upwards.

Prediction Updating During Novel Word Learning: Evidence from Cerebellar TMS

Agnieszka Konopka, Evita Ahmed Hashmi, Martina Italia, Keir Lawley, Joost Rommers, Brian Mathias
University of Aberdeen
agnieszka.konopka@abdn.ac.uk

Background: Successful vocabulary learning involves associating novel word forms with meanings and storing them in long-term memory. New words are often encountered in predictive sentence contexts that help learners infer their meanings (e.g., "David smoked his last *impet*"). While prediction is widely recognized as a core mechanism of human cognition (e.g., [1]), including in language comprehension [2], its role in supporting word learning remains underexplored [3]. Recent cognitive neuroscience evidence has suggested a causal role of the cerebellum in the refinement and updating of predictions in language and other domains [4,5]. Here, we combine visual-world eye-tracking with transcranial magnetic stimulation (TMS) to assess learning success for novel words in predictive sentence contexts: specifically, we examine how generating predictions – and updating *incorrect* predictions – influences word learning outcomes, and whether these processes critically depend on cerebellar function.

Method: Twenty native English speakers completed a two-session, eye-tracked word learning study. In each session, they first underwent continuous theta burst stimulation (cTBS) targeting either cerebellar Crus I/II [6] or a control site (vertex; counterbalanced) and then completed the word learning task. The study phase used the visual world paradigm: participants heard strongly or weakly constraining sentences ending in novel words (e.g., "The man smoked / bought the last *impet*"; 38 novel words per session) and guessed which of four objects shown on the screen (one target, three distractors) the new word might refer to. Feedback was provided after each guess. The test phase consisted of a four-alternative forced choice task.

Results: During study, participants were more likely to fixate target objects when listening to strongly compared to weakly constraining sentences, particularly after receiving cerebellar cTBS (Constraint x Stimulation interaction: z=4.70, p<.001; see Figure 1). As expected, initial guesses were also more accurate in strongly than in weakly constraining sentences (94% vs. 42%). Critically, cerebellar cTBS differentially affected performance in the test phase (see Figure 2). Participants exhibited similar accuracy for words that they had guessed *correctly* during study (i.e., trials on which they had received confirmatory feedback) across TMS conditions. However, they were less likely to update *incorrect* guesses (i.e., trials on which they had received disconfirmatory feedback) after receiving cerebellar cTBS: when they guessed an incorrect object while hearing strongly constraining sentences at study, cerebellar stimulation impaired subsequent retention of the correct object identity (Constraint x Stimulation x Guess Accuracy interaction: z=2.02, p=.04).

Discussion: Cerebellar stimulation did not disrupt predictive eye movements themselves (cf. [5]) but selectively impaired participants' ability to revise predictions in response to disconfirmatory information. This supports theories positing that the cerebellum refines rather than generates predictions [4]: under typical conditions, prediction updating allows learners to revise incorrect interpretations, but cerebellar disruption may hinder this revision process, leading to the persistence of incorrect word-meaning associations. These findings link cerebellar prediction updating mechanisms to long-term memory formation and suggest broader implications for the role of prediction in learning.

Figure 1. Study phase: mean proportions of fixations to the target and distractor objects during the analysis window (~2000 ms from verb offset to target offset), shown separately for strongly and weakly constraining sentences under cerebellar and vertex stimulation.

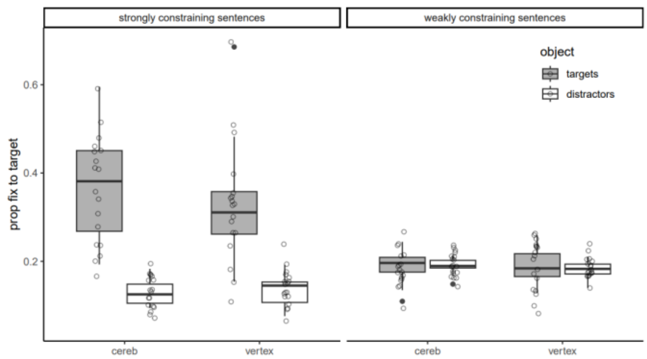
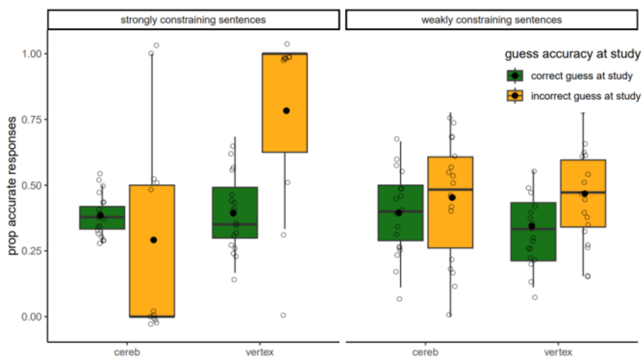


Figure 2. Test phase: mean accuracy at test for words initially presented in strongly and weakly constraining sentences, shown separately for correct and incorrect guesses (made during study), under cerebellar and vertex stimulation.



References

[1] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

[2] Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002-1044.

[3] Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition*, *211*, 104650.

[4] Moberget, T., & Ivry, R. B. (2016). Cerebellar contributions to motor control and language comprehension: Searching for common computational principles. *Annals of the New York Academy of Sciences, 1369*(1), 154-171.

[5] Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS disrupts predictive language processing. *Current Biology*, *22*(18), R794-R795.

[6] Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B., & Endestad, T. (2014). Generalized role for the cerebellum in encoding internal models: Evidence from semantic processing. *Journal of Neuroscience, 34*(8), 2871-2878.

Neural Decoding of Pragmatic Inferential Processing in First and Second Language

Bohyun Tak[1, 2], Jihun Im[1], Ha-a-yan Jang[1], Jungmin Moon[1], Sung-Eun Lee[1, 2, 3]
[1]Artificial Intelligence Institute of Seoul National University; [2]Interdisciplinary Program in Cognitive Science, Seoul National University; [3]Department of German Language and Literature, Seoul National University
cristlo5@snu.ac.kr

Background: Understanding how the brain decodes pragmatic inferential processing remains a challenge, as it involves nonliteral meaning and difficulty in prediction. While neural decoding of other linguistic elements has been widely studied [1], the pragmatic inferential processing remains largely unexplored. This study aims to address this gap by investigating neural decoding of pragmatic inferential processing. In addition, it remains unknown how neural decoding patterns differ between first-language (L1) and second-language (L2) processing. Few studies have investigated neural decoding of pragmatic inferential processing in L1 and L2 contexts, though pragmatic processing has been studied at the neural level. This study further aims to explore differences in neural decoding patterns between L1 and L2 conditions focusing on irony and indirect speech as representative types of pragmatic inference.

Method: We employ a support vector machine (SVM) [2] as a single model and a Conformer model—which combines CNNs with a Transformer [3]—as a combined model to examine whether L1 and L2 processing can be distinguished and predicted from EEG data. Starting from 38 original 64-channel EEG recordings (19 participants × 2 conditions; control vs. target(irony or indirect speech)), we apply five rounds of cumulative data augmentation, resulting in 1,216 samples. Each sample contains a 1,200 ms EEG recording, from which we extract ERP segments based on time windows that show significant condition differences in ERP analyses (600–800 ms for irony, 500–700 ms for indirect speech, see figure 1) for classification.

Results: Across both models, classification accuracies are higher for second-language (L2) than first-language (L1) conditions. In L1, the SVM yields higher accuracy for indirect speech than irony (87.0% vs. 73.9%), whereas the Conformer shows comparable accuracies for the two speech acts (80.4% vs. 80.2%). In L2, both speech acts are classified with similarly high accuracy in both models (SVM: 90.3% for indirect speech, 89.3% for irony; Conformer: 83.0% for indirect speech, 82.7% for irony)(See table 1 and figure 2).

Discussion: While the SVM shows higher but more variable accuracy across conditions, the Conformer maintains stable accuracy, suggesting that the Conformer offers greater robustness and reliability for EEG classification. In the SVM results, indirect speech showed higher classification accuracy than irony, suggesting greater sensitivity to variations in decoding difficulty across speech acts, whereas the Conformer exhibited relatively stable performance. Across both models, classification accuracies were higher for L2 than L1 conditions, suggesting more clearly distinguishable neural representations in L2. This may reflect differences in automaticity or cognitive effort between L1 and L2 processing.

Tables

Table 1. Average Accuracy and Standard Deviation over 20 runs

Model	Condition	Mean Accuracy (%)	Standard Deviation (%)
SVM	L1 Irony	73.9	5.7
	L1 Indirect Speech	87.0	2.8
	L2 Irony	89.3	2.0
	L2 Indirect Speech	90.3	2.3
Conformer	L1 Irony	80.4	2.5
	L1 Indirect Speech	80.2	2.5
	L2 Irony	82.7	1.6
	L2 Indirect Speech	83.0	1.6

Figures

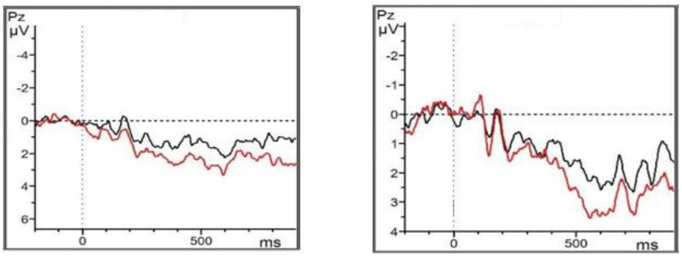


Figure 1. ERP waves for L1 processing: *Irony* (left) and *Indirect speech* (right)

(black line: control, red line: target; irony or indirect speech)

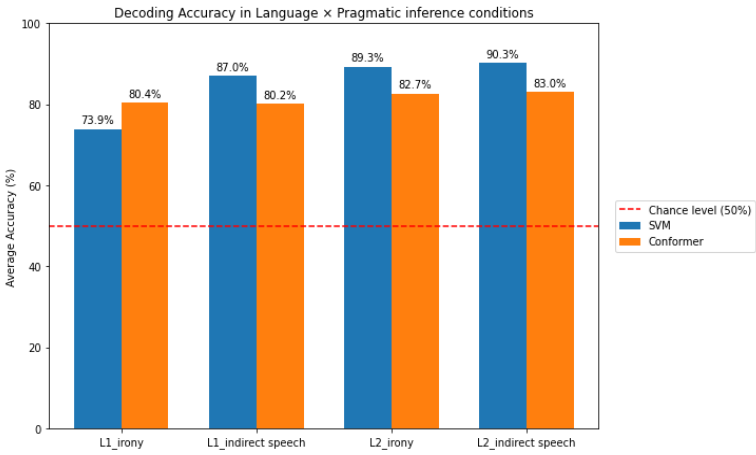


Figure 2. Decoding Accuracy in Language × Pragmatic inference conditions

References

[1] Rybář, M., & Daly, I. (2022). Neural decoding of semantic concepts: A systematic literature review. *Journal of Neural Engineering*, *19*(2), 021002.

[2] Trammel, T., Khodayari, N., Luck, S. J., Traxler, M. J., & Swaab, T. Y. (2023). Decoding semantic relatedness and prediction from EEG: A classification method comparison. *NeuroImage*, *277*, 120268.

[3] Song, Y., Zheng, Q., Liu, B., & Gao, X. (2022). EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 710-719.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(0670-20240045).

World knowledge without world knowledge: Winograd meets the Jabberwocky

Joshua K. Hartshorne
Communication Sciences & Disorders, MGH Institute of Health Professions
jkhartshorne@fas.harvard.edu

1. The city council denied the protesters a permit because they feared violence. [*they* = city council]
2. The city council denied the protesters a permit because they advocated violence. [*they* = protesters]

As these examples show, a great deal of world knowledge can go into something as basic to language as interpreting a pronoun [1]. For this reason, such sentences have been used as a benchmark for artificial intelligence [2]. What about when the semantic context after the pronoun is bleached?:

3. The city council denied the protesters a permit because they were daxes.
4. The city council denied the protesters a permit because they blicketed.
5. The city council denied the protesters a permit because they gorped them.

Because listeners have no specific world knowledge about daxes, blicketing, or gorping to bring to bear, one might expect them to take into account only the material prior to the pronoun. If so, the task reduces to the well-studied implicit causality phenomenon [3, 4, 5, 6], where the verb in the first clause of an explanatory discourse systemtically affects initial interpretation of the pronoun:

6. Alfred frightened Bart because he... [*he* = Alfred]
7. Alfred liked Bart because he... [*he* = Bart]

The verb *deny* is object-biased [4], so under this hypothesis *they* should refer to the protesters in (3-5). Alternatively, listeners may still take into account the bleached, abstract differences in (3-5). This could involve heuristics, learned cues, or even averaging over all possible meanings of *daxes*, *blicket*, and *gorp*. Some initial evidence comes from [7], who tested hundreds of the same verbs as [4] but found an overall shift of biases towards the object. The only major methodological difference is that the former used sentences concluding with a verb like in (e.g., 4), whereas the latter used sentences ending in a noun (e.g., 3).

To test this possibility systematically, we presented 365 English-speaking participants with 32 implicit causality trials. On each trial, they viewed a sentence fragment with one of 503 verbs (drawn from 8 Levin classes [8, 9]) and one male and one female character like *Alfred frightened Beatrice because...*. Participants were asked to continue the sentence using one of two options. There were four conditions: **truncated condition** (*he...* vs. *she...*), **noun condition** (*he is a dax.* vs. *she is a dax.*), **verb condition** (*he blicketed.* vs. *she blicketed.*), and **two referents condition** (*he gorped her.* vs. *she gorped her.*). The choice of pronoun gender indicates pronoun reference bias. All the obvious things were randomized across participants and items. 62 subjects were excluded for failing any of 4 catch trials.

We used a logistic mixed effects regression with random intercepts for verb and condition (random slopes were computationally intractable). Compared with the truncated condition, the noun condition was non-significantly subject-biased ($p = .07$); the verb condition, object-biased ($p = .02$); and two referents, subject-biased ($p = .05$) (Table 1). Interactions between Levin class and condition were minimal. Thus, even semantically bleached conclusions systematically drive Winograd Schema behavior. Implications for theory are discussed.

Levin Class	truncated	noun	verb	two referents
30.1	−0.12	−0.03	−1.63	0.14
31.1	1.34	1.32	1.16	1.08
31.2	−0.78	−0.82	−0.77	−0.44
31.3	−0.28	−0.37	−0.91	−0.05
31.4	1.70	1.03	0.41	0.07
33	−0.45	−0.32	−0.95	−0.49
45.4	0.39	0.60	−0.02	0.42
59	0.94	0.80	0.46	0.70

Table 1: Log-odds that pronoun refers to the subject, by Levin verb class (using the VerbNet numbering convention) and sentence condition. Positive values indicate a subject bias; negative values, an object bias. English verbs can be classified into around 500 groups (Levin classes) based syntactic behavior, and implicit causality biases have been showed to vary systematically by verb class [4, 7], making Levin class a convenient way of summarizing data for 503 verbs.

References

[1] Winograd, T. (1972). Understanding natural language. *Cognitive psychology*, *3*(1), 1–191.

[2] Levesque, H. J., Davis, E., & Morgenstern, L. (2012). The winograd schema challenge. *KR*, *2012*, 13th.

[3] Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic inquiry*, *5*(3), 459–464.

[4] Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474–1508.

[5] Bott, O., & Solstad, T. (2021). Discourse expectations: Explaining the implicit causality biases of verbs. *Linguistics*, *59*(2), 361–416.

[6] Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in english: A corpus of 300 verbs. *Behavior Research Methods*, *43*, 124–135.

[7] Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, cognition and neuroscience*, *30*(6), 716–734.

[8] Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

[9] Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2006). Extending verbnet with novel verb classes. *LREC*, 1027–1032.

Interpreting Plural Predication in Visual Contexts: Cover-Based Resolution of NP Structures

Hening Wang*[1 1], Ren Jia*[2], Michael Franke[1]

[1] University of Tübingen; [2] University of Massachusetts Amherst

hening.wang@uni-tuebingen.de

Introduction: Plurals allow for different interpretations, for instance the distributive and the collective interpretations [1]. Understanding how plurals are internally structured and mentally represented is a central issue in explaining the various interpretations and the semantics and pragmatics of plurals in general. [2] influentially proposes to formalize plurals using covers. Despite the formal elegancy, the mechanism of cover selection is not specified. Experimental studies show that cover selections or the interpretation of plural sentences are influenced by various factors. [3] observed that visual contexts influence the interpretation of plurals. For example, for predicates that are known to be stubbornly distributive, collective readings become available with the help of visual context. [4] provides evidence showing that the availability of structured plurality readings depends on the form of plurals. In this study, we further explore the idea of cover selection via pragmatic reasoning by using visual-context-based experiments to disentangle which covers are available for specific NP structures and what covers are considered more accessible. Specifically, we ask: 1. Are different covers equally available for definite plurals and conjunctions; 2. What pragmatic factors influence the accessibility of covers.

Experiments: We conducted two web-based experiments using a picture-sentence slider rating task. Examples of items are visualized in Fig. 1. The experiments differed in the between-subject factor NP structure. In one experiment, we investigated the NP structure Definite Plural (e.g., *the circles*), and Conjunction (e.g., *the red circles and the blue circles*) in the other. We treated NP structure as a between-subject factor. Both experiments shared a common design: a within-subject factor of Visual Context, which included three levels — *notGrouped*, *randomGrouped*, and *colorGrouped*. The *notGrouped* condition serves as a baseline that is semantically false, while the *randomGrouped* condition involves grouping but does not support a symmetric reading. In the Conjunction condition, we further manipulated a factor called Comparison, indicating whether or not a comparison class (featuring objects of a distinct shape and color) was present.

Results: The results of the experiments (see Fig. 2 for details) show that: 1. Legitimate covers are available for both definite plurals and conjunctions; 2. Speakers find covers with low variation more accessible.
The results conform to Schwarzschild. It interestingly contradicts Buccola et al., in that no fundamental differences were observed between plurals with different forms with the support of visual contexts.

References

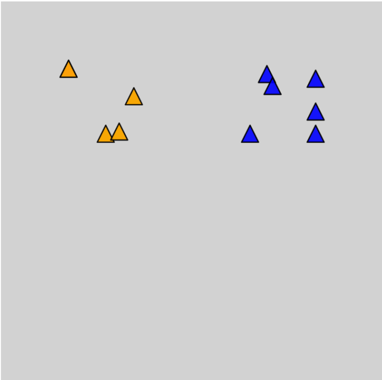
[1] Scha, R. J. (1984). Distributive, collective and cumulative quantification. In J. Groenendijk, T. M. V. Janssen, & M. Stokhof (Eds.), *Selected papers from the third amsterdam colloquium* (pp. 131–158). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110867602.131

[2] Schwarzschild, R. (1996). *Pluralities*. Springer Netherlands. https://books.google.si/books?id=rFk8Ts7RwDYC

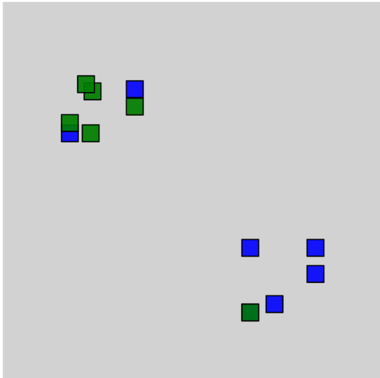
[3] Scontras, G., & Goodman, N. D. (2017). Resolving uncertainty in plural predication. *Cognition*, *168*, 294–311. https://doi.org/10.1016/j.cognition.2017.07.002

[4] Buccola, B., Kuhn, J., & Nicolas, D. (2021). Groups versus covers revisited: Structured pluralities and symmetric readings. *Natural Language Semantics*, *29*(4), 509–525. https://doi.org/10.1007/s11050-021-09179-x

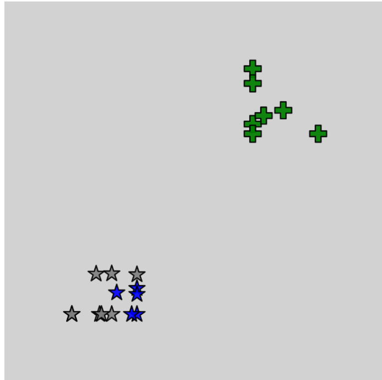
[1]* means equal contribution



(a) **colorGrouped:** Objects are grouped based on color; each cluster contains items of the same color.

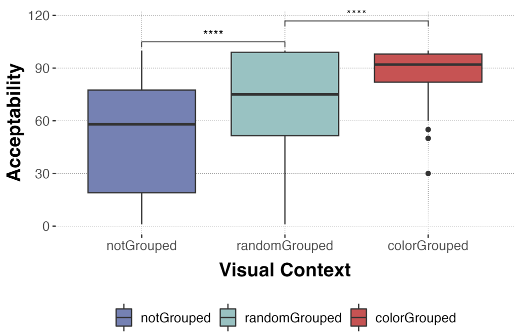


(b) **randomGrouped:** Objects are grouped into two spatial clusters, but colors are randomly distributed within each cluster.

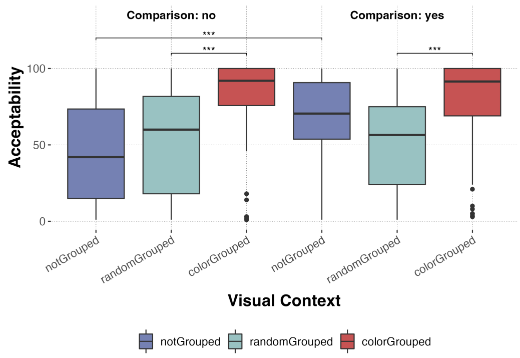


(c) **notGrouped with comparison:** Objects form a single unstructured cluster. An additional cluster introduces an comparison object class with distinct colors and shapes.

Figure 1: Three example items from the experiment, illustrating the overall design. The examples demonstrate the three levels of the factor Visual Context and the two types of NP structure manipulated in the study.



(a) Results of Definite Plural, exemplified by the form *The circles*



(b) Results of Conjunction, exemplified by the form *The red circles and the blue circles*

Figure 2: Two boxplots from the experiments, each corresponding to a different NP structure. Left: Definite Plural; Right: Conjunction. The X-axis represents the three levels of the factor Visual Context, which are also color-coded. The Y-axis shows acceptability ratings, ranging from 0 (completely unacceptable) to 100 (completely acceptable). For the NP conjunction (right), we additionally manipulate the presence of a comparison class, visualized via facet wrap. Significant pairwise differences between groups from t-tests are indicated by asterisks (*)

Processing of novel metaphors in an intergroup context

Kuperwasser, I. [a], Shetreet, E. [a,b]
[a] Sagol School of Neuroscience, Tel Aviv University [b] Department of Linguistics, Tel Aviv University
inbalk@mail.tau.ac.il

Intergroup interactions (i.e., when interlocutors are not part of the same social group) have been shown to negatively affect abilities linked to the facilitation of pragmatic processing, like Theory of Mind (ToM) [1,2]. This suggests intergroup interactions are likely to interfere with pragmatic processing, which is supported by previous work on scalar implicatures [3]. Novel metaphors lack established interpretations, so their processing likely depends on the listener's cognitive ability to infer the speaker's intended meaning, namely ToM [4; although see 5]. Moreover, recent research suggests an association between ToM and metaphor type, differentiating between mental metaphors, requiring inference on mental attributes, and physical ones, requiring inference on physical attributes. It has been shown [6,7] that ToM abilities were more relevant for interpreting mental novel metaphors. Therefore, we examine whether processing of novel metaphors, and specifically, their type, is affected by intergroup settings. We hypothesize intergroup settings would lead to lower rates of successful metaphoric comprehension than non-group settings, and more physical (compared to mental) interpretations.

Participants (N = 170; American native English speakers identified as Democrats) were divided into two group conditions: intergroup settings, and non-group settings. Participants read vignettes which included a short conversation (Table 1). In critical trials (N=5), the conversation ended in a novel nominal metaphor, whereas filler trials (N=5) ended with a conventional one. All novel metaphors were similarly unfamiliar and neutral in content in relation to US politics. We chose metaphors and vignettes that gave rise to both mental *and* physical interpretations, to allow a larger variation in interpretation. In the intergroup settings condition, speakers using novel metaphors were indirectly identified as Republicans (Table 1), while in non-group settings, none of the speakers were politically affiliated. Participants were then asked to give an interpretation for the metaphor.

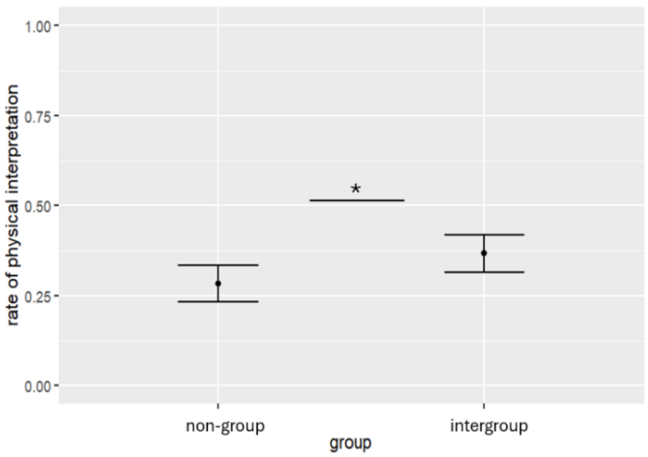
Responses were coded for both metaphoric interpretation (articulating the connection between the topic and vehicle) and for the interpretation type (physical\mental). We carried out mixed-effects models (with fixed effect of group settings and random intercepts of items and participants), for the rate of metaphoric interpretations and for the rate of physical interpretations. The model testing metaphoric interpretations did not reveal an effect of group settings (p = 0.37). However, the model testing physical interpretations revealed an effect of group (p < 0.05) where participants in intergroup settings were significantly more likely to give a physical interpretation to novel metaphors than those in the non-group settings (Fig. 1).

This study revealed a different pattern in interpretation of novel metaphors in intergroup context. Specifically, participants were more likely to interpret a metaphor to refer to physical properties of the metaphor's topic if the speaker was an outgroup member. This adds to growing evidence that intergroup settings affect patterns of pragmatic interpretation. Additionally, the stronger tendency towards physical interpretations in the intergroup settings may be the result of an impairment in ToM in intergroup interactions, as previous research link ToM abilities to the comprehension of mental metaphors. We plan to investigate this link more directly in the future with a ToM task.

Table 1. An example of a critical trial in the **intergroup/non-group** settings. Use of bold is to highlight differences between conditions and did not appear in the original task. Note the use of a comprehension question verifying the recognition of the speaker's political affiliation (when appropriate).

Vignette	You just settled into your seat at a family member's school talent show. Beside you a teenage girl with a **'Make America Great Again'/purple** phone case is looking at her phone with her female friend. Friend: 'He looks kinda cute didn't he take French with you last semester?" Girl: "Him? What you can't see in that picture is that he is an insect"
Comp. Q1	Where did you see the girl? *Possible answers:* Concert; School talent show; Baseball game
Comp. Q2	Which is true for the girl? *Possible answers:* **She is a Republican/She has a purple phone case**; She has green eyes; She has a dog
Critical Q	What did the girl mean when she said, "He is an insect"? *Open textbox response*

Figure 1. Plot of predicted physical interpretations in by group settings



References:

[1] Fairchild, S., & Papafragou, A. (2021). *Cognitive Science*, *45*(2). [2] Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). *JESP*, *52*. [3] Kuperwasser, I., & Shetreet, E. (2025). *Journal of Pragmatics, 242*. [4] Happé, F. G. (1993). *Cognition*, *48*(2). [5] Norbury, C. F. (2005). *British journal of developmental psychology*, *23*(3). [6] Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). *Journal of child language*, *46*(2). [7] Canal, P., Bischetti, L., Bertini, C., Ricci, I., Lecce, S., & Bambini, V. (2022). *Brain and Cognition*, *161*.

The role of information structure for subject position: evidence from Romance languages

Fabian Istrate[a], Laia Mayol[b], Gabriela Bîlbîie[c] & Barbara Hemforth[a]
[a] Université Paris Cité/CNRS, [b] Universitat Pompeu Fabra, [c] University of Bucharest
istratemariusfabian@gmail.com, laia.mayol@upf.edu, gabriela.bilbiie@lls.unibuc.ro, barbara.hemforth@u-paris.fr

Studies on Romance languages have shown cross-linguistic variation in subject position, with some languages exhibiting greater word order flexibility (Leonetti, 2017; Cruschina & Mayol, 2024). This variation is often linked to information structure: broad focus contexts tend to favor pre-verbal subjects, while narrow focus contexts allow post-verbal ones - a pattern attested in Italian, Spanish, and Catalan (Zubizarreta, 1998; Belletti, 2001; Cardinetti, 2004). Romanian, however, may diverge by allowing more post-verbal subjects even in broad focus (thetic) contexts, possibly due to its unmarked VS word order (Dobrovie-Sorin, 1994; Giurgea, 2017). To investigate this, we conduct fully parallel experiments in Italian, Spanish, Catalan and Romanian to examine how information structure and verb semantics influence subject position. Prior work suggests that post-verbal subjects typically occur in narrow focus answers, where new information is encoded sentence-finally (Marandin, 2003). Romanian's frequent use of post-verbal subjects in broad focus may therefore contrast with the SV preference found in other Romance languages. Verb semantics also plays a role: unergative verbs tend to favor pre-verbal subjects, as their subjects are external arguments, unlike unaccusatives, whose subjects are internal arguments integrated into the predicate (Benincà & Poletto, 2004; Belletti et al., 2007).

We ran a series of fully parallel experiments in which native speakers recruited on Prolific (46 Romanian, 49 Italian, 48 Catalan and 48 Spanish) had to choose between an answer with a pre-verbal or a post-verbal subject. The binary forced-choice task we created on Ibex Farm consisted of 20 items, 20 fillers and 10 control questions. Each item was constructed as a dialogue between two speakers, and contained three elements: a statement, a question (either with broad focus or with narrow focus) and and two possible answers (either with a pre-verbal or a post-verbal subject). All answers in our items had intransitive verbs (half unaccusative (1), half unergative (2)).

Interestingly, the focus type seems to play a crucial role in subject position (Fig. 1), however not entirely as expected. In broad focus contexts, participants showed a strong preference for pre-verbal subjects in Spanish (70.2%), Italian (70.8%) and Catalan (70.6%), but for post-verbal subjects in Romanian (58.7%). Equally surprisingly, we found a preference for pre-verbal subjects in narrow focus contexts in Romanian (69.6%) while post-verbal subjects are preferred in Spanish (66.4%), Italian (68.3%) and Catalan (53.7%) (*glmer*, $\beta = -0.89$, sd = .07, z = -11.67, $p < .001$). Subject position was also influenced by verb type: unergative verbs allowed more pre-verbal subjects, as predicted in the literature (glmer, $\beta = 0.77$, SD = .33, z = 2.28, $p < .01$; Fig. 2).

We conclude that both information structure and verb semantics strongly influence subject position in Romance languages. In line with previous findings, broad focus contexts in Romanian allow more post-verbal subjects compared to other Romance languages which tend to prefer pre-verbal subjects. Moreover, Romanian also shows an unexpected pattern by permitting pre-verbal subjects in narrow focus contexts. One possible explanation is that Romanian might be syntactically less restrictive, allowing focus fronting to the left periphery even without mirative or contrastive interpretations (see Cruschina, 2022). Additionally, different strategies for encoding focus - prosodic and/or syntactic - may be involved, as Romanian relies mostly on prosody in the absence of cleft constructions.

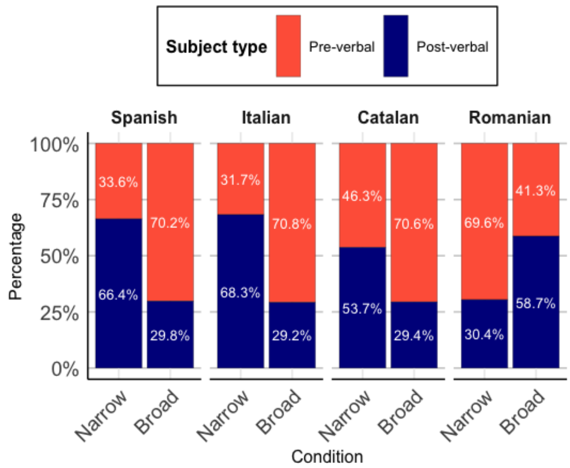


Fig. 1. Subject position & focus type (overall)

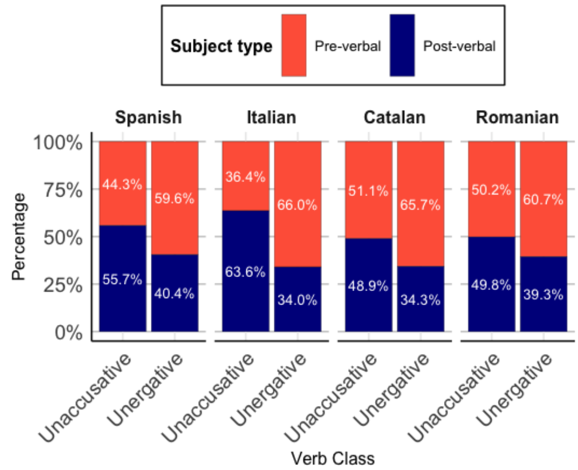


Fig. 2. Subject position & verb semantic class

Selected references

[1] Belletti, A. (2001). “Inversion” as focalization. *Subject inversion in Romance and the theory of Universal Grammar*, 60-90.

[2] Leonetti, M. (2017). 24. Basic constituent orders. *Manual of Romance morphosyntax and syntax*, 887-932.

[3] Giurgea, Ion. (2017). Preverbal subjects and topic marking in Romanian. *Revue roumaine de linguistique* 63(3). 279–322.

[4] Cruschina S, Mayol L. (2024). The distribution of focus and the mapping between syntax and information structure. *Journal of Linguistics*. Published online 2024:1-35. doi:10.1017/S0022226724000264

Examples of experimental items

(1) Persoana A: - Am făcut o cursă cu bicicletele, dar nu prea am fost atenți când am trecut pe lângă terenul de fotbal.

‘Speaker A : - We went for a bike ride, but we weren't paying much attention when we passed the soccer field.’

[Broad focus]:

Persoana B: - *Ce s-a întâmplat până la urmă?* Îmi amintesc că zona aceea nu prea e ideală pentru mersul cu bicicleta.

‘Speaker B: - *What happened in the end?* I remember that area is not really ideal for cycling.’

[Narrow focus]:

Persoana B: - *Cine s-a rănit tocmai lângă teren?* Îmi amintesc că zona aceea nu prea e ideală pentru mersul cu bicicleta.

‘Speaker B: - *Who got injured near the field?* I remember that area is not ideal for cycling.’

Ce răspuns va alege persoana A?

‘What answer will speaker A choose?’

[Unaccusative construction]

Pre-verbal subject: Raluca s-a rănit. ‘Raluca got hurt.’

Post-verbal subject: S-a rănit Raluca. ‘Raluca got hurt.’

(2) Persoana A: Am găsit un club de karaoke și, deși aveam chef să ne distrăm, parcă ne era teamă să urcăm pe scenă.

‘Speaker A: We found a karaoke club and although we wanted to have fun, we were afraid to go on stage.’

[Broad focus]:

Persoana B: - *Ce s-a întâmplat până la urmă?* Eu n-am fost niciodată la karaoke.

‘Speaker B: - *What happened in the end?* I've never been to karaoke.’

[Narrow focus]:

Persoana B: - *Cine a cântat primul?* Eu n-am fost niciodată la karaoke.

‘Speaker B: - *Who sang first?* I've never been to karaoke.’

Ce răspuns va alege persoana A?

‘What answer will speaker A choose?’

[Unergative construction]

Pre-verbal subject: Filip a cântat. ‘Filip sang.’

Post-verbal subject: A cântat Filip. ‘Filip sang.’

The role of distance on pronoun resolution: Evidence for a two-stage model

Quesada, Teresa[1], Torregrossa, Jacopo[2], Lozano, Cristóbal [1]

[1] University of Granada; [2] Goethe University Frankfurt am Main

teresaquesada@ugr.es

We investigate how the processing of overt pronouns (OPs) and repeated names (RNs) in written discourse is influenced by their distance from an antecedent. In short two-sentence discourses such as (1), OPs are typically more felicitous than RNs, reflected in greater processing difficulty for RNs [see the repeated name penalty RNP in 1]. However, the role of distance has been underexplored. If intervening sentences are inserted between a referring expression (RE) and its antecedent (2), RNs may become more felicitous than OPs [2-4].

(1) Luke$_i$ was watering the plants in the garden. **Luke$_i$** / **He$_i$** / was completely happy.

(2) Luke$_i$ was watering the plants in the garden. It was a beautiful and warm autumn day. Leaves were moving softly in the breeze. Kites were flying across the shiny blue sky. Luke$_i$ / He$_i$ was completely happy.

Comparing short (1) vs. long (2) distance can shed light on the processing of OPs. Some models propose that OPs are resolved immediately upon encounter [5], while others posit a two-stage process: initial access to syntactic features (e.g., gender, number), followed by delayed antecedent linking [6,7]. We conducted three experiments to test the effect of RE form (OP vs. RN) and distance (0 vs. 3 intervening clauses) on processing. **Experiment 1** employed a speeded forced-choice task to investigate participants' preferences for OPs vs. RNs at varying distances, and the RTs of their choice. After reading either a short (1-sentence) or long (3-sentence) context, participants pressed a key to reveal two versions of a continuation sentence (one with an OP, one with a RN), then selected their preferred option. Each of two lists contained 32 experimental items and 32 fillers. 100 monolingual English speakers (mean age: 28.5) participated. Results showed a clear preference for OPs over RNs at short distances, which weakened at longer distances (Fig. 1). However, reaction times (RTs) revealed that selecting RNs was consistently more costly than selecting OPs, regardless of distance (Fig. 2). **Experiment 2** and **Experiment 3** employed a sentence-by-sentence self-paced reading task. Participants read either a one- or a three-sentence context, followed by a target sentence containing either an OP or a RN. In Experiment 2, the target sentence was followed by a comprehension question unrelated to referential resolution (*Was Luke watering the plants in the living room?*), allowing us to examine whether the presence of an OP or RN affected overall comprehension. In Experiment 3, by contrast, the question directly probed referential resolution (*Who was completely happy?*), allowing for a more direct test of processing predictions from models of pronoun resolution. Each comprehension question offered two response options (one correct and one incorrect) and RTs for the responses were recorded. The stimuli for Experiment 2 and 3 were identical to those used in Experiment 1 but adapted to the task at hand. We tested 100 (mean age: 29.08) and 80 participants (mean age: 29.3), respectively. In both experiments, target sentences containing RNs elicited longer reading times than those with OPs, regardless of distance (Fig. 3). Crucally, in Experiment 3, comprehension question RTs were longer for OPs than RNs in the three-sentence context, suggesting increased cognitive demands in referential resolution for OPs in more complex discourse (Fig. 4). Our findings suggest that although participants are more likely to choose RNs over OPs as distance increases, this preference appears driven by explicit discourse-level expectations rather than implicit ones. Indeed, RNs consistently incurred higher processing effort across conditions except for the comprehension question RTs in Exp. 3. These results support a two-stage model of OP processing: initial fluent access to syntactic features (gender, number), followed by a delayed antecedent linking to achieve full resolution, incurring greater effort under certain task demands.

Figures

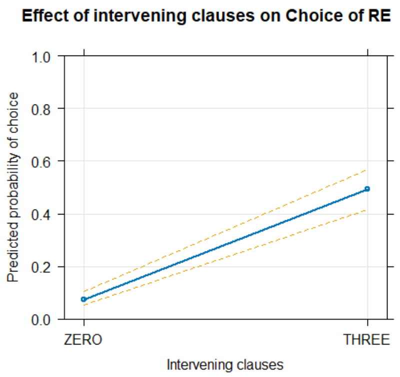


Figure 1

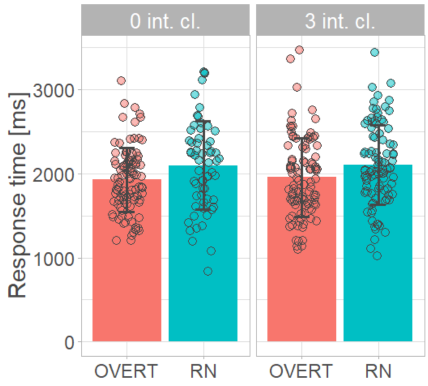


Figure 2

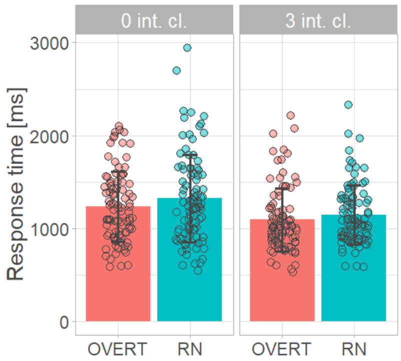


Figure 3

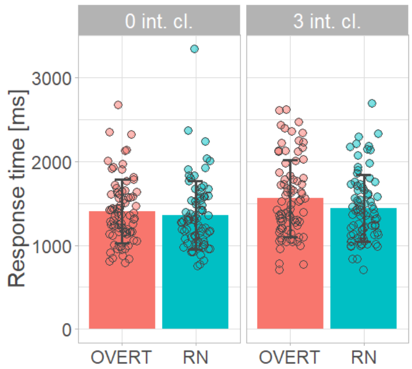


Figure 4

References

[1] Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*, *17*(3), 311-347.

[2] Arnold, J. (1998). *Reference Form and Discourse Patterns*. Stanford University.

[3] Kibrik, A., Khudyakova, M. V., Dobrov, G. B., Linnik, A., & Zalmanov, D. A. (2016). Referential Choice: Predictability and Its Limits. *Frontiers in Psychology*, *7*.

[4] Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641-655.

[5] Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(2), 266-283.

[6] Rigalleau, F., Caplan, D., & Baudiffier, V. (2004). New arguments in favour of an automatic gender pronominal process. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *57*(5), 893-933.

[7] Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology (2006)*, *60*(12), 1680-1696.

When girls are pearls, are they pleasant or are they beautiful? Capturing variation in metaphor interpretation via Intersubject Representational Dissimilarity

Chiara Battaglini[1], Federico Frau[1], Veronica Mangiaterra[1], Luca Bischetti[1], Paolo Canal[1], and Valentina Bambini[1]
[1]University School for Advanced Studies IUSS, Laboratory of Neurolinguistics and Experimental Pragmatics (NEP), Pavia – Italy
chiara.battaglini@iusspavia.it

Background: It is well known that metaphors are open to different interpretations, and that interpretations vary across individuals, even within the same cultural context [1]. Pioneering research has shown that people interpret metaphors differently depending on their cognitive styles [1–3]. However, little attention has been devoted to investigating which individual factors drive differences in preferred interpretation. Capitalizing on previous evidence highlighting individual differences in the recruitment of sensory and affective systems in metaphor understanding [4,5], here we aimed to test whether different metaphor interpretative profiles can be linked to the recruitment of sensory-motor vs. mentalizing processes. To this purpose, we applied a novel data-driven approach combining (i) Natural Language Processing (NLP) analysis on metaphor verbalization to extract a series of variables that are indicative of mindreading, emotional, and sensory-based processes [6] and (ii) Intersubject Representational Dissimilarity (IS-RDM). The latter technique allows to quantify individual differences in high-dimensional multimodal profile space, describing and comparing participant-level geometrical representations [7,8], and, while popular in cognitive neuroscience [9,10], has never been used to study individual differences in language interpretation.
Method: We analyzed the interpretations, automatically tokenized and lemmatized, provided by 76 native Italian speakers (Age: 22.08±1.52, Education: 15±0.76) asked to explain the meaning of 129 nominal metaphors. For each participant, we computed mean values of the occurrence of words denoting affective, cognitive, and sensory processes (with the LIWC software [11]), concreteness, imageability, valence, and dominance (extracted from existing resources [12,13] and type-token ratio. Concatenated values were used to generate vectorized interpretative styles on which we computed Intersubject RDMs and applied an unsupervised hierarchical agglomerative clustering, using the farthest neighbor metric, validated with a repeated hold-out cross-validation Linear Discriminant Analysis (LDA) [14].
Results: Two clusters of participants emerged (Figure 1): the first cluster – mentalizers, includes participants whose interpretations were more verbose (p<.001), relied more on words referring to cognitive mechanisms (p=.022), characterized by higher valence (p<.001) and dominance (p<.001). On the other hand, participants assigned to the second cluster – imagers, produced interpretations that exhibited higher values of concreteness (p<.001), imageability (p<.001), and arousal (p<.05). For instance, participants in cluster 1 interpreted Girls are pearls as being pleasant whereas participants in cluster 2 interpreted the same metaphors as being beautiful (see Table 1 for further example).
Discussion: These findings show that - when interpreting metaphors - participants preferentially recruit different modalities, resulting in distinct interpretative profiles. This highlights the importance of considering individual differences as well as multimodal components of figurative language understanding. Also, our results offer a ground to reconcile contrasting theories of metaphor, as the two clusters may tap into different mechanisms postulated by different theoretical models: analogy for mentalizers and simulation for imagers. As argued by recent proposals within the Relevance Theory framework, there might be multiple alternative routes to solving metaphorical meaning [15].

References

[1] Littlemore & Li Metaphor Symb 2019;34:258-261. [2] Birdsell J Cogn Sci 2018;19:35-67. [3] Boers et al Metaphor and Symbol 2009;15:177-187. [4] Canal et al *Brain Cogn* 2022;161. [5] Bambini et al *Discourse Process* 2024;61:44-68. [6] Bambini et al *Schizophrenia* 2022;8(1):1-12. [7] Nguyen et al *Neuroimage* 2019;184:161-170. [8] Hu et al *Neuroimage* 2022;247. [9] Kriegeskorte et al *Front Syst Neurosci* 2008; 2(249). [10] Kriegeskorte et al *Trends Cogn Sci* 2013;17:410-412. [11] Boyd et al *University of Texas at Austin* 2022;10:1-47. [12] Ljubešić et al *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1018; 217–222. [13] Buechel et al *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 2020; 1202-1217. [14] Bambini et al *Clin Linguist Phon* 2025:1-23. [15] Carston *Proceedings of the Aristotelean Society* 2010;110:295-321.

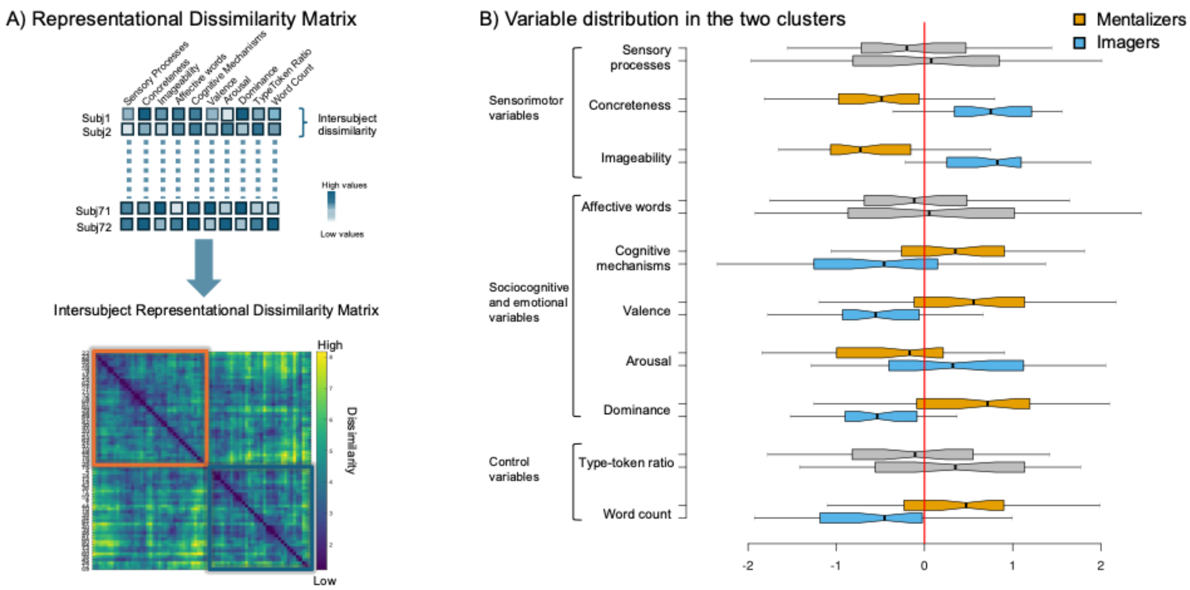


Figure 1. A) Top: Vectorization of the psycholinguistic variables per participant. Bottom: 72x72 intersubject dissimilarity matrix computed on the variables: rows represent participants, each cell represents the Euclidean distance between participants. Cold colors:low dissimilarity, warm colors:high dissimilarity. The two clusters are highlighted in orange and blue. B) Whisker plots of the distribution of normalized values of each psycholinguistic measure in the two clusters (vertical lines correspond to the median value). For each cluster, the defining features are highlighted in orange (mentalizers) and blue (imagers).

Prompt	Some professors are lanterns *Certi professori sono lanterne*	Some children are storm *Alcuni marmocchi sono tempeste*	Some girls are pearls *Alcune ragazze sono perle*	Some policemen are mastiff *Certi poliziotti sono mastini*
Mentalizers Interpretation	They are able to make others understand whatever they want to *Sono in grado di far capire quello che vogliono*	They are un-controllable *Sono incon- trollabili*	They have such pleasant characteristics, they are rare exemplars *Hanno caratteristiche tanto gradevoli da renderle esemplari rari*	They are though in character *Hanno un carattere duro*
Imagers Interpretation	They are luminous *Sono luminosi*	They are noisy *Sono rumorosi*	They are beautiful *Sono belle*	They are massive *Sono grossi*

Table 1. Examples of interpretations from participants in the two clusters, for the same metaphors. The original Italian version is reported in Italics.

Referential Resolution in Naturalistic Contexts: Audio-Visual Integration in the N400/P300 Window

Ingmar Brilmayer[1], Petra B. Schumacher
[1] Institute for German Language and Literature I, University of Cologne, Cologne, Germany
ingmar.brilmayer@uni-koeln.de

Spontaneous conversation engages multiple sensory channels: visual signals – eye gaze, facial movements, gestures, and environmental context – merge with acoustic speech to inform linguistic interpretation. Successfully identifying referents therefore requires listeners to process and integrate auditory and visual sensory input. Previous electrophysiological studies have pinpointed the 300–500 ms interval, spanning the N400 and P300 components, as critical for integrating these modalities (e.g. [1],[2],[3],[4]). In the present work, we validate these markers in a highly naturalistic setting by recording EEG alongside eye-tracking from freely-behaving participants during a referential communication task.

Using a modified director task paradigm [5], we induced referential uncertainty via identical object pairs, one obscured from the director's viewpoint (privileged ground) and one mutually visible (common ground). Participants heard instructions such as „Move the candle up," compelling them to infer the intended referent by accounting for the different perspectives based on visual cues (hidden or not). We applied regression-based deconvolution (rERP; [6,7,8]) alongside fixation-related potentials (FRPs) to disentangle overlapping neural events tied to both linguistic and ocular responses.

Gaze behavior revealed that under ambiguity, participants fixated the shared object more often than without a referential conflict, scarcely attending to the hidden competitor (Figure 1, solid red vs. dashed red; competitor: goldish colour). We consider this as evidence of "cognitive offloading" onto the visual scene (Risko & Gilbert, 2016). Electrophysiologically, noun-locked rERPs in unambiguous trials showed a robust posterior P300-like component (Figure 2A; blue), indicative of fast attentional shifts from percpetion and referent identification to task-related behavior (moving the object). On ambiguous trials (Figure2A, red), by contrast, participants exhibited a diminished parietal positivity and emerging frontal negativity within the N400/P300 epoch, signaling increased effort connected with the integration and identification of the ambiguous referent. Furthermore, the magnitude of this effect varied with the timing of target object fixations relative to noun onset: later fixations amplified the response, while pre-noun fixations attenuated it (Figure 2B).

FRP analyses corroborated these findings: target-oriented fixations during ambiguous trials elicited heightened occipital positivities in the P300/N400 time window (Figure 2C, red) modulated by target gaze latency (Figure 2D). This continuum of neural responses highlights the dynamic relation of bottom-up and top-down information in audio-visual linguistic tasks.

In sum, our results demonstrate that simultaneous EEG and eye-tracking recordings can successfully be applied to tap into audio-visual integration within the N400/P300 time window, both replicating and extending previous findings from controlled laboratory studies, advancing both methodological approaches and theoretical frameworks in psycholinguistic research.

References

[1] Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-related Brain Potentials. Journal of Cognitive Neuroscience, 19(4), Article 4. https://doi.org/10.1162/jocn.2007.19.4.605

[2] Peeters, D., Chu, M., Holler, J., Hagoort, P., & Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. Journal of Cognitive Neuroscience, 27(12), Article 12.

[3] Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and Hearing Meaning: ERP and fMRI Evidence of Word versus Picture Integration into a Sentence Context. Journal of Cognitive Neuroscience, 20(7), 1235–1249. https://doi.org/10.1162/jocn.2008.20085

[4] Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. Proceedings of the Royal Society B: Biological Sciences, 288(1955), 20210500. https://doi.org/10.1098/rspb.2021.0500

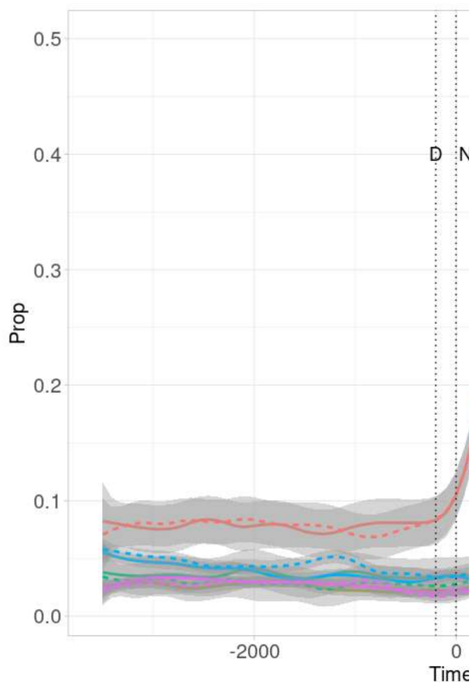
[5] Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological science*, *11*(1), 32-38.

[6] Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. Psychophysiology, 52(2), 157–168. https://doi.org/10.1111/psyp.12317

[7] Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. Psychophysiology, 52(2), 169–181. https://doi.org/10.1111/psyp.12320

[8] Dimigen, O., & Ehinger, B. V. (2021). Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. Journal of Vision, 21(1), 3–3.

[9] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in cognitive sciences, 20(9), 676-688.



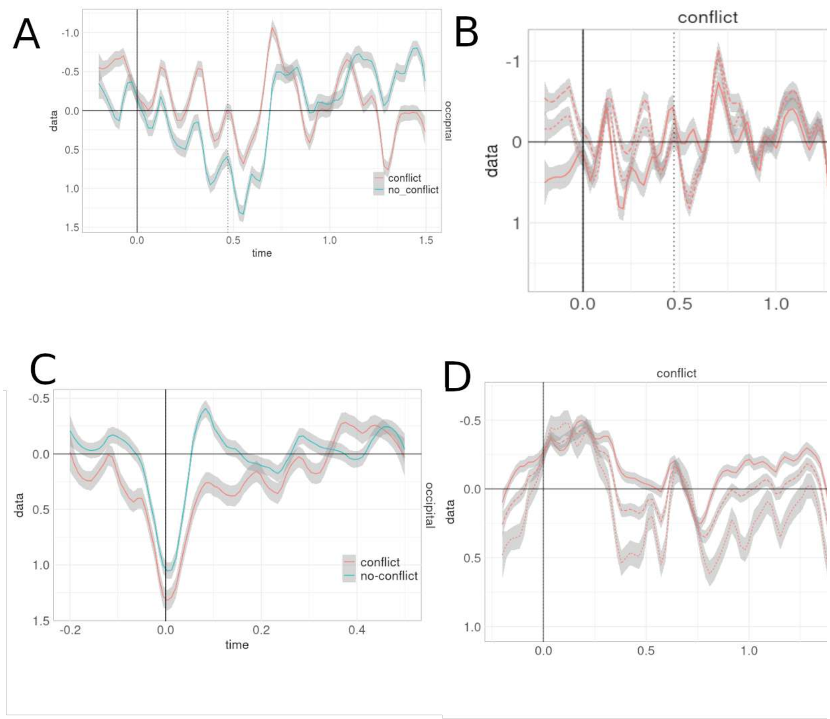


Figure 1: A) Language-related potentials by condition at right occipital ROI; B) Language-related potentials by condition and mean target

Blame it on the verb: Implicit causality verbs and its incidence in relative clause attachment

Noelia A. Stetie[1,2] & Gabriela M. Zunino[1,2]
[1]Universidad de Buenos Aires [2]CONICET
nstetie@conicet.gov.ar

Ambiguous relative clauses with two potential antecedents (Table 1) offer a valuable window into syntactic processing, as they reveal how attachment decisions are made [1] and show both interlinguistic [2, 3] and intralinguistic variability [1, 3, 4]. While Spanish was traditionally considered a high attachment language, attachment preferences remain unclear and recent research suggests that no consistent preference for high or low attachment can be established in Spanish [5], raising further questions about the underlying factors involved.

One key factor influencing attachment appears to be the semantic bias of the main clause verb [3]. Evidence from English suggests that verbs with object-biased implicit causality (e.g., *criticise*) promote high attachment towards the first noun phrase [6]. However, whether similar effects occur in Spanish remains an open question, as little empirical data are available.

In the present research, we examine whether verbs with object-biased implicit causality influence the attachment preferences of ambiguous relative clauses in Rioplatense Spanish. To address this question, we conducted two experiments –a completion task and a lexical maze– with a previous normative study. Our goal was to test whether object-biased verbs would increase the likelihood of high attachment, as observed in studies conducted in English [6].

First, we conducted a norming study (Table 1) with 127 native speakers of Rioplatense Spanish to validate the implicit causality biases of verbs selected from previous Peninsular Spanish norms [7]. Results revealed notable dialectal differences, reinforcing the view of implicit causality as a continuous rather than categorical phenomenon. Based on these findings, we selected a subset of 24 verbs –12 with object-biased implicit causality and 12 with no implicit causality bias– to construct the experimental materials for the subsequent tasks.

Experiment 1 employed a sentence completion task to assess spontaneous attachment preferences. Participants (N = 127) completed sentences containing either an implicit causality verb or a neutral verb (Table 1). Overall, we observed a preference for high attachment and a tendency for implicit causality verbs to elicit more high-attachment completions (77.2%) than neutral verbs (62.9%). This difference was statistically significant (GLMM: $\beta = 0.889$, $SE = 0.407$, $z = 2.183$, $p = 0.029$).

Experiment 2 implemented a lexical maze task with 137 participants to assess real-time processing. Verbs and attachment types were factorially manipulated (Table 1). As shown in Figure 1 and contrary to our hypotheses, we found no significant effects of verb type or attachment type on reaction times at the disambiguating word (LMM: $\beta = -0.011$, $SE = 0.034$, $t = -0.318$, $p = 0.752$) or the spillover word (LMM: $\beta = 0.014$, $SE = 0.042$, $t = 0.323$, $p = 0.748$).

Taken together, these findings suggest that while implicit causality verbs may influence attachment preferences during sentence production, their effect is less evident in online comprehension under conditions of high cognitive demand. This pattern highlights the importance of task demands and processing constraints in the interaction between semantic biases and syntactic parsing.

References

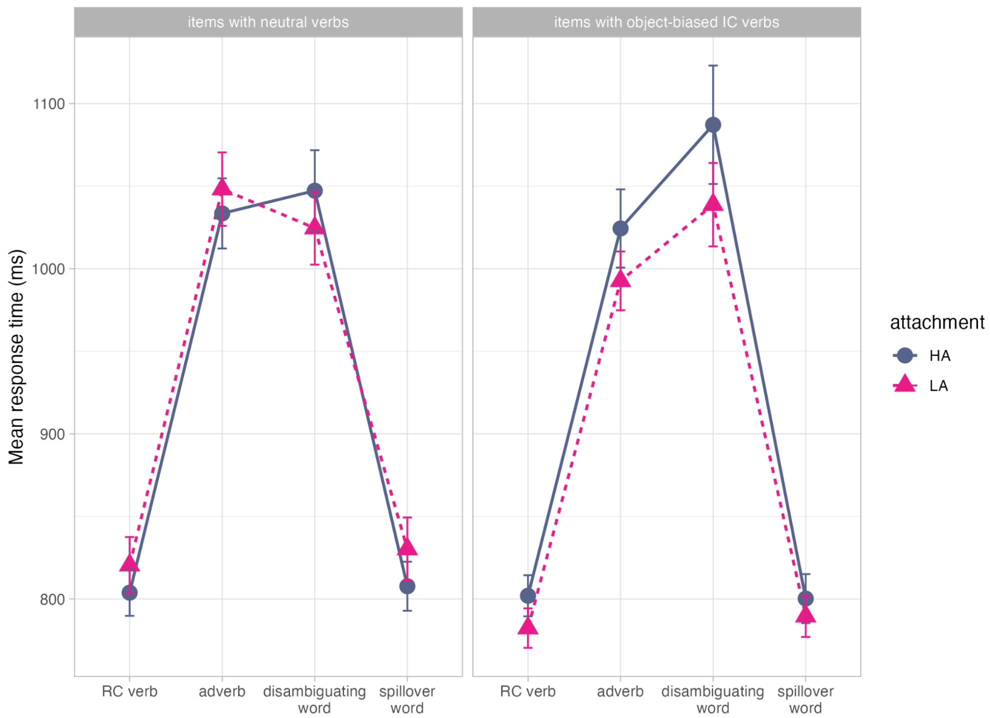
1. Frazier, L., & Clifton, C. (1996). *Construal*. MIT Press.
2. Gibson, E., Pearlmutter, N., Canseco-González, E., & Hickok, G. (1996). Cross-linguistic attachment preferences: Evidence from English and Spanish. *Cognition*, 59.
3. Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, *133*, 156-187.
4. Gilboy, E., Sopena, J. M., Clifton, C., & Frazier, L. (1995). Argument structure and preferences in the processing of Spanish and English complex NPs. *Cognition*, *54*, 131-167.
5. Stetie, N. A. (2021). Cláusulas relativas con doble antecedente nominal en español: un recorrido experimental. *Cuadernos de Lingüística Hispánica*, (38), 1-17.
6. Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*(3), 339-358.
7. Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods*, *40*(3), 760-772.

Table 1. Example items

Experiment	Condition	Example item (in Spanish)
Normative study	IC verb / neutral verb	Diego **eligió / conoció** a Lucía porque…
Experiment 1	IC verb / neutral verb	Lucía **defendió / saludó** a la pasante del funcionario que…
Experiment 2	IC verb HA / LA	Lucía defendió al pasante de la funcionaria que estaba muy **convencido/a** de su opinión.
	neutral verb HA / LA	Lucía conoció al pasante de la funcionaria que resultaba demasiado **antipático/a** en el trato.

Abbreviations. IC verb: object-biased implicit causality verbs. HA: high attachment. LA: low attachment.

Figure 1. Mean response time by verb type and attachment type (Experiment 2)



English translation of the examples presented in Table 1:

Experiment	Condition	Example item (translated to English)
Normative study	IC verb / neutral verb	Diego **chose / met** Lucía because…
Experiment 1	IC verb / neutral verb	Lucía **defended / greeted** the intern-F of the official-M who…
Experiment 2	IC verb HA / LA	Lucía defended the intern-M of the official-F who was very **convinced-M / convinced-F** of the opinion.
	neutral verb HA / LA	Lucía met the intern-M of the official-F who was too **unpleasant-M / unpleasant-F** in their interactions.

The Effect of Adjective Position on Information Recall in Czech

Ondřej Drobil[1], Jan Chromý[1]
Faculty of Arts, Charles University
odrobil@seznam.cz

Background: Previous research has demonstrated that noun modification by adjectives enhances the memory representation of the modifier noun phrase [1]. However, it remains less understood how the adjective position influences memory and processing [2], and to our knowledge, no previous study has directly examined how adjective position affects information recall and language processing. The present study investigates immediate post-sentential recall of information conveyed by adjectives based on their position, taking advantage of the fact that Czech allows both pre- and postnominal modification (for details, see the third page). Based on the Feature-Label-Order principle [3], which assumes that nouns referring to objects narrow down the set of potential features that their modifiers may denote and pre-activate their representations, we hypothesized that this would result in an advantage in the processing of postnominal adjectives, i.e., they would be recalled more accurately.

Method: We conducted four reading experiments using a self-paced reading paradigm and one listening experiment. In Exp. 1, sentences were presented at once; Exp. 2 and 4 employed word-by-word presentation; and Exp. 3 used phrase-by-phrase presentation mode. Exp. 5 contained spoken stimuli generated by AI. In Exp. 1 and 2, we manipulated only the position of the adjective (prenominal vs. postnominal) modifying a locative adjunct. In Exp. 3–5, we introduced a second factor – modifier type (adjective vs. genitive; see Table 1 for an example from Exp. 5). After reading or hearing each sentence, participants answered open-ended questions targeting the information conveyed by the adjective/modifier. We recorded response accuracy and reading times and analysed them using mixed-effects models.

Results:

Recall accuracy: In four experiments (except for Exp. 3) postnominal modifiers were recalled significantly better than prenominal ones. Genitive modifiers were recalled slightly better than adjectives in the written Exp. 3 and 4, compared to no effect of modifier type in the spoken Exp. 5. There was no significant effect of the interaction between modifier position and type in Exp. 3, 4 and 5.

Reading times: In word-by-word Exp. 2 and 4, adjective position had no significant effects on reading times on the manipulated NP. However, processing postmodified NPs resulted in a significant spillover effect. In contrast, in the whole-sentence-at-once Exp. 1, no significant differences in reading times were observed.

Discussion: The recall accuracy results support the prediction based on the Feature-Label-Order principle. (The absence of the effect of modifier position on recall in Exp. 3 can be attributed to the specific presentation mode.) Since the position of genitives is canonically postnominal, the absence of a position-by-type interaction indicates that the effect is not caused by possible syntactic unusualness of postnominal adjectives. Analysis did not reveal a consistent relation between recall accuracy and reaction time patterns, suggesting that recall advantage may be instead rooted in processing mechanisms not captured by reading times. The findings contribute to our understanding of how word order may impact sentence comprehension in languages with relatively free word order, such as Czech.

<table>
<tr><th>Modifier position X Modifier type</th><th>Stimulus</th><th>Comprehension question</th></tr>
<tr><td rowspan="2">Prenominal – Adjective</td><td>Zjistila jsem, že do toho horkýho moštu taky potom nepřidali zázvor s rakytníkem a hřebíčkem.</td><td rowspan="4">Jaký mošt to byl?

What cider was it?</td></tr>
<tr><td>I found out they also didn't add ginger with sea buckthorn and cloves to the hot cider afterwards.</td></tr>
<tr><td rowspan="2">Postnominal – Adjective</td><td>Zjistila jsem, že do toho moštu horkýho taky potom nepřidali zázvor s rakytníkem a hřebíčkem.</td></tr>
<tr><td>I found out they also didn't add ginger with sea buckthorn and cloves to the cider hot* afterwards.</td></tr>
<tr><td rowspan="2">Prenominal – Genitive</td><td>Zjistila jsem, že do toho Renaty moštu taky potom nepřidali zázvor s rakytníkem a hřebíčkem.</td><td rowspan="4">Čí mošt to byl?

Whose cider was it?</td></tr>
<tr><td>I found out they also didn't add ginger with sea buckthorn and cloves to Renata's cider* afterwards.</td></tr>
<tr><td rowspan="2">Postnominal – Genitive</td><td>Zjistila jsem, že do toho moštu Renaty taky potom nepřidali zázvor s rakytníkem a hřebíčkem.</td></tr>
<tr><td>I found out they also didn't add ginger with sea buckthorn and cloves to cider Renata's* afterwards.</td></tr>
</table>

Table 1 Example item (2x2) from Exp.5, 6 and 7 with comprehension questions. Exp.1 and 2 had only the first two conditions. English non-literal translations preserve the Czech word order within the NP.

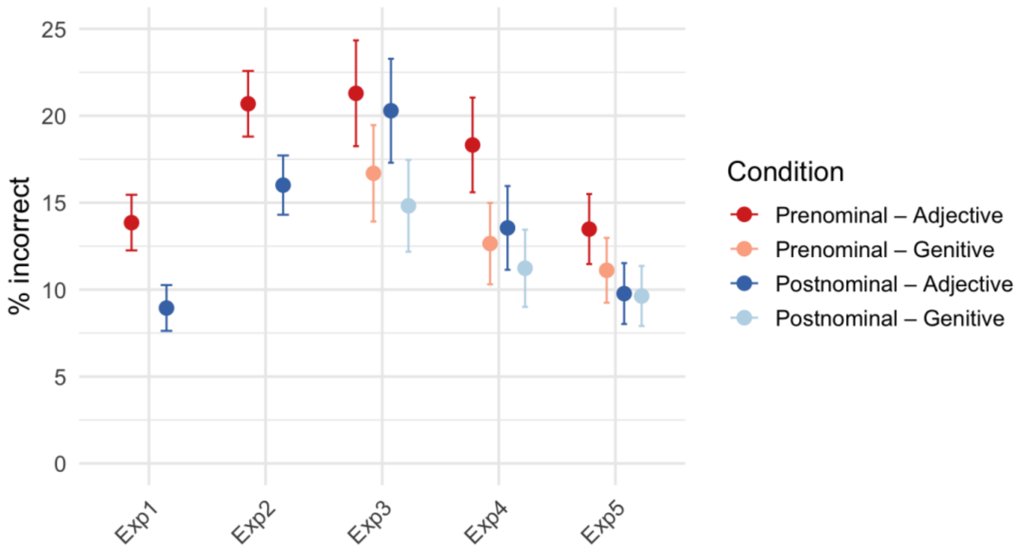


Figure 1 Recall accuracy (% incorrect) based on the adjective position and type. Exp.1 (whole-sentence-at-once), 2 and 4 (word-by-word), 3 (phrase-by-phrase), 5 (spoken). Error bars denote 95% confidence intervals.

References

[1] Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. Language and Cognitive Processes, 26(3), 376–405. https://doi.org/10.1080/01690965.2010.492642

[2] Karimi, H., Diaz, M., & Ferreira, F. (2019). "A cruel king" is not the same as "a king who is cruel": Modifier position affects how words are encoded and retrieved from memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 45(11), 2010–2035. https://doi.org/10.1037/xlm0000694

[3] Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The Effects of Feature-Label-Order and Their Implications for Symbolic Learning. *Cognitive Science, 34*(6), 909–957. https://doi.org/10.1111/j.1551-6709.2009.01092.x

[4] Štícha, F. (2013). *Akademická gramatika spisovné češtiny*. Academia.

[5] Lukeš, D., et al. (2024). *ORTOFON v3: Korpus neformální mluvené češtiny s víceúrovňovým přepisem.* Ústav Českého národního korpusu FF UK.

Additional information about the position of adjectives in Czech

Czech is generally considered a language with flexible word order, but grammars describe the adjective position as relatively fixed, with bare adjectives canonically occurring prenominally and postnominal placement being restricted to specific contexts such as emphasis, contrast, or terminological usage. [4] However, a small-scale probe from ORTOFON version3 [5] (a corpus of spontaneous informal spoken corpus) suggested that postnominal adjectives occur quite systematically (in approximately 10% of cases) and often without an evident function. These findings challenge the assumption that postnominal position is always marked and suggest that positional variability of adjectives in spoken Czech may be more neutral than previously thought.

Irony Processing in Reading: Eye-Tracking Evidence on the Predictive Role of Mentalizing and Vocabulary

Background: Irony is a complex form of figurative language in which intended meanings often contrast with the literal interpretation of words. While most theoretical models have been developed based on spoken language, understanding irony in written texts also presents significant cognitive challenges. Different frameworks have been proposed to explain irony comprehension. The graded salience hypothesis [1] posits that the most salient meaning of an utterance is accessed first, requiring reanalysis when the literal interpretation proves incongruent with the context. In contrast, building on models of parallel constraint satisfaction, Pexman [2] proposed that irony comprehension involves the simultaneous integration of multiple linguistic, contextual, and social cues. Eye-tracking research has shown that ironic statements often elicit increased rereading and longer reading times, reflecting greater processing demands [3], but see [4]. Because irony comprehension requires readers to go beyond literal meanings and infer the speaker's communicative intent, individual differences in mentalizing—the ability to attribute mental states and recognize pragmatic intentions—have been consistently linked to successful irony interpretation [5]. Recent evidence shows that higher mentalizing abilities predict better performance in irony comprehension tasks, even when controlling general language abilities and executive functioning [6]. Other linguistics skills, such as vocabulary, has been less explored to irony comprehension. Yet, there is evidence that vocabulary supports inference making, potentially facilitating the recognition of lexical incongruities and supporting the flexible mapping between linguistic expression and general context [7]. In the present study we examine whether mentalization and vocabulary can modulate reading times during irony understanding. **Method:** A total of 118 monolingual Spanish-speaking adolescents aged between 16 and 18 participated in an eye-tracking experiment in which they read 45 short texts presented on a computer screen. Texts were distributed across two conditions, namely 15 ironic and 15 literal meaning (plus 15 fillers sentences). Each target sentence appeared in two versions (ironic or literal), but participants only read one version, assigned through a counterbalanced Latin Square. The structure of the texts followed a standardized format adapted from Olkoniemi et al. [3], including: Introduction, Critical Context (literal or ironic), Target Utterance, Spillover Region, and Ending. After each text, participants answered a comprehension question assessing whether the intended meaning had been interpreted correctly. Eye movements were recorded using an EyeLink 1000 system. Dependent variables included first pass reading times, regression path duration, and total reading time on the target utterance, as well as comprehension accuracy. Predictor variables included scores from the Multidimensional Mentalizing Questionnaire (MMQ, [8]) and the TEVI (a vocabulary test using picture-based word recognition. [9]). **Results:** We found an effect of vocabulary on irony comprehension, but not of mentalizing (see Table 1): higher TEVI scores were linked to greater accuracy in identifying ironic sentences (see Figure 1) and longer reading times in ironic contexts in regression path duration. Literal items showed uniformly high accuracy and fast reading, accelerated by vocabulary. **Discussion:** Our findings suggest that vocabulary facilitates efficient processing of literal language, while in ironic contexts it may enhance sensitivity to incongruity. These findings align with the graded salience hypothesis in showing that vocabulary skills modulate the reanalysis of ironic statements.

Table 1. Results of the generalized linear mixed effect regression on participants' accuracy.

	Estimate	se	z	P	
(Intercept)	1,650	0,172	9,621	<0,001	***
Condition (ironic vs literal)	-2,135	0,180	-11,859	<0,001	***
TEVI	0,333	0,133	2,492	0,013	*
MMQ	0,028	0,137	0,205	0,838	
Condition x MMQ	-0,016	0,130	-0,121	0,904	
Condition x TEVI	0,134	0,129	1,039	0,299	

***=p<.001; *=p<.05

Table 2. Results of the linear mixed effect regression on participants' regression path duration.

	Estimate	se	t	P	
(Intercept)	825,090	37,610	43,834	<0,001	***
Condition (ironic vs literal)	26,944	15,824	28,960	0,099	#
MMQ	6,766	18,268	107,467	0,712	
TEVI	-30,714	18,574	108,106	0,101	
Condition x MMQ	0,121	6,917	413,221	0,986	
Condition x TEVI	14,598	6,977	241,757	0,037	*

***=p<.001; *=p<.05; #=p<1

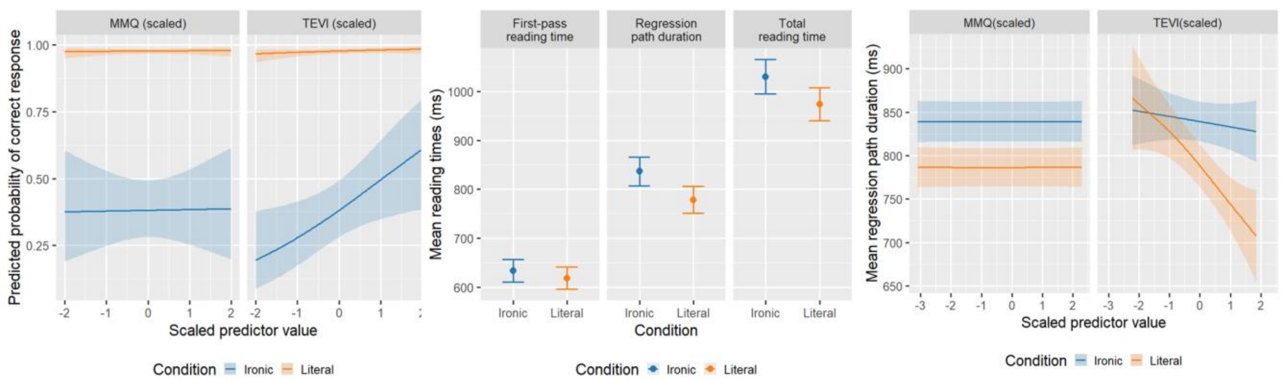


Figure 1. Summary of comprehension accuracy and reading-time measures for ironic and literal sentences as a function of scaled predictors and experimental conditions. The left panel shows the model-predicted probability of a correct response plotted against scaled MMQ and TEVI values for ironic (blue) and literal (orange) sentences, with 95 % confidence bands. The center panel displays mean first-pass reading time, regression-path duration and total reading time for each condition, with 95 % confidence intervals. The right panel depicts mean regression-path duration as a function of scaled MMQ and TEVI values for both conditions with mean standard error as ribbons.

References

[1] Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8(3), 183–206. [2] Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, 17(4), 286–290. [3] Olkoniemi, H., Johander, E., & Kaakinen, J. K. (2019). The role of look-backs in processing written verbal irony. *Quarterly Journal of Experimental Psychology*, 72(2), 422–436. [4] Ronderos, C. R., Tomlinson Jr, J. M., & Noveck, I. (2024). When irony is faster than its literal control: The role of mindreading during irony comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(3), 509. [5] Happé, F. G. E. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbol*, 10(4), 275–295. [6] Saban-Bezalel, R., Mashal, N., & Alfi-Yitzhak, N. (2019). Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, 67, 101426. [7] Guerra, E., & Kronmüller, E. (2024). Are Adolescents with a Wider Vocabulary Faster at Inference Making During Reading? Evidence from Self-Paced Reading. *Education Sciences, 14*(12), 1368. [8] Aldunate, N., López-Silva, P., Brotfeld, C., Guerra, E., & Kronmüller, E. (2024). Spanish version of Multidimensional Mentalizing Questionnaire (MMQ): Translation, adaptation and psychometric properties in a Chilean population. *Plos one*, *19*(3), e0296691. [9] Echeverría, M. S.; Herrera, M. O. & Segure, J. T. (2002). Test de Vocabulario en Imágenes TEVI-R. Concepción, Chile: Ediciones Universidad de Concepción.

Detecting foreign rhythm in native-language speech at birth

Martina Dvořáková, Natálie Kikoťová, Josef Urbanec, Antonia Goetz, Kateřina Chládková
Institute of Psychology, Czech Academy of Sciences; Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University; Department of Medical Biophysics, Faculty of Medicine in Hradec Králové, Charles University; Paediatrics Department, Havlíčkův Brod Hospital; MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
martina.dvorakova@ff.cuni.cz

Humans tune in to the native language prosody before they are even born. Prior findings with newborns reported language-specific processing of forward versus backward speech and intonational contours [3, 4, 5, 7] and also indicated language-specific processing of iambic versus trochaic patterns in non-linguistic tone stimuli [1]. The present experiment tested newborns' processing of temporal rhythm patterns in naturalistic native-language speech.

Czech-learning newborns (N=48, 1–6 days old) were played naturally recorded well-formed Czech utterances with native Czech rhythm (virtually lacking cues to word-level stress) and with non-native rhythm (prolonged foot-initial syllables).Their hemodynamic activity was recorded with fNIRS while they were asleep in a quiet room at a maternity ward. Data of 27 newborns were included in the final analysis.

Grand-average HbO and HbR curves per channel and per condition are plotted in Figure 1. The results showed larger hemodynamic responses to the non-native than to the native rhythm in a late window (Condition * Window: $\beta = 0.484$, $SE = 0.127$, $t[1932] = 3.810$, $p < 0.001$), attributable to a double-peak response shape in the non-native condition. This finding is discussed in terms of suprisal-induced resonating activity after hearing familiar native speech paired with an unfamiliar rhythm pattern [2]. Further, there was an overall attenuated response to the native rhythm localized in the right frontal region (Condition * Hemisphere * Region: $\beta = 0.368$, $SE = 0.127$, $t[1932] = 2.89$, $p = 0.004$), evidencing right-lateralized processing of speech rhythm.

Traditional language development theories claimed that only coarse between-class rhythm differences between languages are processed at birth [6]. Having demonstrated that newborns differentially process non-native vs. native rhythmic patterns within natural native-language speech, even in a language outside of the traditional rhythm classes, the present findings disprove some of the early theories and substantially deepen our understanding of early speech development.

References

[1] Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. Brain and Language, 162, 46–59. https://doi.org/10.1016/j.bandl.2016.08.002

[2] Boersma, P., Chládková, K., & Benders, T. (2022). Phonological features emerge substance-freely from the phonetics and the morphology. Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 67(4), 611–669. https://doi.org/10.1017/cnj.2022.39

[3] Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. Proceedings of the National Academy of Sciences, 100(20), 11702–11705. https://doi.org/10.1073/pnas.1934290100

[4] Martinez-Alvarez, A., Benavides-Varela, S., Lapillonne, A., & Gervain, J. (2023). Newborns discriminate utterance-level prosodic contours. Developmental Science, 26(2), e13304. https://doi.org/10.1111/desc.13304

[5] May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. Developmental Science, 21(3), e12564. https://doi.org/10.1111/desc.12564

[6] Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human perception and performance, 24(3), 756.

[7] Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Uchida-Ota, M., Konishi, Y., & Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. Human Brain Mapping, 33(9), 2092–2103. https://doi.org/10.1002/hbm.21350

Seeing the Little Things: Context Effects on the Processing of Inflectional Affixes on Novel Words in Late Second Language Learners

Vera Heyer
TU Braunschweig
v.heyer@tu-braunschweig.de

Inflectional affixes have been shown to cause particular difficulties for late second language (L2) learners, with lower-level as well as some advanced learners applying these inconsistently (e.g., Lardiere, 2006) and showing less automatic processing in priming studies (e.g., Jacob et al., 2018). When it comes to inconsistent application of inflectional markers such as past tense -*ed* or plural -*s*, one can distinguish between omission and commission errors. In omission errors, the appropriate inflectional marker is not produced even though it is grammatically required in the sentence context (e.g., *two cake*). In commission errors, a rarer type of error, affixes are supplied where they are not licensed by the sentence context (e.g., *one cakes*). The present study used omission and commission errors in the reading of novel words in singular and plural contexts to investigate if late L2 learners are sensitive to how morphological endings fit into the sentence context.

In an eye-tracking during reading study, 48 advanced L2 learners of English read novel words with plural -*s* (e.g., *fruttles*) or the agentive suffix -*er* (e.g., *fruttler*) in matched plural and singular contexts, differing only in the numeral used (e.g., *ten* vs. *one* in (1)). The affix -*er* was chosen because it denotes people or gadgets (e.g., *baker*, *printer*), thus resulting in a countable (compared to a mass) noun that should receive plural marking in a plural context. Therefore, the appearance of an -*er* word in a plural context (e.g., *ten fruttler* in (1a)) constitutes an omission error (under the assumption of regular inflection being required rather than less common unmarked forms) and a novel word ending in -*s* presented in a singular context is an example of a commission error (e.g., *one fruttles* in (1b)). LMER models revealed that participants took longer to read overtly plural-marked forms such as *fruttles* in incongruent singular contexts (t=2.867; dotted red line vs. solid red line in Fig. 1) and regressed to earlier parts of the sentence from the novel word in this condition more often (t=2.062). For singular forms with -*er*, on the other hand, context did not have an effect on reading times, with missing plural markers (omission) not slowing down reading times in the plural compared to the singular context (t=-0.311; dotted blue line vs. solid blue line in Fig. 1).

The results show that late L2 learners' processing is slowed down when plural marking is present in non-licensed singular contexts (commission error), while missing plural markers in plural contexts (with numerals) do not disrupt reading (omission error). This suggests that late L2 learners compute the plural affix's function when actively encountering it but do not penalise the absence of this marking in obligatory contexts.

Materials

(1) The children saw many rare animals at the zoo last month. In the Australian section, they counted eight koalas and…

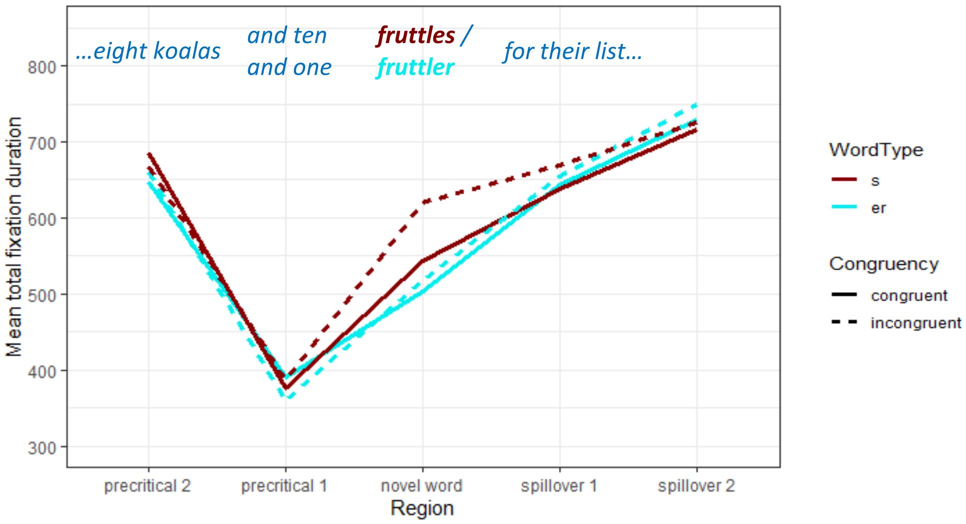
a. …**ten** ***fruttles***/ ***fruttler*** ...

b. …**one** ***fruttles***/ ***fruttler*** …

…for their list of Australian animals.

Figure 1

Total reading times in the different areas of interest. For omission error, compare blue lines; for commission error, compare red lines.



References

Jacob, G., Heyer, V., & Veríssimo, J. (2018). Aiming at the same target: A masked priming study directly comparing derivation and inflection in the second language. *International Journal of Bilingualism, 22*(6), 619–637.

Lardiere, D. (2006). *Ultimate attainment in second language acquisition: A case study*. Routledge.

Animacy cues and word order in language acquisition and dialogue corpora

David Pagmar[1], Yuval Marton[2], Asad Sayeed[1]
[1] Dept of Philosophy, Linguistics, and Theory of Science, University of Gothenburg; [2] University of Washington
david.pagmar@gu.se

Background: This study investigates the role of animacy in child-directed speech (CDS) and its potential influence on children's syntactic development. Specifically, we examine the distribution of animacy in transitive constructions and how adaptations in CDS may act as scaffolding for syntactic development.

The sentences (1) *the child reads the book* and (2) *the book scares the child* follow the same syntactical structure in terms of word order, i.e. subject-verb-object. In sentence 1, the child is the subject and the book is the object, while in sentence 2, it is the other way around. There is a difference between the nouns the child and the book in terms of animacy, that is how sentient or alive the referent of the noun is. In general, animate referents are more likely to be a subject in a sentence [1]. Two primary research questions are addressed: (1) Does the use of animacy in transitive constructions in CDS differ from adult language use, and if so how? (2) How might these differences provide a foundation for children's acquisition of argument structure and thematic roles?

Method: The study draws on corpora of naturalistic language use to compare transitive constructions in CDS (CHILDES, four parent-child dyads [2]) and the naturalistic adult dialogue-part of the British National Corpus [3]. 200 randomly selected transitive constructions from each corpus, defined as clauses with a subject, verb, and direct object, were extracted and coded for the animacy of both subjects and objects through a qualitative examination of conversational contexts. We identified patterns in how caregivers and adult conversational partners position animate and inanimate entities in discourse. The analysis incorporates quantitative measures of animacy distribution [4].

Results: Our findings reveal that caregivers disproportionately use animate entities as subjects and inanimate entities as objects in CDS, creating highly predictable animacy patterns (Figure 1). These patterns differ significantly from those observed in ADS, where the animacy of subjects and objects is more varied. By aligning animacy configurations with children's cognitive and perceptual biases, CDS appears to provide implicit cues that facilitate mapping syntactic roles to semantic and thematic properties. Such cues may play a critical role in helping children acquire argument structure and understand the relationships between subjects, objects, and verbs.

Discussion: By isolating animacy as a specific cue, this research highlights the dynamic interplay between input pragmatics and linguistic structure, offering insights into how caregivers adapt their speech to meet the developmental needs of young language learners. The study also examines potential implications for broader theories of language acquisition. Current frameworks emphasize either innate grammatical knowledge or the statistical properties of input. This research contributes to an emerging perspective that underscores the importance of pragmatic adjustments as the bridge between these domains. The animacy patterns observed in CDS suggest that caregivers strategically shape their speech to align with children's developmental capacities, providing a reliable source of information for language learning.

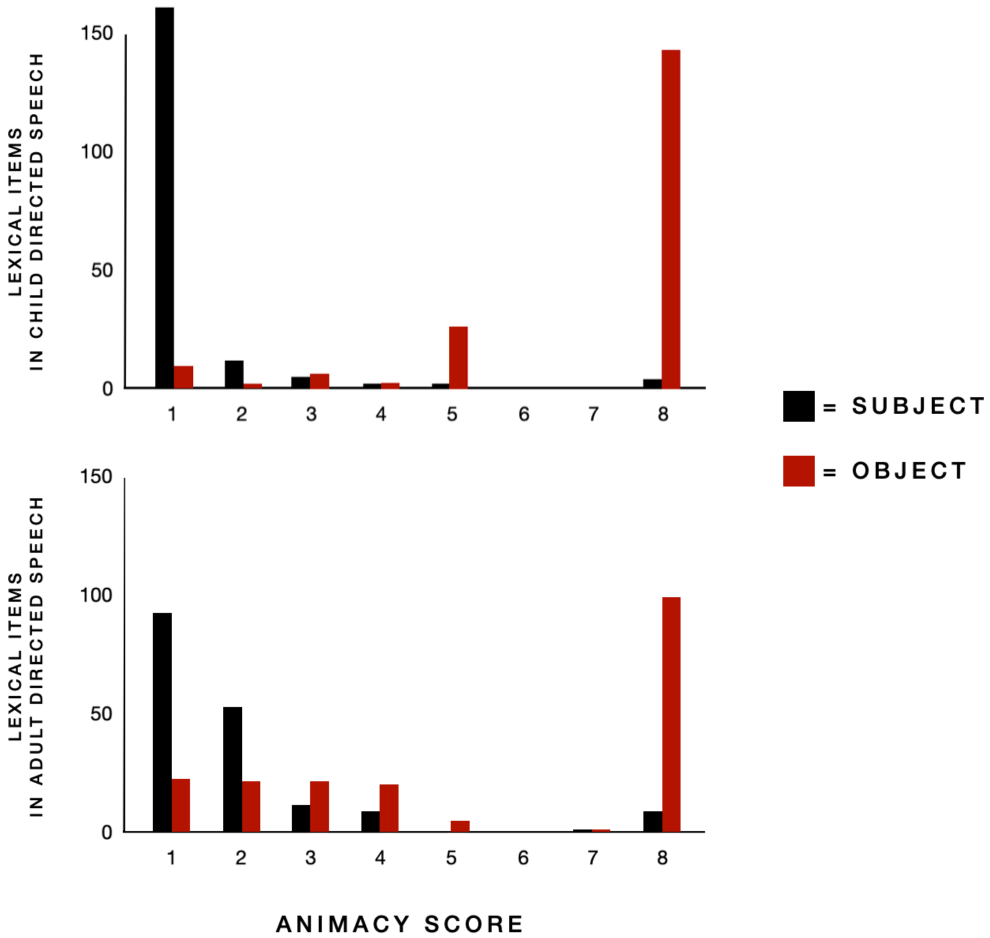


Figure 1: Lexical usage in child-directed speech (CHILDES) and adult dialouge (BNC) vs. animacy ratings from [4] given subject or object role.

References

[1] De Swart, P., & De Hoop, H. (2018). Shifting animacy. *Theoretical Linguistics*, *44*(1-2), 1–23.

[2] MacWhinney, B., & Erlbaum, L. (2000). *The childes project . volume i: Tools for analyzing talk: Transcription format and programs*. Mahwah, NJ: Lawrence Erlbaum.

[3] BNC Consortium. (2007). British national corpus.

[4] Ji, J., & Liang, M. (2018). An animacy hierarchy within inanimate nouns: English corpus evidence from a prototypical perspective. *Lingua*, *205*, 71–89. https://doi.org/https://doi.org/10.1016/j.lingua.2017.12.017

Reanalysis as Last Resort: Coercion in Tense Harmony Violations

Buhan Guo[1], Andrea Santi[2], Shayne Sloggett[1], Giuseppina Turco[3], Sven Mattys[1], Nino Grillo[1]

[1] University of York; [2] University College London; [3] Université Paris Cité

buhan.guo@york.ac.uk

Background: We tested for a preference of coercion over structural reanalysis when faced with grammatical violations in English it-clefts. Clefts obey *Tense Harmony* [1]: the tense of Connected Clauses (CCs) must semantically align with its matrix copula (1-a). However, since Tense is referential in Relative Clauses (RCs), (1-a) can be rescued if the embedded clause is reanalyzed as an RC (1-b). Recent work [2] shows that Harmony violations reduce acceptability and slow reading times, suggesting incremental processing of Tense Harmony and a preference for CCs over RCs. Yet these results cannot confirm that reanalysis ensued. Our Experiment 1 shows that Tense Harmony violations do *not* lead to reanalysis, revealing the parser's strong bias towards maintaining its initial interpretations [3, 4] and providing a case where coercion [5, 6] might be employed also when an alternative grammatical parse is available. Experiment 2 demonstrates a strong preference for coercion over reanalysis under Tense Harmony violations, despite contextual bias for RC parses.

Exp.1–Method: A G-Maze study (n=74) tested whether Tense Harmony violations trigger reanalysis. A second embedded clause–also interpretable as CC or RC–was added to the stimuli from [2] and disambiguated via animacy: animate-subject predicates (*was hired*) forced CC parses (subject: clefted DP *the scientist*), while inanimate-subject predicates (*was collected*) forced RC parses (subject: closest DP *the data*). The 2 TENSE (Match/Mismatch) * 2 STRUCTURE OF CP2 (CC/RC) design was shown in Table 1. If readers reanalyze Harmony violations as RCs, they should readily integrate a second clause as the missing CC. If instead they maintain their initial parse (e.g., via coercion; as in (2)), a second CC should be unexpected and thus increase RTs.

Results: For CP1, RTs at the disambiguating auxiliary (*was/will*) were significantly longer when Tense Harmony was violated (β=0.18, p<.001, Fig 1), consistent with [2]. At CP2, a significant TENSE*STRUCTURE interaction emerged at the disambiguating verb (*hired/collected*): crucially, CC disambiguation of CP2 generated longer RTs under Mismatched Tense (β=0.15, p<.001, Fig 2), contrary to predictions based on syntactic reanalysis.

Exp.2-Method: A G-Maze study with acceptability judgment (n=54) tested whether contextual support influences CC/RC interpretation under Tense Harmony violations. We manipulated CONTEXT (CC/RC-supporting) and TENSE (Match/Mismatch) as in (3). RC contexts introduced two competing referents (*two scientists*), prompting an RC parse as a modifier; CC contexts contrasted two roles (*scientist* vs. *technician*) that were paired with temporally flexible tasks (e.g. *was asked to*), supporting a CC parse and compatible with coercion.

Results: At clefted nouns (e.g. *scientist*), RTs were significantly longer under RC than CC contexts (β=0.27, p<.001, Fig 3). This validates the experiment and shows that participants paid attention. At auxiliaries (*was/will*), RTs were significantly longer for Mismatched than Matched tenses (β=0.31, p<.001) and for RC than CC contexts (β=0.06, p<.001), with a significant interaction between TENSE and CONTEXT (β=-0.07, p=.021). Across TENSE, sentences under RC contexts had lower acceptability than under CC ones (β=-0.24, p<.001, Fig 4).

Discussion: Exp.1 reveals that Tense Harmony violations do not trigger successful RC reanalysis without contextual support, likely due to a need for joint structural-prosodic revisions (cf. [7, 8]). Exp.2 shows that even supporting context fails to override the CC preference: persistent online effects of Tense reflect costs of both coercion and reanalysis. Paradoxically, offline judgment suggests that RC contexts yielded lower acceptability, reinforcing the parser's bias for CCs and the relatively lower costs of coercion. An ongoing eye-tracking study replicates experiment 2 to provide time-sensitive evidence for coercion vs. reanalysis.

(1) a. It was [the boy] [$_{CC}$that left/*will leave]. ✓Past / Past; *Past / Future

b. Q: Who called? A: It was$_{past}$ [the boy [$_{RC}$ that will leave]] ([$_{CC}$that called$_{past}$]).

(2) Q: Was it Max that was going to fix it? A: No, it was Bill that **will** fix it ⇝ **was supposed to** fix it.

T-Harmony	CP2	Example item
Match	CC	It was the scientist that$_{CP1}$ **was** studying the data that$_{CP2}$ was **hired** last month.
Mismatch	CC	It was the scientist that$_{CP1}$ **will** study the data that$_{CP2}$ was **hired** last month.
Match	RC	It was the scientist that$_{CP1}$ **was** studying the data that$_{CP2}$ was **collected** last month.
Mismatch	RC	It was the scientist that$_{CP1}$ **will** study the data that$_{CP2}$ was **collected** last month.

Table 1: Experimental design and examples of stimuli in Experiments 1 (ROIs marked in bold)

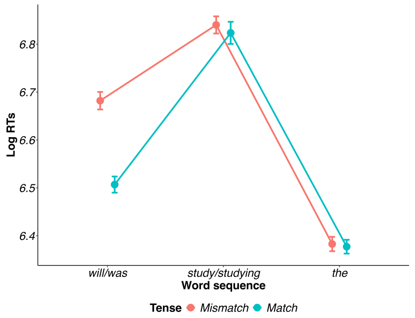


Figure 1: Log RTs near ROI-1 (will/was)

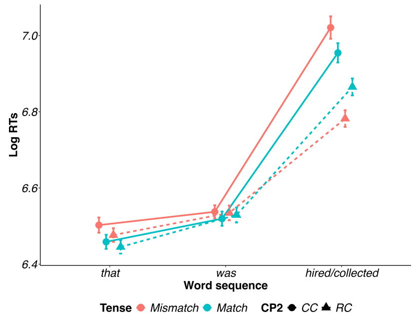


Figure 2: Log RTs near ROI-2 (hired/collected)

(3) **Experiment 2: RC/CC Contexts and Questions, and example of target sentences**

Richard's lab launched a new research project and assigned roles to two scientists$_{RC}$ / a scientist and a technician$_{CC}$. One$_{RC}$ / The scientist$_{CC}$ was tasked with analysing the collected measurements, the other$_{RC}$ / the technician$_{CC}$ was asked to prepare samples. (Richard spoke with one of them about their role)$_{RC}$.

Qs: Which one did Richard talk to?$_{RC}$ / Who was tasked with analysing the collected measurements?$_{CC}$

→ It was the scientist that was studying$_{Match}$ / will study$_{Mismatch}$ the data, according to the lab's report.

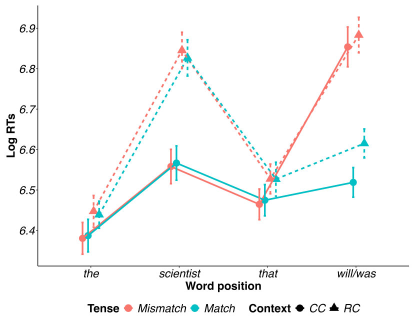


Figure 3: Log RTs near ROI (will/was)

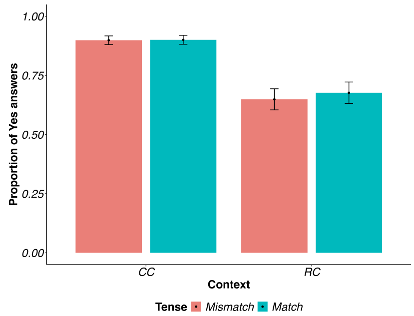


Figure 4: Acceptability judgment for Exp 2

References

[1] Higgins. (1973). *The pseudo-cleft construction in English* [PhD Thesis]. MIT.

[2] Guo, Grillo, Mattys, Santi, Sloggett, & Turco. (2024a). Clefted garden-paths. *AMLaP 30*.

[3] Fodor & Inoue. (1998). Attach anyway. In *Reanalysis in sentence processing*. Springer.

[4] Sturt, Pickering, Scheepers, & Crocker. (2001). Preservation of structure in lang. comprehension. *JML*, *45*.

[5] Gibson, Bergen, & Piantadosi. (2013). Rational integration of noisy evidence. *PNAS*, *110*.

[6] Villata & Tabor. (2022). A self-organized sentence processing theory of gradience. *Cognition*, *222*.

[7] Bader. (1998). Prosodic influences on reading. In *Reanalysis in sentence processing*. Springer.

[8] Guo, Grillo, Mattys, Santi, Sloggett, & Turco. (2024b). Prosody of clefted relatives. *Speech Prosody*.

The interplay between bilingualism and sleep quality in modulating executive performance

Liliia Terekhina[1, 2], Andriy Myachykov[1,3], Jason G. Ellis[4], Jubin Abutalebi[1,2,5], Yury Shtyrov[1,6], Federico Gallo[1,5]
[1] Higher School of Economics; [2] Vita-Salute San Raffaele University; [3] University of Macau; [4] Northumbria University; [5] UiT-The Arctic University of Norway; [6] Aarhus University
l.terekhina@studenti.unisr.it

Background: Various lifestyle factors and everyday activities such as nutrition, physical exercise, memory and technology training have been shown to affect cognitive processing. Two of such everyday activities – sleep and bilingualism – are known to influence similar cognitive and neural mechanisms [1]. Nonetheless, the joint effects of sleep and bilingualism on cognition have never been addressed together. This study is the first attempt to fill this gap. Since research in both fields has produced inconsistent results, which may be due to the failure to control other factors that influence EF, resulting in specific effects potentially being obfuscated or insufficiently manifested. Thus, it can be hypothesized that the points of intersection between sleep and bilingualism may be significant at a theoretical and functional level. Specifically, it can be assumed that these two life-time experiences may modulate each other's impact on cognitive functioning. Based on the evidence from the two research strands reviewed above, our primary hypothesis, therefore, is that bilingualism may overcome the detrimental effects of poor sleep quality upon EF. Conversely, poor sleep quality could potentially exert a mitigating effect on beneficial consequences of bilingualism.
Method: We investigated the interactive effects of bilingualism and sleep on executive functioning at the behavioral level. We conducted two experiments using two independent samples of bilingual young adults (L1 Russian, L2 English), the Flanker task to assess executive performance, the Pittsburg Sleep Quality Index to measure retrospective sleep quality over a 1-month period, the Insomnia Severity Index to assess insomnia-related symptoms, Cambridfe Test and LEAP-Q to evalute bilingual experience.
Results: In experiment 1, we registered bilingualism effects on executive performance in poor, but not in good sleepers. In experiment 2, the magnitude of bilingual effects increased with increasing severity of insomnia symptoms.
Discussion: We conclude that when poor sleep quality and insomnia negatively affect cognitive resources, bilingualism-related cognitive effects emerge more prominently. This suggests higher degrees of bilingualism may compensate detrimental effects of poor sleep quality and insomnia on executive functioning. We suggest that cognitive research in bilingualism and sleep could benefit from controlling for interindividual variability in sleep quality and vice versa. Different aspects of bilingual experience (e.g., proficiency vs. exposure) showed distinct effects depending on sleep quality, likely due to sociolinguistic context. These findings highlight the need to account for both sleep and language background in cognitive research and emphasize the value of using objective sleep measures in future studies.

References

[1] Gallo, F., Myachykov, A., Abutalebi, J., DeLuca, V., Ellis, J., Rothman, J., & Wheeldon, L. R. (2025). Bilingualism, sleep, and cognition: An integrative view and open research questions. *Brain and Language*, *260*, 105507. https://doi.org/10.1016/j.bandl.2024.105507

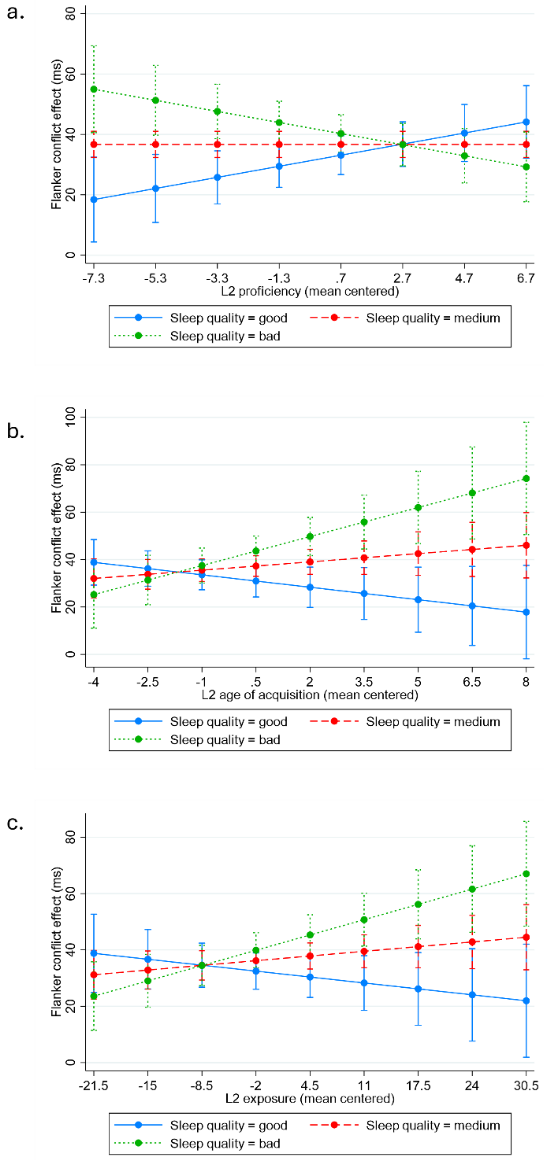


Figure 1. Interaction plot for the two-way interaction between bilingual experience factors*sleep quality predicting Flanker conflict effect (in ms). The conflict effect was calculated as the difference between the average individual RTs in the incongruent and congruent conditions. *a.* = L2 proficiency, *b.* = L2 age of acquisition, *c.* = L2 exposure. For graphical representation purposes, here we selected three representative values of PSQI score, i.e., 1 SD below the mean, mean, and 1 SD above the mean, to represent three levels of sleep quality in our sample, i.e., good, medium and bad, respectively. Please note that PSQI score was inserted as a continuous variable in the statistical model.

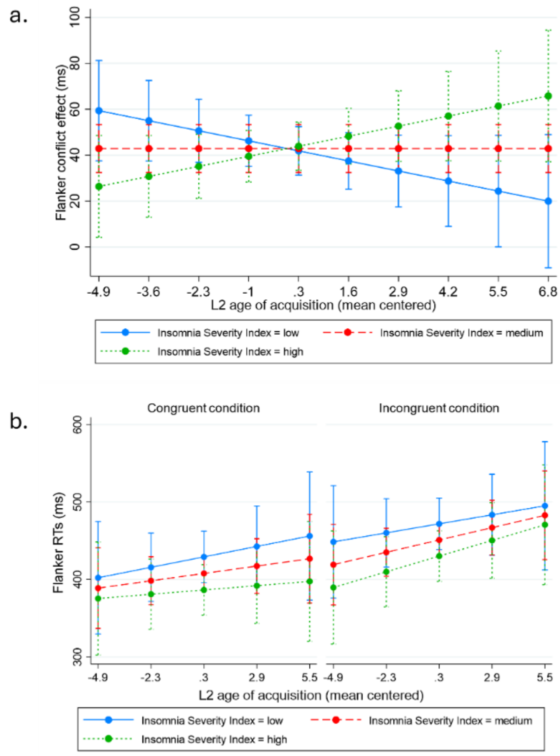


Figure 2. a. Interaction plot for the two-way interaction between L2 age of acquisition*insomnia-related symptoms predicting Flanker conflict effect (in ms). The conflict effect was calculated as the difference between the average individual RTs in the incongruent and congruent conditions. For graphical representation purposes, the three levels of ISI selected for plotting (low, medium, high) are represented as 1 SD below the mean, mean, and 1 SD above the mean of our sample, respectively. Please note that ISI score was inserted as a continuous variable in the statistical model. b. Interaction plot for the three-way interaction between L2 age of acquisition*insomnia-related symptoms*task condition predicting Flanker RTs (in ms). For graphical representation purposes, the three levels of ISI selected for plotting (low, medium, high) are represented as 1 SD below the mean, mean, and 1 SD above the mean of our sample, respectively. Please note that ISI score was inserted as a continuous variable in the statistical model.

Letter computation in Russian: further exploration

Yana Nikonova[1], Svetlana Alexeeva[1]
[1] Center for language and brain, HSE University
nikonova0312@gmail.com

Background: Two main strategies are distinguished at the early stages of letter recognition during reading: sequential and parallel. Sequential scanning involves recognizing letters within a word one by one. Parallel scanning implies the simultaneous processing of multiple letters. It has been hypothesized that the scanning strategy may depend on the orthography of a language [3, 4]. For languages with transparent orthography, sequential scanning is expected to dominate. In contrast, sequential and parallel scanning are used for languages with deep orthography. This hypothesis has been confirmed for English and Greek [3, 4]. However, for Russian, whose orthography is considered relatively transparent, sequential and parallel processing was observed [1]. The limitations of the methodology may have influenced the result. In the study involving Russian, a visual search paradigm was used: participants were first shown a letter, followed by a sequence of letters. Their task was to indicate whether the presented letter appeared in the given sequence. The sequence remained on the screen until the participant responded. Consequently, high-level cognitive processes might have affected the processing pattern.

Method: A total of 102 adult native Russian speakers (Mean age = 23.83, SD = 7.17) participated in this study. In the present study, we aim to verify findings of previous research in Russian using two different methodologies. Pseudorandomized five-letter sequences were used as stimuli. Two blocks of the experiment were created: one replicating the previous study [1] and the other employing more sensitive methodology [2, 5] in which the sequence was shown first and followed by the choice from two letters. Thus, 200 stimuli were developed: 100 were presented in the first block of the experiment, and another 100 in the second block. Our dependent variable was the position of the letter (from 1 to 5). We measured reaction time and eye movement data, such as landing position.

Results: Preprocessing and data analysis were conducted in the RStudio environment separately for each block. First, we included letter position as a polynomial predictor and factor of letter in the model with reaction time as the dependent variable, and intercepts for participants and stimuli. We also used letter position as a factor predictor, allowing us to compare reaction times between positions. For the letter/sequence block we found significant quadratic and quartic trends indicating parallel processing. Importantly, reaction time at the fifth position was significantly shorter than at the fourth ($p < 0.001$). The last letter was processed significantly faster than the penultimate letter, also suggesting a parallel processing component. For sequence/letter block we found significant linear and quadratic trends as well as a significant difference in reaction time between the 4 and 5 positions (5 was processed faster, $p < 0.001$).

Discussion: We observed indications of different scanning strategies across the experimental blocks. In the letter/sequence block, scanning appeared to be predominantly parallel. In the sequence/letter block, both parallel and serial processing features were present. Thus, parallel processing was evident in both blocks. The elements of serial processing observed in the sequence/letter block may be explained by a leftward shift in the attentional field. The parallel scanning strategy is likely universal, which supports the view proposed by Tydgat and Grainger [5] and contradicts the assumption that the type of orthography influences the scanning strategy.

Figure 1

Partial effects of reaction time on letter position in letter/sequence block

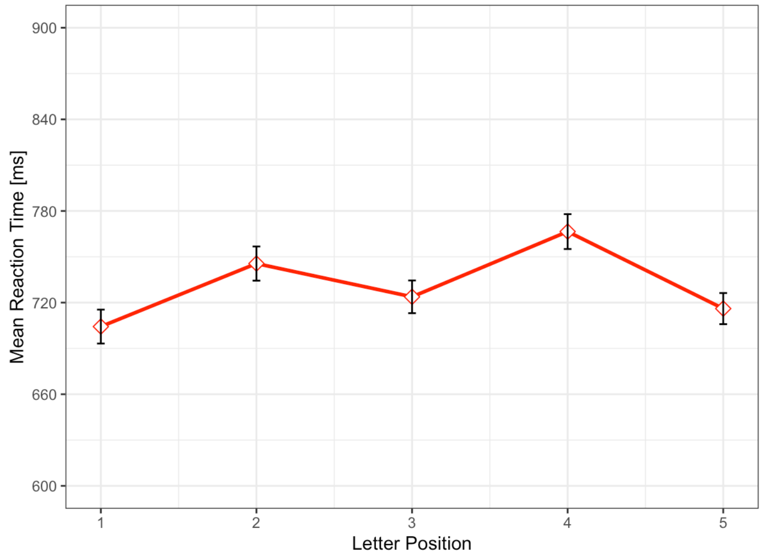
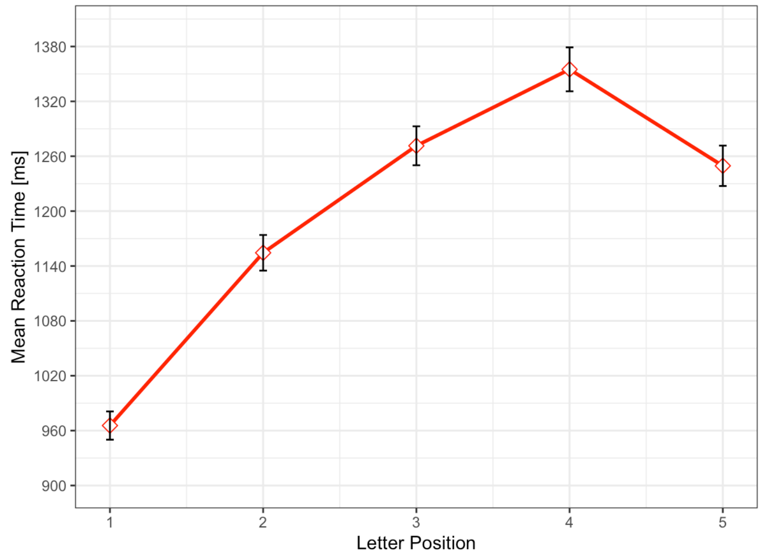


Figure 2

Partial effects of reaction time on letter position in sequence/letter block



References

[1] Alexeeva, S., & Dobrego, A. (2021). Letter processing in Russian: Does orthography matter?. *Acta Psychologica*, *218*, 103355.

[2] Grainger, J., Bertrand, D., Lété, B., Beyersmann, E., & Ziegler, J. C. (2016). A developmental investigation of the first-letter advantage. *Journal of Experimental Child Psychology*, *152*, 161-172.

[3] Ktori, M., & Pitchford, N. J. (2008). Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers. *Language and Cognitive Processes*, *23*(2), 258-281.

[4] Ktori, M., & Pitchford, N. J. (2009). Development of letter position processing: effects of age and orthographic transparency. *Journal of Research in Reading*, *32*(2), 180-198.

[5] Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. *Journal of Experimental Psychology: Human perception and performance*, *35*(2), 480.

Modelling Temporal Connective Processing with LLMs: Insights from English & Norwegian

Hannah S. Rognan[1], Shohini Bhattasali[1,2]
[1]Linguistics, University of Toronto; [2]Language Studies, University of Toronto Scarborough
hannah.rognan@mail.utoronto.ca

Background: Temporal connectives like *before* and *after* play an important role in sentence comprehension by linking events in sentences to real-world event order. Their position can produce either chronological event order (1 & 3), where the order of mention matches the real-world event sequence, or reverse event order (2 & 4), where the order of mention does not match the real-world event sequence. Prior psycholinguistic studies have observed that reverse sentences reflect higher cognitive load [1, 2].

(1) **After** blew out the candles, the boy cut the cake.
(2) The boy cut the cake **after** he blew out the candles.
(3) The girl poured the juice **before** she ate the pancakes.
(4) **Before** she ate the pancakes, the girl poured the juice.

Goal: Large Language Models (LLMs) are being widely used in cognitive modelling but we do not fully understand the event knowledge representations they learn. The current study investigates whether LLMs can recognize event order and reflect human-like processing of temporal connectives cross-linguistically, using data from English and Norwegian. Expanding on [3], where they found a human-like chronological preference in English only when the connective appeared sentence-initially, we examine preferences with sentences that are both ambiguous and unambiguous and examine an understudied language in computational psycholinguistics.

Method: To investigate model prediction patterns, we ran masked modeling experiments with the English LLM, BERT [4] and the Norwegian LLM, NorBERT [5]. We use bidirectional models to reflect offline sentence processing, analogous to a cloze task which has been used to study connectives in human comprehension [6]. For each sentence, we extracted the prediction scores for each of the two connectives *before* and *after*, *før* and *etter* in Norwegian, across three experiments. The first replicates the study by [3] and examines the prediction scores for the connectives that produce chronological and reverse order with ambiguous sentences. The second does the same, however, using unambiguous sentences where there is a logical temporal relationship between the two events in the sentence (see Table 1). The third experiment uses unambiguous sentences, and quantifies how accurate the model predictions are for each of the two positions of the connective.

Results: Our results with ambiguous sentences corroborated prior work [3]. In contrast to ambiguous sentences, the model predicts the connective that reflects chronological order over the connective that reflects reverse order in both languages and both positions (Fig 1 & 3). Furthermore, model accuracy is higher when the target connective represents the chronological order (Fig 2 & 4).

Discussion: When there is a clear temporal relationship between events in a sentence, the LLM is able to predict the correct connective in sentence-initial and sentence-medial positions, in contrast to prior work with ambiguous sentences. This suggests that LLMs are able to learn event knowledge representations and are sensitive to event order. In addition, the LLM's preference for the temporal connective which reflects chronological order mirrors human-like processing patterns. The similarity of results across English and Norwegian indicates that such preferences could generalize across languages and contributes to the cross-linguistic understanding of temporal connective processing.

Position	Ambiguity	Target	Example item
Initial	Ambiguous	*after* > *before*	***[MASK]*** the diver reached the wreck, the fishing boat cast its nets.
Medial	Ambiguous	*before* > *after*	The fishing boat cast its nets ***[MASK]*** the diver reached the wreck.
Initial	Unambiguous	*before*	***[MASK]*** he went to bed, Jack watched TV.
Medial	Unambiguous	*after*	They got on the bus ***[MASK]*** it arrived.

Table 1: Stimuli for English masked modelling experiments: Ambiguous sentences from [1] and unambiguous sentences from [7]

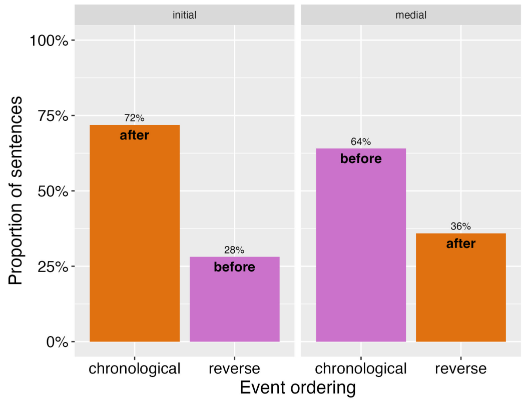


Figure 1: Proportion of English unambiguous sentences in which predictions prefer each connective.

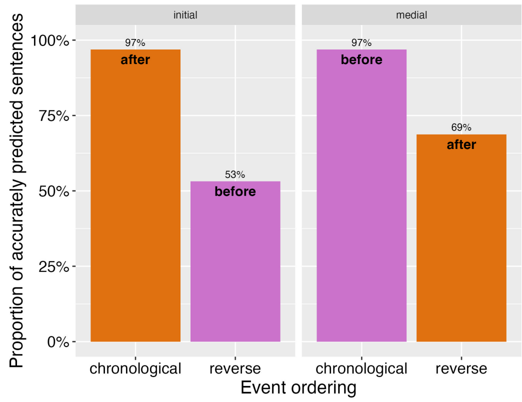


Figure 2: Proportion of English unambiguous sentences that are correctly predicted.

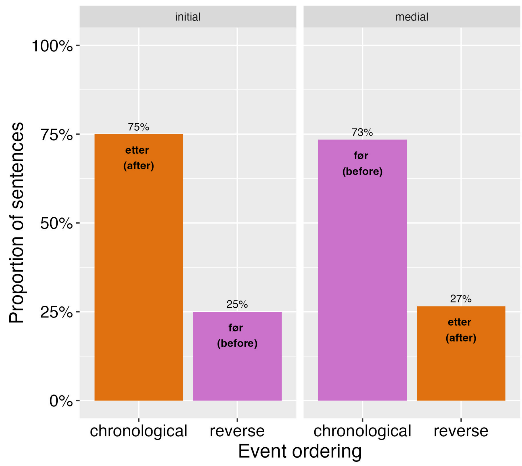


Figure 3: Proportion of Norwegian unambiguous sentences in which predictions prefer each connective.

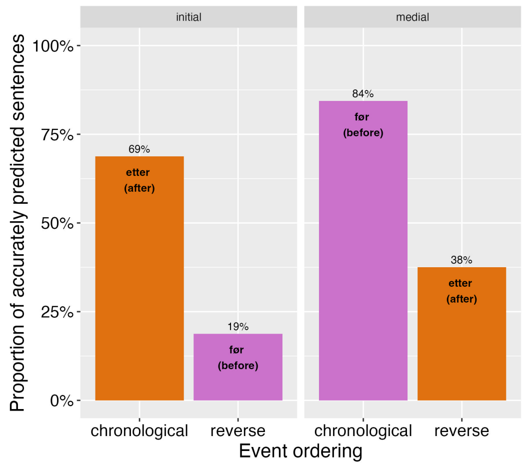


Figure 4: Proportion of Norwegian unambiguous sentences that are correctly predicted.

References

[1] Politzer-Ahles, S., Xiang, M., & Almeida, D. (2017). ”before” and ”after”: Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PLOS ONE*, *12*(4), e0175199. https://doi.org/10.1371/journal.pone.0175199

[2] Blything, L. P., & Cain, K. (2016). Children’s processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses. *Developmental Psychology*, *52*(10), 1517–1529. https://doi.org/10.1037/dev0000201

[3] Pandia, L., Cong, Y., & Ettinger, A. (2021). Pragmatic competence of pre-trained language models through the lens of discourse connectives. *Proceedings of the 25th Conference on Computational Natural Language Learning*, 367–379.

[4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

[5] Samuel, D., Kutuzov, A., Touileb, S., Velldal, E., Øvrelid, L., Rønningstad, E., Sigdel, E., & Palatkina, A. (2023, May). NorBench – a benchmark for Norwegian language models. In T. Alumäe & M. Fishel (Eds.), *Proceedings of the 24th nordic conference on computational linguistics (nodalida)* (pp. 618–633). University of Tartu Library. https://aclanthology.org/2023.nodalida-1.61/

[6] Tskhovrebova, E., Zufferey, S., & Tribushinina, E. (2022). French-speaking teenagers’ mastery of connectives: The role of vocabulary size and exposure to print. *Applied Psycholinguistics*, *43*(5), 1141–1163. https://doi.org/10.1017/S0142716422000303

[7] Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers’ processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. https://doi.org/10.1037/a0022824

Additional information about Norwegian

Norwegian is a V2 language, meaning that the finite verb typically appears in the second position of main clauses, even when the sentence begins with a temporal connective clause. This syntactic regularity may reduce ambiguity compared to English where the verb is not in the second position in the same sentences.

When translating English sentences with the connective *after* into Norwegian, we encountered structural variation: the connective *etter* (“after”) can appear in several forms, including *etter at* followed by a subject, *etter å* followed by a verb, or simply *etter* followed by a subject. In contrast, the connective *før* (“before”) occurs only in the *før* + *subject* structure. To ensure consistency and maintain ambiguity in the masked connective position, we constrained all Norwegian stimuli to use the *[MASK]* + *subject* structure. This allowed the model to choose between *etter* and *før* within the same syntactic context, eliminating structural bias.

Investigating crosslinguistic processing constraints on preposition omission under ellipsis

Robin Lemke[1]

[1] Saarland University

robin.lemke@uni-saarland.de

Background: Empirical studies revealed crosslinguistic differences with respect to the acceptability of P(reposition)-omission under ellipses like sluicing (1a) and fragments (1b) [e.g. 1, 2]. In particular in languages allowing for P(reposition-)stranding (2) like English, P-omission is acceptable (1), while it seems to be degraded in many non-P-stranding languages like German. Using four acceptability rating experiments, I systematically investigate the crosslinguistic pattern across two languages (German, English) and constructions (fragments/sluicing) and test a processing-based hypothesis on the intralinguistic variation proposed by [1].

Accounts of P-omission: From a syntactic perspective, [3, 4] claims that fragments and sluicing are derived from left dislocation (2), which predicts P-omission to occur only in P-stranding languages, and to a similar extent in fragments and sluicing. Taking a processing approach, [1] predict P-omission to be more strongly preferred if the preposition semantically depends on the verb (3) than when it does not (1), since processing (3) requires accessing the verb in the question. While [1] attribute the effect to dependency length, my experiments test this explicitly: Longer dependencies should be more difficult [5] and reduce acceptability (4).

Method I conducted four rating studies, each of which addressed one type of ellipsis in one language and crossed the binary variables OMISSION, DEPENDENCY (present/absent) and DISTANCE (short/long) between antecedent and target in a 2×2×2 design. The items (n = 24, semantically parallel between experiments) and 64 fillers were rated by 72 subjects per study on a 7-point Likert scale (7=fully natural).

Results: The data (See Fig. 1) were analyzed for each experiment separately with LMMs [6]. For English sluicing, OMISSION interacts (marginally) with DEPENDENCY ($z = 1.76, p = 0.08$) and with DISTANCE ($z = 2.21, p < 0.05$), indicating a weakened preference for P-omission in the presence of a dependency and with a longer distance. However, there is no three-way interaction, which would show that lengthening a semantic dependency is particularly degraded ($z = 0.14, p > 0.8$). For fragments, there is a marginal three-way interaction ($z = -1.9, p = 0.058$), which however suggests that P-omission is more strongly preferred in the short distance condition, unlike the dependency account would predict. In the case of German sluicing, there are no relevant interactions. For fragments, there might be a trend favoring P-omission in the presence of dependencies (OMISSION:DEPENDENCY, $z = 1.4, p > 0.1$), which is not affected by DISTANCE ($z = 0.26, p > 0.7$).

Discussion: The data show that P-omission is less acceptable in German than in English, but within German, it is more acceptable in fragments than under sluicing. This might indicate that sluicing is derived from *wh*-questions [3], whereas fragments do not involve movement (*contra* [4]). As for the dependency account, there is a tendency for P-omission to be crosslinguistically more acceptable when the preposition depends on the verb. However, this effect is not modulated by the length of the potential dependency as [1] predict. This indicates that the relationship between preposition and verb might play a role, but there is no evidence that this results in a dependency which is hard to process. Prepositions depending on the verb might be more salient and easier to retrieve during ellipsis processing. This could also explain the OMISSION:DISTANCE interactions in the English data in terms of a greater difficulty to retrieve prepositions due to memory decay.

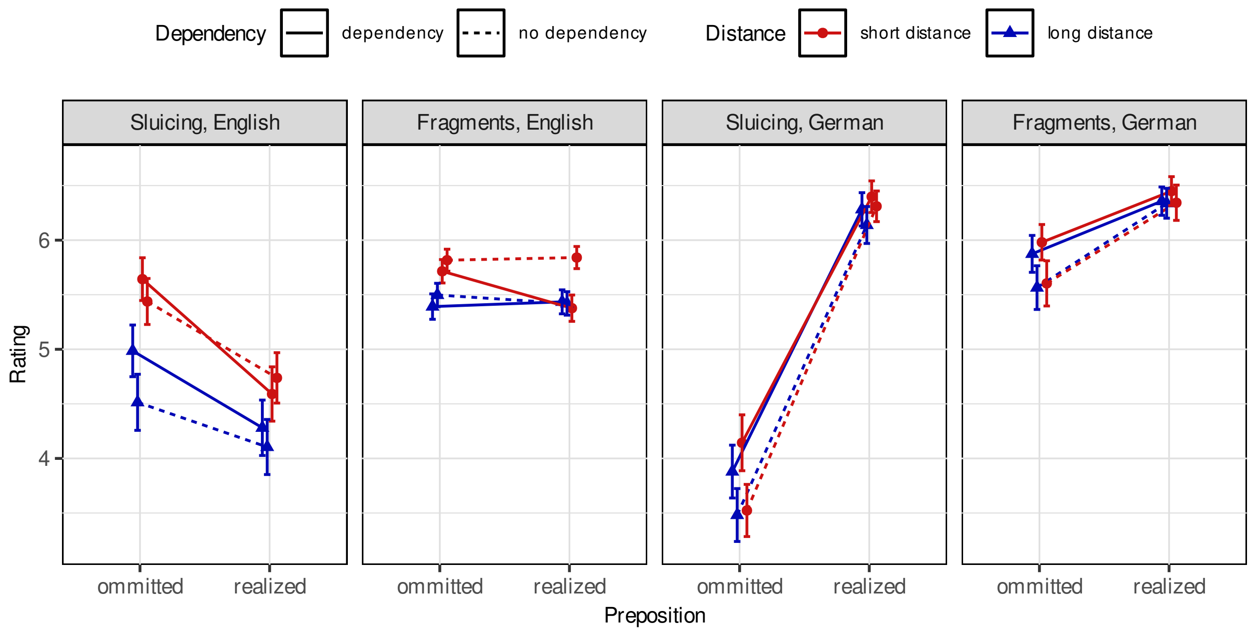


Figure 1: Mean ratings and standard errors for each of the experiments and conditions

(1) a. Nicole is playing board games with somebody today, but I don't know (with) who(m). (Sluicing)

b. With whom is Nicole playing board games today? – (With) a friend. (Fragment)

(2) a. Who$_i$ is Nicole playing board games with t_i?

b. *Wem$_i$ spielt Nicole Brettspiele mit t_i
who plays Nicole board.games with

(3) Who does the decision depend on? – (On) the director. (Fragment, P depends on V)

(4) Nicole is playing board games with somebody. *Since I did not talk to her*, I don't know (with) who(m).
(Sluicing, long condition, no dependency; lengthening expression is italicized)

References

[1] Nykiel, J., & Hawkins, J. A. (2020). English fragments, Minimize Domains, and Minimize Forms. *Language and Cognition*, *12*(3), 411–443.

[2] Lemke, R. (2021). *Experimental investigations on the syntax and usage of fragments*. Language Science Press.

[3] Merchant, J. (2001). *The syntax of silence. Sluicing, islands, and the theory of ellipsis*. Oxford University Press.

[4] Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy*, *27*(6), 661–738.

[5] Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

[6] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

When Half a Word Is Enough: How Lexical Expectations Modulate Visual Uncertainty in English and Chinese

Cui Ding[1], Shan Gao[1], Ethan Wilcox[2], Lena Ann Jäger[1]

[1]University of Zurich [2]Georgetown University

cui.ding@uzh.ch

Background: Reading is a dynamic process involving the integration of bottom-up visual input with top-down linguistic expectations. We introduce a novel *visual entropy manipulation* — selectively revealing the upper or lower half of words — to probe how readers compensate when visual information is degraded. This method tests how top-down lexical properties, derived from naturalistic texts (e.g., frequency, surprisal, entropy), modulate language processing under visual uncertainty. We apply this paradigm in both English (alphabetic script) and Chinese (logographic script) to seek cross-linguistic, script-specific insights into the mechanisms that link visual word recognition and predictive processing in reading comprehension, as well as how these processes vary across individuals.

Method: We adapted Mouse Tracking for Reading (MoTR) [1] to present naturalistic texts from the OneStopQA dataset [2] under three conditions: whole word, upper half, or lower half revealed **(Fig. 1)**. L1 speakers of English (n = 54) and Chinese (n = 57) were recruited online to read the texts and rate perceived readability. To probe the effects of visual entropy, we analyzed how partial visibility impacted word readability, its modulation by lexical variables (Zipf frequency, GPT-2 surprisal, Rényi entropy), and how individual differences in visual adaptation related to reliance on predictive cues. In parallel, we estimated the mutual information (MI) between word segments and whole words by training large language model-based optical character recognition models (LLM-OCR models) to quantify the visual redundancy across scripts.

Results: Most participants found the upper half easier to read in both languages **(Fig. 2)**. Reading-time analyses confirmed this preference: reading the lower half led to greater slowdowns than the upper ($p < .05$), especially in English **(Fig. 3)**. The lower half was less disruptive for Chinese than for English, indicating larger visual redundancy in the Chinese lower word contour, likely due to the compact, non-linear structure of Chinese characters. This aligns with a larger MI between the lower half and the whole words for Chinese, as estimated by the LLM-OCR models, compared to English. Yet, surprisingly, unlike humans, LLM-OCR models exhibit a smaller gap in word recognition ability between the upper and lower halves.

Analyzing reading data with the two-task hierarchical model from [3], we found significant interactions between visual uncertainty and lexical properties. High-frequency, low-surprisal words were more robust to visual uncertainty, whereas low-frequency, high-surprisal words had greater cost, especially in English **(Fig. 4)**. Correlation estimates revealed that individuals' sensitivity to degraded visual input was weakly associated with surprisal and entropy effects (~0.1), but more strongly with frequency (~-0.4), suggesting that adaptation to visual uncertainty relates more to experience with common words than to contextual prediction.

Discussion: These findings indicate that upper word segments convey more perceptual information across languages, and the lower half is more informative in Chinese than in English. Script-specific visual structures likely shape how readers compensate under degraded input. The interaction between visual uncertainty and lexical properties highlights the dynamic integration of bottom-up and top-down processes. Individual differences suggest that readers vary in their reliance on lexical distributional properties versus contextual predictability.

Fig. 1: Visualization of the three conditions (whole word, upper or lower half revealed) in the MoTR experimental setup for Chinese (top) and English (bottom). Visual entropy manipulation of the words in reading texts was randomized on a trial-by-trial basis.

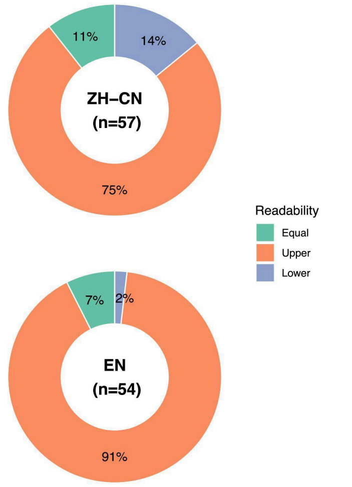


Fig. 2: Self-reported readability judgments of each word segment.

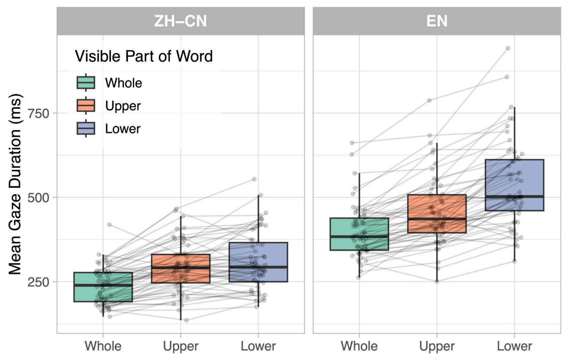


Fig. 3: Mean gaze duration visualization. Boxes show the middle 50% of values, center lines indicate the median, and whiskers represent the overall data spread. Grey lines trace each participant's mean across conditions.

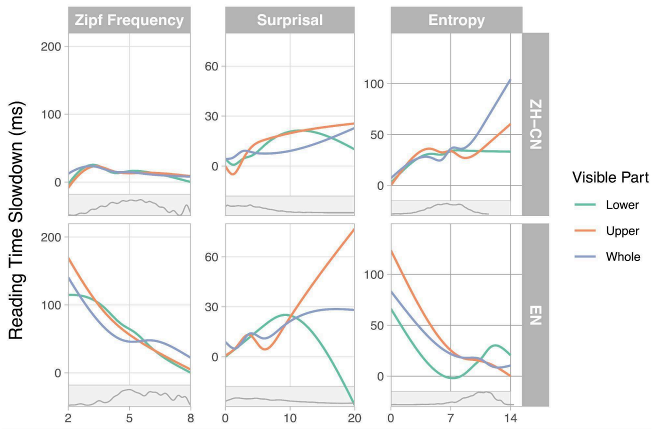


Fig. 4: Interaction between visual masking and lexical variables in Chinese (top row) and English (bottom row). Lines are GAM fits allowing for non-linear trends. They show the reading time slowdowns due to each lexical property under the three visual entropy manipulations. Grey density plots show the distribution of each lexical property.

References

[1] Ethan G Wilcox, Cui Ding, Mrinmaya Sachan, and Lena A Jäger. 2024. Mouse Tracking for Reading (MoTR): A new naturalistic incremental processing measurement tool. Journal of Memory and Language 138 (2024).

[2] Yevgeni Berzak, Jonathan Malmaud, and Roger Levy. 2020. STARC: Structured annotations for reading comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics.

[3] Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. Psychonomic bulletin review, 26(2), 452–467.

A unified threshold for individual learning and convergence across populations

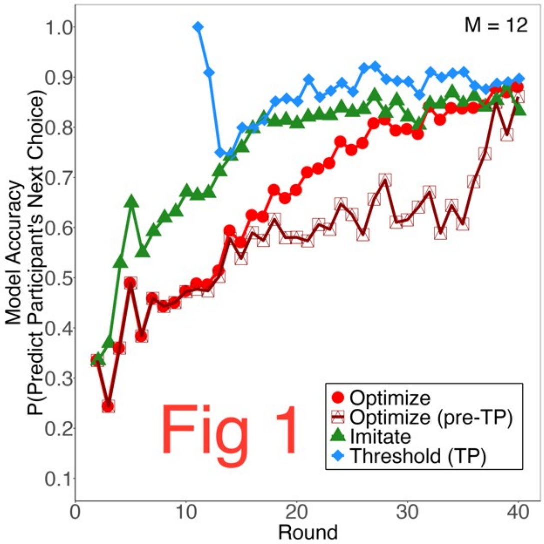
Spencer Caplan[1], Douglas Guilbeault[2], Charles Yang[3]
[1]CUNY Graduate Center, [2]Stanford University, [3]University of Pennsylvania
scaplan@gc.cuny.edu

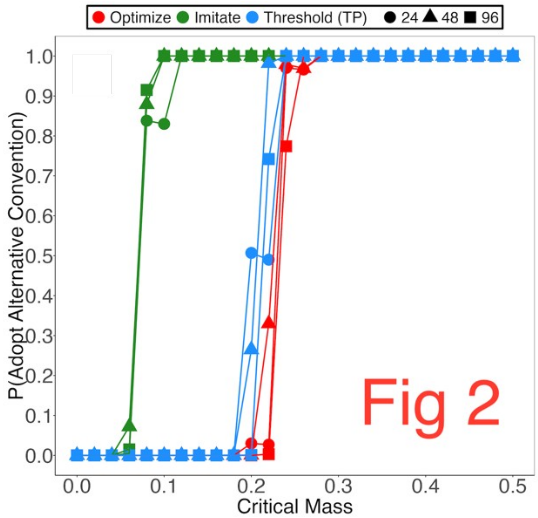
A foundational challenge in language is explaining how learners acquire stable linguistic patterns in the face of noise, variation, and exceptions [1, 2]. Under the Tolerance Principle (TP) [3], a pattern defined over N unique items is treated as productive only if the number of exceptions does not exceed the threshold of N/lnN. While TP has been successfully applied to a range of problems in language acquisition [6], its implications for social or population-level learning remain underexplored. Here, we investigate whether this same threshold can explain how shared conventions emerge from decentralized learning in dynamic social environments—where each learner receives and contributes to highly variable input over time.

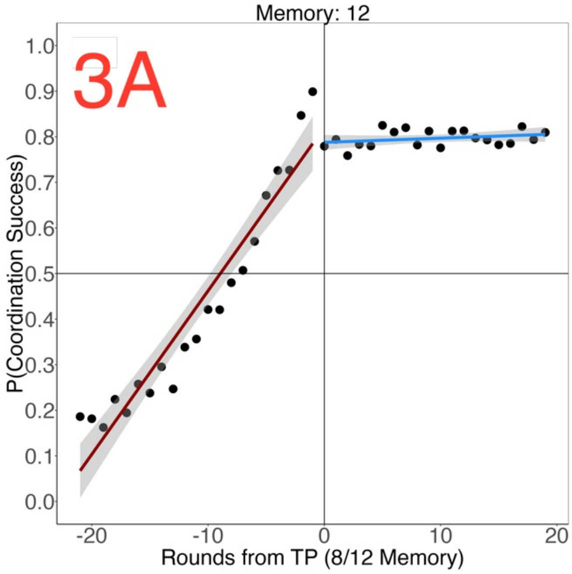
Name Game: We analyze behavioral data from the "name game" paradigm [7]—a paradigmatic coordination task in which participants embedded in a social network are repeatedly paired at random and tasked with coordinating to assign a name to a shared referent (e.g., an unfamiliar face). Each participant is incentivized to match their partner's name choice, creating pressure toward consensus. Prior work has modeled both the emergence of conventions and their disruption by a sub-population of dedicated defectors [8]. Two dominant approaches guide this literature: **imitation**—where agents repeat successful behaviors from past interactions [7], akin to usage-based learning [4], and **optimization**—where agents attempt to make statistically optimal predictions of each other's choices [8], similar to Bayesian approaches to acquisition [5]). We compare both to an extension of the TP framework applied to social learning.

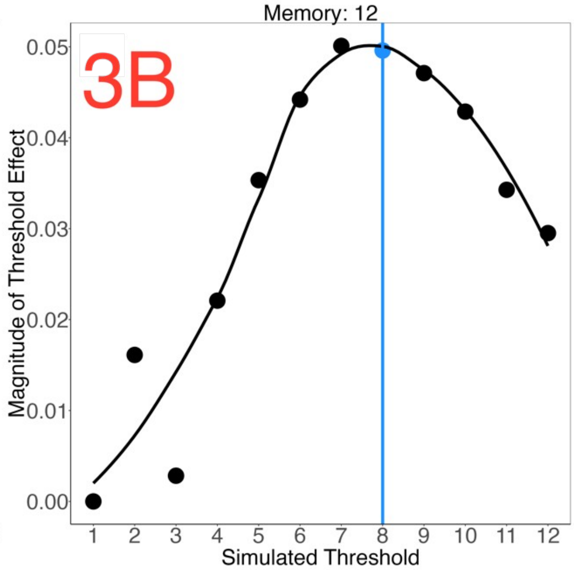
Models: All models in question assume that agents maintain a bounded memory queue of size M, storing the most recent names encountered. When paired agents agree on a name, it enters memory; if they disagree, each agent adds their partner's choice while removing their least recent self-produced entry. In the TP model, a name becomes **productive** if it occupies at least M-M/lnM slots in memory (e.g., 8/12). When this threshold is reached, agents consistently produce that name; if no form satisfies the threshold, agents sample probabilistically from memory. TP is compared against an imitation model which always repeats a successful name back, and an optimization model which always selects the most frequent name in memory.

Results: We first evaluate how well each model predicts participants' round-by-round name choices. The TP achieves significantly higher accuracy (87.4%) than imitation (76.3%) or optimization (68.7%); see Fig 1. TP is also the only model that captures the tipping point dynamics of convention switching in response to dedicated defectors: imitation reacts too quickly; optimization, too slowly (Fig 2). Fig 3A shows participants' likelihood of coordination before and after a name crosses the TP threshold in memory—consistent with a two-stage learning process. Fig 3B tests alternative thresholds by simulating a range of cutoff values and measuring the regression discontinuity in coordination success. Strikingly, TP (i.e., M/ln(M)) reliably identifies the inflection point with the strongest discontinuity. Our findings demonstrate that an individual-level threshold scales to explain the emergence and change of shared conventions in a population. Compared to models grounded in imitation or optimization, the TP provides a more accurate and cognitively plausible account of how people coordinate under uncertainty. More broadly, these results suggest that simple, rule-based mechanisms—rather than exhaustive statistical inference—may underlie both language acquisition and population-level communicative dynamics.









References

[1] Labov, W. (1989). The child as linguistic historian. *Language variation and change*, *1*(1), 85-97.

[2] Newport, E. L. (2020). Children and adults as language learners: Rules, variation, and maturational change. *Topics in cognitive science*, *12*(1), 153-169.

[3] Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

[4] Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive science*, *27*(6), 843-873.

[5] Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306-338.

[6] Schuler, K. D., Yang, C., & Newport, E. L. (2016). Testing the Tolerance Principle: Children form productive rules when it is morecomputationally efficient to do so. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 38).

[7] Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, *112*(7), 1989-1994.

[8] Centola, D., Becker, J., Brackbill, D., & Baronchelli, A. (2018). Experimental evidence for tipping points in social convention. *Science*, *360*(6393), 1116-1119.

Chinese Compound Word Production in Typing to Dictation

Shiyu Li[1 4] *, Jordan Gallant[2] *, Gary Libben[3 4], Gonia Jarema[1 4]

[1]Université de Montréal, [2]McMaster University, [3]Brock University, [4]CRIUGM

shiyu.li@umontreal.ca, gallaj20@mcmaster.ca

Background: Typing has recently been established as a valid paradigm to address whether compounds (e.g., "blackboard"), are decomposed into their constituents ("black" and "board"), a central issue in compound processing [1]. In English, inter-keystroke intervals (IKIs), i.e., the time between the onset of typing one letter and the next, lengthen at syllabic and morphological boundaries, indicating decomposition at both syllabic and morphemic levels [2]. This lengthening effect is more pronounced at morpheme boundaries [2], suggesting the boundary effect is additive, as morpheme and syllable boundaries overlap. Chinese speakers type in Pinyin, a Romanized phonetic transcription of logographic characters (cf. p.3). This study aims to examine whether typing is an effective paradigm for investigating Chinese word processing (methodological goal), and whether Chinese compounds are decomposed into morphemes in typing production (theoretical goal). Investigating Chinese compound typing not only extends the paradigm to a non-Indo-European language but also highlights how writing systems can shape compound production cross-linguistically.

Exp.1 [3]: ***Method.*** Sixty-six Chinese speakers saw 80 disyllabic two-character Chinese nouns (half compounds, half monomorphemic) and typed them in Pinyin on Pavlovia (Fig. 1). Word frequency and Pinyin syllable length were matched across the two stimulus types. No stimuli shared the same Pinyin disyllable. Due to character homophony, constituent character frequency was not matched but modeled as a random effect. ***Results & Discussion.*** Linear mixed-effects modeling[1] showed significantly longer IKIs at character boundaries (β = 0.193, p < .001, Fig. 2), without interaction with morphological structure (β = 0.003, p = .85). As characters represent syllables, this effect suggests syllabic but not morphological decomposition. However, higher initial character frequency predicted faster typing onsets (β = -0.019, p < .04)[2], possibly indicating character-level planning, as multiple characters can map to the same syllable, showing that character and syllable frequencies differ. Alternatively, character boundaries might provide visual segmentation cues that triggered these character-related effects without genuine decomposition driven by linguistic units. **Exp.2:** ***Method.*** To disentangle visual vs. linguistic sources of the boundary effect, we created Exp.2 following the same procedure and stimuli as Exp.1, except that 58 participants typed upon hearing the stimuli, synthesized by a text-to-speech tool, with a female Mandarin voice. ***Results & Discussion.*** The character boundary effect on IKIs was again observed (β = 0.244, p < .01), ruling out visual segmentation as its sole source. Crucially, initial character frequency interacted with morphological structure (β = 0.065, p < .001), facilitating overall typing speed in compounds but inhibiting it in monomorphemic words. This character effect confirmed that character activation is not task-dependent but a general mechanism in Chinese word processing across modalities. Morphological structure also modulated the boundary effect (β = 0.054, p < .01), underscoring the role of morphology in planning.

General Discussion: Two typing experiments revealed linguistic processing beyond mere motor execution, confirming the value of typing in probing Chinese word processing. While stronger evidence for morphological decomposition is needed, morphemes may nevertheless influence typing behaviour. The salience of characters as cross-modal planning units may obscure morphemic boundary effects, especially under visual presentation. Beyond examining the Chinese typing paradigm across stimulus presentation modalities, this study contributes to clarifying the interplay between characters and other sublexical units in written production. This provides cross-modal insights into Chinese compound processing and clarifies whether morphological decomposition observed in languages like English, reflects universal or language specific processes.

*Shiyu Li and Jordan Gallant contributed equally and share first authorship.

References

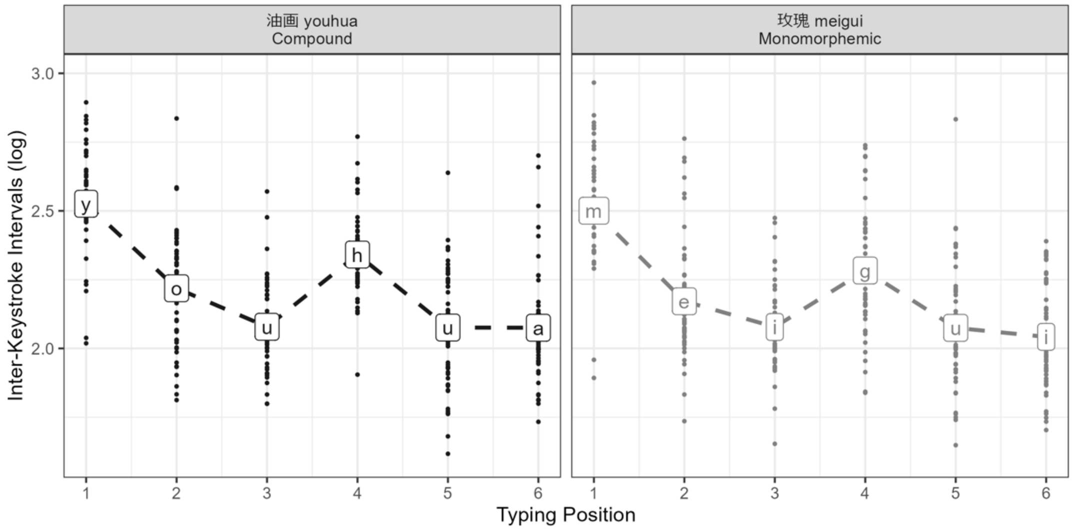
[1] Taikh, A., Gagné, C., & Spalding, T. (2023). Accessing the semantic and lexical information of constituents while typing compounds. *The Mental Lexicon*, *18*(2), 265–299. https://doi.org/10.1075/ml.21005.tai

[2] Will, U., Nottbusch, G., & Weingarten, R. (2006). Linguistic units in word typing: Effects of word presentation modes and typing delay. *Written Language & Literacy*, *9*(1), 153-176. https://doi.org/10.1075/wll.9.1.10wil

[3] Gallant, J., Chen, Y. & Li, S. (in press). Understanding the processing of Chinese compounds : Insights from Pinyin typing.

Figure 1

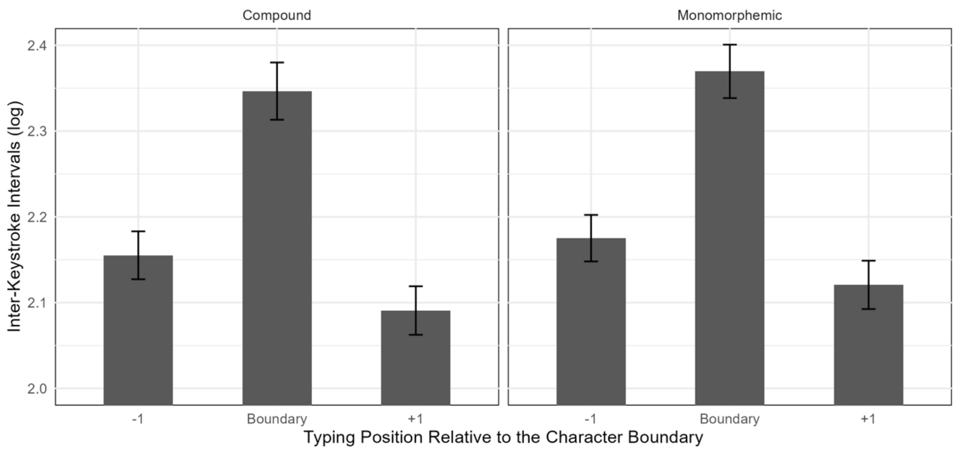
Log-transformed IKIs for All Keystrokes of Two Sample Stimuli



Note. This figure demonstrates the log-transformed IKIs for the compound word 油画 ('oil' + 'painting' = 'oil painting', typed as "youhua") and the monomorphemic word 玫瑰 ('rose', typed as "meigui").

Figure 2

Effects of Character Boundary Position on Log-transformed IKIs in Monomorphemic and Compound Words



Note. Error bars show 95% confidence intervals.

Model Notes

[1] iki ~ typing proficiency + bigram frequency + word frequency + morphological structure + whether the IKI occurred at the character boundary + (1 | participant) + (1 | word) + (1 | bigram)

[2] iki (onset) ~ typing proficiency + word frequency + initial character frequency + (1 | participant) + (1 | word); analysis limited to the first keystrokes

Chinese Characters and Pinyin Typing

The Chinese writing system can be described as morphosyllabic [1], meaning that its basic writing unit, the character, generally corresponds to one morpheme and one syllable. Chinese characters are uniformly monosyllabic, and most morphemes in the language are also monosyllabic [2]. For example, the character 油 represents the morpheme 'oil' and is pronounced /yóu/. Thus, bimorphemic Chinese compounds such as 油画 ('oil painting') contain two morphemes ('oil' and 'painting') and two syllables (/yóu/ and /huà/), and are written in two characters (油 and 画).

However, this one-to-one correspondence between a character, a syllable and a morpheme does not always hold. In polysyllabic monomorphemic words, the constituent characters are purely syllabic, without separate morphemic meaning. For example, 玫瑰 ('rose') is a disyllabic monomorphemic word, whose two characters (玫/méi/ and 瑰/gui/) lack compositional meaning. The same applies to transliterated loanwords such as 捷克 (/jiékè/, 'Czech').

Each character is composed of strokes and occupies an equally sized square space [3], making character boundaries visually salient. In dissyllabic Chinese words—compounds or monomorphemic—the two constituent Chinese characters are always written with uniform spacing and no separators, unlike English compounds (e.g., "long-term" and "hot dog").

Due to the large number of characters and relatively limited number of syllables [4], a syllable may correspond to multiple characters. For example, the tonal syllable /huà/ may correspond to morphemes such as 'painting', 'change', 'speech', 'birch', etc., written as 画, 化, 话, and 桦. Disyllabic or polysyllabic words tend to reduce this ambiguity, as fewer are homophonic.

Chinese speakers type in Pinyin, a phonographic input method based on Latin letters. Although Mandarin has four tones indicated by diacritics, tones are not typed during input (nor in this study). For example, 油画 (/yóuhuà/) and 玫瑰 (/méigui/) are typed as "youhua" and "meigui".

References

[1] DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. University of Hawaii Press.

[2] Myers, J. (2019). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. Routledge. https://doi.org/10.4324/9781315265971

[3] Cheung, H., McBride-Chang, C., & Chow, B. W.-Y. (2005). Reading Chinese. In R. Malatesha Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 421–438). Routledge. https://doi.org/10.4324/9780203448526

[4] Duanmu, S. (2011). Chinese Syllable Structure. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–24). John Wiley & Sons. https://doi.org/10.1002/9781444335262.wbctp0115

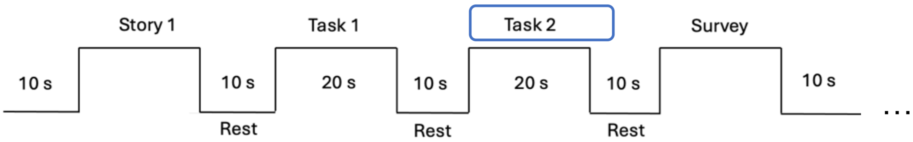
Dilemmas and Language: Observing the decision-making process between first and second language using fNIRS

Miyuki Rachel Oshima[1], Yasunori Morishima[1]
[1]International Christian University
g269005r@icu.ac.jp

Background: The Foreign Language Effect (FLE) is a phenomenon where in a foreign language (L2), individuals are more accepting of utilitarian judgments (favoring outcomes over moral rules) over deontological judgments (belief that outcomes do not outweigh the costs) [1]. Although previous studies have observed differences in utilitarian and deontological judgments in native language (L1) context [2], the cognitive mechanisms underlying this effect remains unclear for L2 context. The present study used functional near-infrared spectroscopy (fNIRS) to examine brain activation during the evaluation of moral dilemmas in both L1 and L2 conditions. Specifically, the study focused on whether those who considered the utilitarian action as “acceptable” or “unacceptable” differed in their brain activity in L1 and L2 contexts.
Methods: Twenty Japanese university students (L1: Japanese; L2: English) completed four moral dilemma blocks in Japanese and four in English. To maintain language immersion, dilemmas were presented in separate session by language. The order of language session (Japanese first or English first) was counterbalanced across participants, with a 5-minute break in between. Each story was followed by a dilemma question, acceptability question, and survey with 10 sec breaks (Figure 1). They rated the acceptability of utilitarian actions on a 6-point Likert scale with 1 being strongly unacceptable and 6 being strongly acceptable. Brain activity was recorded using a 16-channel functional near-infrared spectroscopy (fNIRS) system (Spectratec OEG-16), observing changes in oxygenated (HbO) and deoxygenated hemoglobin (HbR). English proficiency was assessed with the Total English Placement Test. Arrangement of the channels was based on the international 10-20 system. Channel locations were estimated reference charts mapping fNIRS placements to cortical regions, as provided by the Dan Chuo Lab and align with prior studies [2]. Data were band-pass filtered (0.01-0.8 Hz) and baseline-corrected using the 5 seconds before task onset. Motion artifacts were minimized through standard preprocessing. HbO changes during the task and subsequent rest period were analyzed, with peak responses identified through peak search.
Results: Participants showed similar rates of utilitarian judgments acceptance rating across language conditions (L1: 54%, L2: 56%), suggesting no clear behavioral evidence of the FLE. However, for judgments where participants rated the utilitarian action unacceptable higher activation was observed in L1 (Figure 3) than in L2 (Figure 2) over channels 11 ($t(10) = 2.78$, $p = .027$) and 12 ($t\,(11) = 2.34$, $p = .039$), with a trend in channel 10 ($t\,(7) = 1.94$, $p = .081$). For acceptable judgments, only a moderate difference was found in channel 12 ($t\,(11) = -1.98$, $p = .073$).
Discussion: Although the behavioral data did not show the FLE, the neural data suggest that rejecting utilitarian outcomes in L1 may involve greater emotional engagement than in L2. Channels 10 and 12 are estimated to potentially overlap with areas near the ventromedial prefrontal cortex (vmPFC) [3], a region associated with emotion-based decision-making. These results offer tentative support for the idea that emotional responses are dampened in L2 contexts, possibly contributing to the FLE observed in other studies. Future research using larger samples and more rigorous statistical approaches, such as cluster-based permutation testing, would help clarify these effects.

Figures

Figure 1. *Illustration of block design for one story*



Note. For this study, we will be looking at task 2, the acceptability rating for dilemmas

Figure 2. *Brain Activity during Not Acceptable Decision in English Condition*

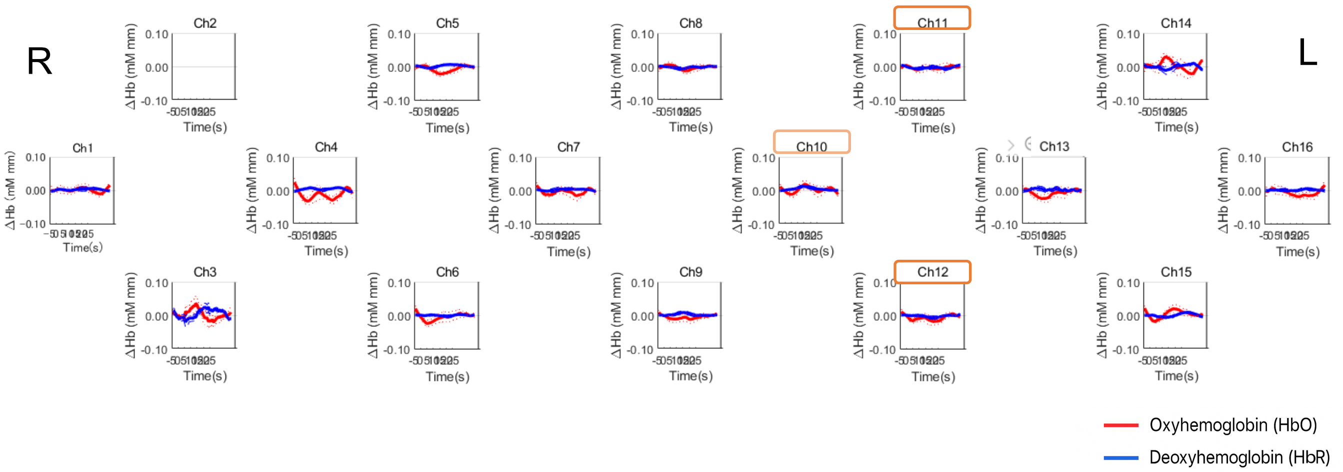
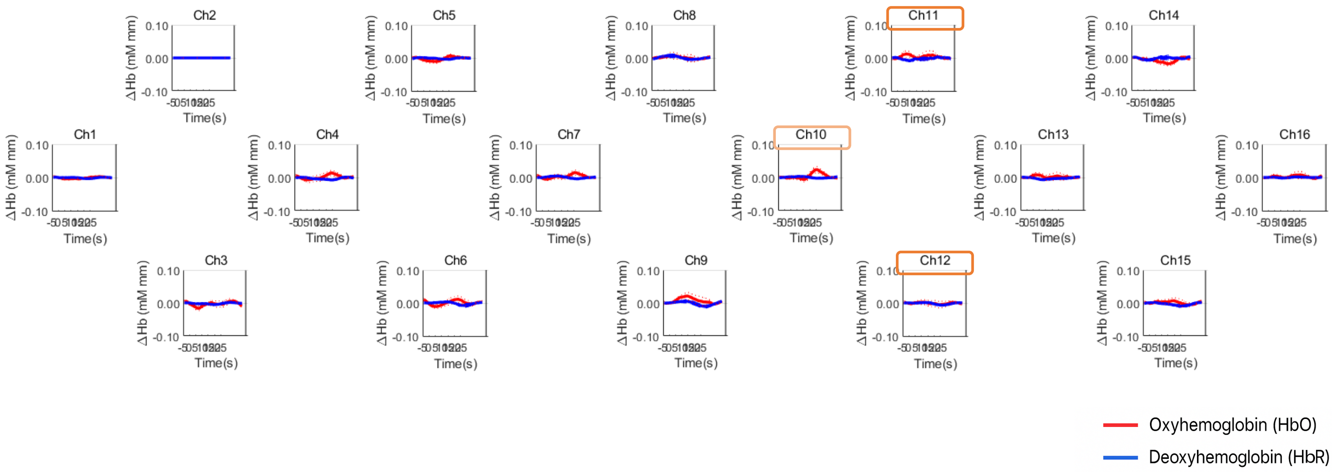


Figure 3. *Brain Activity during Not Acceptable Decision in Japanese Condition*



References

[1] Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. Cognition, 173, 8-15.

[2] Dashtestani, H., Zaragoza, R., Kermanian, R., Knutson, K. M., Halem, M., Casey, A., Shahni Karamzadeh, N., Anderson, A. A., Boccara, A. C., & Gandjbakhche, A. (2018). The role of prefrontal cortex in a moral judgment task using functional near-infrared spectroscopy. Brain and Behavior, 8(11), e01116.

[3] Moretto, G., Làdavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, 22(8), 1888–1899.

Earlier and stronger effects of prediction through production

Miriam Schulz[1], Masato Nakamura[1], Matthew W. Crocker[1]
[1]Saarland University
mschulz@lst.uni-saarland.de

The prediction-by-production hypothesis proposes that comprehenders predict by recruiting their own production system during language comprehension [5, 6], with studies suggesting facilitation of highly predicted words by modulating the involvement of production processes [1, 2, 5]. It remains less clear, however, whether the engagement of production benefits all predictable words (high and low cloze), or only those that are most expected (high cloze, in a non-linear “winner-takes-all” fashion). In a German self-paced reading (SPR) study (N=120 participants) we investigate the strength and locus of expectancy effects, contrasting standard “Reading-for-Comprehension” (RfC: SPR with comprehension questions) with a novel “Reading-for-Production” paradigm (RfP: SPR during a sentence completion task, such that readers are ready for production during comprehension of the target word; Fig. 1). Trials began with a context sentence displayed in full and continued with an SPR target sentence containing a target noun manipulated for expectancy (*HC*: high cloze, *LC*: low cloze but attested, *IM*: implausible; Table 1), followed by at least two spillover words prior to either the production prompt (in RfP) or a comprehension question (in RfC). Compared to RfC, if covert production during comprehension generally benefits predictable words, we should observe a greater facilitation (i.e., faster reading times) in RfP for *HC vs. IM*, but only a small, if any, processing advantage for *HC* over *LC*. By contrast, if covert production leads to the rapid demotion of all but the best target candidates, we should observe greater facilitation in RfP for both *HC vs. IM* and for *HC vs. LC*, shifting RTs for *LC* more towards those for *IM*.

An LMER analysis reveals a strong main effect of task in all regions, with longer RTs in RfP (all $p<.001$; Fig. 2). For *HC vs. IM*, we observe a main effect in all critical regions, showing a facilitation for *HC* (all $p<.01$), as well as interactions with task, with faster RTs for *HC vs. IM* in RfP in the target ($p=.004$) and first spillover regions ($p=.024$), but longer RTs in the second spillover region ($p<.001$). For *HC vs. LC*, however, we only observe a main effect showing faster RTs for *HC* in the first spillover region ($p<.001$), and an interaction with task in the second spillover region showing a greater facilitation for *HC vs. LC* in RfC ($p<.001$). These findings suggest that the covert production task – while increasing RTs overall – attenuates RTs for both highly- and moderately-expected target words (*HC* and *LC*), rather than just highly-expected ones (“winner-takes-all”).

The early effect of implausible words in RfP but not RfC might arise from a non-linear relationship between expectancy and RTs in the RfP task, disproportionately penalizing only implausible words in RfP. To further investigate this, we replicated the LMER analysis replacing our three-level categorical predictor with continuous target surprisal from four German LLMs, confirming both earlier and stronger effects of predictability on RTs in RfP (Fig. 3). A follow-up analysis using GAMMs further shows that the relationship between target noun surprisal and RTs remains linear in both tasks across most regions, with stronger and earlier effects in RfP in all models (Fig. 4). Interestingly, the one exception where the non-linear model outperforms the linear one reveals a superlinear trend in RfP, reflecting precisely the equal facilitation of both *HC* and *LC* words in RfP in the target region observed in our first analysis. Taken together, these findings show that comprehenders generate earlier and enhanced predictions across all levels of expectancy – but no support for a “winner-takes-all” strategy – when they recruit the production system for prediction.

Context sentence (displayed in full): Dennis möchte Yvonne heute Abend bitten, ihn im nächsten Sommer zu heiraten.
Dennis wants to ask Yvonne this evening to marry him next summer.

Target sentence (SPR):

Condition	Start	Pre-critical	Target	Spill 1	Spill 2	Continuation	Production/End
HC: High cloze	Er hat bereits einen *He has already a*	wunderschönen *beautiful*	**Ring** ***ring***	für *for*	die *the*	Verlobung *engagement*	ausgesucht. *picked.*
LC: Low cloze	Er hat bereits einen *He has already a*	wunderschönen *beautiful*	**Strauß** ***bouquet***	für *for*	die *the*	Verlobung *engagement*	ausgesucht. *picked.*
IM: Implausible	Er hat bereits einen *He has already a*	wunderschönen *beautiful*	**Teppich** ***rug***	für *for*	die *the*	gemeinsame Wohnung *shared apartment*	ausgesucht. *picked.*

Table 1: Example item. Materials were partially adapted from [4] and [7]. Target nouns were selected to have a cloze probability ranging from 50-95% for *HC* (mean = 78%), 5-16% for *LC* (mean = 6%), and zero cloze for *IM.* Plausibility ratings were collected from N=45 participants in a norming pretest (7-point Likert scale) presenting target sentences up to and including the target noun, mean ratings per condition were: *HC* = 6.76, *LC* = 5.14, *IM* = 1.56.

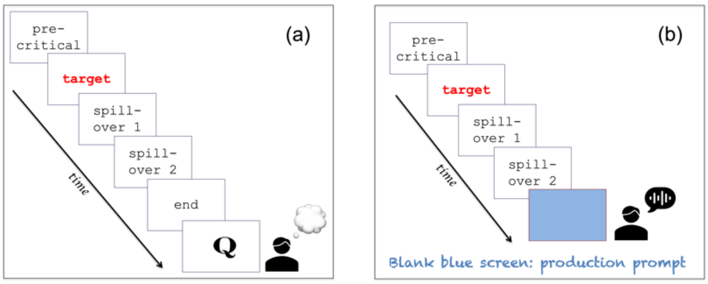


Figure 1: SPR task regimes: (a) *Reading-for-Comprehension (RfC):* SPR with a comprehension question following every trial. (b) *Reading-for-Production (RfP):* SPR with participants instructed to read silently until the screen turned blue, and then quickly complete each sentence aloud with one to two words. Their answers were recorded. All participants completed 90 trials in one and then the other task. Task order was counterbalanced across participants. Short fillers were used to vary the locus of the production prompt in the target sentence.

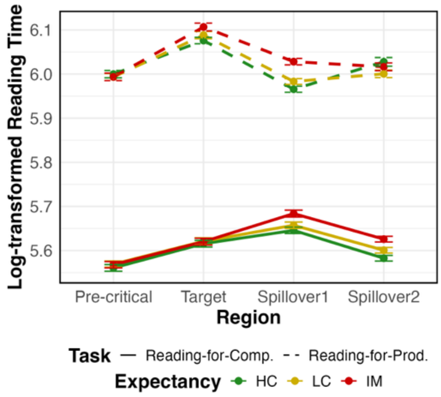


Figure 2: Reading time results. Error bars show ±1SE.

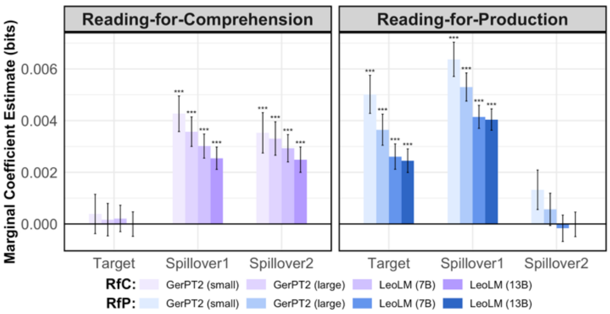


Figure 3: LMER surprisal model coefficient estimates in logRTs/bits (marginal means). One LMER model was run for each LLM in each critical region. Error bars show ±1SE.

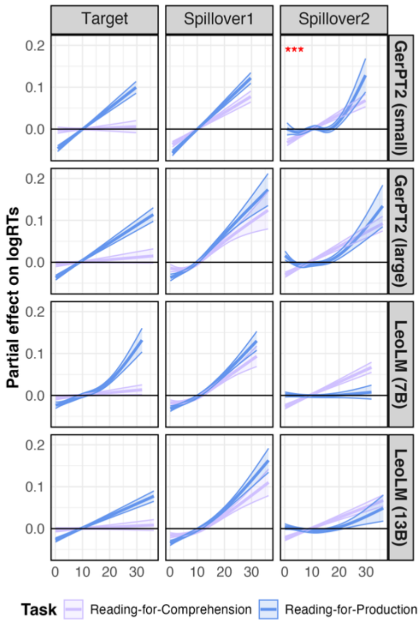


Figure 4: GAMM fits for the relation between target word surprisal and log-transformed reading times (±1SE). Asterisks indicate a significantly improved model fit for the non-linear model (plotted) compared to its linear counterpart (not plotted) through a permutation test.

References

[1] Hintz, Meyer, & Huettig. (2016). *Q. J. Exp. Psychol.*
[2] Lelonkiewicz, Rabagliati, & Pickering. (2021). *Q. J. Exp. Psychol.*
[3] Martin, Branzi, & Bar. (2018). *Sci. Rep.*
[4] Nicenboim, Vasishth, & Rösler. (2020). *Neuropsychologia.*
[5] Pickering & Gambi. (2018). *Psychol. Bull.*
[6] Pickering & Garrod. (2013). *Behav. Brain Sci.*
[7] Stone, Nicenboim, Vasishth, & Rösler. (2023). *Neurobiol. Lang.*

Relative Activation of Competing Event Roles in Mandarin Discourse Development

Fang Yang, Holly P. Branigan & Martin J. Pickering
University of Edinburgh
fang.yang@ed.ac.uk

BACKGROUND. Previous research shows that adding a character in scene descriptions makes speakers less implicit when referring to the main character, suggesting referents compete for activation during sentence formation [1, 2]. However, the nature of the interplay between referent activations remains unclear: Is it *inhibitory*, where enhanced activation of one reduces the other, or *facilitative*, where one increases the other? According to psycholinguistic assumptions, more activated referents appear earlier and with more implicit expressions [3-5]. Thus, if it is *inhibitory*, greater implicit reference to the patient should *reduce* (i) early mention or (ii) implicit reference to the agent. If it is *facilitative*, it should *increase* these tendencies instead. We tested this by examining how Mandarin speakers continue agent–patient transitive sentences.

METHODS. Participants viewed scenes where one character acted on another with consequences (e.g., Chan kicking Beckham, causing him to fall), alongside a scripted sentence describing the event. They read the prompt sentence and, while viewing the scene, freely articulated a continuation they felt naturally followed, with no other constraints. Experiment 1 (N=36) used a competition-boosting context where half of the prompt sentences highlighted the patient morpho-syntactically without diminishing the agent, using the *Ba*-structure, and the other half used canonical SVO structure. Experiment 2 (N=32) used a competition-mitigating context where two-thirds of the prompt sentences marked the patient as the topic through information structure (33% Topicalisation and 33% Left-dislocation) and the remaining third used SVO (see Table 1).

We analysed whether participants mentioned both the agent and patient or just one in the continuations. For continuations mentioning both, we examined mention order (agent-patient vs. patient-agent; treated as binomial) in a GLMER model, and the implicitness of reference (Null Pronoun > Pronoun > Name; treated as ordinal) for both the agent and the patient[1] in two separate GAM models. All analyses included prompt sentence structure and coherence relation between the prompt and the continuation (Cause-Effect vs. Other) as control variables. We excluded 12.0% and 22.4% of responses in Experiments 1 and 2, respectively, as they either mentioned both referents together (e.g., 'they') or did not mention either referent at all.

RESULTS. Participants preferred referring to one referent (70.0% and 60.0% in Experiments 1 and 2), typically the patient (65.5% and 59.9%), rather than both. Crucially, when both referents were mentioned, greater implicit reference to the patient was associated with a higher likelihood of early mention of the agent (Experiment 1: 39.4%, 97.1%, and 100% when patient reference was Name, Pronoun and Null Pronoun; β = 2.17, z = 3.91, p < .001; Experiment 2: 35.3%, 89.5% and 57.1%; β = 1.17, z = 3.92, p < .001). Moreover, in the competition-boosting context (Experiment 1), greater implicitness with one referent was also associated with greater implicitness with the other (Patient effect on Agent: edf = 0.96, z = 62.14, p < .001; Agent effect on Patient: edf = 1.22, z = 65.26, p < .001). By contrast, in the competition-mitigating context (Experiment 2), this association disappeared (edf = 0.00, z = 0.00, p = .524 and edf = 0.00, z = 0.00, p = .817).

DISCUSSION. Taken together, our results support the *facilitative* hypothesis, suggesting that referent activation is not considered independently, but rather relatively and interactively, within the speaker's situation model during sentence formation.

---

[1] Agent or Patient refers to the referent's thematic role in the prompt sentences. We did not annotate what thematic role each referent played in the continuations as it is irrelevant to our research question.

Table 1. Prompt Sentence Structure in Experiments 1 and 2.

Experiment	Prompt Sentence Structure	Example				
Experiment 1	Ba	成龙 Jackie Chan	把 Ba	小贝 Beckham	踢倒 kick-fall	了。 LE$_{aspect}$
	SVO	成龙 Jackie Chan	踢倒 kick-fall	了 LE$_{aspect}$	小贝。 Beckham	
Experiment 2	Topicalisation	小贝， Beckham	成龙 Jackie Chan	踢倒 kick-fall	了。 LE$_{aspect}$	
	Left-dislocation	小贝， Beckham	成龙 Jackie Chan	踢倒 kick-fall	了 LE$_{aspect}$	他。 him
	SVO	成龙 Jackie Chan	踢倒 kick-fall	了 LE$_{aspect}$	小贝。 Beckham	

Notes. Prompt sentence structure was sum-contrast coded as follows: In Experiment 1, Ba = +0.5 and SVO = -0.5; in Experiment 2, contrast 1 coded TOP = 1, LDT = 0, SVO = -1, and contrast 2 coded TOP = 0, LDT = 1, SVO = -1.

References

[1]. Fukumura, K., van Gompel, R. P., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, 65, 331–344. https://doi.org/10.1016/j.jml.2011.06.001

[2]. Arnold, J. E., & Zerkle, S. A. (2019). Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience*, 34(9), 1152–1175. https://doi.org/10.1080/23273798.2019.1636103

[3]. Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

[4]. Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge university press.

[5]. Von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, 154, 117-127. https://doi.org/10.1016/j.pragma.2019.07.025

Learning morphological rules across typical and atypical development

Joana Miguel[1,2], Susana Cardoso[1], Catarina Barbosa[3], João Veríssimo[3]
[1] Research and Development Department CAIDI (NID-CAIDI); [2] Center of Linguistics NOVA University of Lisbon (CLUNL); [3] Center of Linguistics, School of Arts and Humanities, University of Lisbon (CLUL)
joana.miguel@caidi.pt

Background: Learning a language entails learning both regular patterns that can be productively extended to novel cases (e.g., adding -s to form noun plurals) and exceptions that apply only to specific items (e.g., child–children). Evidence from natural and artificial language learning shows that children can rapidly acquire regular rules, even when exposed to inconsistent input [1,2]. In contrast, adult learners are less likely to regularize and tend to match the probabilities of the patterns in their input [2,3]. However, the extent and causes of these developmental differences remain contested [4,5]. Moreover, little is known about rule learning in atypical cognitive profiles such as developmental language disorder (DLD).

To address these gaps, we investigated how children—both typically developing (TD) and with DLD—and adults learn and generalize plural suffixes in an artificial language. The input was manipulated to promote or disfavour rule abstraction, on the basis of a type-frequency criterion [6]. Drawing on the Declarative/Procedural model, which associates rule learning with procedural memory and exception learning with declarative memory, we hypothesized that adults and children with DLD would be less likely to learn regular rules, due to maturational changes or procedural memory impairments [7].

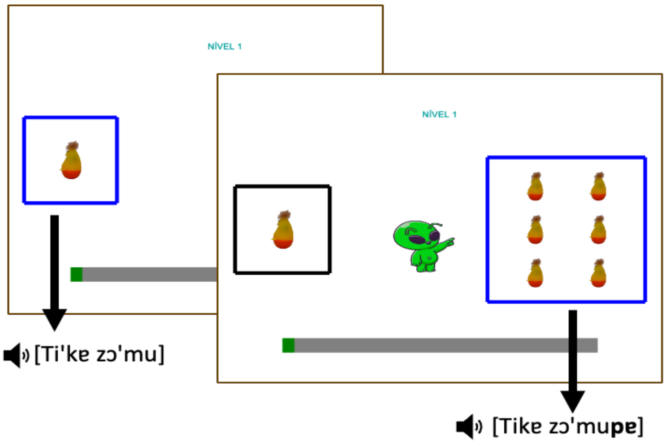
Method: We tested 216 native speakers of European Portuguese: 72 TD children 72 children with DLD (aged 5–10 years), and 72 adults (aged 22–44 years). Participants were exposed to a miniature artificial language, where plural formation involved suffixation. They were trained on 10 singular– plural pseudoword pairs (e.g., zɔmu → zɔmupɐ), accompanied by visual referents from the NOUN database (Fig. 1). Two input conditions manipulated the distribution of suffixes: (i) in condition 6R4E, six stems shared a regular suffix and four had distinct suffixes (exceptions); (ii) in condition 4R6E, there were four regulars and six exceptions, disfavoring rule abstraction [6]. Token frequency was held constant across conditions, with regular suffixes comprising 50% of tokens. Suffix generalization was tested on 24 novel items, and participants were prompted to produce the plural form orally. Responses were analysed using mixed-effects logistic regression models contrasting (i) learned vs. nonlearned suffixes and (ii) regular vs. exception suffixes (within learned suffixes).

Results: We have obtained significant interactions between group and condition. In the 6R4E condition (rule-promoting), TD children produced significantly more learned (vs. non-learned) and regular (vs. irregular) suffixes than adults and children with DLD (Fig. 2). In the 4R6E condition (rule-disfavoring), adults showed higher retention of learned suffixes than the other groups, but proportions of regular suffixes produced did not differ across groups. TD children—but not adults nor children with DLD—showed a significant difference between conditions, producing more learned and regular suffixes in the 6R4E than in the 4R6E condition.

Discussion: Only TD children extended the regular suffix beyond its input frequency, and only when the distribution of types promoted rule learning, suggesting that they were able to acquire a morphological rule. Adults retained more trained suffixes, but did not extend the regular suffix beyond its input (type or token) frequency. Children with DLD performed similarly to adults, with their regular suffix use reproducing the probabilistic properties of the input. Thus, although all groups generalized to novel items, only TD children showed evidence of productive rule learning. These findings are compatible with the Declarative/Procedural model and highlight the roles of maturation and learner profile in shaping morphological generalization.

Supplementary material

Fig. 1. Training and generalization phases in the artificial language task. In each training trial, participants saw a singular object and heard an auditory prompt *tikɐ* [noun] (e.g., *tikɐ zɔmu*), with *tikɐ* serving as a fixed introductory frame ("this is"). This was followed by the plural display and its corresponding plural form (e.g., *tikɐ zɔmupɐ*). Visual referents provided number cues. In the generalization phase, participants saw a novel singular item (e.g., *tikɐ gɛbu*), followed by its plural display, and were prompted to complete the sentence aloud (e.g., tikɐ…) and to apply a suffix based on prior exposure to the novel name.



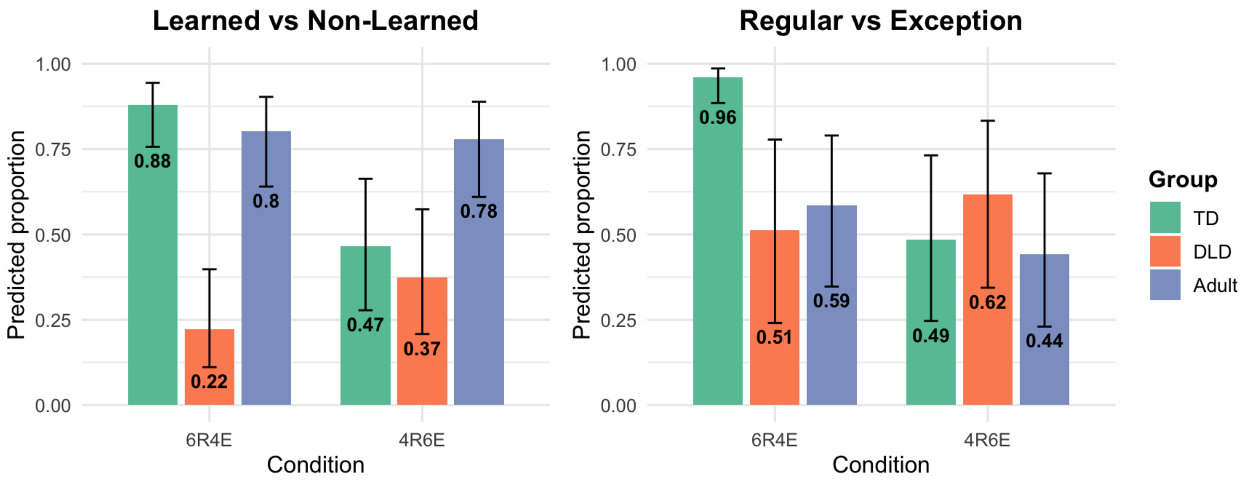


Fig. 2. Model-estimated probabilities of producing (A) a learned suffix and (B) a regular suffix, by group and input condition. Error bars represent 95% confidence intervals; numbers indicate the estimated proportions.

References

[1] Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

[2] Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*, 151–195.

[3] Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30–66.

[4] Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development*, *12*, 138–155.

[5] Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*, 1–14.

[6] Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.

[7] Ullman M. T., Pierpont E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex*, 399-433.

Saturday Morning Posters

Morphological processing in Modern Greek: A form-then-meaning, dual-route account

Nikolaos Ntagkas, Despina Papadopoulou
Aristotle University of Thessaloniki
nntagkas@lit.auth.gr

Background: Experimental research on the role of semantics in morphological processing has been quite prolific over the years, yielding inconclusive results that have given rise to different accounts of morphological processing [1, 2]. Form-then-meaning accounts suggest that morphological processing is form-based at the early/peripheral stages and semantically-informed at the later/central stages [3], in contrast to form-with-meaning accounts assuming that form and meaning are processed concurrently and interdependently already from the very early stages of morphological processing [4]. Dual-route accounts postulate decompositional processing for semantically transparent complex words and non-decompositional processing for semantically opaque complex words in an all-or-none fashion [5, 6], unlike emergentist accounts positing graded semantic effects on morphological processing [7, 8]. Furthermore, full-decomposition accounts put forward a strict dissociation between morphology and semantics, so that morphological processing is entirely independent of meaning [9].

Method: These accounts were tested with respect to the processing of derived words by typical adults in Modern Greek, a fusional Indo-European language with a very rich and highly productive morphological system. Participants were 144 (48 per experiment) university students, skilled native speakers of Modern Greek; 72 female and 72 male; mean age 21.9 (2.8) years, range [18–30]. We designed a morphological priming lexical decision task comprising 90 critical prime-target pairs, equally distributed in 3 semantic transparency conditions, i.e., transparent, semitransparent, and opaque [7, 8, 10, 11]. Each critical pair consisted of a suffixed verb or adjective as prime and a base noun as target. For each target we selected a morphologically, formally, and semantically unrelated control prime, appropriately matched to the corresponding critical prime (see Table 1). The same task was implemented in 3 reaction time experiments with different priming protocols considered to tap into different stages of morphological processing: forward masked priming (SOA = 50 ms) for the early/peripheral stages [1, 12]; unmasked visual priming (SOA = 250 ms) and immediate cross-modal priming (SOA = auditory prime duration) for the late/central stages [1, 13].

Results: Modeling reaction time data with linear mixed-effects regression in R revealed statistically significant forward masked priming of the same magnitude among the 3 semantic transparency conditions (transparent = semitransparent = opaque; see Figure 1). On the other hand, unmasked visual priming as well as immediate cross-modal priming was statistically significant and equivalent in the transparent and semitransparent condition, yet completely absent in the opaque condition (transparent = semitransparent ≠ opaque; see Figures 2 and 3).

Discussion: Our findings indicate that meaning is not significantly engaged in the early/peripheral stages of morphological processing, but plays a determinant role in the later/central stages of morphological processing. These results can be readily accommodated within a form-then-meaning, dual-route account, supposing that, faced with a morphologically complex stimulus, decompositional processing has initially a head start over holistic processing; later though, as semantics come into play, decompositional processing predominates in the case of transparent and semitransparent derived words, while holistic processing predominates in the case of opaque derived words.

References

[1] Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology and skilled reading: An empirical overview. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 311–332). Cambridge University Press.

[2] Marelli, M. (2023). The role of semantics in the processing of complex words. In D. Crepaldi (Ed.), *Linguistic morphology in the mind and brain* (pp. 65–79). Routledge, Taylor and Francis Group.

[3] Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*(7-8), 942–971.

[4] Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience, 9*, Article 111. https://doi.org/10.3389/fnhum.2015.00111

[5] Clahsen, H., Sonnenstuhl, I., & Blevins, J. P. (2003). Derivational morphology in the German mental lexicon: A dual mechanism account. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 125–155). Mouton de Gruyter.

[6] Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Lawrence Erlbaum Associates.

[7] Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General, 136*(2), 323–345.

[8] Jared, D., Jouravlev, O., & Joanisse, M. F. (2017). The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(3), 422–450.

[9] Smolka, E., Preller, K. H., & Eulitz, C. (2014). 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language, 72*, 16–36.

[10] Kielar, A., & Joanisse, M. F. (2011). The role of semantic and phonological factors in word recognition: An ERP cross-modal priming study of derivational morphology. *Neuropsychologia, 49*(2), 161–177.

[11] Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes, 23*(3), 394–421.

[12] Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 3–37). Psychology Press.

[13] Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*(1), 3–33.

Table 1. Experimental conditions and examples of materials (between-target design)

Semantic Transparency Condition Ratings Mean (SD) [95% CI] (7-point scale, min = 1, max = 7)	Target	Critical Prime	Unrelated Control Prime
Transparent 6.40 (0.299) [6.26 – 6.55]	πλήθος /ˈpliθos/ ‘crowd’	πληθαίνω /pliˈθeno/ ‘to multiply’	ζεσταίνω /zeˈsteno/ ‘to warm’
Semitransparent 4.17 (0.458) [4.02 – 4.31]	λίπος /ˈlipos/ ‘fat’	λιπαίνω /liˈpeno/ ‘to lubricate’	βαθαίνω /vaˈθeno/ ‘to deepen’
Opaque 1.83 (0.423) [1.69 – 1.98]	χόρτο /ˈxorto/ ‘grass’	χορταίνω /xorˈteno/ ‘to feel full’	μακραίνω /maˈkreno/ ‘to lengthen’

Figure 1. Forward masked priming

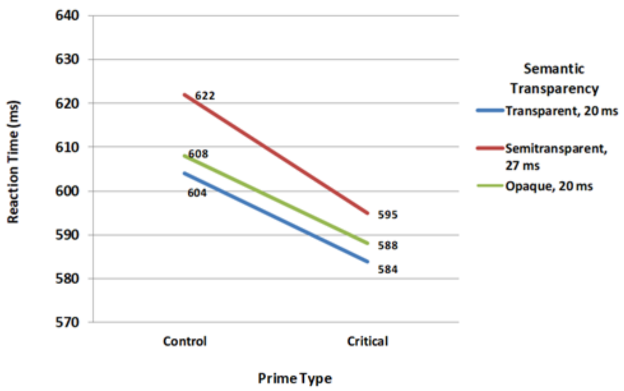


Figure 2. Unmasked visual priming

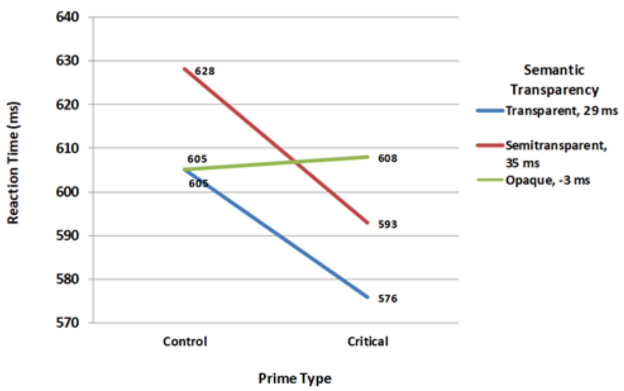
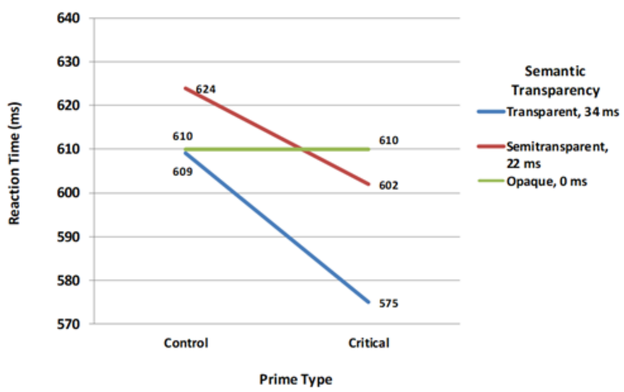


Figure 3. Immediate cross-modal priming



Language Processing Insights from Average Phonemic Bigram Surprisal

Alexander Kilpatrick[1], Rikke Bundgaard Nielsen[2]

[1]Univeristy of Aizu, [2]University of Melbourne

alex@u-aizu.ac.jp

The human brain processes predictable events faster and more accurately than unpredictable ones. This fundamental aspect of cognitive processing is quantifiable through metrics such as Shannon's surprisal [1], which measures information expression in communication systems based on probability. Improbable events are more surprising and carry more information. Cross-linguistic studies of surprisal have shown that languages tend to converge in how much information they express over time, suggesting that speech rates are constrained by shared processing limitations [2]. Average phonemic bigram surprisal is the negative log probability (square) of one phoneme following another, averaged out for each bigram in a word. Our recent research reveals that phonemic surprisal predicts a range of psycholinguistic behaviours and provide novel insights into language processing and lexeme evolution. The present study synthesises these findings. Our research is based on a very large corpus (~50 million words) of spoken American English [3], cross referenced with an American English pronouncing dictionary [4] to calculate average phonemic bigram surprisal. We cross-referenced the results of existing experiments to include as dependent and independent variables in a series of multiple linear regression models, examples of which can be found in Table 1. Included in these models are iconicity scores [5] which are important because iconic words—those with a sound/meaning resemblance—are processed with a cognitive advantage [e.g., 6] and can therefore provide a benchmark.

Our first experiments measured the **cognitive load** associated with average phonemic bigram surprisal. [7] showed that increased surprisal was correlated with decreased accuracy ($p < 0.01$) and increased response times ($p < 0.01$). In a lexical decision task, decreased accuracy ($p < 0.001$) and increased response times ($p < 0.001$). In a read aloud experiment, increased age of acquisition in two age of acquisition experiments ($p < 0.001$ in both cases), and increased accuracy ($p < 0.001$) in a memory recognition experiment. These models (Table 1) suggest that high-surprisal words are more difficult to process but also more efficiently retained in long-term memory. **Iconicity** scores were also highly predictive in these models which uncovered an interesting area of investigation considering the iconic treadmill hypothesis [8] which posits that iconic words typically evolve towards arbitrariness through a series of de-iconization stages. Using an etymological dictionary that assigns de-iconization stages [9], we showed a correlation ($p < 0.001$) between stages and iconicity scores suggesting that as iconic words evolve towards arbitrariness, surprisal also declines, offsetting the loss of processing advantage associated with iconicity [10]. Iconic words are said to be phonologically marked, and we considered that surprisal might be a quantifiable measure of phonological markedness. We tested this using the **negativity bias** which is the tendency for negative information to be processed more intensely and remembered more vividly than positive information. Using valence and emotion experiment results, [11] showed that words associated with negative valence carry more information than neutral or positive words ($p < 0.001$). Indeed, negative emotions like disgust ($p < 0.05$) were associated with high surprisal while positive emotions like anticipation ($p < 0.01$) and joy ($p < 0.05$) were associated with low surprisal. **Humour**; however, presents an interesting case as it was shown [12] to be associated with both positive valence ($p < 0.001$) and high surprisal ($p < 0.001$). **In conclusion**, average phonemic bigram predicts both processing difficulty and lexical retention, offering insight into how languages balance efficiency and expressiveness. As we expand this work cross-linguistically, surprisal may help uncover universal strategies languages use to balance efficiency, expressiveness, and memorability.

Test	Lexical Decision		Reading Task		Age of Acquisition		Memory
Variable	Accuracy	RT	Accuracy	RT	Dataset 1	Dataset 2	Accuracy
Intercept	88.98***	67.69***	244.53***	122.09***	44.76***	20.84***	30.25***
F statistic	11.05	28.4	44.79	417.6	261.8	70.83	58.71
DF	15:10324	15:10324	15:13093	15:13093	15:12499	14:4086	11:4561
Adjusted R2	0.016	0.04	0.049	0.3228	0.239	0.195	0.122
Average Surprisal	**-3.02****	**3.05****	**-6.10*****	**9.37*****	**8.79*****	**6.62*****	**30.25*****
Iconicity	**3.1****	**-2.17***	**16.07*****	**-11.77*****	**-20.26*****	**-6.89*****	**8.09*****
Phonemic Length	6.46***	13.25***	-13.40***	56.65***	33.40***	21.38***	3.32***
Morphemic Length	1.97*	2.169*	5.81***	-4.44***	1.42	0.17	-2.36*
PoS_Adverb	1.33	-0.6	3.30***	-5.64***	-11.47***	-3.60***	-3.87***
PoS_Article	-4.64***	1.33	0.92	0.10	-2.57*	-1.54	
PoS_Conjunction	-1.39	1.02	0.27	-1.06	-5.09***		
PoS_Determiner	-0.36	0.42	1.98*	-1.56	-5.95***	-2.35*	
PoS_Interjection	-0.35	2.80**	-1.17	0.33	-3.81***	-2.54*	0.28
PoS_Name	0.68	2.49*	1.08	-0.32	-0.80	-1.01	2.57*
PoS_Noun	2.63**	0.73	5.42***	-3.53***	-8.40***	-2.24*	5.47***
PoS_Number	0.48	0.82	2.03*	-1.56	-8.48***	-2.44*	0.49
PoS_Preposiiton	-1.6	0.77	1.92	-1.56	-6.53***	-2.71**	-1.06
PoS_Pronoun	1.22	0.016	2.56*	-3.93***	-10.06***	-1.97*	
PoS_Verb	2.13*	2.94**	4.84***	-0.60	1.16	0.37	-10.39***

Table 1: Series of Multiple Linear Regression analyses testing the relationship between average surprisal and a series of psycholinguistic tests. RT represents Response Time. Asterisks represent statistical significance.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379-423.

[2] Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. Science Advances, 5(9), eaaw2594.

[3] Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

[4] Carnegie Mellon University. (n.d.). CMU Pronouncing Dictionary. Retrieved April 1, 2025, from https://www.speech.cs.cmu.edu/cgi-bin/cmudictCMU Speech

[5] Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2024). Iconicity ratings for 14,000+ English words. Behavior research methods, 56(3), 1640-1655.

[6] Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. Language and cognition, 12(1), 164-181

[7] Kilpatrick, A., & Bundgaard-Nielsen, R. L. (2025). Exploring the dynamics of Shannon's information and iconicity in language processing and lexeme evolution. PLoS One, 20(4), e0321294.

[8] Flaksman, M. (2017). Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In Dimensions of iconicity (pp. 15-38). John Benjamins Publishing Company.

[9] Flaksman, M. (2024). An Etymological Dictionary of English Imitative Words. Peter Lang.

[10] Flaksman, M., & Kilpatrick, A. (2025). Against the tide: How language-specificity of imitative words increases with time (as evidenced by Surprisal). In M. Flaksman & P. Akumbu (Eds.), SKASE Journal of Theoretical Linguistics, 22(2), 78-93.

[11] Kilpatrick, A. (2025). The negativity bias is encoded in language. JASA Express Letters, 5(2).

[12] Kilpatrick, A., & Flaksman, M. (2025). The exception of humour: Iconicity, Phonemic Surprisal, Memory Recall, and Emotional Associations. arXiv preprint arXiv:2502.01682.

Experimental approach to advice-giving in French L1 and Japanese L1

Emma Corbeau[1], Céline Pozniak[2], Heather Burnett[1]
[1] Laboratoire de Linguistique Formelle, CNRS, Université Paris Cité; [2] Structurelle Formelle du Langage, CNRS, Université Paris 8
emma.corbeau@u-paris.fr

The literature on advice-giving highlights the epistemic asymmetry intrinsic to this speech-act in which the advisor is perceived as more knowledgeable than the advice-receiver (e.g.: [4]). Studies adopting the conversational analysis (CA) framework investigated contexts where the advisor's expertise on the issue at hand was obvious for both participants (e.g. for English L1 interactions: [2]). To a lesser extent, studies in Japanese also used the CA framework in contexts of intrinsic epistemic asymmetry (university professors advising their students) or where the advisor established the epistemic asymmetry (advice between friends) (e.g.: [1]). Experimental studies on advice-giving investigated, among other things, the role of social distance (e.g. for Japanese L1: [5]). To our knowledge, no studies have looked at the influence of the advisor's lack of expertise. Advice in French has not been studied much from either approach.

Moreover, as far as we know, advice in French L1 and Japanese L1 have never been compared.

We present the results of a forced-choice experiment conducted online with 148 French L1 speakers (mean age: 36) and 160 Japanese L1 speakers (mean age: 41). We tested the influence of two variables on the type of advice chosen: the advisor's expertise (or lack of), and the social distance (close or far) between interlocutors. Following a 2x2 design, participants had to read situations (where the level of expertise and social distance varied) and to choose the most appropriate answer between three phrasings of the same advice. To design the answers, we took inspiration from a corpus study we conducted beforehand on advice in French L1 and Japanese L1. We also used Hinkel's [3] typology which was the framework for the corpus analysis (see Table 1 for an example of an item).

Using hedged advice (advice including tools of epistemic mitigation) as our reference level, results showed an influence of all independent variables individually (Bayesian multinomial model). Even though expertise and social distance did not interact with each other, we found high probabilities of having an interaction between the language and expertise ($\hat{\beta} = -0.80$, 95% CrI = [-1.57, -0.05], $P(\beta < 0) = 0.98$ and $\hat{\beta} = -0.46$, 95% CrI = [-1.13, 0.20], $P(\beta < 0) = 0.91$), and language and social distance ($\hat{\beta} = -1.40$, 95% CrI = [-2.28, -0.61], $P(\beta < 0) = 1$). As seen in Figure 1 and confirmed in the model, there seems to be a preference in French rather than in Japanese for direct or indirect advice over hedged advice with expertise rather than without, and no matter the social distance. Our model also yielded weak evidence ($\hat{\beta} = 0.35$, 95% CrI = [-0.31, 1.06], $P(\beta > 0) = 0.84$) for another effect of language and social distance: we observed more indirect comments than hedged advice in French than in Japanese when advising a stranger compared to a friend.

It seems that "hedged advice" ([3]) is the least compatible type when lacking expertise in French. In contrast, Japanese participants seem to favor this type across all conditions (see Figure 1 below). This result does not seem to align with the literature in regards to social distance. Indeed, [5]'s findings suggest that Japanese speakers are sensitive enough to the social distance to alter the way they are speaking. Given this tendency for hedged advice in all conditions, Japanese participants may be showing a higher degree of ritualization in the situations tested than French participants ([6]).

Situation:		
Condition 1	Expertise	No expertise
	Vous êtes médecin avec 15 ans d'expérience dans un cabinet médical. *You are a doctor in a health center with 15 years of practice.*	Vous travaillez à l'accueil depuis 3 semaines dans un cabinet médical. *You've been working at the front desk of a health center for 3 weeks.*
Context	Le téléphone de votre bureau sonne. *The phone on your desk is ringing.*	
Condition 2	Social distance: close	Social distance: far
	Il s'avère que la personne à l'autre bout du téléphone est votre meilleure amie. *The person on the phone is your best friend.*	Il s'avère que la personne à l'autre bout du téléphone est une nouvelle patiente. *The person on the phone is a new patient.*
Hearer's problem	Elle tousse depuis hier et souhaite votre avis sur la chose à faire. *They are coughing since yesterday and would like your opinion on what to do.*	
Instruction:		
Choisissez la réponse qui vous semble la meilleure : *Choose the answer you deem the best:*		
Possible answers:		
Direct Advice	<u>Va</u> // **Allez*** à la pharmacie. *Go to the pharmacy.*	
Hedged Advice	Je pense que <u>tu peux</u> // **vous pouvez** aller à la pharmacie. *I think you can go to the pharmacy.*	
Indirect Comment	La pharmacie peut fournir des médicaments. *Pharmacists can give out medecine.*	

Table 1: Example of an item in all conditions

*When talking to a friend, the participant saw the answers with the underlined parts. When talking to a stranger, the participant saw the answers with the bold parts.

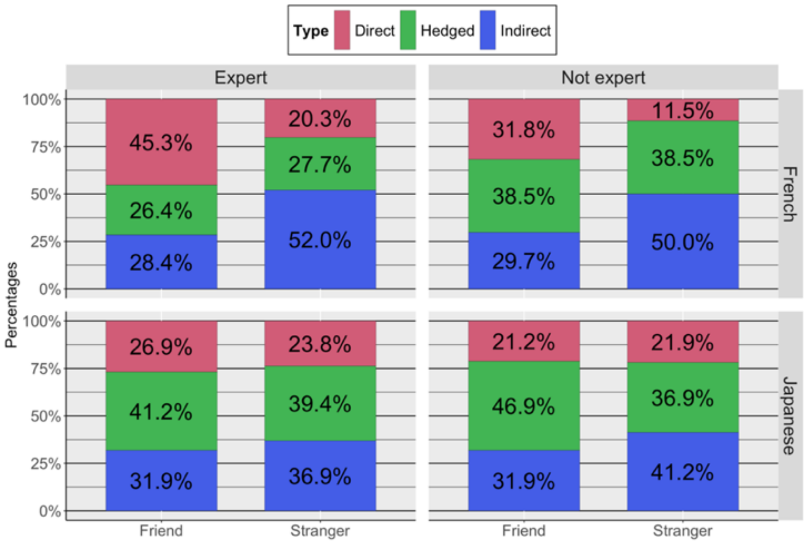


Figure 1: Distribution of types of advice according to experise, social distance and language

References

[1] Hatsuda, A. (2022). Advice Giving in Japanese Conversations [PhD Thesis, The University of Arizona].

[2] Heritage, J., & Sefi, S. (1992). Dilemmas of advice: Aspects of the delivery and reception of advice in interactions between health visitors and first-time mothers. *Talk at work: Interaction in institutional settings*, *359*, 417.

[3] Hinkel, E. (1997). Appropriateness of Advice : DCT and Multiple Choice Data1. Applied Linguistics, 18(1), 1–26. https://doi.org/10.1093/applin/18.1.1

[4] Hutchby, I. (1995). Aspects of recipient design in expert advice-giving on call-in radio. Discourse Processes, 19(2), 219–238. https://doi.org/10.1080/01638539509544915

[5] Song, L. (2023). "Nihongo no jogen bamen ni okeru nigenkankei no kanri – rapôru manejimento ni modozuita bogowasha to chûgokujin nihongo gakushûsha no bunseki wo tsûjite 29-". [Relationship Management of Spoken Japanese in Advising Situations]. *Nihongo kyôiku hôhô kenkyûkaishi*, 29(2): 70–1. https://doi.org/10.19022/jlem.29.2_70

[6] Spencer-Oatey, H. (Ed.). (2004). *Culturally speaking: Managing rapport through talk across cultures*. A&C Black.

Bridging models of linguistic planning and speech production: The case of lexical stress in English

Julia Chauvet[1], Andrea E. Martin[1,2], Ardi Roelofs[3,4], Frank H. Guenther[5], Antje S. Meyer[1,3]

[1] Max Planck Institute for Psycholinguistics, [2] Donders Center for Cognitive Neuroimaging, Radboud University, [3] Radboud University, [4] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, [5] Departments of Speech, Language, & Hearing Sciences and Biomedical Engineering, Boston University

Speaking involves both the linguistic planning of an utterance and its motor execution. While some aspects of speech emerge dynamically at the sensorimotor level, others—such as prosody—are shaped by high-level planning. Lexical stress, in particular, influences speech duration, intensity, and pitch. Theories of word production have addressed the planning of lexical stress within psycholinguistic frameworks [1] or, independently, its execution in speech production [2, 3]. Yet, no proposal formally captures the trajectory from an abstract prosodic specification to a motor command. To formalize this trajectory, we align the representational assumptions and neural mappings of the WEAVER++/ARC model of lexical access [4] and the GODIVA model of speech sound sequencing [5]. We focus specifically on the aspects of these models that address phonological planning, which we propose involves two key functional demands: (i) managing serial order, movement initiation and timing, and (ii) generating prosody and realizing syllable stress in the speech signal.

First, managing serial order involves coordinating segmental and metrical information. The Lexical Output Form in left pSTG activates the phonemic segments of a word, its metrical pattern, and their relative order. These are buffered in working memory: phonemic segments in the Segmental Content Buffer (left pIFS, IPS), and the weak-strong metrical frame in the Sequential Structure Buffer (preSMA). We propose that phonemic content in left pIFS is assigned to syllable positions in the metrical frame through rule-based interactions with preSMA, mediated by subcortical structures. Movement initiation and timing, in turn, depend on interactions between the Sequential Structure Buffer and motor planning regions. Contents of the Segmental Content Buffer are read out to left vPMC, where the most appropriate motor program(s) for the sequence is selected. Their initiation and intensity is gated by the Sequential Structure Buffer via the Initiation Map in SMA. Second, generating prosody involves coordination between the buffers and the frontal operculum region. These regions may jointly shape utterance prosodic contour, modulate intensity, duration, and pitch to realize syllable stress in the speech signal.

We plan to evaluate the framework through computer simulations of multisyllabic word production, comparing simulated speech acoustics with recordings of human speakers producing the same utterances under controlled conditions. By interfacing the models, we provide a neurobiologically grounded account of how prosodic specification such as lexical stress can shape motor output. This approach extends our understanding of how high-level linguistic planning is translated into continuous motor behavior, and bridges a key gap between psycholinguistics theory and speech motor control.

References

[1] Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–75. https://doi.org/10.1017/s0140525x99001776

[2] Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594–621. https://doi.org/10.1037/0033-295X.102.3.594

[3] Guenther, F. H. (2015). *Neural control of speech*. Cambridge, MA: MIT Press.

[4] Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48. https://doi.org/10.1016/j.cortex.2014.07.001

[5] Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, 22(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

Online and offline pronoun comprehension by German-speaking children and adults

Nevena Klobučar[a], Esther Rinke[b], Raffaella Folli[a], Chrisina Sevdali[a], and Juliana Gerard[a]
[a]Ulster University, [b]Goethe University Frankfurt
j.gerard@ulster.ac.uk

Background: In sentences like *John$_1$ likes him$_{*1}$/himself$_1$*, the antecedent to the pronoun *him* must be non-local, while the anaphor *himself* requires a local antecedent. However, in some languages, such as English and Dutch [1-2], children up to age 7 accept ungrammatical antecedents for pronouns referents, while correctly interpreting anaphors. The same asymmetry is observed online: pronouns seem harder to process for both adults and children [3]. German-speaking children, however, show fully adultlike comprehension of both pronouns and anaphors from the age of 4 years, in offline measures [4, 5]. This could mean (a) that pronouns and anaphors are equally easy to process in German, or (b) that the asymmetry does not emerge through offline tasks, similarly to English-speaking adults (i.e., with difficulty in processing but recovery before an offline interpretation) or (c) they might differ entirely, with difficulty online nor offline.

Method: 28 German-speaking children (4;0 to 6;8) and 33 adults participated in a simultaneous offline (Picture Verification Task) and online (eye-tracking) task. Participants were shown one picture at a time with two characters (e.g. Figure 1 or 2) and heard (1) an introduction of the characters, (2) an introduction of the main event taking place, and (3a/b) the experimental sentence. This contained either a pronoun (*sie/ihn*) or an anaphor (*sich*). We further manipulated the truth-value of the sentences (true/false). Hence, half the stimuli were true (i.e., sentence and picture matched), and the other half were false (picture and sentence mismatch). Fillers with no pronouns and anaphors were interspersed between experimental stimuli. Participants were asked to press a smiley face if they thought that the sentence matched the picture, or a sad face if it did not match.

Results: In the offline task, adults performed at ceiling with both true and false pronouns and anaphors (95%, 97%), while children's accuracy was lower overall with both pronouns and anaphors (58%, 61%), but no significant difference between the two (glmer, p=.51), nor true and false stimuli (p=.09). Online, however, **both adults and children show significant differences between conditions** (GAMM) [figures 3-5]. Within true stimuli and false stimuli, **pronouns yield fewer looks to Target (grammatical antecedent) than anaphors**.

Discussion: Online data shows an asymmetry in processing comparable to that observed in English-speaking children [3], as well as English and German-speaking adults. This asymmetry is not observed offline, which is in line with previous findings [4, 5]. However, we find low accuracy, which could be due to differences in experimental methodologies (judgment vs preference). The online results suggest that pronouns are more difficult to process independently of the language, but offline tasks do not fully capture this in all languages, confirming the need for more fine-grained measures (e.g., eye-tracking). The absence of a difference in the offline comprehension task suggests that German-speaking children recover from the pronoun-specific processing difficulty, contrary to what has been observed in English. Thus, while pronoun processing might be similar across languages, differences exist in children's acquisition of the relevant constraints on coreference (syntactic and/or extra-syntactic).

Stimuli example.

1. Hier sind ein Junge und ein Großvater. *(Here are a boy and a grandfather).*

2. Das passiert gleich: Mit einer Bürste bürsten. *(Something is happening: brushing with a brush).*

3a. Der Großvater bürstet ihn mit einer Bürste. *(The grandfather is brushing him…)*

3b. Der Großvater bürstet sich mit einer Bürste. *(The grandfather is brushing himself…)*

Figure 1. Grandfather brushing boy.

Figure 2. Grandfather brushing himself.

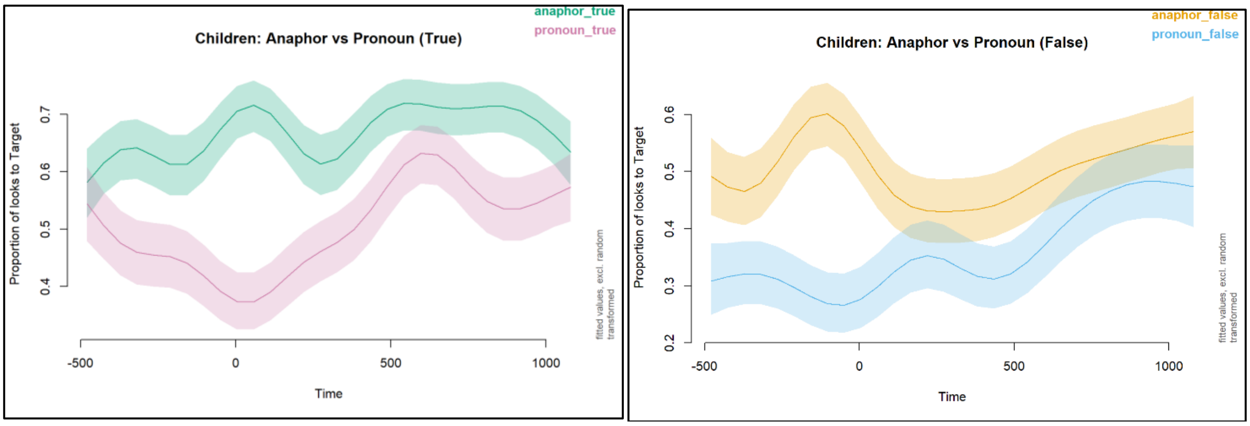


Figure 5. Children’s proportion of looks to Target across True and False conditions. Time 0 = critical word onset.

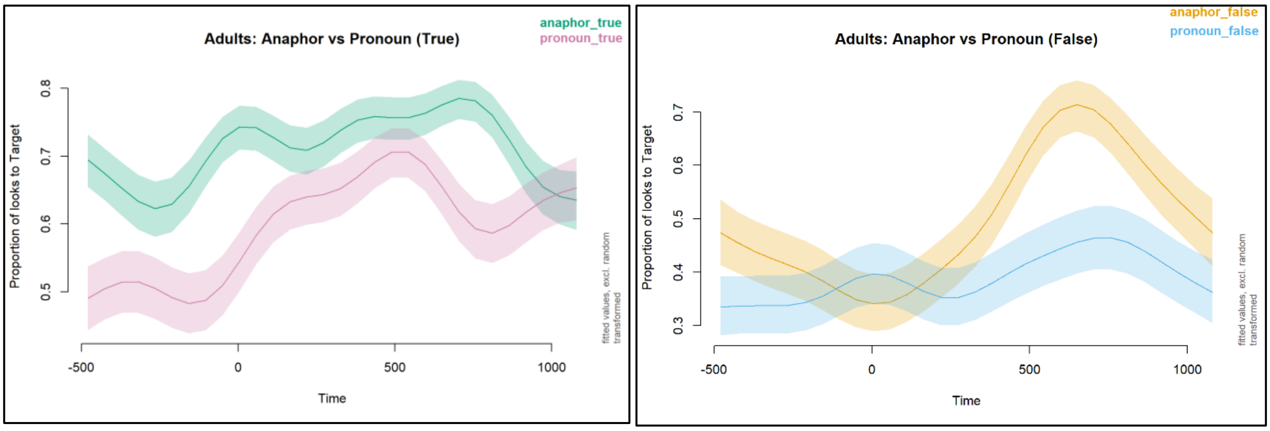


Figure 3. Adults proportion of looks to Target across True and False conditions. Time 0 = critical word onset.

References

[1] Chien, Y-C, & Wexler, K. (1990). Language Acquisition, 1 (3).

[2] Philip, W., & Coopmans, P. (1996). Proceedings of NELS 26, 241–255.

[3] Clackson, K., Felser, C., & Clahsen, H. (2011). Journal of Memory and Language, 65 (2): 128–144.

[4] Ruigendijk, E. (2008). In Gavarró, A.and Freitas, J., editors, Language Acquisition and Development: Proceedings of GALA 2007.

[5] Ruigendijk, E., Friedmann, N., Novogrodsky, R., & Balaban, N. (2010). Lingua,120(8):1991–2005.

Predicting Reading Comprehension from Eye-Tracking Measures with Random Forests

Diane Mézière[1], Titus von der Malsburg[2]
1 University of Turku; 2 University of Stuttgart
Corresponding author: Diane Mézière, diane.meziere@utu.fi

Background: In recent years, researchers have examined the possibility of using eye-tracking measures to predict performance on a variety of reading comprehension tasks[1–4]. Research has shown that it is possible to accurately predict performance on reading comprehension tasks, but that the relationship between eye movements and comprehension is mediated by the comprehension task being used[3,4]. In previous work, we used simple logistic regression to predict performance on reading comprehension assessments from eye tracking measures and found that the models could explain on average around 37% of the variance in comprehension scores. However, these models were limited to capturing monotonic effects, linear in log-odds, and no interactions and therefore may not have been capable of capturing important patterns in the data. Unlike the linear regression models used in previous work, random forests can account for interactions between predictors, handle predictor multicollinearity, and avoid overfitting even with small datasets – all of which are common challenges in typical eye-tracking datasets[6].In this study, we used random forests to predict performance on a range of reading comprehension tasks to examine whether this method leads to more accurate predictions compared to linear regression models.[5]

Method: We used random forests to predict reading comprehension scores across five reading comprehension tasks by re-analyzing two datasets from Mézière et al.[2,4] We examined two datasets that include eye movements and comprehension data. Dataset 1 contains data from 79 participants while they were administered three commonly-used standardized reading comprehension tests: The GORT, the YARC, and the sentence comprehension subtest of the WRAT. Dataset 2 contains data from 62 participants while they read passages of text silently according to two reading conditions: a recall condition, and a read-only condition in which they had no over comprehension task. We fit random forests for each comprehension task in two ways: once with data aggregated at the participant level, and once with data aggregated at the test item level, using eight canonical eye-movement measures, one scanpath measure, and reading speed as our predictors.

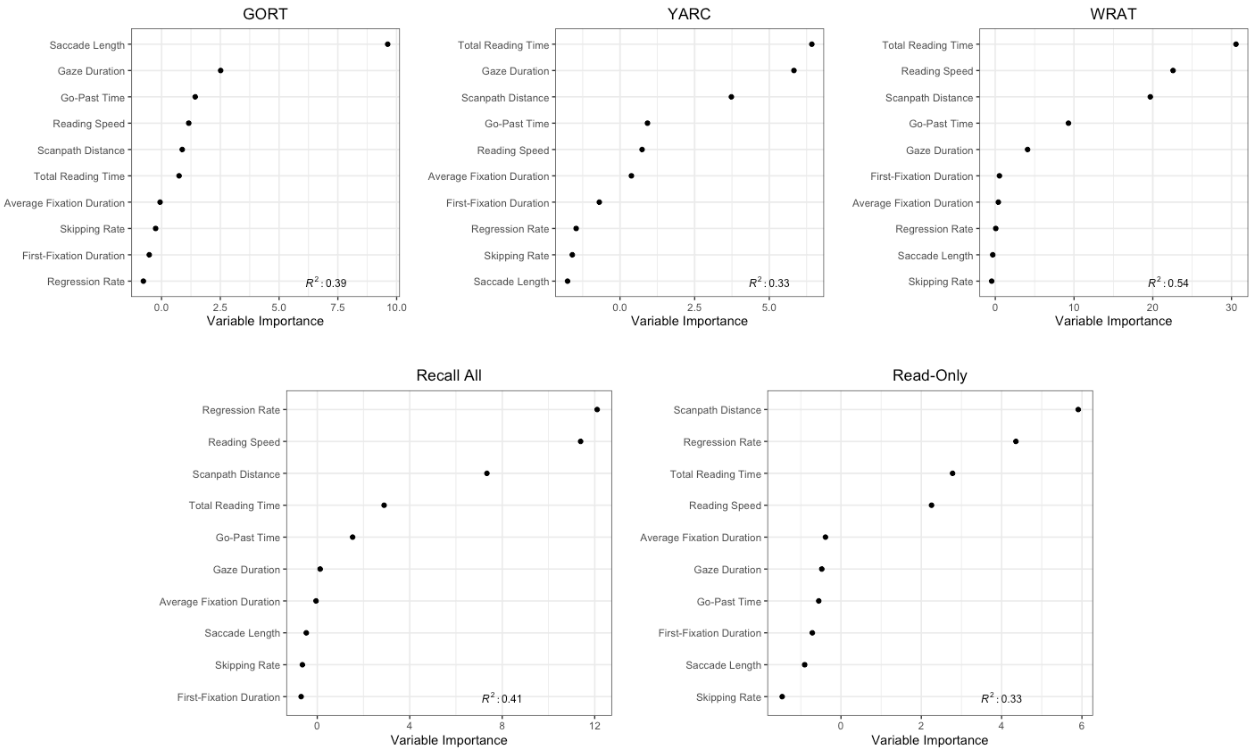
Results: Results from the random forest analyses are shown in Table 1 and Figure 1. Random forest showed only a small increase in R^2 for participant-level data (8% increase on average) but showed a significantly larger R^2 for the item-level analysis as they explained up to 70% of the variance in the data, and 53% of the variance in the data on average. Hence, the results are in line with previous work suggesting that random forests can make more accurate predictions compared to regression models. In addition, we found that, across reading tasks, the most important predictors were much more similar in the random forest analysis, as gaze duration, go-past time, and total reading time appeared as useful predictors for nearly all comprehension tasks.

Discussion: Overall, these results provide further support for the idea that eye movement measures can be used to predict performance on a variety of reading comprehension tasks. However, results are also in line with previous work showing that the predictive relationship between eye movements and reading comprehension depend at least in part on the type of reading comprehension task under scrutiny, as readers adapt their reading behaviour and cognitive processes to the task at hand.

Table 1. Variance Explained by Model Type and Analysis Level

	Regression Models	Random Forests	
	Participant-Level	Participant-Level	Item-Level
GORT	0.42	0.39	0.53
YARC	0.29	0.33	0.70
WRAT	0.46	0.54	0.49
RECALL	0.29	0.41	0.48
READ-ONLY	0.36	0.33	0.46
Average R^2	36.4%	40%	53.2%

Figure 1. Variable Importance per Comprehension Task



References

1. Copeland, L., & Gedeon, T. (2013). Measuring reading comprehension using eye movements.*CogInfoCom 2013.* 2. Mézière, Diane., C., Yu, L., Malsburg, T., Reichle, E. D., & McArthur, G. (2024). *Using Eye Movements from a 'Read-Only' Task to Predict Text Comprehension.* PsyArXiv. 3. Mézière, D., Yu, L., McArthur, G., Reichle, E., & von der Malsburg, T. (2023). Scanpath Regularity as an Index of Reading Comprehension. *SSR.* 4. Mézière, D., Yu, L., Reichle, E. D., Von Der Malsburg, T., & McArthur, G. (2023). Using Eye-Tracking Measures to Predict Reading Comprehension. *RRQ.* 5. Martínez-Gómez, P., & Aizawa, A. (2014). Recognition of understanding level and language skill using measurements of reading behavior. *IUI* 6. Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. *SSR.*

The Cost of L2 Fluency: Eye-Movement Evidence for Reduced L1 Reading Automaticity

Shiyu He[1], Petar Milin[2], Dagmar Divjak[2]

[1]University of Bristol [2]University of Birmingham

celia.shiyu.he@bristol.ac.uk

Background Word frequency and length are well-established predictors of reading fluency, in both L1 and L2, influencing eye-movement measures such as first fixation duration and total reading time [1, 2, 3]. However, the influence of these factors on reading fluency in bilinguals remains underexplored. This study investigates whether L2 English proficiency and exposure modulate the effects of word frequency and length on L1 Chinese.

Method Forty native Chinese speakers with university-level education were classified into Advanced and Upper Intermediate groups based on a Principal Component Analysis and non-hierarchical Cluster Analysis of English vocabulary size and IELTS scores, along with years of residence in the UK [4]. Participants read 24 short Chinese texts (38–71 words each, M = 55.9, SD = 10.3), drawn from online Chinese news articles, while their eye movements were recorded. This study builds on the eye-tracking data and analysis framework presented in He [5].

Results Linear mixed-effects models examined first fixation duration (FFD) and total reading time (TRT) at word N, as well as spill-over (N-1) and parafoveal preview (N+1) effects. Advanced L2 English readers showed significantly longer FFD and TRT than the Upper Intermediate group, particularly for low-frequency and longer words during L1 Chinese reading. These effects were strongest at the fixated word (N), with weaker but consistent influences in adjacent regions. Greater sensitivity to lexical properties in the Advanced group suggests reduced automaticity and efficiency in L1 word recognition [6].

Discussion Increased L2 proficiency and exposure may reduce L1 reading fluency, potentially due to decreased lexical automaticity or early stages of L1 attrition [7]. By showing how L2 proficiency and exposure reshape sensitivity to lexical properties in L1 reading, this study contributes to models of bilingual language processing across domains and supports the continuing need for more balanced bilingual education.

Conclusion: Our research contributes a pivotal perspective to bilingualism literature, indicating that L2 English proficiency and exposure can lead to subtle changes in L1 Chinese reading fluency. These insights compel a re-examination of bilingual reading models and pedagogical approaches, highlighting the potential cost of neglecting L1 fluency in the pursuit of advanced L2 skills.

References

[1] Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

[2] Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

[3] Gao, X.-Y., Li, M.-F., Chou, T.-L., & Wu, J.-T. (2016). Comparing the frequency effect between the lexical decision and naming tasks in Chinese. *JoVE (Journal of Visualized Experiments), (110)*, e53815. https://doi.org/10.3791/53815

[4] He, S., Divjak, D., & Milin, P. (2025). Optimising participant grouping methods in bilingualism studies: Insights from eye-tracking data. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.24019.he

[5] He, S. (2024). *Learning to optimise reading: An exploratory study of eye movement transfer across languages* [Doctoral dissertation, University of Birmingham].

[6] Lahoud, H., Share, D. L., & Shechter, A. (2023). A developmental study of eye movements in Hebrew word reading: The effects of word familiarity, word length, and reading proficiency. *Frontiers in Psychology, 14*, 1052755. https://doi.org/10.3389/fpsyg.2023.1052755

[7] Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism, 10*(3), 210–223. https://doi.org/10.2167/beb441.0

What Word-Guessing Reveals About Your Brain:
Patterns of Lexical Storage and Processing

Kamila Kuishibekova[1], Valentina Apresyan[2]
[1] Nazarbayev University; [2] Dartmouth College
kamila.kuishibekova@nu.edu.kz, Valentina.Yuryevna.Apresyan@dartmouth.edu

Our study investigates how lexical information is strategically retrieved and communicated under cognitive pressure, using the word-guessing game Alias to test competing theories of semantic representation. Attributional theories (Collins & Quillian, 1969) propose that words are organized by shared semantic features, while distributional theories (Harris, 1954) argue that meaning arises from statistical patterns of co-occurrence.

We used a naturalistic, game-based approach (see Allen et al., 2024), recording speaker-listener pairs playing the Russian version of Alias. While not mirroring all aspects of natural language, this goal-oriented task provides a window into how speakers retrieve information from their mental lexicon and comprehend clues in a communicative setting. In the game, the speaker receives a card with a list of 8 target words and must provide clues to help their partner guess them within 1 minute, free to shuffle the words within a card and skip some of them. 2 games, each with 4 fluent Russian speakers (aged 18-25), were conducted. The stimuli comprised 96 words, balanced across grammatical categories (32 nouns, 32 verbs, 32 adjectives), concreteness (48 concrete, e.g., *jacket*, 48 abstract, e.g., *inflation*), and frequency. We recorded 180 explanations (link to the dataset), annotated for retrieval strategies: definitions (73 cases), real-life examples (49), collocations (23), synonyms (16), antonyms (8), hyponyms (7), co-hyponyms (3), and hypernyms (1). Strategy choice did not vary with concreteness but did depend on part of speech and frequency. Lower-frequency words were more often explained through definitions, while adjectives, unlike nouns or verbs, were more frequently explained using antonyms and synonyms.

Success rates varied by strategy and word properties. The most effective strategies were co-hyponymy (100 % success rate), antonymy and collocations (75%). Synonymy was the least effective (31%), while real-life examples and definitions showed moderate success (56-57%.) Higher-frequency words were generally more likely to be guessed correctly, though the role of frequency differed by strategy. The frequency gap between guessed and not guessed words was smaller for definitions, real-life examples, and collocations, but larger for semantic relation-based strategies (e.g., synonyms, antonyms) (See Table 1.)

These findings support hybrid models of the mental lexicon (Andrews et al., 2014; Riordan & Jones, 2011). In lexical production, there was a slight preference for attributional strategies (definitions, semantic relations like antonyms, synonyms, hyponyms), though distributional strategies (collocations, real-life examples) were also prominent. On the other hand, in comprehension, both attributional (e.g., antonyms) and distributional (e.g., collocations) strategies were highly effective (See Fig. 1). The low success rate of synonyms aligns with lexical access theories, where higher-frequency competitors dominate. Our results suggest that mental representations reflect both attributional and distributional patterns, influenced by lexical properties (part of speech, frequency) and individual experience. The limited success of real-life examples highlights variation in personal knowledge. This study makes three contributions: (1) introducing a naturalistic gameplay method, (2) providing evidence for hybrid semantic organization, and (3) showing that word-level features modulate retrieval. Limitations include sample size and the focus on Russian only; future research should broaden language scope and apply neuroimaging to examine neural mechanisms.

Figure 1. *Success rate by strategy and word properties (POS, Concreteness)*

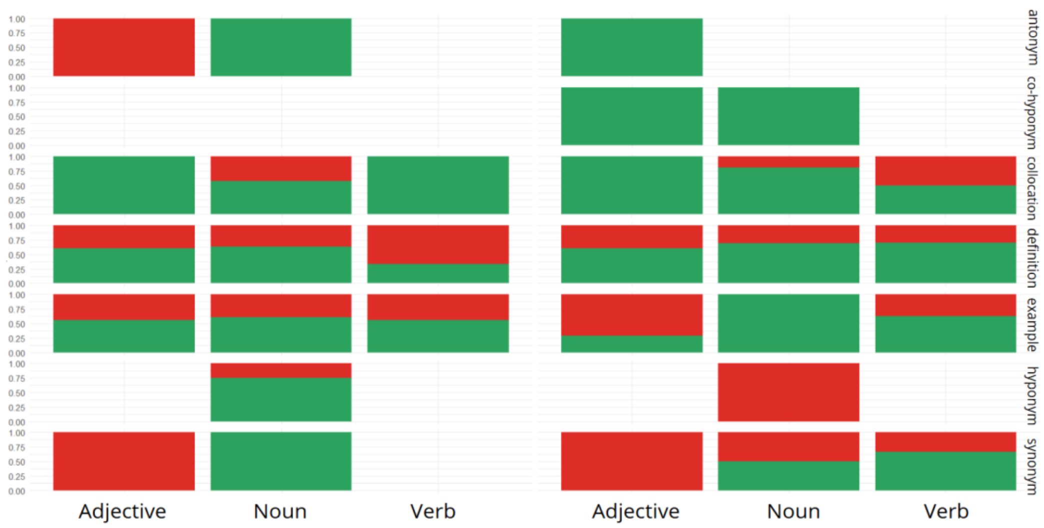


Table 1. *The mean and median frequencies of guessed and not guessed words across strategies*

Production Strategy	#	Guessing Success	Mean frequency			Median frequency		
			Overall	Not guessed	Guessed	Overall	Not guessed	Guessed
Antonymy	8	6 (75%)	29537	`20838	32437	26900	20838	29514
Synonymy	16	5 (31%)	19087	14971	34935	13251	12173	26172
Co-hyponymy	3	3 (100%)	29513	---	29513	32775	---	32775
Hyponymy	7	3 (43%)	15014	13892	21472	8931	8931	21472
Hypernym	1	0 (0%)	29138	29138	---	29138	29138	---
Collocation	23	17 (74%)	16867	14184	15730	10375	15135	6340
Definition	73	41 (56%)	10187	7663	9054	5658	6741	8568
Example	49	28 (57%)	13681	13036	11376	6428	6785	4914

References

[1] Allen, K., Brändle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., ... & Schulz, E. (2024). Using games to understand the mind. *Nature Human Behaviour*, 1-9.

[2] Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in cognitive science*, *6*(3), 359-370.

[3] Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240-247.

[4] Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146-162.

[5] Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature- based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303-345.

Interference Triggered by Syntactic and Semantic Similarities in L1 and L2 Japanese

Yoko Nakano[1], Chunxia Hu[2], Atsushi Yuhaku[3]
[1&2] Graduate School of Kwansei Gakuin University; [3] Ritsumeikan University
[1] y-k.nakano@kwansei.ac.jp, [2] HuCX@kwansei.ac.jp, [3] yuhaku@fc.ritsumei.ac.jp

Background: The cue-based retrieval model [1] predicts that when a sentence contains a constituent (i.e., attractor) similar to the argument (i.e., target) of a head (e.g., a verb), it causes interference; and that different information types (e.g., syntactic and semantic) also trigger interference. Some studies have reported facilitatory effects in several experiments with native speakers (NSs) of a language—*facilitatory interference effects* [2, 3]. On the other hand, a hypothesis for the second language (L2), the shallow structure hypothesis (SSH [4]), assumes that unlike NSs non-native speakers (NNSs) tend to rely more heavily on surface-level and semantic information than on abstract hierarchical syntactic information. Consequently, it is predicted that NNSs will show facilitatory interference effects elicited by semantic similarity, but weaker or no such effects elicited by syntactic similarity. A different hypothesis, [5], argues that NNSs are more susceptible to similarity-based memory interference than NSs. Therefore, it predicts that NNSs show larger interference effects than NSs. The predictions are mixed, and the sentence structures and tasks used in previous studies differed between syntactic and semantic conditions. We tested these hypotheses through an eye-tracking experiment with the sentences that had the same structure for both syntactic and semantic similarities in L1 and L2 Japanese.

Method: Thirty-seven Japanese NSs and 28 advanced, L1 Chinese-speaking Japanese learners participated. We constructed 48 sets of 8 Japanese sentences, in which the congruencies and incongruencies of targets and attractors with the head were syntactically and semantically manipulated, e.g., (1) and (2). In the experiment, a sentence was presented on a PC monitor, followed by a content question. Participants' eye movements were recorded by Eyelink 1000 Plus. Data were analyzed with LME models with Information (syntactic vs. semantic), Language (NS vs. NNS), Target (congruent vs. incongruent), and Attractor (congruent vs. incongruent) as fixed factors, and dwell times per character as dependent variable.

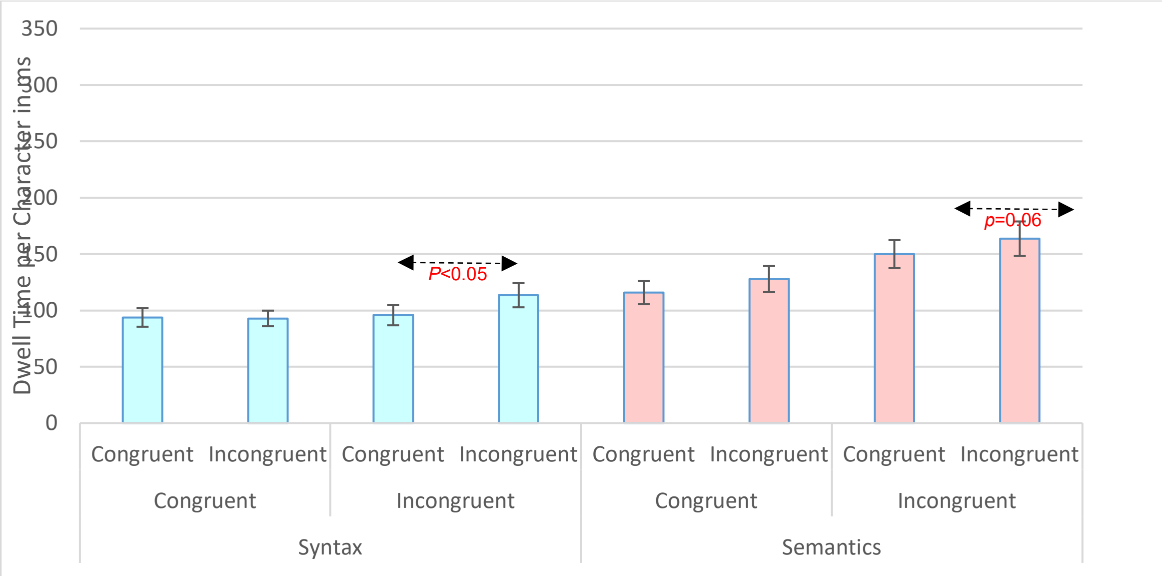
Results: The main effects of language and information, and the interaction of information, language and target were significant. Further analyses with ls-means indicated that NSs read sentences faster than NNSs; that dwell times were longer for incongruent targets than for congruent targets; and that for incongruent targets dwell times were shorter for congruent attractors than for incongruent attractors. These results indicate facilitatory interference. Further, the facilitatory interference effects were observed only in syntactic conditions but not in semantic conditions for NSs, and they were observed only in semantic conditions but not in syntactic conditions for NNSs.

Discussion: The different results between NSs and NNSs seem to be compatible with the SSH. But the difference might have been caused by the simple structure of the stimulus sentences, which allowed the NSs to check syntactic and semantic features in the very early stage. Our NNS participants' L1 was Chinese, whose sentences are written only in *kanji*. In Japanese, content words are written in *kanji,* and affixes are written in *kana*. Therefore, it is conceivable that the L1-speaking Chinese participants could easily spot *kanji*-written content words and thus rely more on semantic information.

To conclude, NSs and NNSs may differ in how they recognize content words in a sentence, which could influence the different use of syntactic and semantic information.

Figure 1

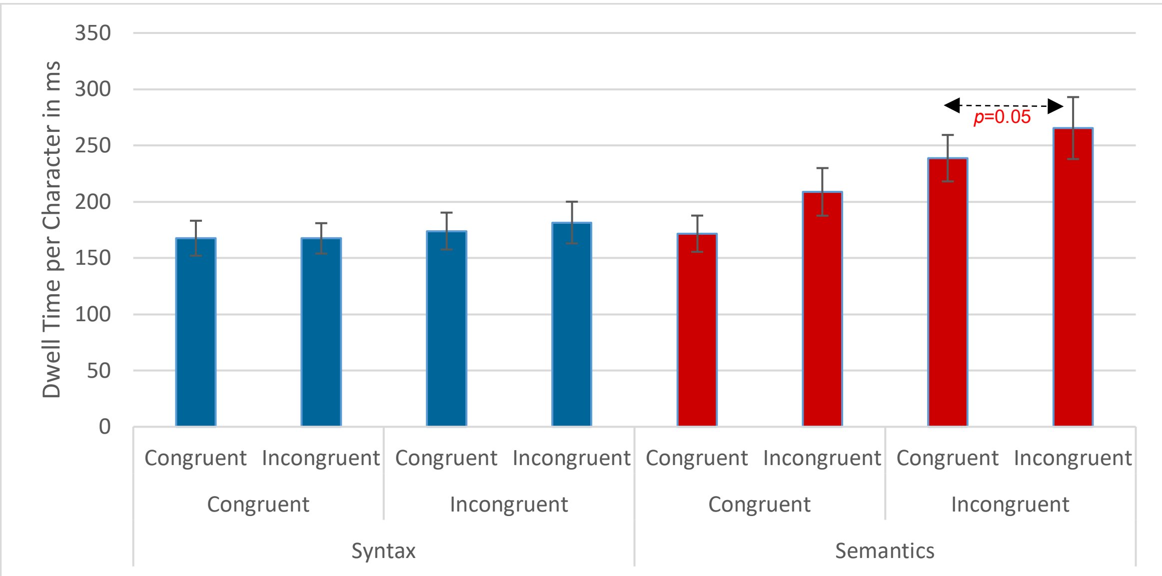
Native Speakers' Dwell Times per Character in ms



Error bars indicate 95% confidence intervals.

Figure 2

Non-Native Speakers' Dwell Times per Character in ms



Error bars indicate 95% confidence intervals.

[1] Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences, 10*(10), 447-454.

[2] Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206-237.

[3] Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language, 102*, 16-27.

[4] Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*(1), 107-126.

[5] Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism*: *Language and Cognition, 20*(4), 659-678.

Sample Stimulus Sentences:

For syntactic condition, in Japanese grammatical sentences that contain honorific expressions require that the subject (target) and the verb are marked with honorific affixes and syntactically agree, sharing the feature [+honorable] as in (1a & b). In (1c & d), the subject (target) is not marked with honorific affix and has no [+honorable], it is syntactically incongruent with the verb (1c & d). And the NP (attractor) that modifies the object is syntactically congruent with the verb when it has [+honorable] (1a & c) but it is syntactically incongruent with the verb when it has no [+honorable] (1b & d). For semantic condition, when the object (target) holds the feature [+openable], it is semantically congruent with the verb (head) (1a & b), but when it has no [+openable] feature, it is semantically incongruent with the head (1c & d). And when the NP (attractor) that modifies the object holds [+openable], it is semantically congruent with the verb (2a & c), but when it has no [+openable] it is semantically incongruent with the verb (2b & d).

(1) Syntactic conditions:

NP1 (target)-nom NP2 (attractor)-gen NP3-o o-Verb-ni naru …

a. Tanaka-sensei-ga Ueda-sensei-no tana-o o-ake-ni-natte mita-youda.
Tanaka-honorific([+h])-nom Ueda-honorific ([+h])-gen cabinet-acc open-honorific-([+h]) tried seemed

b. Tanaka-sensei-ga Ueda-no tana-o o-ake-ni-natte mita-youda.
Tanaka-honorific([+h])-nom Ueda([-h])-gen cabinet-acc open-honorific-([+h]) tried seemed

c. ??/*Tanaka-ga Ueda-sensei-no tana-o o-ake-ni-natte([+h]) mita-youda.
Tanaka([-h])-nom Ueda-honorific ([+h])-gen cabinet-acc open-honorific-([+h]) tried seemed

d. ??/*Tanaka-ga Ueda-no tana-o o-ake-ni-natte mita-youda.
Tanaka([-h])-nom Ueda([-h])-gen cabinet-acc open-honorific-([+h]) tried seemed

#[+h]: holding an “honorable” feature, [-h]: holding no “honorable” feature.

(2) Semantic conditions:

NP1-nom NP2 (attractor)-gen NP3 (target)-o Verb

a. Tanaka-ga shorui-no tana-o akete mita-youda.
Tanaka-nom document ([+opn])-gen cabinet([+opn])-acc open ([+opn]) tried seemed

b. Tanaka-ga shokki-no tana-o akete mita-youda.
Tanaka-nom dish(es) ([-opn])-gen cabinet ([+opn])-acc open ([+opn]) tried seemed

c. ??/*Tanaka-ga shorui-no size-o akete mita-youda.
Tanaka-nom document ([+opn])-gen size ([-opn])-acc open ([+opn]) tried seemed

d. ??/*Tanaka-ga shokki-no size-o akete mita-youda.
Tanaka-nom dish(es) ([-opn])-gen size ([-opn])-acc open ([+opn]) tried seemed

“Tanaka seemed to have tried to open the cabinet (openable) / size (unopenable) / of documents (openable) / dishes (unopenable).”

#[+opn]: holding an “openable” feature, [-opn]: holding no “openable” feature.

The causal chain in English conditionals depends on event structure

Elena Marx, Zofia Kordas, Hannah Grobauer, Eva Wittenberg*
Central European University, Vienna
*wittenberge@ceu.edu

People effortlessly understand cause and effect in sentences like “If it rains today, the street gets wet”: Rain causes wetness – an example of the common inference from antecedent to consequent [e.g., 1, 2]. People also understand that in sentences like “If you’re hungry, the biscuits are on the shelf”, hunger does not cause biscuits to appear [e.g., 3]. However, as researchers, we still struggle to understand what makes conditional sentences either a classic conditional or a biscuit conditional, or what reverses the common causal inference, such that the consequent is a cause for the antecedent [4].

Here, we propose, similar to [5], that a crucial element in turning the common causal inference (antecedent-causes-consequent) into its reverse is dynamicity – specifically, whether the consequent is a dynamic event or a static state. Previous research on temporal inference has shown that people expect states to happen before events (states-before-events inference): Example (1a) evokes a linear order interpretation where marriage precedes pregnancy, but (1b), which frames the pregnancy as a state, elicits a pregnancy-before-marriage interpretation [e.g., 6, 7, 8]. Since causes always happen before effects, we would expect that a robust state-before-events inference will also result in more consequent-before-antecedent interpretations in conditional sentences. We tested this prediction in two pre-registered forced-choice experiments.

Exp. 1: 100 English native speakers read conditional sentences like (2a/b) and then answered the question “What comes first?” in a two-alternative forced choice task: antecedent (e.g., *makeup*) or consequent (e.g. *pimple*; n=20; 30 fillers/controls, Latin-square design, Fig. 1A). Results (both Exps.: mixed-effects logistic regression models) are shown in Fig.2A: The common inference (antecedent first) prevailed when the consequent encoded a dynamic event, but the reverse (consequent first) gained ground when the consequent encoded a static state (β=1.60, z=15.95, p<0.001).

Exp. 2: We used the same design and stimuli as Exp.1 (N=100, n=20; 30 fillers/controls, Latin-square design, Fig. 1B) but asked “What does this mean?”, with either the classic antecedent-causes-consequent (e.g., *makeup causes pimple*) or the reverse consequent-causes-antecedent (e.g., *pimple causes makeup*) as choice options. Fig.2B shows that participants reliably interpreted sentences as antecedent-causes-consequent when the consequent encoded a dynamic event, but they were more likely to reverse causality when the consequent encoded a static state (β=1.20, z=16.53, p<0.001).

Discussion: These results contribute to two topics: First, we extend findings on temporal inferences through dynamicity [e.g., 6, 7, 8] to the causal domain, showing that the state-before-event inference impacts causal reasoning. Specifically, we show that dynamicity encoding affects the processing of conditionals, turning common antecedent-causes-consequent interpretations into the reverse. And second, we can explain this effect by tying together Figure-Ground explanations and the semantics of predicates, in which stative predicates do not carry causal meaning, whereas dynamic predicates do [e.g., 5].

EXAMPLES:

(1) a. Mary got married to John. She got pregnant.

b. Mary got married to John. She was pregnant.

(2) a. If Daria uses makeup, she gets a pimple.

b. If Daria uses makeup, she has a pimple.

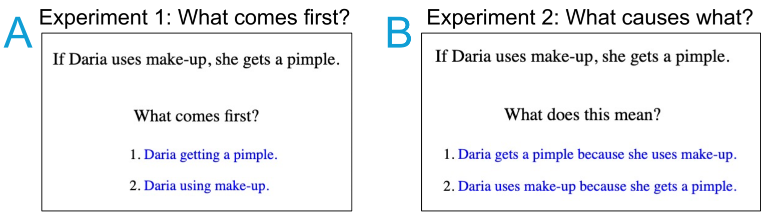


Fig.1: *Example trial of the forced-choice task (A: Exp.1, B: Exp. 2). Performance was at ceiling for filler and control trials which checked for causal and temporal reasoning in non-conditional sentences.*

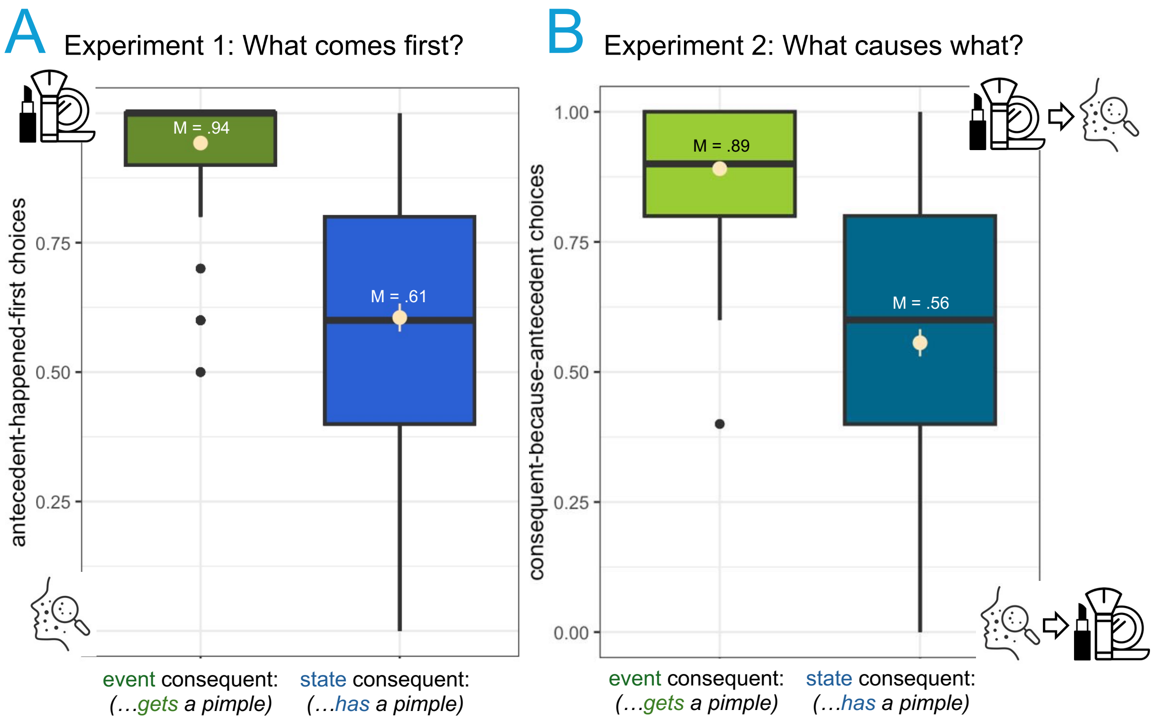


Fig.2: *Results (in proportions) for **A**. Exp. 1 and **B**. Exp. 2, depending on dynamicity type in the consequent. Higher boxes indicate more common antecedent-causes-consequent interpretations, lower boxes indicate more reverse interpretations. In green-tinted bars, the consequent is encoded as event, in blue-tinted bars, it is encoded as a state. Yellow dots and whiskers represent distribution of means and Standard Errors per participant, and black horizontal lines medians. Black dots are outlier participant means.*

References: [1] Johnson-Laird & Byrne, *Psychological Review*, 2002. [2] Bhatt & Pancheva, *Conditionals.* The Blackwell companion to Syntax, 2006. [3] Franke, in *16th Amsterdam Colloquium*. 2007. [4] Lassiter, *Mind & Language*, 2023. [5] Copley, in *Conditionals: Logic, Linguistics and Psychology*. 2023. [6] Marx & Wittenberg, *Glossa Psycholinguistics*, 2024. [7] Marx & Wittenberg, *Cognition*, 2025. [8] Marx et al., *Cognitive Science*, 2025.

The effect of the focus particle 'only' on discourse expectations and discourse marking

Katharina Spalek[1], Merel Scholman[2], Vera Demberg[3]
[1] Heinrich-Heine-Universität Düsseldorf; [2] Universiteit Utrecht; [3] Universität des Saarlandes
katharina.spalek@hhu.de

Background: We investigated the effect of the focus particle "only" on discourse expectations and discourse marking in German and English. In keeping with previous findings [1, 2], we expect *only* to increase reference to focus alternatives in the subsequent discourse (H1). Based on the assumption that *only* keeps alternatives active (creating a need to justify why they don't occur) and previous findings that non-connective cues can trigger expectations about discourse relations and content [3], we hypothesize that *only* serves as a cue for causal discourse relations. Specifically, we expect more causal relations to follow *only* than in the control condition (H2a). Given that we expect this causality based on the need to explain why the alternatives were rejected, we also predict that the increase in causal discourse relations should occur specifically when the alternatives are referred to in the continuation of the discourse (H2b). Since discourse relations are less likely to be marked explicitly when they are highly expected, we hypothesize that fewer connectors will be used for causal relations in the presence of the cue *only* (H3). Finally, we expect these effects to arise both for English "only" and for the German equivalent "nur" (H4).

Method: *Participants.* 32 German speaking participants (18 female, mean age 34) took part in Exp. 1 and 33 English speaking participants with residence in the UK or Ireland (17 female, mean age 39) in Exp. 2.

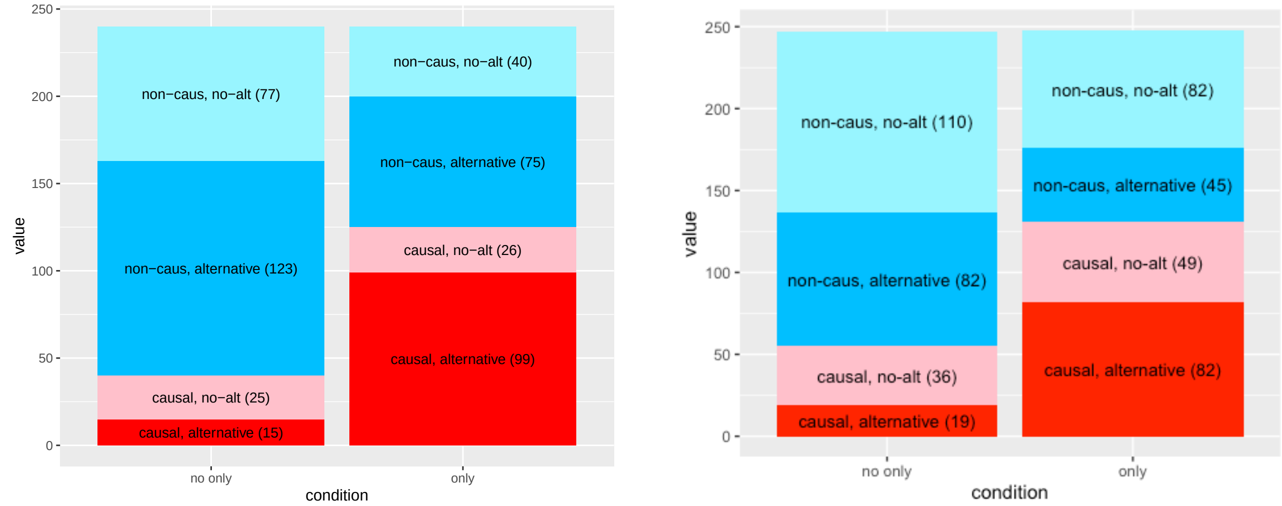
Materials. 30 experimental items and 10 fillers were created in German (Exp. 1) and English (Exp. 2), following the structure in [1]. The crucial manipulation was presence/absence of *only* in the third sentence.

1. 'Charles reached into a basket with peaches, cherries and bananas. He considered what he would like to eat. He (only) took out the peaches.'

Procedure. Participants, recruited via Prolific, wrote two complete and grammatical sentences to continue each of the written stories. These completions were annotated with discourse relation labels [PTDB, 4].

Results and Discussion: All analyses were carried out with generalized linear mixed effects models. H1 was confirmed by a significant increase of alternatives after *only*, in German ($\beta = 0.76, z = 3.53, p < .001$) and English ($\beta = 0.60, z = 2.28, p < .05$). H2 was also confirmed by a main effect of *only* on the number of causal continuations (H2a) which interacted with the presence of alternatives in the continuation (H2b), in German ($\beta = 0.72, z = 1.97, p < .05$, interaction: $\beta = 1.14, z = 2.95, p < .01$) and English ($\beta = 0.85, z = 2.94, p < .01$, interaction: $\beta = 1.40, z = 3.11, p < .01$). H3 could only be investigated when pooling the data of both Experiments, due to very low numbers of explicit markings. A model with marking of the relation (implicit / explicit) as the response variable and presence of *only* and presence of an alternative in participants' continuations as predictors, as well as their interaction, showed a significant interaction ($\beta = 2.46, z = 2.33, p < .05$), indicating that the causal relation is more likely to be implicit when *only* is present in the first argument of the discourse relation and its second argument mentions the alternatives, thereby confirming hypothesis H3. Finally, the results for the German and English participants showed similar effects, confirming H4. These results suggest that focus-sensitive particles can be interpreted as cues for discourse relations. These findings provide insight into the interface between sentence-level meaning and discourse-level structure.

This page serves for Tables, Figures, and References.



(a) Distribution of causal relations based on the presence or absence of *only* in Exp. 1

(b) Distribution of causal relations based on the presence or absence of *only* in Exp. 2

Figure 1: Comparison of causal relation distributions in German (Exp. 1) and English (Exp. 2) with or without *only*.

References

[1] Gotzner, N., & Spalek, K. (2022). Expectations about upcoming discourse referents: Effects of pitch accents and focus particles in german language production. *International Review of Pragmatics*, *14*(1), 77–94.

[2] Spalek, K., & Zeldes, A. (2017). Converging evidence for the relevance of alternative sets: Data from nps with focus sensitive particles in german. *Language and Cognition*, *9*(1), 24–51.

[3] Scholman, M. C., Rohde, H., & Demberg, V. (2017). “on the one hand” as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, *97*, 47–60.

[4] Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*, *35*, 108.

The good-enough listener: A visual world paradigm reveals the interaction between prediction and bottom-up input

Jiaxuan Li*[1], Kayla Keyue Chen*[2], Anne Wang[2], Yuhan Shen[2], Yijia Luo[2], Richard Futrell[1], Wing-Yee Chow[2]
[1] University of California Irvine; [2] University College London
jiaxul19@uci.edu, keyue.chen.19@ucl.ac.uk (* = equal contribution)

Introduction: Efficient language comprehension involves anticipating upcoming input, but unexpected input can be costly to process. Studies show that this processing cost is reduced when the unexpected word is phonologically related to the predicted word, compared to when it is unrelated [1,2]. The mechanisms behind this facilitation remain unclear. According to the Spreading Activation Hypothesis, activation of a predicted word may spread to phonologically related alternatives, facilitating the retrieval of unexpected but related input [3,4]. In contrast, the Misperception Hypothesis suggests that high phonological overlap between the predicted and unexpected word may lead comprehenders to retrieve the predicted word instead of the actual unexpected input, forming a “good-enough” representation [5,6]. The present study tested these accounts by examining responses to phonologically related versus unrelated unexpected words in a visual-world paradigm. We found evidence for the Misperception Hypothesis.

Method: Participants (n = 44) listened to Chinese sentences while viewing a display of six Chinese words (Fig. 1). The sentences (e.g., (1)) strongly predicted a specific word (“book” [书 shu1]) but continued with an unexpected, anomalous word that was either phonologically related to the predicted word through rhyme overlap (“pig” [猪 zhu1]; Experimental condition) or unrelated (“cotton” [棉 mian2]; Control condition). The visual display contained the three critical words and three unrelated distractors (Fig. 1). The target word in the Experimental condition was used as the Control in another sentence item, ensuring that the same set of words were used across both conditions. The Spreading Activation Hypothesis predicts facilitated recognition of phonologically related words and thus *more* fixations on the target in the Experimental condition than in the Control condition. In contrast, according to the Misperception Hypothesis, listeners are expected to persist with their top-down prediction until sufficient evidence overrides it, so it predicts *fewer* fixations on the target in the Experimental condition than in the Control condition.

Results: More fixation on the predicted word prior to target word onset indicated prediction based on prior contextual information. Following target word onset, participants shifted their gaze away from the predicted word towards the target. Critically, this shift was *slower* in the Experimental condition than in the Control condition (Fig. 2). A cluster-based permutation test was conducted in the post-target-onset (0-1500 ms) time window, using the log ratio of fixations on the target versus the predicted word as the dependent variable. The analysis revealed a significant main effect of condition in the 300-1500 ms interval (red segment in Fig. 3; cluster mass = 406, $p < .001$), with *lower* log ratios observed in the Experimental condition than in the Control condition. No significant differences in log ratios between conditions were found prior to target onset.

Conclusion: Using a visual world paradigm, we found that listeners were less likely to fixate on the target (and more likely to fixate on the predicted word) when the presented target was phonologically related to the predicted word than when it was unrelated. This finding suggests that strong top-down predictions can override bottom-up input, leading to inaccurate representations when the input is confusable with a more predictable alternative.

(1)

阅览室的桌子上摆着一些……

On the reading room's table are placed some……

Non-presented predicted word: book [书 shu1]

Experiment condition: pig [猪 zhu1] (phonologically related)

Control condition: cotton [棉 mian2] (unrelated)

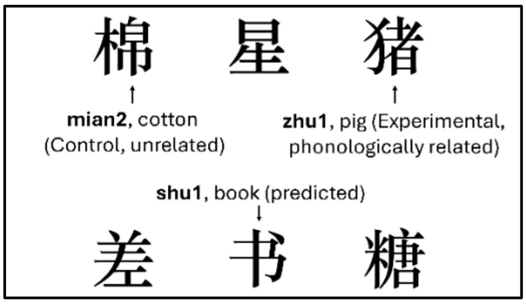
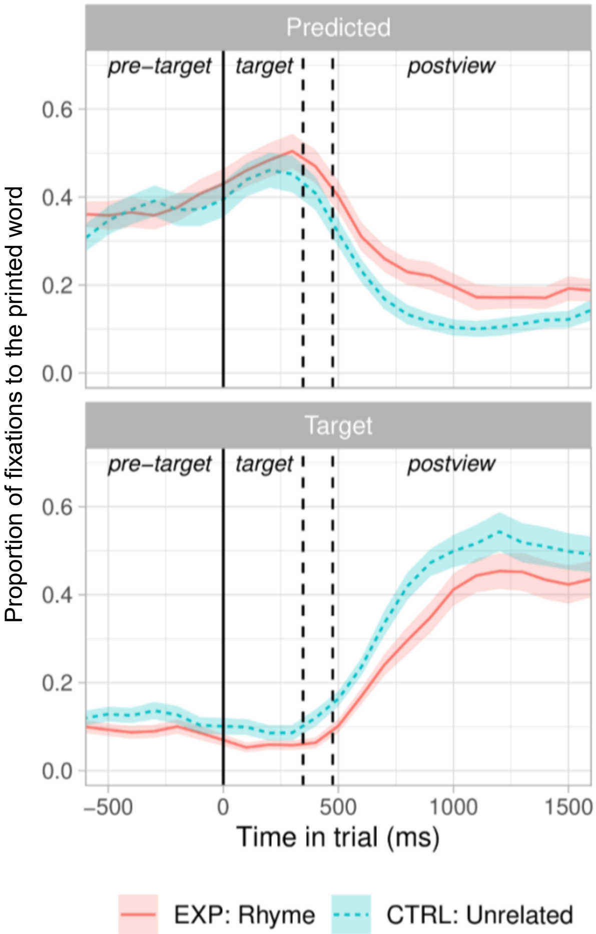


Fig. 1. Sample visual display.



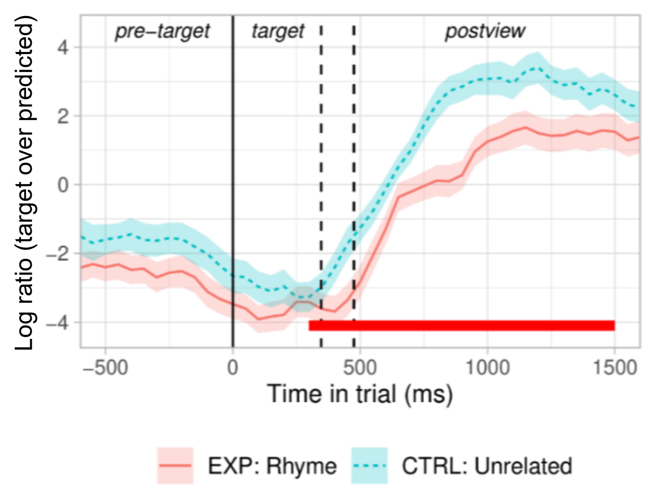


Fig. 3. **Log ratio** of fixations to the target versus the predicted word. The red segment indicates a significant cluster showing a main effect of condition in the cluster-based permutation tests.

Fig. 2. **Proportion** of fixations to the printed words. The upper panel shows fixations to the predicted word; the lower panel shows fixations to the target word. The solid line indicates target word onset. The first dashed line marks the average target word offset; the second dashed line marks the average sentence offset. Shaded areas represent standard errors of the mean.

References

[1] Kamide, Y. (2008). Anticipatory Processes in Sentence Processing: Anticipatory Processes in Sentence Processing. *Language and Linguistics Compass*, *2*(4), 647–670.

[2] Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311–327.

[3] Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495.

[4] Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2153–2162.

[5] Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological review*, *113*(2), 327.

[6] Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and linguistics compass, 1*(1‐2), 71-83.

Cue weighting in prediction: context and classifier effects in English-Chinese bilinguals

Danning Sun, Aine Ito
National University of Singapore
danning_sun@u.nus.edu

People often encounter multiple cues that guide or constrain predictions [1], yet how these cues jointly contribute to prediction remains unclear. While second language processing often relies on broad semantic context [2], early bilinguals' non-dominant language processing may differ, as early language exposure can bias the way how cues are weighed for optimal prediction mechanisms [3]. In this regard, classifiers (CL) in Chinese provide an interesting test case because early bilinguals may rely more on CLs than contextual cues due to CLs' stronger semantic constraints to nouns they modify [4]. To better understand cue weighting in prediction, this study uses the visual-world eye-tracking paradigm and investigates how early English-Chinese bilinguals use contextual constraints and CL for prediction in their non-dominant language (Chinese) when these cues align or conflict, and whether cue proximity (e.g., cue-target distance) affects prediction.

We will recruit 60 English-Chinese bilinguals who speak Mandarin Chinese as a non-dominant language. They will listen to 24 sentences with two predictability types (i.e., highly predictable and less predictable) in Chinese (Table 1) while viewing four images. All sentences include two predictive cues (i.e. context and a CL) with the CL appearing either late or early. In the CL late condition, the contextual constraints (relative clause; RC) precede the CL (i.e. *this is*+ RC+ numeral+ **CL** +noun). In the CL early condition, the CL precedes the contextual constraints (RC) (i.e. *this is*+ numeral+ **CL**+RC+ noun). Each sentence is paired with a scene depicting a highly predictable object, a less predictable object, and two distractors (Fig.1). The context is always constraining towards the highly predictable object (e.g., coffee). The CL is only compatible with one of the two critical objects. We predict that in the CL early condition, participants can generate immediate and strong predictions based on the classifier and will maintain their preference for the target object even when the context suggests the alternative. In the CL late condition, we expect that the classifier will trigger rapid changes in eye movements if participants use classifier to confirm their initial prediction (when it aligns with the preceding context) as well as to revise prediction (when the two cues conflict).

Our offline plausibility rating test asked 40 English-Chinese early bilinguals to read the sentences up to and before the sentence-final target word (e.g., before the word "coffee" as in Table 1) and view a corresponding scene (Fig.1). They were asked to rate (from 0-100) how plausible each object was to be mentioned following the given sentence fragments. The results of the plausibility test (Fig.2) show that the target object was rated as less plausible and the competitor object more plausible when the CL appeared early vs. late. This indicates that participants assign greater weight to cues that are linearly closer to the target noun, regardless of cue type. Based on these pre-test results, this planned study thus is interested in whether cue proximity also affects fixation patterns in the eye-tracking experiment.

We hypothesize that in real-time language processing, CL may serve as more reliable cues and guide predictions more effectively than context, due to their strong and clear constraint on the target noun. CL position can affect predictions, with early CLs facilitating immediate and robust predictions, and late CLs leading to prediction confirmation or revision. The eye-tracking data are expected to reveal how early bilinguals flexibly adjust cue weighting to optimize sentence comprehension.

Table 1 Example of critical sentence and experimental design

Sentence where the target is highly predictable from the context (the two cues align with each other)								
CL late	这是	小明	在	星巴克	买的	**一杯**	美味的	咖啡。
	This is	Xiaoming	at	Starbucks	bought	CL-bei	tasty	coffee.
	This is a cup of tasty coffee that Xiaoming bought at Starbucks.							
CL early	这是	**一杯**	小明	在	星巴克	买的	美味的	咖啡。
	This is	CL-bei	Xiaoming	at	Starbucks	bought	tasty	coffee.
	This is a cup of tasty coffee that Xiaoming bought at Starbucks.							
Sentence where the target is less predictable from the context (the two cues conflict)								
CL late	这是	小明	在	星巴克	买的	**一块**	美味的	蛋糕。
	This is	Xiaoming	at	Starbucks	bought	CL-kuai	tasty	cake.
	This is a slice of tasty cake that Xiaoming bought at Starbucks.							
CL early	这是	**一块**	小明	在	星巴克	买的	美味的	蛋糕。
	This is	CL-kuai	Xiaoming	at	Starbucks	bought	tasty	cake.
	This is a slice of tasty cake that Xiaoming bought at Starbucks.							

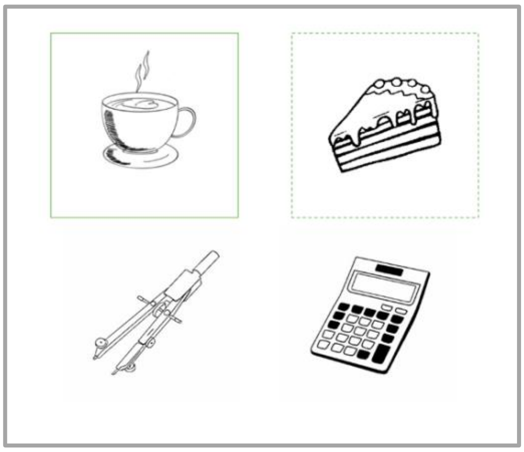


Figure 1 | An example of the visual stimuli. The two objects in the green boxes fit the contextual constraints, and the distractors do not. 咖啡 'coffee' is the target and 蛋糕 'cake' is the competitor in the highly predictable condition. 蛋糕 'cake' is the target and 咖啡 'coffee' is the competitor in less predictable condition. The classifier is only compatible with one of the two critical nouns (**杯** **'bei'** is compatible with coffee but not cake, and **块** 'kuai' is compatible with cake but not coffee).

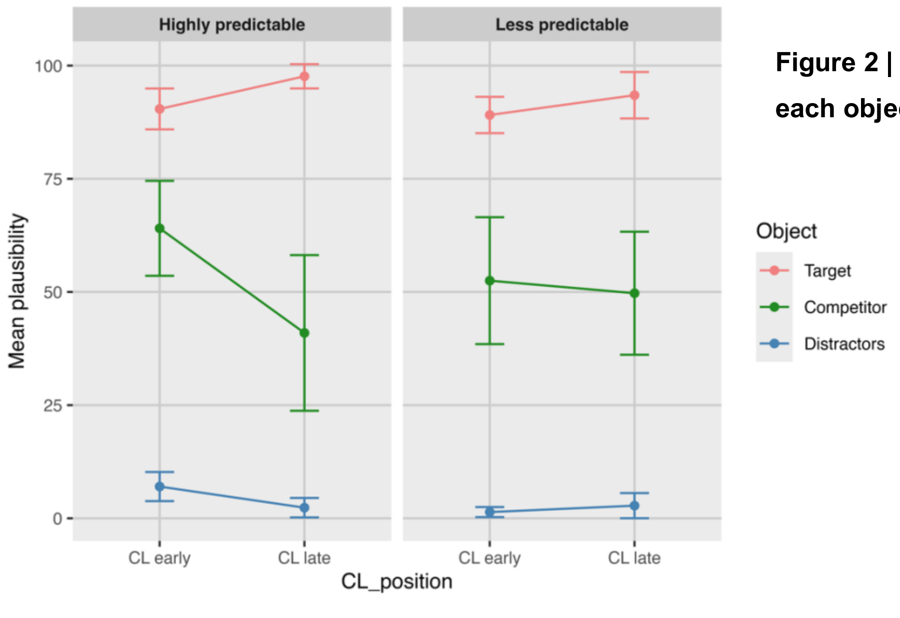


Figure 2 | The mean plausibility rating with ±1SD for each object in each condition.

References

[1] Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience, 32*(10), 1229–1249.
[2] Kaan, E., & Grüter, T. (2021). Chapter 1. Prediction in second language processing and learning: Advances and directions. In Kaan, E., & Grüter, T. (Eds.), *Prediction in Second Language Processing and Learning* (pp.1–24). John Benjamins Publishing Company.
[3] Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition*, *164*, 61–73.
[4] Qian, Z., & Garnsey, S. (2016). An ERP study of the processing of Mandarin classifiers. In H. Tao (Ed.), *Studies in Chinese Language and Discourse* (Vol. 7, pp. 59–80). John Benjamins Publishing Company.

The effect of language distance on bilingual lexical processing

Lara Kelly-Iturriaga, Mitsuhiko Ota, Martin Pickering
School of Philosophy, Psychology and Language Sciences, The University of Edinburgh
lara.kelly-iturriaga@ed.ac.uk

There is ample evidence for language non-selective lexical access, i.e., bilinguals experience language co-activation [1]. Still, most evidence comes from bilinguals who speak *close* languages, e.g., Dutch-English [2, 3, 4], and it is unclear how generalisable this finding is across different types of bilinguals. Previous research has suggested that bilinguals of *distant* languages may be less prone to experience language co-activation than bilinguals of *close* languages [5].

To address this issue, this study examines the effect of language distance in bilinguals and trilinguals on their language-selectivity during word recognition in an L2 context. Three comparisons will be carried out: (1) 36 Turkish-English (*distant*) and 36 Spanish-English (*close*) bilinguals, (2) 36 Basque-Spanish (*distant*) and 36 Catalan-Spanish (*close*) bilinguals, and (3) 36 Azeri-Turkish-English trilinguals' performance in their L2s, English (*distant*) and Turkish (*close*). Such comparisons enable comparison of early bilinguals from bilingual societies/communities (Study 2) and late bi/trilinguals from monolingual societies/communities (Studies 1 and 3), as well as trilingual lexical processing (Study 3).

Participants will complete L2 semantic relatedness tasks, deciding whether two words are related to each other in meaning. Critical items are interlingual homographs (IHs) – words with the same spelling but different meanings across languages (e.g., *pie* in English and Spanish, meaning *foot*). IHs will be presented with words related to their L1 meaning (e.g., *toe-pie*) or unrelated words (e.g., *pie-eve*) matched in frequency, orthographic length and number of syllables. Participants' L2 proficiency will be tested with Lextale [6] for Studies 1 and 3, Lextale-Esp [7] for Study 2, and a newly-created Turkish Lextale for Study 3.

Cross-language activation should give rise to interference from the L1 meaning of IHs, leading to lower accuracy and longer latency for L1-related IHs compared to unrelated IHs. We expect bilinguals of *close* languages to experience more cross-language activation than bilinguals of *distant* languages, and trilinguals to experience more language co-activation when performing in their *close* L2 in comparison to their *distant* L2.

Data collection is in process, with preliminary data available for Studies 1 and 2, please see Table 1. Two linear mixed-effects models were fitted to log-transformed reaction times (RTs) to explore the effects of L1 group (Study 1: Spanish vs. Turkish; Study 2: Catalan vs. Basque), condition (IH related vs. IH unrelated), and proficiency (centred Lextale score) and their interactions on RTs. The model for Study 1 revealed a main effect of L1 (b = -0.24, SE = 0.07, t(29.55) = -3.24, p = .001) and condition (b = 0.05, SE = 0.01, t(13.95) = 3.2, p = .002). L1-Spanish participants' responses (M = 1428, SD = 896) were significantly faster than L1-Turkish participants' responses (M = 3261, SD = 4430), and all participants responded more quickly to unrelated IHs (M = 2176, SD = 2996) than to related IHs (M = 2458, SD = 3540). The model for Study 2 did not reveal a significant main effect of L1 (b = 0.07, SE = 0.04, t(51) = 1.83, p = .07) or condition (b = 0, SE = 0.01, t(59.2) = 0.18, p = .99) on RTs.

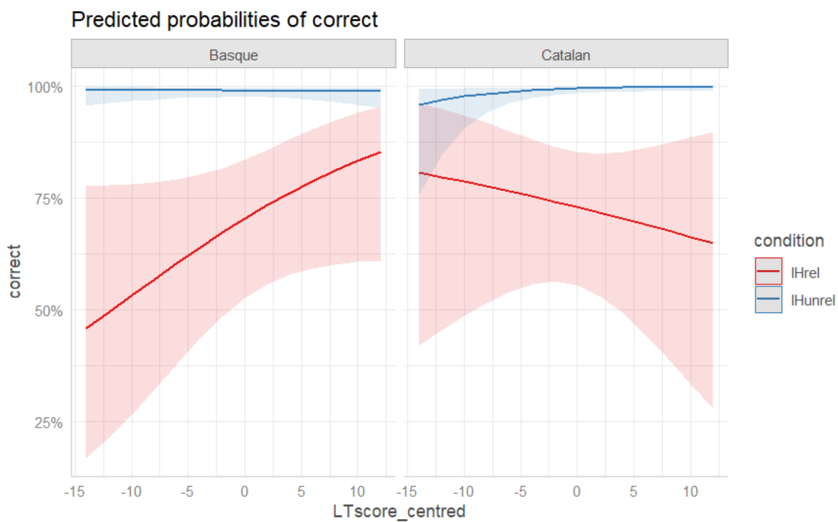
Two generalised linear mixed-effects models were employed to examine the same effects and their interactions (L1 group, condition, and proficiency) on accuracy. Both models revealed a main effect of condition, with all participants displaying lower accuracy in the IH related condition compared to the IH unrelated condition (Study 1: β = -1.08, *SE* = 0.15, $p < .001$; Study 2: β = -2.04, *SE* = 0.24, $p < .001$). The L1 × Condition × Proficiency interaction reached significance in Study 2 (β = 0.07, *SE* = 0.02, *p* = .007), suggesting that an increase in Basque participants' proficiency in a *distant* L2 may lead to less L1 activation whereas increasing Catalan participants' proficiency in a *close* L2 may lead to more L1 activation (see Figure 1).

Available data suggests that bilinguals of *close* and *distant* languages experience cross-language activation when encountering IHs presented with an L1-related L2 word. Early and late bilinguals from monolingual and bilingual societies/communities with differing levels of proficiency and age of acquisition (AoA) seem to experience language co-activation to different extents and levels. Moreover, improvements in L2 proficiency might modulate language co-activation differently depending on a bilinguals' L1-L2 distance.

Table 1. Mean reaction times (in milliseconds), proportion of accurate responses by condition and L1, mean proficiency scores and age of acquisition. Standard deviations are presented in brackets.

Study	L1	N	Condition	RTs (ms)	Accuracy	Proficiency	AoA (year)
1	Spanish	17	IH related	1496 (1002)	0.75 (0.43)	82.5 (10.8)	5.5 (2.3)
			IH unrelated	1360 (772)	0.96 (0.18)		
	Turkish	16	IH related	3480 (4774)	0.72 (0.45)	71.5 (10)	9.2 (2.9)
			IH unrelated	3043 (4055)	0.91 (0.29)		
2	Catalan	26	IH related	2023 (1410)	0.62 (0.49)	92.3 (4.7)	1.7 (2.6)
			IH unrelated	1993 (2120)	0.98 (0.16)		
	Basque	26	IH related	2323 (1699)	0.62 (0.48)	88 (6.22)	1.8 (2.2)
			IH unrelated	2299 (1429)	0.96 (0.19)		

Figure 1. Predicted probabilities of a correct response by L1, condition and proficiency (centred Lextale score) for Study 2.



References

[1] Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. http://dx.doi.org/10.1017/S1366728902003012.

[2] Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533–550. https://doi.org/10.3758/BF03195845.

[3] van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language Conflict in the Bilingual Brain. *Cerebral Cortex*, *18*(11), 2706–2716. https://doi.org/10.1093/cercor/bhn030

[4] Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301. https://doi.org/10.1016/j.jml.2009.12.003.

[5] Cenoz, J. (2003). The Role of Typology in the Organization of the Multilingual Lexicon. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *The Multilingual Lexicon* (pp. 103–116). Springer, Dordrecht. https://doi.org/10.1007/978-0-306-48367-7_8.

[6] Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior research methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0.

[7] Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica, 35*(1), 49–66.

Not All Vowels Are Learned Alike: The Limits of L2 Experience in Cross-Linguistic Vowel Perception

Annett Jorschick, Yuan Zhou
Faculty of Linguistics and Literary Studies, Bielefeld University
annett.jorschick@uni-bielefeld.de

Background: How do Mandarin L1 speakers perceive German vowels that do not exist in their native phonology? According to the Perceptual Assimilation Model (PAM) [1, 2], non-native sounds are perceived by mapping them onto the closest native (L1) categories. As a result, when listeners are asked to discriminate two sounds of a non-native language, their performance depends on whether the two non-native sounds are mapped to (1) different L1 categories (high accuracy), (2) the same category (low accuracy), or (3) no L1 category at all (accuracy depends on how acoustically distinct the sounds are). Previous studies (e.g., [3]) have confirmed that discrimination performance tends to decrease as perceptual overlap between non-native and native categories increases. Faris and colleagues [3] further refined the distinction by proposing: (i) non-overlapping contrasts (as in case 1) are most easily discriminated, followed by (ii) partially overlapping contrasts (where the two non-native vowels share similarity with the same L1 vowel but also relate to different additional L1 categories). These are, in turn, better discriminated than (iii) completely overlapping contrasts (as in case 2), where both non-native vowels are strongly associated with a single L1 category. This study investigated how Mandarin L1 listeners perceive German vowel contrasts classified according to PAM assimilation types (cf. Methods). Crucially, we extend previous work by comparing two listener groups: those with and without prior experience in German.

Method: Fifty-one native Mandarin speakers participated, including 26 with German experience and 25 without. Participants completed an online AXB discrimination task with seven German vowel pairs, each produced by three female native speakers of German to model natural variation in speech. German vowels were classified based on phonetic distance and perceived similarity to Mandarin vowel categories, as judged by a native Mandarin speaker on a scale from 0 (no overlap) to 4 (high similarity). One vowel pair was selected for each of the seven PAM assimilation types (see Table 1). Data were analyzed using GLMMs, incorporating random slopes where appropriate. Custom contrasts were defined to examine differences between assimilation types based on their degree of perceptual overlap: non-overlapping (TC, UC-N, UU-N) > partially overlapping (CG, UC-P, UU-P) > completely overlapping (UC-C).

Results: Systematic differences emerged across assimilation types (Figure 1): non-overlapping contrasts were discriminated best, followed by partially overlapping and completely overlapping contrasts. However, CG and UU-P deviated from predictions, showing unexpectedly low accuracy. L2 German speakers generally outperformed L1-only participants, but this advantage was absent for CG and UU-P resulting in a significant interaction between group and assimilation type.

Discussion: Our findings support PAM's predictions but also indicate limits to the benefits of L2 experience. Poor performance on CG and UU-P suggests that subtle contrasts with shared category mappings remain difficult to acquire, even with exposure. This is consistent with PAM and the Speech Learning Model (SLM) [4, 5], both of which argue that L2 learning is constrained by initial L1-L2 mappings and phonetic similarity. Overall, the results highlight that the perceptual learning of L2 vowels depends on the nature of the contrast and the structure of the L1 phonological system.

	Assimilation Type	Vowel Pairs	Description	Predicted Accuracy
TC	Two-Category, Non-overlapping	/y:/–/u:/	Each vowel maps onto a distinct L1 category	high
UC-N	Uncategorized–Categorized, Non-overlapping	/ɛ:/–/ø:/	One vowel uncategorized, one clearly categorized	high
UU-N	Uncat.–Uncat., Non-overlapping	/ɐ/–/ɛ:/	Both vowels uncategorized but acoustically distinct	high
CG	Category-Goodness, Partial Overlap	/y:/–/ø:/	Both vowels map to the same L1 category, but differ in category fit	moderate
UC-P	Uncategorized–Categorized, Partial Overlap	/ə/–/ɐ/	Partial similarity to the same or different L1 categories	moderate
UU-P	Uncat.–Uncat., Partial Overlap	/ɛ:/–/e:/	Both vowels map partially to the same and to different L1 vowels	moderate
UC-C	Uncategorized–Categorized, Complete Overlap	/i:/–/e:/	Both vowels perceived as equally good examples of the same L1 category	low

Table 1: PAM assimilation types for seven German vowel pairs as materials in an AXB discrimination task

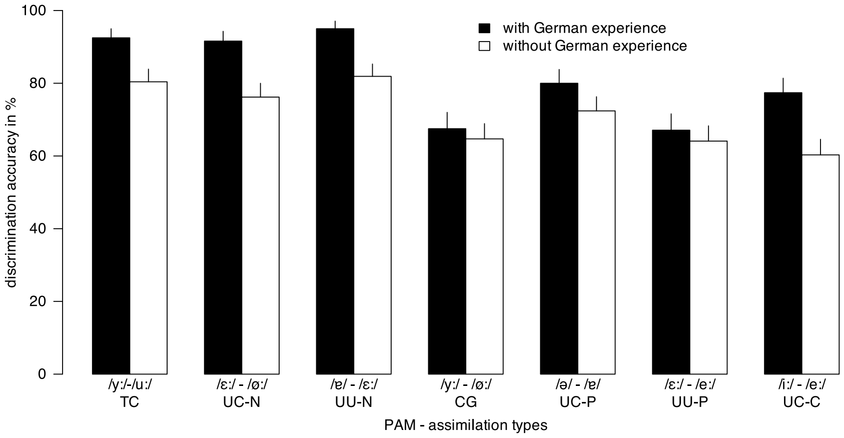


Figure 1: Mean discrimination accuracy and their 95%-CIs across PAM assimilation types for Mandarin L1 listeners with (black bars) and without (white bars) German experience. Assimilation types are grouped by degree of perceptual overlap between German and Mandarin vowels: non-overlapping (TC, UC-N, UU-N), partially overlapping (CG, UC-P, UU-), and completely overlapping (UC-C).

References

[1] Best, C. T. (1995). A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research***,** pp. 171–204.

[2] Best, C. T., & Tyler, M. D. (2008). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). John Benjamins Publishing Company.

[3] Faris, M. M., Best, C. T. & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, 70, 1–19.

[4] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*(1), 233-277.

[5] Flege, J. E., Aoyama, K., & Bohn, O. S. (2021). The revised speech learning model (SLM-r) applied. Second language speech learning: Theoretical and empirical progress, 84-118.

Formal and semantic cues in gender assignment to novel words in Italian

Valeria Galimberti[1], Beatrice Giustolisi[1], Caterina Donati[2], Francesca Foppolo[1]
[1] Department of Psychology, University of Milano-Bicocca; [2] Université Paris Citè
valeria.galimberti@unimib.it

Background: Italian is a highly gender transparent language in which almost all nouns end with a vowel and are thus grammatically gender-marked. Typically, nouns ending in -a are grammatically feminine (F), those ending in -o are masculine (M), with the latter being more frequent overall [1]. Nouns ending in -e are opaque and can be M or F; nouns ending in consonant (-c) tend to be borrowings, typically anglicisms. *Which gender is assigned to loan words?* Corpus research shows that borrowings tend to be assigned either the gender of the Italian equivalent (semantic analogy) or 'default' M, although some degree of variation is attested [2]. Experimental studies with novel words in Spanish, which has a grammatical system similar to Italian but more frequent consonant endings, reveal a preference for morpho-phonological (over semantic) cues by native speakers and Spanish-dominant early bilinguals [3]. We investigate the role of morpho-phonological traits and semantic cues in gender assignment by L1 (and L2, planned) Italian speakers during first exposure to novel words, manipulating word ending (to test the *formal* strategy) and the noun's gender of the objects associated with these words to serve as translation equivalents (to test the *semantic* strategy).
Method: We ran a production study in E-Prime 3 in which a set of images of inanimate objects (N = 180, from the MultiPic database [4]; 50% F in Italian) and pseudo-objects (N = 90, NOUN Database [5]), were associated with pseudowords ending in -a/-o/-c (N = 270). The picture-pseudoword association was random to control for semantic associations [6], and fully counterbalanced across lists in 3x3 conditions: congruent, incongruent, or opaque, considering *word ending* of the same root (dimaba/dimabo/dimab) and the gender of the *object* (or non-object, N) in the picture (M/F/N). Participants (N = 45) were asked to produce a sentence starting with 'è un/una…' (it's $a_{F/M}$) + pseudoword. We coded the gender assigned to the word on the basis of the article produced (*un/uno*$_{[M]}$, *una*$_{[F]}$) and recorded speech onset time.
Results: Results of a logistic mixed model regression (Table 1) on the article produced (Figure 1) show a prevalence of the formal strategy for words ending in vowels and default M for words ending in consonant: -a endings were more likely assigned F, -o endings M, compared to -c (1st z = -55.69; 2nd : z = -26.26). Semantic effects emerged in association with F objects, as words ending in -c were more likely assigned F compared to words ending in -o (z = -5.83). Exploratory analysis of speech onset latency (Table 2) showed longer RTs for words ending in -a vs. -c when associated with an M (vs. N) object (z = 2.69), supporting a possible interference of the object's gender (Figure 2).
Discussion: In line with previous research on novel words, we show evidence for (i) a prevalence of the morpho-orthographic strategy in gender assignment to transparent novel word; (ii) M as the "default" for opaque novel words ending in -c. In addition, we find (iii) limited but consistent evidence for an influence of the gender of the associated object. The emergence of semantic effects may have been facilitated by an increase in statistical power compared to previous studies [e.g., 7], while controlling for semantic associations with randomized picture/word pairs. Further planned studies with L2-Italian speakers of L1 French (gender-marked language), and L1-Dutch (no gender assigned to inanimate nouns) will allow us to explore the effects of cross-linguistic influence and L2 proficiency on gender assignment strategies.

Tables 1-2. Outputs of the logistic regression (1) and linear model (2) on gender assignment to nonwords (M/F) and speech onset latencies, with Ending (-c, -a, -o) and Object (N/F/M), contrasts set at (0.5, -0.5, 0) and (0.5, 0, -0.5), OLD20 as fixed effect, Participants, Items, Word root and Trial order as random intercepts.

Table 1. *glmer*

	Est.	S.E.	z val.	p
(Intercept)	-1.537	0.112	-13.731	0.000***
Ending_contrast1	-7.068	0.127	-55.689	0.000***
Ending_contrast2	5.114	0.195	26.263	0.000***
Object_contrast1	-1.911	0.128	-14.942	0.000***
Object_contrast2	1.051	0.163	6.454	0.000***
old20_cent	-0.291	0.126	-2.307	0.021*
Ending_contrast1:Object_contrast1	-1.770	0.304	-5.826	0.000***
Ending_contrast2:Object_contrast1	1.990	0.430	4.624	0.000***
Ending_contrast1:Object_contrast2	0.774	0.354	2.187	0.029*
Ending_contrast2:Object_contrast2	-0.321	0.579	-0.554	0.580

Table 2. *lmer*

	Est.	S.E.	t val.	d.f.	p
(Intercept)	6.883	0.047	146.701	35.410	0.000***
Ending_contrast1	-0.025	0.006	-4.435	9451.046	0.000***
Ending_contrast2	0.026	0.006	4.657	9475.359	0.000***
Object_contrast1	-0.009	0.006	-1.506	265.841	0.133
Object_contrast2	0.007	0.006	1.146	265.231	0.253
old20_cent	0.026	0.006	3.967	371.339	0.000***
Ending_contrast1:Object_contrast1	-0.030	0.016	-1.922	9461.215	0.055.
Ending_contrast2:Object_contrast1	0.008	0.016	0.539	9469.598	0.590
Ending_contrast1:Object_contrast2	0.042	0.016	2.686	9463.433	0.007*
Ending_contrast2:Object_contrast2	-0.028	0.016	-1.769	9463.648	0.077.

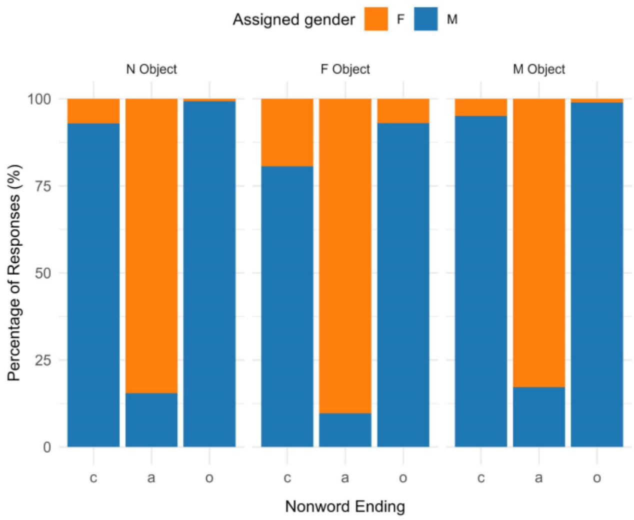


Figure 1. Percentage of masculine (M) and feminine (F) articles produced before novel words.

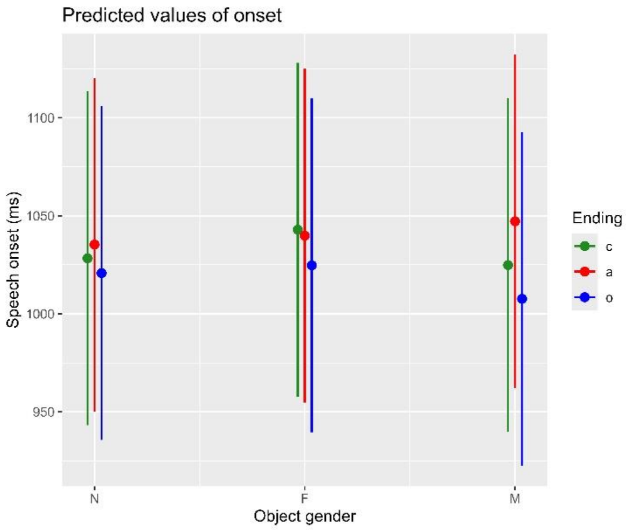


Figure 2. Interaction plot of Model 2.

References

[1] Padovani, R., and Cacciari, C. (2003). Il ruolo della trasparenza morfologica nel riconoscimento di parole in Italiano. *Giornale italiano di psicologia.* 30, 749–772.
[2] Pulcini, V. (2023). *The Influence of English on Italian: Lexical and Cultural Features*. De Gruyter.
[3] Pérez-Tattam, R., Ezeizabarrena, M. J., Stadthagen-González, H., & Mueller Gathercole, V. C. (2019). Gender Assignment to Spanish Pseudowords by Monolingual and Basque-Spanish Bilingual Children. *Languages*, *4*(3), 58.
[4] Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly journal of experimental psychology*, *71*(4), 808-816.
[5] Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*, *48*, 1393-1409.
[6] Gatti, D., Marelli, M., & Rinaldi, L. (2023). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *Journal of Experimental Psychology: General*, 152(3), 851–863.
[7] Clegg, J. H. (2010). Native Spanish speaker intuition in noun gender assignment. *Language Design*, 12, 5-18.

Investigating prediction error cost during natural reading in young and older adults: Evidence from eye movements and fixation-related potentials

Rupali Limachya[1,2], Steven Frisson[1], Federica Degno[3], Simon P. Liversedge[4], Kevin B. Paterson[2], & Ascensión Pagán[2]

[1] School of Psychology, University of Birmingham; [2] School of Psychology and Vision Sciences, University of Leicester; [3] School of Psychology, Bournemouth University; [4] School of Psychology, University of Central Lancashire

rxl025@student.bham.ac.uk

Predictive processing is thought to support efficient language comprehension, enabling readers to anticipate upcoming words based on contextual information [1]. Prior research using eye movements (EMs) and event-related potentials (ERPs) shows that word predictability influences reading behaviour, with highly predictable words more likely to be skipped and fixated for shorter durations - especially in younger compared to older adults [2]. ERP research similarly shows reduced N400 amplitude for predictable words, with stronger effects in young adults [3]. However, few studies have examined the consequences of prediction violations across age groups, specifically when a highly predictable word is replaced by a less predictable but plausible word in a strongly constraining context, potentially eliciting a processing cost known as a prediction error cost [4]. While no such cost has been consistently observed in EMs with young adults, ERP studies have reported post-N400 late negativity, indexing prediction error costs, in young but not older adults [5], though this pattern has not always been replicated [6]. Eye movement and N400 findings converge in young adults but diverge in older adults, highlighting a gap in our understanding of how cognitive aging impacts prediction in reading.
We investigated age differences in prediction and prediction error during natural sentence reading using simultaneous EM and fixation-related potential (FRP) recordings. Forty-one young adults (18–30 years) and 36 older adults (60–80 years) read 294 sentences containing high- or low-predictability target words in either strongly constraining or neutral contexts. Materials were normed for cloze probability and plausibility. EMs were analysed using (generalised) linear mixed-effects models across skipping, first fixation duration, gaze duration, go-past time, and total reading time.
Results revealed robust effects of contextual constraint and predictability across both age groups (see Figure 1). Strongly constraining contexts led to higher skipping rates and shorter fixation durations. Older adults showed generally longer fixation times, consistent with age-related differences in reading behaviour.
Prediction error costs defined as increased reading time for low-predictability words in constraining versus neutral contexts, were not observed in either group. Instead, facilitation was observed for low-predictability words in constraining contexts in all EM measures, suggesting that semantically rich contexts may ease integration of plausible, though unexpected continuations. These findings support a view of prediction in which readers generate graded, semantic-level expectations rather than anticipating a specific word form. Ongoing analyses of FRPs will clarify whether neural responses align with the eye movement findings or instead reflect prediction error signals that are not evident in behaviour. The integration of EM and ERP techniques in this study allows for a richer characterisation of prediction and aging, with implications for understanding predictive language processing across the lifespan.

Figure 1. Mean fixation durations and skipping rates across early and late measures by predictability condition and age group

References

[1] Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044.

[2] Zhang, J., Warrington, K. L., Li, L., Pagán, A., Paterson, K. B., White, S. J., & McGowan, V. A. (2022). Are older adults more risky readers? Evidence from meta-analysis. *Psychology and Aging*, *37*(2), 239–259.

[3] Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141.

[4] Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214.

[5] Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785.

[6] Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135-147.

Structure dependent differences in the persistence of syntactic priming

Salma Gilani[1], Katrien Segaert[1], Linda Wheeldon[2], Evelien Heyselaar[3]

[1] School of Psychology, University of Birmingham; [2] Department of Foreign Languages and Translation, University of Agder; [3] Behavioural Science Institute, Radboud University

sng201@student.bham.ac.uk; k.segaert@bham.ac.uk; linda.r.wheeldon@uia.no; evelien.heijselaar@ru.nl

Syntactic priming, the tendency to repeat syntactic structures across utterances, is a widely used tool for investigating the mechanisms of sentence production [1]. It has been shown to be both cumulative and long-lasting, key characteristics that align with implicit learning theories [2]. While these effects are well-established for transitive sentence structures, such as actives and passives [3], there is limited evidence for ditransitive structures (double-object and prepositional-object) [4]. The current study investigates both cumulative and long-term syntactic priming in transitive (Experiment 1) and ditransitive (Experiment 2) sentence production using a picture-description task. To reduce potential bias and increase generalisability, we created a new set of stimuli using non-human, geometric shape characters (e.g., triangle, circle; Figure 1). Both experiments consisted of 40 native English-speakers each, who completed two sessions spaced one week apart. Each session included a Baseline Phase (30 trials) followed by a Priming Phase (100 trials). One trial consisted of a color-coded prime stimulus followed by a greyscale target stimulus, and responses were manually coded using a binary system. In Experiment 1 (transitive setup), a mixed-effects model revealed a significant decrease in passive sentence production from Session A to Session B (with a one-week interval), indicating a lack of long-term persistence (Figure 2). However, a generalised additive mixed model (GAMM) analysis showed clear cumulative growth in passive usage within each session, consistent with previous findings of implicit learning (Figure 3). In contrast, Experiment 2 (ditransitive setup) revealed no cumulative or long-term priming effects for ditransitive structures; instead, production of double object constructions decreased across sessions, suggesting a reversal of previously primed preferences (Figure 3). These findings challenge the assumption that syntactic priming consistently results in stable, long-term learning. Rather, the results suggest that structural persistence may be structure-specific, potentially modulated by differences in semantic support, thematic role mapping, and cognitive load. The lack of accumulation for the ditransitive sentence structure, despite clear within-session learning for transitives, highlights the need for further research into how stimulus design and construction type influence the dynamics of syntactic learning. In contrast to prior work showing persistent priming for some structures [2], our findings suggest that long-term priming may be more fragile and structure specific than previously assumed.

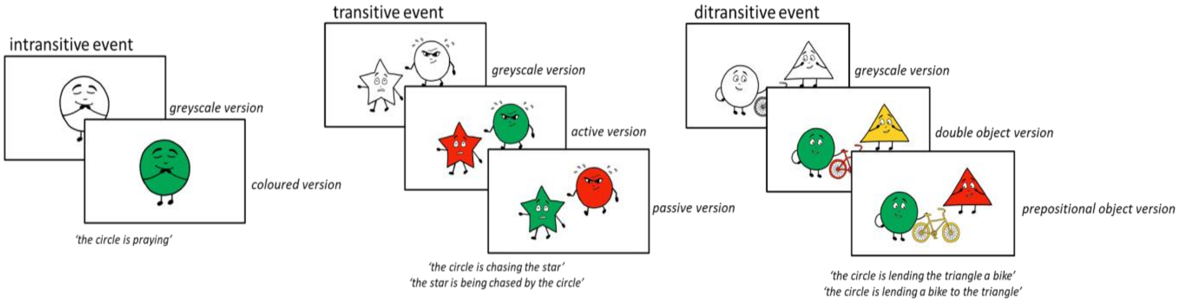


Figure 1. *Example of the newly created stimuli; intransitive, transitive and ditransitive event.*

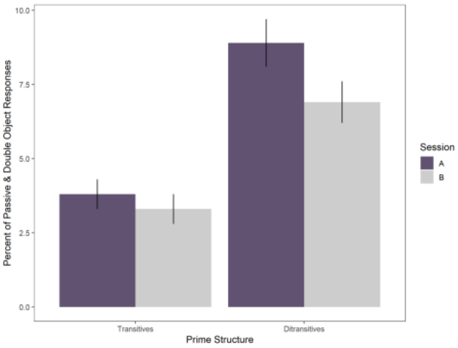


Figure 2. *Percent of passive (Exp 1) and double object (Exp 2) responses in the Baseline Phase, across Session A and B. Error bars represent standard error.*

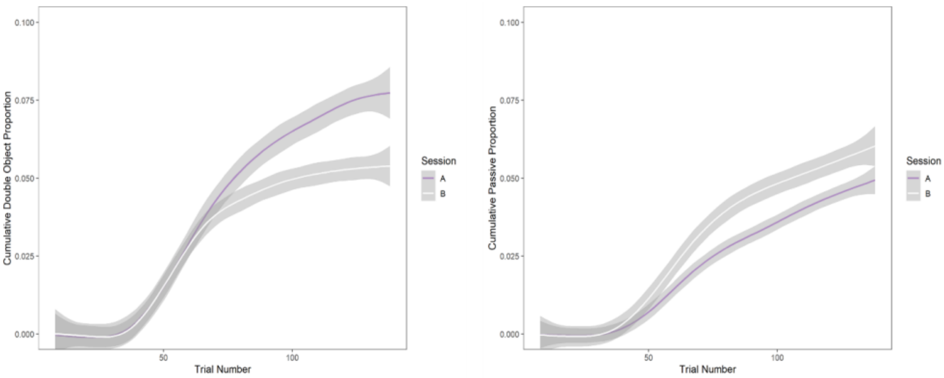


Figure 3. *Cumulative passive (right) and double object (left) proportion per trial. Error clouds represent standard error.*

References

[1] Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355–387.

[2] Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*(2), 177–192.

[3] Heyselaar, E., & Segaert, K. (2022). Structural priming persists for (at least) one month in young adults, but not in healthy older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(8), 1219.

[4] Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition, 39*(3), 381–388.

Tracking Affixation: ERP and Behavioural Insights into the Suffixing Bias

Inês Cardoso Ferreira[1,2], Leona Polyanskaya[1,2], Mikhail Ordin[3]
[1]Faculty of Psychology and Educational Sciences, University of Coimbra, [2] Language, Metacognition and Decision-Making Laboratory, University of Coimbra, [3]Faculty of Medicine, University of Coimbra
cafe@uc.pt

Background: Language acquisition relies on the ability to detect structural regularities, including morphological patterns like affixation [2]. A robust typological trend is the suffixing bias, a widespread tendency across languages to mark grammatical information using suffixes rather than prefixes [1,2]. This bias has been attributed either to domain-general processing preferences [1] or to language-specific experience [3,4]. To contribute to this debate, the present study examines how native Portuguese speakers (a suffix-predominant language [2]) learn affixation patterns in an artificial language. More details on the suffixing bias and how it affects the Portuguese language can be found on page 3.

It is important to note that the artificial units used in our study do not carry meaning and thus differ from real morphemes. However, we aim to investigate whether an evolutionarily ancient statistical learning mechanism—one capable of segmenting continuous streams of sensory input into discrete, recurring units, regardless of semantic content—can contribute to the suffixing bias observed across natural languages.

Methods: Thirty-nine Portuguese-speaking adults completed an artificial language learning task recorded with electroencephalography (EEG). Pseudowords were created using the UniPseudo generator [5], modified with either a prefix (/li/) or suffix (/pu/) and paired with images from the RealPic database [6]. During learning, syllables were presented sequentially, separated by filler syllables to reduce reliance on transitional probabilities. Participants then completed a word recognition (yes/no) test and a picture-matching task.

Behavioural Results: Both tests confirmed above-chance learning, with overall accuracies of 56.7% and 63.5%, respectively. While suffix words showed slightly higher accuracy than prefixes, this difference was not statistically significant, suggesting a mild cognitive suffixing bias.

ERP Results-Learning Phase: P2 amplitudes were significantly larger for suffixes (M=2.70 μV, SD=2.45) than for prefixes (M=1.91 μV, SD=1.67), t(38)=2.414, p=.021, d=0.361, indicating enhanced attentional engagement with final-position units, likely shaped by Portuguese's suffixing structure.

ERP Results-Test Phase: In this phase, we split the possible answers for these specific analysis into hits for suffixed words, hits for prefixed words and false alarms. A RM-ANOVA revealed a significant effect of the item type, F(2,76)=4.319, p=.017. False alarms elicited significantly larger P2 amplitudes (M=7.03 μV, SD=3.70) than both suffix (M=5.67 μV, SD = 4.86), t(38)=2.216, p=.033, d=0.355, and prefix items (M=5.62 μV, SD=4.79), t(38)=2.911, p=.006, d=0.466. No significant difference was found between suffixes and prefixes, t(38)=0.087, p=. 932.

Discussion: The finding that suffixes elicited significantly larger P2 amplitudes than prefixes during the learning phase supports the hypothesis that positional learning biases—potentially shaped by language experience—facilitate enhanced attention to final elements. In fact, between 180 to 280ms, participants had already identified the presence of suffixes and prefixes, and we observe differences in their neural responses, with suffixes generating a more positive wave. This is consistent with prior studies linking the P2 component to early attentional engagement and stimuli categorization [7]. During the word recognition test phase, no P2 differences were observed between affix types, however, false alarms elicited significantly greater P2 amplitudes. This pattern suggests that greater neural effort was required when trying to match unfamiliar stimuli with stored representations.

References

[1] Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference. *Language and Cognitive Processes, 24*(6), 876–909. https://doi.org/10.1080/01690960902719267

[2] Dryer, M. S. (2013). Prefixing vs. suffixing in inflectional morphology. In M. S. Dryer & M. Haspelmath (Eds.), The world atlas of language structures online. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.7385533

[3] Martin, A., & Culbertson, J. (2020). Revisiting the Suffixing Preference: Native-Language Affixation Patterns Influence Perception of Sequences. *Psychological Science, 31*(9), 1107–1116. https://doi.org/10.1177/0956797620931108

[4] Ordin, M. (2025). Affixation patterns in native language and sequence processing by statistical learning mechanisms. *Evolutionary Human Sciences*, *7*, e11. doi:10.1017/ehs.2025.6

[5] New, B., Bourgin, J., Barra, J., & Pallier, C. (in press). UniPseudo: A universal pseudoword generator. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218231164373

[6] Souza, C., Garrido, M. V., Saraiva, M., & Carmo, J. C. (2021). RealPic: Picture norms of real-world common items. *Behavior Research Methods*, *53*(4), 1746–1761. https://doi.org/10.3758/s13428-020-01523-z

[7] Favero, J. D., Luck, C., Lipp, O. V., Nguyen, A. T., & Marinovic, W. (2023). N1-P2 event-related potentials and perceived intensity are associated: The effects of a weak pre-stimulus and attentional load on processing of a subsequent intense stimulus. *Biological psychology*, *184*, 108711. https://doi.org/10.1016/j.biopsycho.2023.108711

[8] Hammarström, H. (2021). Measuring Prefixation and Suffixation in the Languages of the World. Acl Anthology, 81–89. https://doi.org/10.18653/v1/2021.sigtyp-1.8

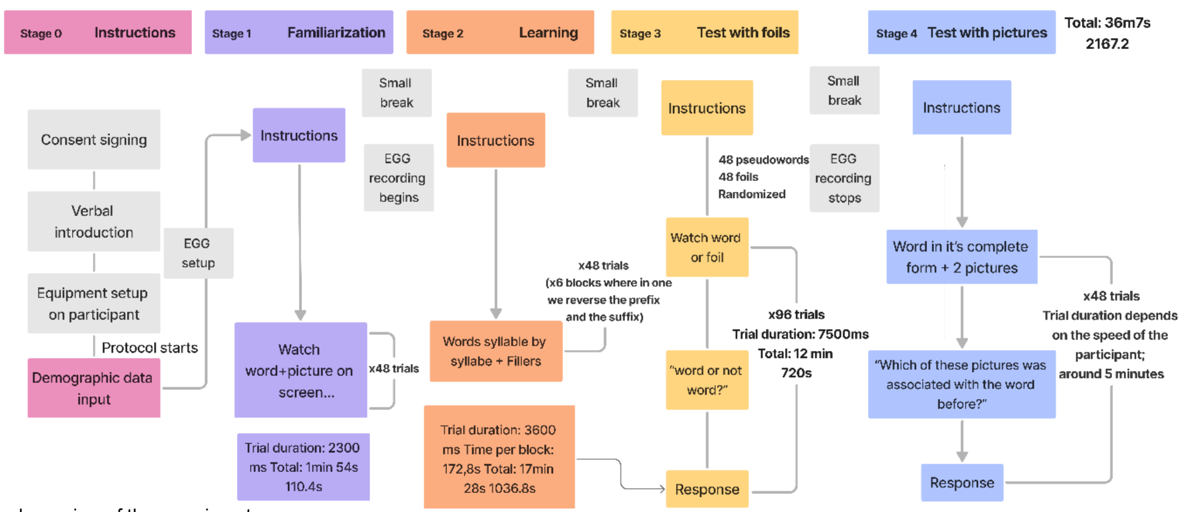


Figure 1. Visual overview of the experiment

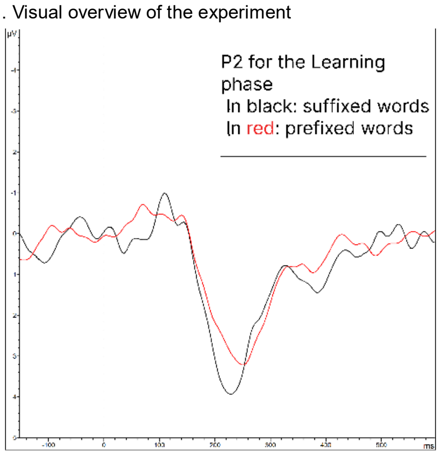


Figure 2. P2 component seen in the Learning Phase

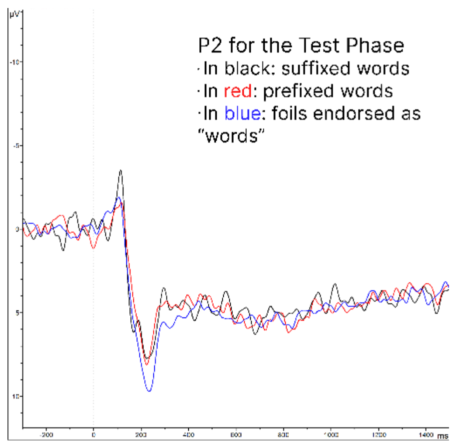


Figure 3. P2 component seen in the Test Phase

The specificities of affixation in the Portuguese language

Before we understand the importance of suffixation in the Portuguese language, we have to analyze the overall affixation tendency that most languages follow. In concrete terms, suffixing languages outnumber prefixing languages by approximately 8:1 [2]. Table 1 provides examples of prefixes and suffixes across various languages, illustrating how grammatical information is typically encoded at the beginning or end of words.

Language	Inflectional prefix	Example	Inflectional suffix	Example
Portuguese	Does not use prefixes for inflection.	---	-s → Marks the plural -eu → Marks the past tense	Menino**s** ("boys") Root: -menino- ("boy") Corr**eu** ("he ran") Root: -correr- ("to run")
Kĩĩtharaka	mû(u)- →Denotes that a word is singular a- → Used to convey the number two	**M**untû ("one person") **A**ntû ("two people") Root: -ntû - ("person")	-ire → Marks the perfective aspect	Nda-rug-**ire** ("I have gone") Root: Nda-rug- ("go")
Basque	n→ Used to convey the first singular of a present tense verb h- → Used to convey the second singular of a present tense verb	**N**oa ("I go") **H**aiz ("you go") Root: -joan- ("to go")	k- → Used to convey the absolutive definite plural	Gizona**k** ("the men") Root: -gizon- ("man)

Table 1. Examples of prefixes and suffixes used in a strongly suffixing language (i.e., Portuguese), a strongly prefixing language (i.e., Kĩĩtharaka) and an equal prefixing and suffixing language (i.e., Basque).

As mentioned on page 1, this global asymmetry has given rise to competing theories. On one side, domain-general accounts suggest that this bias reflects universal cognitive constraints, such as a heightened ability to process or recall sequence-final elements [1]. Supporting this, Hupp and colleagues [1] showed that English speakers more easily perceived sequences with end-based variation.

Alternatively, language-specific theories argue that affix preferences are shaped by linguistic experience. For example, Martin and Culbertson [3] demonstrated that speakers of Kĩĩtharaka, a prefixing Bantu language, preferred onset-based variation, contrary to English speakers. Similarly, Ordin [4] showed that suffixing preferences emerged only in linguistic, not non-linguistic contexts, and were modulated by participants' native morphology. These findings highlight that affix processing may not be governed by universal constraints but by structural features of a speaker's native language.

In this context, Portuguese provides a particularly suitable test case. From a morphological perspective, Portuguese displays a notable preference for suffixation over prefixation, especially through inflectional suffixes marking tense (e.g., cantou — sang), number (e.g., casas — houses), and gender (e.g., menino vs. menina — boy vs. girl). Prefixation, in contrast, is less productive and mostly restricted to derivational morphology (e.g., rever — to review). According to the World Atlas of Language Structures [2], the suffixing index is calculated using the formula Suffixes/(Suffixes + Prefixes), based on lexical entries across ten categories of inflectional morphology. Importantly, this metric excludes derivational morphology and reflects lexicon structure, not real-world frequency. To address this limitation, Hammarström [8] applied machine-driven analysis across 4,437 languages, including Portuguese, using corpus frequency to estimate affixation patterns. The data extracted from the database showed that Portuguese text featured 2,959 suffixes and 2,524 prefixes, which means that the ratio of suffixes to all affixes in Portuguese texts is 0.54. These findings suggest that while Portuguese maintains a suffixing tendency, it is moderate when considering frequency-based corpora rather than lexical structure alone (but including not only inflectional but also derivational morphemes, we are not considering the latter in this study).

Taken together, these conclusions suggest that Portuguese is a suffixing language. Building on this classification, Portuguese provides an appropriate context to investigate whether native language experience facilitates the acquisition of affixation patterns, allowing us to explore whether learners are more attuned to regularities occurring at the beginnings or ends of syllable sequences in an artificial language learning paradigm.

Cross-linguistic influence in bilingual minds: A Visual World eye-tracking study on grammatical aspect processing in Croatian-German and Croatian-Italian children

Sara Košutar[1], Judith Schlenter[2], Natalia Mitrofanova[1], Serge Minor[1]
UiT The Arctic University of Norway[1], University of Cologne[2]
sara.kosutar@uit.no

Background. How languages influence each other in a bilingual's mind has long been a prominent topic of research in the field of bilingual child acquisition [1]. Cross-linguistic influence (CLI) may explain why grammatical aspect has sometimes been found to be vulnerable in heritage bilingual children [2]. However, the findings are inconsistent. Either the acquisition of aspect in the heritage language is facilitated when the societal language encodes the aspect morphologically [3], or there is no evidence of such facilitation [4]. Moreover, most existing findings are based on offline studies and comparisons between bilinguals and monolinguals. To better understand the role of CLI in bilingual language acquisition, more research is needed, particularly studies examining children's automated grammatical knowledge and comparing different bilingual groups. We investigated the role of CLI in the processing of verbal aspect by 6–12-year-old Croatian heritage speakers in Germany and Italy. While German does not mark aspect grammatically, Croatian and Italian exhibit a certain overlap in their use of verbal aspect (Perfective-PF vs. Imperfective-IMPF). We hypothesized that the Croatian-Italian group would outperform the Croatian-German group, as structural similarity between the languages should facilitate the processing of aspect. **Method.** The Croatian-Italian (*N*=29) and Croatian-German (*N*=38) groups were closely matched for age, and the children's overall exposure to Croatian was assessed [5]. We adopted the Visual World eye-tracking paradigm developed by Minor et al. [6]. The children listened to sentences with IMPV or PF aspect and were asked to select the picture that matched the sentence (Fig. 1). We analyzed the selection of the Ongoing event picture in the IMPF condition and the Completed event picture in the PF condition, and the proportion of looks to the Ongoing Event picture in each condition. **Results.** The accuracy scores appear to follow a bimodal distribution in both groups of participants (Fig. 2). We applied finite mixture models to analyze between-speaker heterogeneity [7]. We identified a high-accuracy cluster with a predicted accuracy of 81.3%, and a low-accuracy cluster with a predicted accuracy of 55.9%. Logistic regression analysis revealed a positive effect of cumulative exposure to Croatian (B=0.04, z=1.81, *p*=0.07), and an effect of Group (B=1.81, z=2.25, *p*=0.02), with Croatian-Italian bilinguals being significantly more likely to belong to the high-accuracy cluster. We next fit a Bayesian mixed effects beta regression model to predict the proportion of looks to the Ongoing event picture by Aspect, Group and Exposure [8]. Only the Croatian-Italian group improved in distinguishing the aspect with exposure, as there were more looks to the Ongoing Event picture in the IMPF condition, and fewer looks in the PF condition (Fig. 3). Posterior distributions showed that cumulative exposure has a positive effect on aspect sensitivity in the Croatian-Italian group, but not in the Croatian-German group. This is mostly driven by how exposure affects the performance of the two groups in the PF condition (Fig. 4). **Discussion.** Our findings indicate that the Croatian-Italian group is more likely to achieve high accuracy and show improved performance in processing of aspect with increased exposure to Croatian. These results align with the prediction of facilitative CLI from Italian, which also grammatically encodes the distinction between Completed and Ongoing events. The absence of CLI during real-time processing in the IMPF condition may be attributed to default aspect effects, as the PF aspect is generally acquired earlier and often serves as the default form in the past tense [9]. We conclude that structural similarity between languages may facilitate the processing of grammatical aspect in a heritage language.

Figure 1. Example of experimental stimuli (visual display and audio stimuli)

Figure 2. Distributions of accuracy scores on the picture selection task by Group

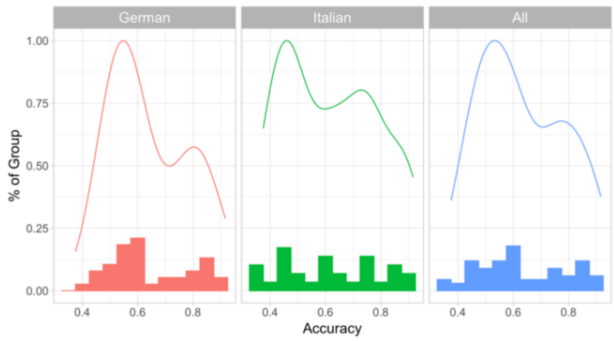


Figure 3. Predicted proportion of looks to ongoing event picture by Aspect, Group and Exposure

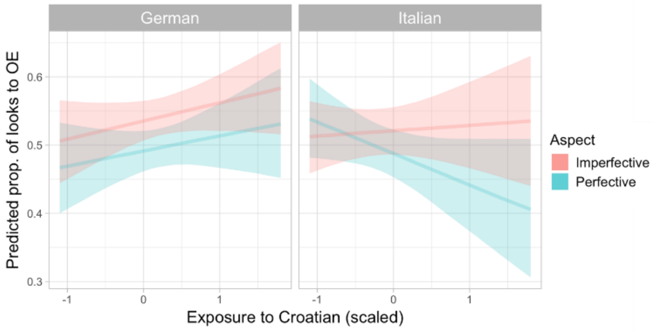
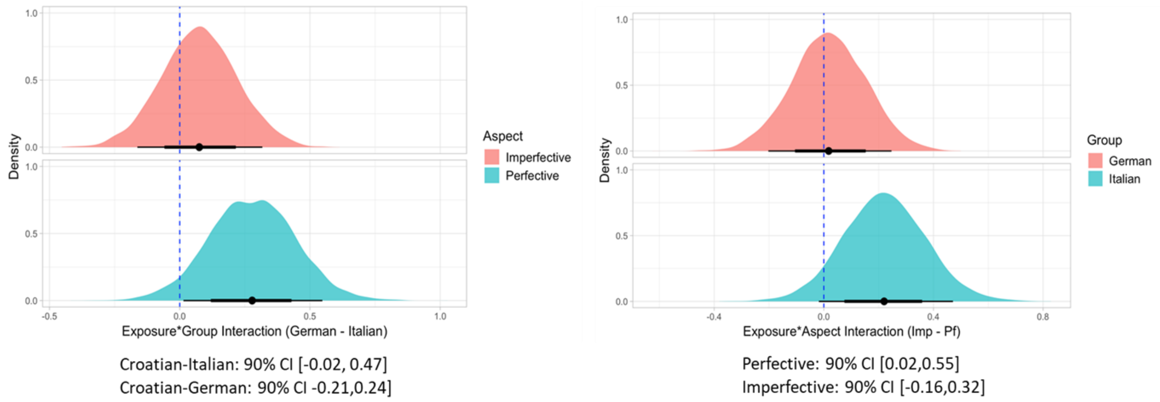


Figure 4. Posterior distributions of the interaction effects of Exposure with Aspect and Group



Model formula: PropOE ~ Aspect*Group*ExpCroation_sc + (1 + Aspect | Participant) + (1 + Aspect*Group+ExpCroation_sc | item)

References

[1] Van Dijk, C., Van Wonderen, E., Koutamanis, E., Kootstra, G. J., Dijkstra, T., & Unsworth, S. (2022). Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis. *Journal of Child Language*, *49*(5), 897–929.

[2] Polinsky, M. (2018). Without aspect. In G.G. Corbet and Noonan (Eds.), *Case and Grammatical Relations* (pp. 263–282). John Benjamins.

[3] Andreou, M. & Tsimpli, I. (2017). Aspectual distinctions in the narratives of bilingual children. International *Review of Applied Linguistics in Language Teaching*, *55*(3), 305–324.

[4] Minor, S., & Mitrofanova, N. (2024, May 23–24). Crosslinguistic influence in child bilingual acquisition: A Visual World eye-tracking study on grammatical case and aspect processing by German-Russian and Spanish-Russian bilinguals [Talk]. International Symposium on Bilingual and L2 Processing in Adults and Children (ISBPAC 2024), Swansea, UK.

[5] De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022). *Quantifying Bilingual EXperience (Q-BEx): questionnaire manual and documentation.*

[6] Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect processing. *PLoS ONE*, *17*(2), e0264132.

[7] McLachlan, G. J., Lee, S.X., & Rathnayake, S.I. (2019) Finite Mixture Models*. Annual Review of Statistics and Its Applications, 6*, 355–78.

[8] Figueroa-Zúñiga, J. I., Arellano-Valle, R. B., & Ferrari, S. L. P. (2013). Mixed beta regression: A Bayesian perspective. *Computational Statistics & Data Analysis*, *61*, 137–147.

[9] Salaberry, R. (2000). *The development of past tense morphology in L2 Spanish*. Philadelphia: Benjamins.

Semi-automatic selection of semantic substitutes in sentence comprehension stimuli

Cassandra L. Jacobs[1], Loïc Grobol[2], Alvin Tsang[1], & Ryan J. Hubbard[3]
1: University at Buffalo, SUNY; 2: Université Paris Nanterre; 3: University at Albany, SUNY
cxjacobs@buffalo.edu

Background: Hierarchical predictions about the linguistic future may support rational language comprehension in cases of uncertainty (Luke & Christianson, 2016). For example, the amplitude of the N400 ERP components following the sentence preamble, "They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of ___", was lowest (most positive) to the **expected** (E) ending *palms*, highest to the **between-category** (BC) word *tulips*, and medial for **within-category** word (WC) *palms*, highlighting the role of semantic category structure on sentence processing. Stimulus creation, however, is typically done by hand, and the determination and validation of category boundaries remains an unsolved problem in sentence stimulus creation. Here we present a data-driven method that identifies semantically related words for sentence processing studies by learning semantic clusters on the basis of patterns of productions in sentence-final cloze tasks.

Method: We produce clusters of sentence-final cloze responses from two datasets of both American English (AmEng; 3085 sentences; Peelle et al., 2020) and Standard Continental French (Fr; 403 sentences; Brunel et al., 2025). For clustering we apply Bayesian Gaussian mixture models over context-specific word vectors for each cloze response using Transformer encoders (Liu et al., 2019; Martin et al., 2020) following our previous work (Jacobs et al., 2025). The inferred clusters partly encode frequency, concreteness, part-of-speech, as well as semantic category structure (e.g., fruits). In Experiment 1, we evaluate the clustering behavior for whether the terms belonged to the same basic-level category or not (Rosch et al., 1976) against the Federmeier and Kutas (*F&K99;* 1999) stimuli and in Experiment 2, we test whether the cloze response category labels can be applied to create new stimulus sets.

Experiment 1: Validation of within- and between-category structure. We follow the stimulus categorization of Federmeier and Kutas (1999) for the AmEng stimulus set for which no category labels have been inferred. As their stimuli were nouns drawn from different semantic categories, we estimate clusters using only noun responses in the AmEng corpus and then assign category labels. Of the F&K99 sentence preambles, 61 of 132 (46%) were correctly distinguished along within- and between-category boundaries. Representative errors are shown in Table 2. This highlights the validity of the method for stimulus selection.

Experiment 2: Semi-automated creation of within- and between-category foils. Given a clustering C, we heuristically selected a word E as the most probable word in the most probable cluster c_1, WC as the second-most probable word in c_1, and BC as the top word from the second-most probable cluster c_2, reasoning that E and WC, but not BC, should belong to the same category. We rated a random sample of 50 preambles from the AmEng and Fr stimulus sets and found that 80% of E-WC pairs were rated as same-category, as opposed to 32% of between-category (E, BC) terms, showing sensitivity that may aid in stimulus creation. We present representative stimuli created by this process in Table 1.

Discussion: In light of the above results, we argue that our method can empirically separate within- and between-category words in stimulus designs that manipulate lexical semantics. Our method shows promise to help researchers identify suitable, data-driven semantic neighbors for sentence processing studies. Furthermore, additional gains stand to be made by integrating linguistic resources and ontologies.

Table 1A. English Sentences

Sentence	Expected	Within-Category	Between-Category
Cara went to the pet store to buy her new	cat	dog	parrot
From the mountaintop, Wyatt could see the whole	town	city	valley
Phil restarted the stalled lawn mower with a strong	pull	tug	hum
The baker said those lemons were especially	sour	tart	good

Table 1B. French Sentences

Sentence	Expected	Within-Category	Between-Category
Le lapin s'est caché dans son	terrier	trou	chapeau
L'origami est un art du	pliage	design	papier
Il est le chef de son gang, c'est le	boss	leader	meilleur
Le thé avait refroidi, il est devenu	froid	tiède	imbuvable

Table 2. Example errors in cluster assignment in Federmeier and Kutas (1999) sentences

Preamble	Expected	Within-Category	Between-Category	Error type
One fell off her blouse and got lost, and she didn't have any extras. She ended up searching all over town to find a matching	button (c_{18})	zipper (c_7)	jar (c_{39})	All Unique (16/132)
Our cat is lazy and fat, and his favorite food is lasagna. Everyone comments on his resemblance to	Garfield (c_7)	Snoopy (c_4)	Batman (c_7)	E and BC match (9/132)
Pablo wanted to cut the lumber he had bought to make some shelves. He asked his neighbor if he could borrow her	saw (c_{32})	hammer (c_7)	rake (c_7)	WC and BC match (16/132)

References

[1] Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22-60. **[2]** Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495. **[3]** Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behavior Research Methods*, *52*, 1795-1799. **[4]** Brunel, J., Dujardin, E., Delord, S., & Mathey, S. (2025). Emotional valence, cloze probability, and entropy: Completion norms for 403 French sentences. *Behavior Research Methods*, *57*, 81. **[5]** Martin, L., Muller, B., Suarez, P. O., Dupont, Y., Romary, L., De La Clergerie, É. V., ... & Sagot, B. (2020, July). CamemBERT: a Tasty French Language Model. In *Proceedings of ACL* (pp. 7203-7219). **[6]** Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. **[7]** Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

The absence of D-linking effects in Cantonese Wh-islands

Chi Hou Pau, Grant Goodall
University of California, San Diego
chpau@ucsd.edu

Background: There has been a debate on what the nature of island constraints is. D-linking effects in wh-islands provide an opportunity to examine this issue. It is observed that the use of D-linked wh-phrases ameliorates wh-island effects as demonstrated in the contrast between (1) and (2). Syntactic accounts ascribe the D-linking effects to Featural Relativized Minimality (fRM) [1] whereas processing accounts ascribe it to reduced similarity-based interference [2]. Alternatively, semantic accounts explain the D-linking effects in terms of semantic incompatibility [3]. While these accounts predict the presence of D-linking effects in wh-ex-situ languages, their predictions about D-linking effects in wh-in-situ languages diverge. For syntactic accounts, assuming a covert movement analysis [4], fRM predicts the presence of D-linking effects as D-linking reduces the degree of morphosyntactic feature overlap as in (3). For processing accounts, processing an in-situ wh-phrase involves a cue-based retrieval to search for its scope position [5] and the scope position of an embedded question induces similarity-based interference. Since D-linking does not alter the retrieval cues associated with the scope positions, it is irrelevant to the degree of similarity-based interference induced as illustrated in (4). Hence, processing accounts predict the absence of D-linking effects. Semantic accounts predict the presence of D-linking effects as the semantic interpretation of wh-phrases is the same regardless of being in-situ or not. The current study examines the D-linking effects in Cantonese wh-islands using an acceptability experiment.

Method: The experiment applies a 2x2x2 (*Dependency Length* x *Structure Type* x *Wh-phrase Type*) design as shown in (5). Question-answer pairs were presented to ensure the wh-question interpretation (Details explained in Appendix). There are four question/answer pairs per condition, resulting in 32 experimental items. Participants were asked to rate each of the question/answer pairs on a scale from 1 - 7.

Results: After filtering inattentive participants, the data collected (N=44) were z-transformed (plotted in (6)) and then fitted with a linear mixed model in (7). The interaction effect between *Dependency Length* and *Structure Type* is estimated to be -0.87992 ($p<2e^{16}$), confirming the presence of wh-island effects. Importantly, the estimation of the three-way interaction does not reach statistical significance (p=0.5096), suggesting the absence of D-linking effects in wh-islands. This finding is also supported by the result of a likelihood ratio test (Chisq=0.4321, p=0.511) comparing models with/without the three-way interaction shown in (8).

Discussion: The results are in favour of processing accounts and against syntactic accounts and semantic accounts. Accounting for wh-island effects in terms of similarity-based inference can explain why D-linking effects are present in wh-ex-situ language but absent in wh-in-situ language, supporting the idea that wh-island effects stemmed from constraints on the retrieval mechanism in working memory. Using wh-islands as an example, this study demonstrates different languages utilize the same mechanism to process wh-dependencies and certain types of island constraints can be reduced to more general constraints on language processing mechanism. This might help to explain why island constraints can be acquired even without negative evidence in the inputs.

(1) *What$_i$ does Tom wonder who repaired _$_i$?

(2) ?Which computer$_i$ does Tom wonder who repaired _$_i$?

[uWH][uN] [+WH][+N] [+WH]

(3) ___ go3 ji1sang1 soeng2zi1dou6 bin1go3 saan3bo3-zo2 bin1zek3 beng6duk6 aa3?

CL doctor wonder who spread-PFV which virus Q

"Which virus does the doctor wonder who spread?"

[+C][+Q] [+C][+Q]

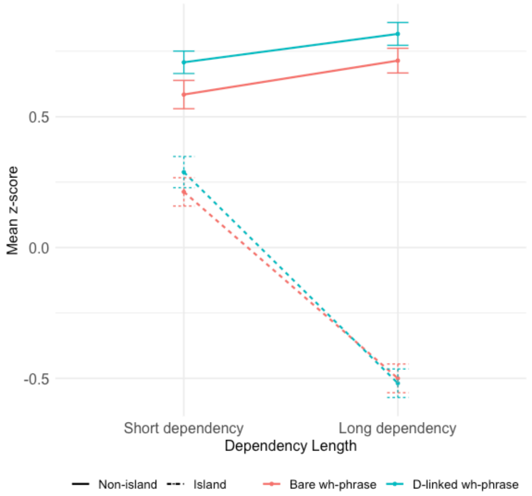
(4) [$_{CP}$go3 ji1sang1 soeng2zi1dou6 [$_{CP}$bin1go3 saan3bo3-zo2 bin1zek3 beng6duk6] aa3]?

CL doctor wonder who spread-PFV which virus Q

"Which virus does the doctor wonder who spread?"

(5)

	Short Dependency	***Long Dependency***
Non-island	{bin1go3 ji1sang1/bin1go3} jing6wai6 siu2waa4 saan3bo3-zo2 zek3 {which doctor /who } think Siuwaa spread -PFV CL beng6duk6 aa3? virus Q "{Which doctor/Who} thinks Siuwaa spread the virus?"	go3 ji1sang1 jing6wai6 siu2waa4 saan3bo3-zo2 {bin1zek3 beng6duk6 CL doctor think Siuwaa spread -PFV {which virus / mat1je5} aa3? /what } Q "{Which virus/What} does the doctor think Siuwaa spread?"
Island	{bin1go3 ji1sang1/bin1go3} soeng2zi1dou3 bin1go3 saan3bo3-zo2 {which doctor /who } wonder who spread -PFV zek3 beng6duk6 aa3? CL virus Q "{Which doctor/Who} wonders who spread the virus?"	go3 ji1sang1 soeng2zi1dou3 bin1go3 saan3bo3-zo2 {bin1zek3 CL doctor wonder who spread -PFV {which beng6duk6 / mat1je5} aa3? virus / what } Q "{Which virus/What} does the doctor wonder who spread?"
Answer	ji1zi6 lau4gam2 go2 go3 ji1sang1 treat flu that CL doctor "The doctor who treats flu."	dou6zi3 tou5o1 go2 zek3 beng6duk6 cause diarrhea that CL virus "The virus which causes diarrhea."



(6)

(7) z-score ~ Wh-phrase Type * Structure Type * Dependency Length + (1 + Structure Type | Participant) + (1 + Dependency Length | Item)

(8) z-score ~ Dependency Length * Structure Type + Dependency Length * Wh-phrase Type + Structure Type * Wh-phrase Type + (1 + Structure Type | Participant) + (1 + Dependency Length | Item)

References

[1] Villata, S. (2017). *Intervention effects in sentence processing* [Ph.D.]. Univ. Genève.

[2] Atkinson, E., Apple, A., Rawlins, K., & Omaki, A. (2016). Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology*, *6*, 167105.

[3] Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics*, *1*(3), 235–284.

[4] Hunag, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Ph.D.]. Massachusetts Institute of Technology.

[5] Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies—The case of Mandarin wh-in-situ dependency. *Journal of Memory and Language*, *84*, 139–166.

Specificities of Cantonese

Cantonese is a wh in-situ language. In a Cantonese wh-question, instead of moving to the sentence initial position like in English, the wh-phrase stays in its original position. This is demonstrated in the following two examples.

1.	keoi5	maai5-zo2	bou6	din6waa2
	he	buy-PFV	CL	phone
	'He bought a telephone.'			
2.	keoi5	maai5-zo2	mat1je5	aa3
	he	buy-PFV	what	Q
	'What did he buy?'			

In Cantonese, wh-questions are indicated by the sentence-final particle *aa3* whereas yes/no questions can be signalled by the sentence-final particle *aa4*. The two sentence-final particles share the same written form "啊". Since the stimuli are presented in Chinese characters, the questions are ambiguous between being a wh-phrase or a yes/no question. Stimuli in *Long Dependency x Island* can be interpreted as ungrammatical wh-questions as in 3 or grammatical yes/no question with an embedded indirect multiple wh-question as in 4.

3.	*keoi5	soeng2zi1dou3	bin1go3	maai5-zo2	mat1je5	aa3
	he	wonder	who	buy-PFV	what	Q
	Intended: 'What does he wonder who bought?'					
4.	keoi5	soeng2zi1dou3	bin1go3	maai5-zo2	mat1je5	aa4
	he	wonder	who	buy-PFV	what	Q
	'Does he wonder who bought what?'					

By pairing the question with a definite NP as the answer, the participants are forced to interpret the stimuli as wh-questions to ilicit the wh-island effects.

EEG time-frequency analysis of syntactic unification in Cantonese and English

Junghwan Maeng, Hyun Kyung Rachel Lee, Samuel Sui Lung Sze, Zhiyin Zhu, Yoonsang Song
The University of Hong Kong
Corresponding Author: Yoonsang Song (yoonsang@hku.hk)

This study investigates how cross-linguistic differences influence the syntax-semantics interface during sentence comprehension. Specifically, we examined whether syntactic unification—the process of combining syntactic units into a sentence representation—operates independently of semantics in Cantonese, a major Chinese language, as it does in Indo-European (IE) languages such as English and Dutch, which feature richer inflection and more rigid word categories (e.g., Bastiaansen & Hagoort, 2015). Research on IE languages has highlighted the independence of syntax from semantics, demonstrating that syntactic unification must be completed irrespective of semantic content before semantic integration can occur. In contrast, extensive research on Chinese suggests that limited morphosyntactic cues may lead to greater engagement of semantic processing during sentence comprehension, with semantic integration proceeding independently of syntactic unification (e.g., Yu & Zhang, 2008). While these findings point to cross-linguistic variance in the degree to which semantics depends on syntax, the question remains whether syntactic unification itself is universally independent of semantic content.

To address this gap, we employed time-frequency analysis to track increased beta-band synchronization in EEG, a neural signature shown to effectively segregate syntactic unification from semantic processes when paired with appropriate experimental designs (e.g., Bastiaansen & Hagoort, 2015). We compared power in the low beta band (15-20 Hz; Segaert et al., 2018) across Correct sentences (both syntactically and semantically well-formed), Prose sentences (syntactically well-formed but semantically anomalous), and Random sentences (non-syntactic word strings) in Cantonese and English (see Table 1). Sentence stimuli were matched for syntactic complexity across languages. In English, an IE language, we predicted increased beta power for both Correct and Prose sentences relative to Random sentences, reflecting syntactic unification independent of semantics. In contrast, in Cantonese, we expected increased beta power only for Correct sentences, indicating greater reliance on semantic information during syntactic unification, due to the limited and less reliable syntactic cues.

We recorded and analyzed EEG data from 29 native Cantonese speakers and 30 native English speakers, who read the three types of sentences for comprehension in their native language. As shown in Figure 1, English results revealed significant beta power increases for both Correct and Prose sentences relative to Random sentences, with no difference between Correct and Prose sentences. This pattern suggests that syntactic unification occurred for both Correct and Prose sentences in English. In contrast, Cantonese results showed reliable beta power increases for Correct sentences relative to both Prose and Random sentences, with no difference between Prose and Random sentences. This pattern reflects that syntactic unification was engaged only for Correct sentences in Cantonese.

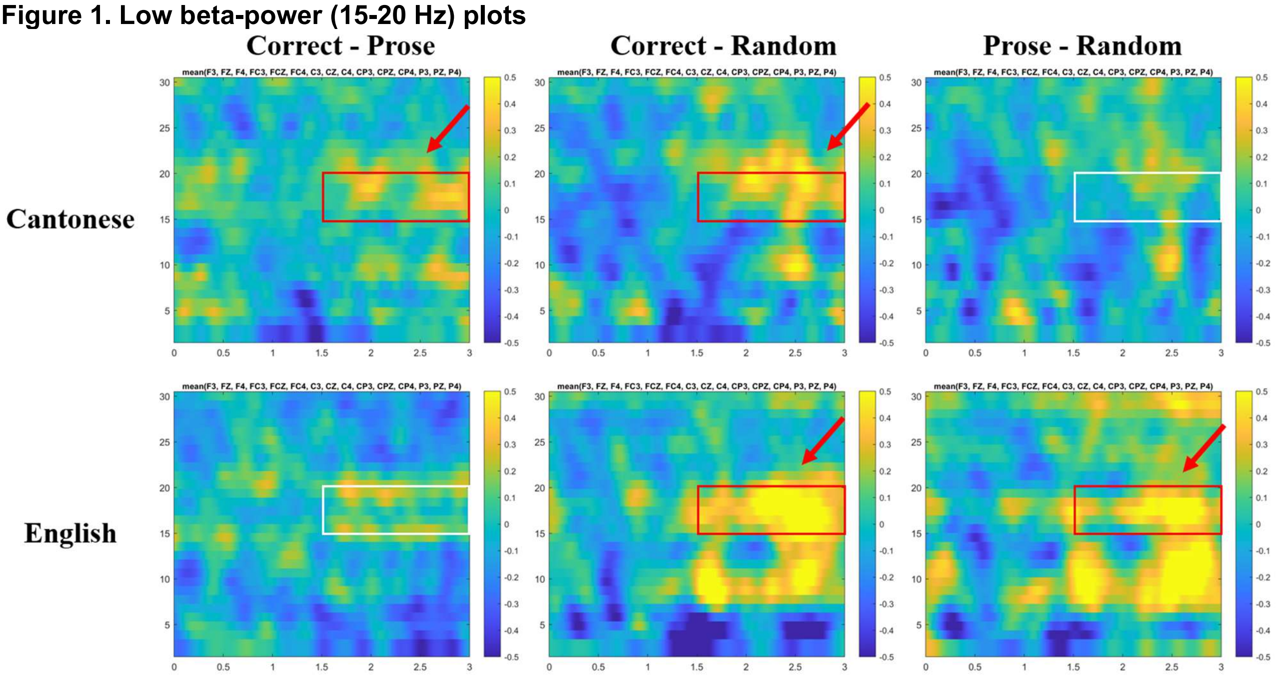
Taken together, these findings suggest that the independence of syntactic unification from semantics is influenced by the amount and reliability of syntactic cues. While parsers for languages with relatively rich and reliable syntactic cues, such as English, can construct syntactic structure independently of semantic content, parsers for languages with sparse or ambiguous syntactic cues, such as Cantonese, appear to rely more heavily on semantic interpretation to achieve successful syntactic unification.

Table 1. Example Stimuli

Condition	Cantonese (Structure: Det. + Adj. + Noun + Prep. + Det. + Noun + Verb + Det. + Adj. + Noun)
Correct	呢位 溫柔嘅 老師 為 嗰班 學生 講咗 呢個 感人嘅 故事 *neiwai wanjauge lousi wai gobaan hoksaang gongzo neigo gamjange gusi* this-CL gentle teacher for that-CL student tell-ASP this-CL touching story 'This gentle teacher told this touching story for the students.'
Prose	呢位 溫柔嘅 理由 為 嗰班 泳池 講咗 呢個 感人嘅 心情 *neiwai wanjauge leijau wai gobaan wingci gongzo neigo gamjange samcing* this-CL gentle reason for that-CL pool tell-ASP this-CL touching mood 'This gentle reason told this touching mood for the pool.'
Random	*呢位 講咗 心情 為 理由 呢個 泳池 溫柔嘅 嗰班 感人嘅 *neiwai gongzo samcing wai leijau neigo wingci wanjauge gobaan gamjange* this-CL tell-ASP mood for reason this-CL pool gentle that-CL touching
	English (Structure: Det. + Adj. + Noun + Verb + Det. + Adj. + Noun + Prep. + Det. + Noun)
Correct	The popular actress met a persistent reporter outside the trailer.
Prose	The popular cotton met a persistent equipment outside the trailer.
Random	*cotton popular the persistent a met outside the equipment trailer

Note. Det. = Determiner; Adj. = Adjective; Prep. = Preposition

Figure 1. Low beta-power (15-20 Hz) plots



Note. Each box in the plot marks a region of interest. Red boxes with arrows indicate significant differences; white boxes indicate insignificant differences.

References

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2095–2107. https://doi.org/10.1162/jocn_a_00829

Segaert, K., Mazaheri, A., & Hagoort, P. (2018). Binding language: structuring sentences through precisely timed oscillatory mechanisms. *European Journal of Neuroscience*, 48(7), 2651–2662. https://doi.org/10.1111/ejn.13816

Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *NeuroReport, 19*(7), 745–749. https://doi.org/10.1097/WNR.0b013e3282fda21d

The Production of Coercion in Japanese: Evidence from Priming

Mikihiro Tanaka
Ritsumeikan University, Japan
miki-t82@fc.ritsumei.ac.jp

Background: To effectively convey a message, a speaker must draw upon both semantic and syntactic information to construct a sentence. Numerous studies have investigated how this information is processed during sentence production in English. For example, Raffray et al. (2014) found that participants were more likely to describe target pictures using coerced expressions (e.g., "The celebrity began the champagne") after describing prime pictures with similar coerced expressions, as opposed to non-coerced ones (e.g., "The celebrity began drinking the champagne"). They also observed that this effect was absent when the verbs were not repeated between the primes and targets. In addition to coercion priming, Raffray et al. identified structural priming (e.g., Bock, 1986): participants were more inclined to describe target pictures with VP structures (e.g., began drinking the champagne) after describing prime pictures with VP structures rather than non-VP structures (e.g., began the champagne). Although these findings are robust, it remains uncertain whether such effects are present in other languages.

Method: The current study presents two psycholinguistic experiments (Ferreira, 2003) employing a recall-based priming paradigm to examine semantic and structural priming effects in Japanese sentence production. In this task, participants memorized a prime and filler sentence, recalled the prime sentence, then memorized the target and filler sentence, and recalled the target sentence. The prime forms were manipulated as follows: (1-a) sentences with coerced expressions, (1-b) non-coerced expressions, and (1-c) event-VP. The target fragments consistently involved coerced expressions (2). Experiment 1 (30 participants) utilized the same coerced verbs between prime and target, whereas Experiment 2 (27 participants) did not (Example (3) was the target fragment).

Results: The results of Experiment 1 demonstrated that participants recalled sentences with coerced expressions more accurately after recalling coerced expressions, compared to non-coerced or event-VP expressions (e.g., "The celebrity began the interview"). However, as Raffray et al. observed, Experiment 2 did not confirm the priming effect of coercion. Furthermore, across all experiments, when controlling for coerced structures, participants were more likely to recall the target sentence with a VP structure (non-coerced expressions) than with non-VP structures (coerced or event-VP structures).

Discussion: In summary, these results support the distinct processing of semantic and syntactic representations during sentence production, indicating that semantic and syntactic information is partially abstract in nature. Moreover, this suggests that the observed effects were not language-dependent, representing instances of conceptual and structural priming in both English and Japanese.

References

[1] Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

[2] Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying “that” is not saying “that” at all. *Journal of Memory and Language*, *48*(2), 379–398. https://doi.org/10.1016/s0749-596x(02)00523-5

[3] Raffray, C. N., Pickering, M. J., Cai, Z. G., & Branigan, H. P. (2014). The production of coerced expressions: Evidence from priming. *Journal of Memory and Language*, *74*, 91–106. https://doi.org/10.1016/j.jml.2013.09.004

Tables

(1) Experiment 1

Coercion priming			
Prime/Target	coerced	non-coerced	errors
coerced	198 (96.1)	8 (3.9)	58
VP	183 (89.7)	21 (10.3)	60
event vp	197 (95.6)	9 (4.4)	58

Syntactic Priming		
Prime/Target	VP	non-VP
coerced	3 (1.4)	203 (98.6)
VP	13 (6.4)	191 (93.6)
event vp	6 (2.9)	200 (97.1)

(2) Experiment 2

Coercion priming			
Prime/Target	coerced	non-coerced	errors
coerced	158 (95.0)	8 (4.8)	53
VP	163 (93.2)	12 (6.8)	45
event vp	167 (94.4)	10 (5.6)	45

Syntactic Priming		
Prime/Target	VP	non-VP
coerced	2 (1.2)	164 (98.8)
VP	6 (3.4)	169 (96.6)
event vp	1 (0.6)	176 (99.4)

Examples

(1) Prime conditions

a.ryourinin-ga hiyashi-chuka-o hajimeta
  cook-nom cold noodle-acc begin-past
The cook began the cold noodle.

b.ryourinin-ga hiyashi-chuka-o uri-hajimeta
  cook-nom cold noodle-acc sell-begin-past
The cook began selling the cold noodle.

c.ryourinin-ga interview-o hajimeta
  cook-nom interview-acc begin-past
The cook began the interview.

(2) Target Condition in Experiment 1

isha-ga chuusha-o hajimeta
Doctor-nom injection-acc begin-past
The chef began the injection. (intended: began giving the injection).

(3) Target Condition in Experiment 2

Jaanaristo-ga kiji-o oeta
Journalist-nom article-o finish-past.
The journalist finished the article. (intended: finished writing the article)

The role of attentional resources on errors and disfluency in speech production across different degrees of speech rate restriction.

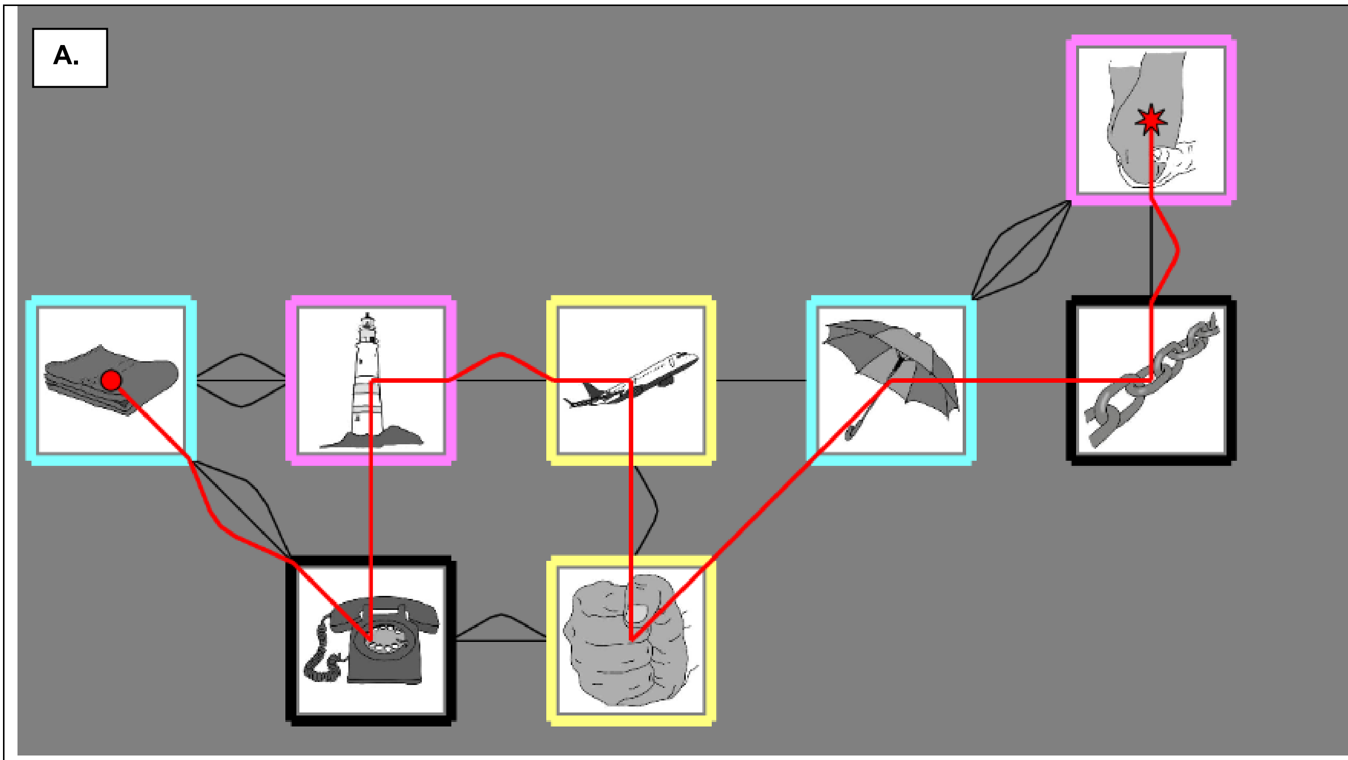
1 Bram De keersmaecker[1], Robert Hartsuiker[1], Aurélie Pistono[2]

[1] Department of Experimental Psychology, Ghent University; [2] University of Toulouse

bram.dekeersmaecker@ugent.be

In two conducted experiments using the network task[1] (analysis is ongoing and will be finalized before the onset of AMLaP 2025), we aim to test two accounts of disfluency and their opposing predictions regarding the influence of attentional resources on speech production. The covert repair hypothesis (CRH)[2] proposes disfluencies as repairs of self-detected speech errors. The CRH is nested in the perceptual loop theory of monitoring[3], which assumes that monitoring requires attention. Therefore, attrition of attentional resources would contribute to a decline in repairs and to an increase in undetected speech errors. On the other hand, the EXPLAN model[4] proposes planning and execution as independent processes. Under this account, disfluency arises when execution of one segment has finished before the plan for the next segment is finalized. By executing existing plans (i.e., stalling disfluencies such as filled pauses, prolongations and repetitions) speakers stall time for the planning process to catch up. Either slow planning, fast execution, or both would contribute to disfluency under this account. As planning is more likely to suffer from attentional load than execution, this account predicts more disfluencies with more load, unless they adapt execution (i.e., speech rate). To investigate the effect of attentional resources on disfluency we collected 10 connected-speech samples by speakers in both divided and undivided attention conditions of the network task. In E1, we instructed 30 speakers to describe each network as well as possible; they were free to choose and vary their speech rate if needed. In E2, 29 speakers received the same instructions but this time they were unable to choose and vary their speech rate as they had to match the pace of a moving indicator. In the divided attention condition, we used a novel dual task that relied solely on tactile perception; participants wore a wrist strap that vibrated intermittently with random intervals. Participants were asked to keep track of the number of buzzes. The secondary task was developed with the goals of reducing stimulus modality interference (tactile, not auditory or visual), and of ensuring that it does not consume resources related to overt production of any sort (perceptive, not productive). CRH predicts fewer repairs and more unrepaired errors when attention is divided, whereas EXPLAN predicts more stalling disfluencies (unless speakers slow down sufficiently). Next to the distinct predictions for the effects of attention, CRH and EXPLAN also differ in their predictions regarding speech rate and the flexibility thereof. According to EXPLAN, speakers are able to modulate their speech rate to keep control of the production process. This means that in E1, participants should slow down in the divided attention condition, and doing so would suppress any increase of stalling disfluencies that would result from the lack of attention. Subsequently, not being able to modulate their speech rate freely in E2 should promote any effects of attention on stalling disfluencies. On the other hand, CRH predicts that when speakers are able to modulate their speech rate, the monitor retains the ability to detect errors, provided that speakers slow down sufficiently. If this is true, we should see more repairs and less unrepaired speech errors in E1 compared to E2. For E1, we plan to test for possible effects of divided attention on speech rate, monitoring accuracy, and fluency. First, using a novel technique that yields a quasi-continuous speech rate measure, we will examine whether and how speakers adjust their speech rate over time as cognitive demand increases. Second, we will test whether the ability to detect and repair speech errors declines when less attentional resources are available. Third, we will test whether fluency is affected by a decrease in attentional resources. For E2, we plan the same analyses and will compare results with the first experiment to test the effect of fixed speed of execution processes on the presence of disfluencies and errors. In sum, we will use this novel approach to investigate how disfluency and speech errors behave in relation to each other and across different contexts of attentional resources and planning time.

Fig. 1. A. Example of a network. **B.** Flowchart of Experiments 1 & 2 procedure.



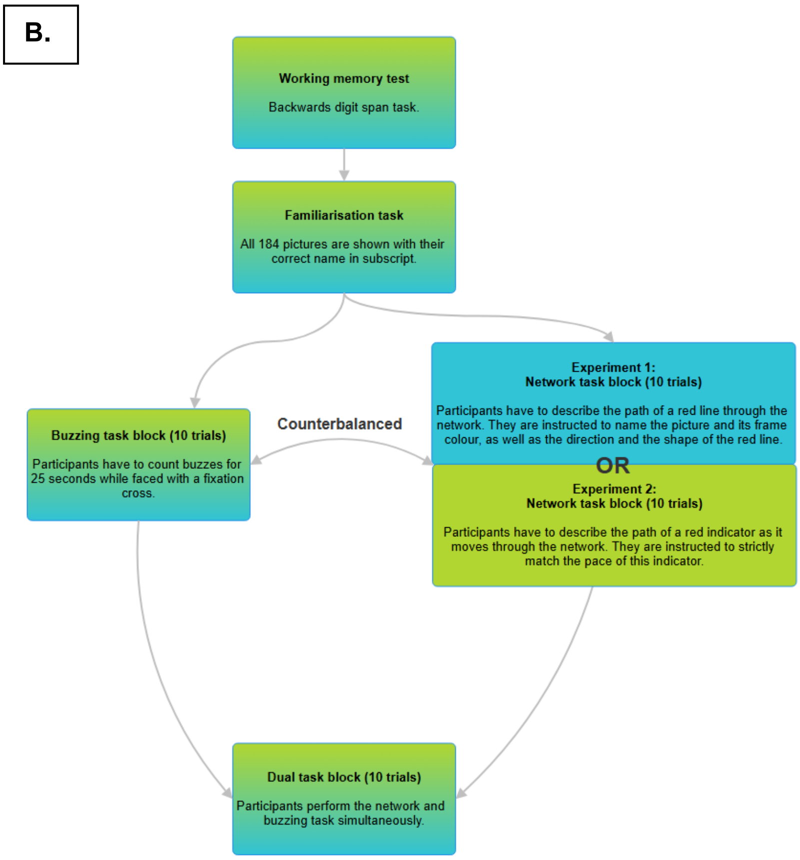
References

[1] Oomen, C. C. E., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. Journal of Speech, Language, and Hearing Research, 44(5), 997–1004.

[2] Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech, Language, and Hearing Research, 36(3), 472-487.

[3] Levelt, W. J. (1993). Speaking: From intention to articulation. MIT press.

[4] Howell, P., & Au-Yeung, J. (2011). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. John Benjamins Publishing Company.



This page may be used only for additional information about the generally less-known language you are targeting in the abstract.

NA

The pervasive role of linguistic knowledge in verbal fluency tests:

How individual differences in language skills shape the mental lexicon

Kyla McConnell* (1), Berit Reise (2) and Antje S. Meyer (1,3)

1 Max Planck Institute for Psycholinguistics, Nijmegen (NL)

2 University of Vienna, Vienna (AT)

3 Donders Centre for Brain, Cognition, and Behavior, Radboud University, Nijmegen (NL)

* kyla.mcconnell@mpi.nl

Background: "Name as many words that start with M as you can in one minute." The verbal fluency (VF) test seems simple, yet a typical university student only manages to produce about 15 words and there is substantial variability around this mean (SD: 5; current dataset, N = 571). What can we learn from these individual differences (IDs) and the cognitive and linguistic profiles that underlie them? We investigate how word retrieval is influenced by linguistic knowledge while controlling for working memory, processing speed, and nonverbal reasoning (IQ) [1,2]. In order to go beyond simple claims of whether linguistic knowledge improves overall VF scores, we zoom in on the differential demands of two common VF task types (semantic/category vs. phonemic/letter) as well as three scoring methods. Together, these factors allow us to look at temporal dynamics and to determine how linguistic knowledge affects lexical retrieval under different task demands.

Method: 571 native Dutch speakers aged 18-30 completed an extensive individual difference test battery, from which we calculated multi-test factor scores across six tests of linguistic knowledge and seven of domain-general skills (see Figure 1). Participants also completed 4 verbal fluency trials: 2 semantic trials (naming animals/foods) and 2 phonemic trials (listing words starting with "M" or "S"). Verbal fluency performance was quantified as the total number of correct words, as well as two temporal variables: the time to first response and the time by which half of the responses had been produced. These proxy initial lexical access efficiency and the declining rate of recall as previous productions are inhibited [3].

Results: There were generally positive correlations between better scores on individual difference measures and VF scores, ranging from weak to moderate ($< r = 0.22$) (see Figure 2). Linear mixed effects models confirmed these patterns. Importantly, although domain-general skills like processing speed and working memory were significant predictors of VF scores, linguistic knowledge consistently emerged as the strongest predictor. This was especially pronounced for letter fluency trials, where there was a significant main effect ($\beta = 0.10$, $p < 0.0001$)[1] and an interaction effect indicating that the effect was larger for sum scores ($\beta = 0.11$, $p < 0.0001$) compared to the time course variables (see Figure 3).

Discussion: Linguistic knowledge clearly has a strong impact on verbal fluency, and not only on the semantic trials (when higher vocabulary might lead to someone simply knowing more cue-appropriate words, e.g. capybara or axolotl), but also when retrieving words based on phonological/orthographic form. This is remarkable as adult native speakers certainly know many more than the small set of appropriate words they retrieve during the response period. The effect is reduced in the temporal variables, thus is not only an initial "burst" at the beginning of a trial when many words are easily accessible. Consistent with the Lexical Quality Hypothesis [4], this pattern suggests that greater linguistic knowledge has a pervasive effect on the mental lexicon, leading to more efficient and detailed associative links between words that share semantic or phonological/orthographic features.

Fig 1: Schematic representing the individual differences tests. One factor score was calculated for each construct using *lavaan*. Not pictured: IQ test (Raven's advanced progressive matrices).

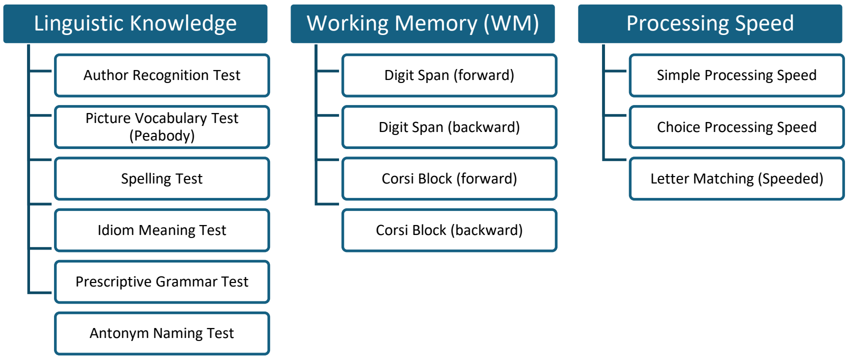
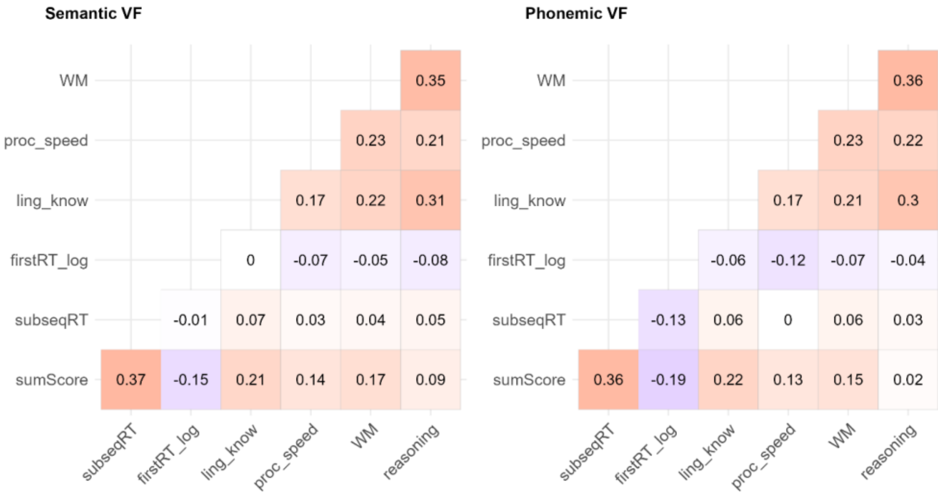


Fig 2: Pearson's correlations between IDs (WM, processing speed and linguistic knowledge) and scoring methods (first RT, subsequent RT and sum scores), in the semantic and phonemic VF trials, respectively.



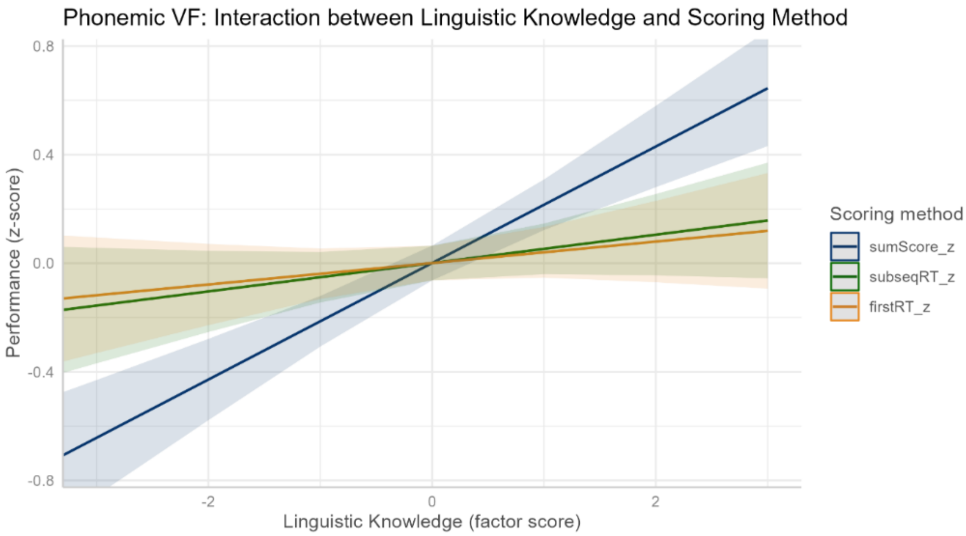
Semantic VF

	subseqRT	firstRT_log	ling_know	proc_speed	WM	reasoning
WM						0.35
proc_speed					0.23	0.21
ling_know				0.17	0.22	0.31
firstRT_log			0	-0.07	-0.05	-0.08
subseqRT		-0.01	0.07	0.03	0.04	0.05
sumScore	0.37	-0.15	0.21	0.14	0.17	0.09

Phonemic VF

	subseqRT	firstRT_log	ling_know	proc_speed	WM	reasoning
WM						0.36
proc_speed					0.23	0.22
ling_know				0.17	0.21	0.3
firstRT_log			-0.06	-0.12	-0.07	-0.04
subseqRT		-0.13	0.06	0	0.06	0.03
sumScore	0.36	-0.19	0.22	0.13	0.15	0.02

Fig 3: Interaction between linguistic knowledge factor score (z-score) and verbal fluency score (z-score), across the three scoring methods for phonemic VF only. (Log) first RTs have been harmonized so that higher scores represent better performance, as with the other scoring methods



References

1: Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. Applied Neuropsychology: Adult, 26(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

2: Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S., & Weis, S. (2020). Executive functions predict verbal fluency scores in healthy participants. Scientific Reports, 10(1), 11141. https://doi.org/10.1038/s41598-020-65525-9

3: Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00772

4: Perfetti, C. A., & Hart, L. (2008). The lexical quality hypothesis. In Precursors of functional literacy (pp. 189-213). John Benjamins Publishing Company.

Marking Aspect in Social Events:
The Hungarian Verbal Prefix *Meg-* Increases Perceived Mutuality

Anna Viola Sáfrány, Anna Kamenetski, Tiziana Srdoc, Attila Balla, Eva Wittenberg
Central European University, Language Comprehension Lab
safranya@ceu.edu

A growing body of psycholinguistic research has shown that grammatical aspect plays a significant role in how we mentally construe events [1–3]. Aspectual cues influence not only the construed temporal dynamics of an event (i.e., whether it is ongoing or completed), but also the salience of event components, such as locations, recipients or instruments [4–6]. While much work has focused on motion events, less attention has been given to how aspect might affect the construal of social events. This pre-registered study investigates whether grammatical aspect in Hungarian, marked with the verbal prefix *meg-*, modulates the interpretation of verbs denoting inherently social events, such as *házasodni* ("to marry") or *beszélni* ("to speak"), and ones that *can* be interpreted as social, such as *látogatni* ("to visit") or *írni* ("to write"). We hypothesized that, by lending telic boundedness to the verb and highlighting the end state of the event [7], *meg-* might increase the saliency of action patients or partners required for the completion of said event. For example, while the base verb *csókolni* ("to kiss", atelic infinitive) can refer to an ongoing or habitual action without a specified end state, the prefixed *meg-csókolni* (telic infinitive) marks the event as complete and requires a clearly individuated patient, thereby strengthening the interpretation of the event as one involving not only the agent (the kisser), but necessarily also the patient (the kissee). Building on these notions, we asked: Does the telic prefix *meg-* reinforce the interpretation of social event verbs as mutual, in comparison with their atelic counterparts?

Method 95 infinitive verbs selected from a Hungarian corpus [8] were paired up with *meg*-prefixed counterparts to study the aspectual effect. 51 native Hungarian speakers completed a web-based forced-choice task, identifying each verb as referring to either a self-directed or a mutual action (Figure 1). A Latin square design was used to ensure that each participant saw only one form of each verb. After compiling participants' ratings, stimuli were divided into four quartiles based on the base verbs' mean mutuality ratings—Quartile 1 (low mutuality), Quartile 2 (moderately low), Quartile 3 (moderately high), and Quartile 4 (high)—to test whether the effect varied with the inherent lexical semantics of base verbs. Specifically, we expected a ceiling effect for base verbs rated as highly mutual. **Results** Mutuality judgments were analyzed using a generalized linear mixed-effects model with fixed effects for verb type (base or prefixed), mutuality quartile, and their interaction. It included random intercepts for participants and verb pairs, and a by-participant random slope for verb type. *Meg*-prefixed verbs were overall significantly more likely to be judged as mutual than their base forms (β=0.8032, SE=0.1088, p<.0001). The verb type by quartile interaction was significant: the addition of *meg-* significantly increased the likelihood of mutual ratings for Quartiles 1-3, but not for Quartile 4, consistent with the predicted ceiling effect. The odds ratios were as follows: Q1=4.98 (SE=1.080, p<.0001), Q2=1.68 (SE=0.278, p=.0017), Q3=4.06 (SE=0.840, p<.0001), and Q4=1.17 (SE=0.299, p=.5423). Figure 2 shows the mean mutuality ratings for base and *meg*-prefixed verbs across quartiles. **Discussion** These results suggest that *meg-* shifts the construal of events toward a more mutual interpretation, especially when the social meaning is ambiguous. We assume that by marking the event as complete and bounded, the prefix *meg-* enhances cognitive focus on the involvement of a clearly individuated and necessary action recipient or partner, thereby reinforcing the implication of mutuality and social directedness.

Figure 1: Screenshot of a sample trial, showing the verb *meg-ünnepelni* ("to celebrate", telic infinitive).

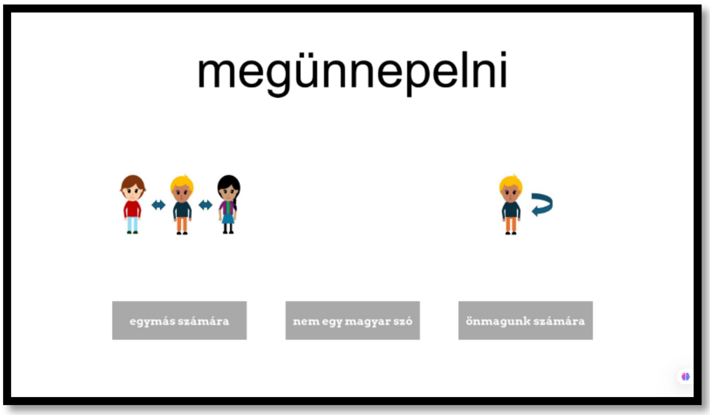
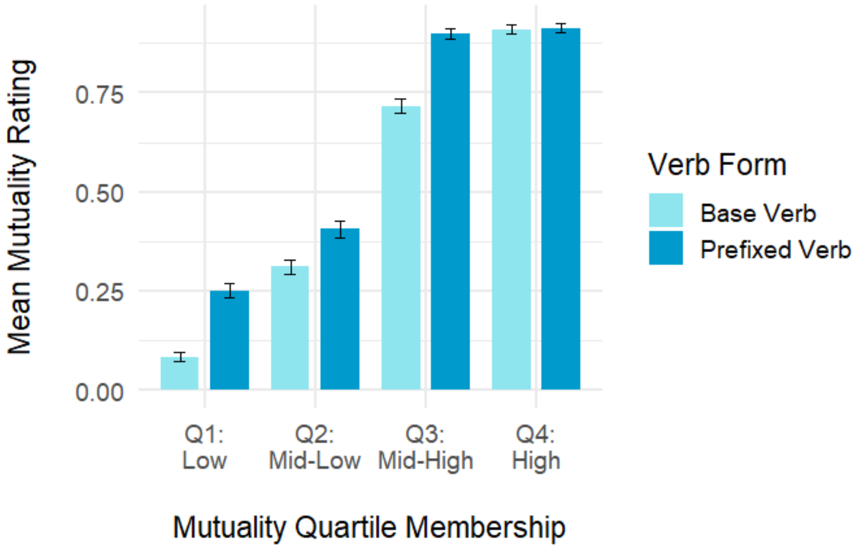


Figure 2: Mean mutuality ratings for base and *meg*-prefixed Hungarian verbs across mutuality quartiles.



References

[1] Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language, 112*(3), 150–158. **[2]** Romagno, D., Rota, G., Ricciardi, E., & Pietrini, P. (2012). Where the brain appreciates the final state of an event: The neural correlates of telicity. *Brain and Language, 123*(2), 68–74. **[3]** Becker, R. B., Ferretti, T. R., & Madden-Lombardi, C. (2013). Grammatical aspect, lexical aspect, and event duration constrain the availability of events in narratives. *Cognition, 129*(2), 212–220. **[4]** Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 182–196. **[5]** Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language, 61*(2), 191–205. **[6]** Madden-Lombardi, C., Dominey, P. F., & Ventre-Dominey, J. (2017). Grammatical verb aspect and event roles in sentence processing. *PLOS ONE, 12*(12), e0189919. **[7]** Szili, K. (2024). *The function of the Hungarian preverb meg- as a marker of lexical aspect / A meg- igekötő funkcióiról másképpen I–II: A meg- mint lexikaiaspektus-képző*. *Magyar Nyelvőr, 142*(2–3), 170–186, 428–435. **[8]** Srdoc, T., Marx, E., Safrany, A. V., Balla, A., & Wittenberg, E. (2025). *Linguistic Impact on Social Action Construal Database (LISADA).* https://osf.io/xzdqc/

Gender stereotype in auditory sentence processing: effects of talker and listener gender

Yun Feng[1], Shinyi Li[2], Ming Xiang[2], Yao Yao[1]
[1] The Hong Kong Polytechnic University; [2] The University of Chicago
yunfeng@polyu.edu.hk

Background: Understanding spoken language involves integrating linguistic messages with social information, such as the talker gender. Prior research [1] has shown that listeners predict and produce content that aligns with gender stereotypes, and that their comprehension can be disrupted when such predictions are violated [2, 3]—for instance, when women are described in traditionally male-dominated roles, or the reverse. However, the effects of talker gender and participant's own gender were not sufficiently separated in previous research, and the confusion could be exacerbated by the use of self-referential pronouns in the stimuli [1–3]. This study used an improved experimental paradigm [1] that aimed to disentangle potential effects of talker gender and participant gender on sentence processing.

Method: Test materials included 160 gender-neutral self-referring sentence prompts in English (Table 1). The prompts were presented auditorily using two AI-generated voices (1F, 1M). The gender typicality of the AI-voices was normed. 119 native English speakers (62F) recruited from Prolific completed the study online. Participants‘ task was to complete the sentences they heard in one or two words from the **talkers'** perspective.To invetigate whether human completions exhibited context-driven gender biases, we quantified the gender bias of responses using semantic embeddings. A set of gender-referential words comprising eighteen masculine (e.g., "gentleman") and eighteen feminine words (e.g., "woman") were used as benchmarks [4]. For each sentence prompt in each voice, cosine similarity was computed between female/male participants‘ completions (averaged embedding) and masculine/feminine reference embeddings. The difference in cosine similarities (masculine − feminine) served as the gender bias index, with positive values indicating masculine bias and negative values feminine bias. Two pre-trained models—Word2Vec[1] [5] and all-MiniLM-L6-v2 ("MiniLM")[2] [6]—were used for calculating semantic embeddings to ensure generalizability. Linear mixed-effects models were constructed on gender bias, with talker gender and participant gender as fixed factors and prompt item as a random factor.

Results: The female voice elicited significantly more feminine-oriented responses (i.e., more negative gender bias values; Table 2) than the male voice, supported by both models (*p*s < .001; Fig 1). Participant gender also influenced responses, but the effect is less strong (p = .003 for Word2Vec; p = .095 for MiniLM). In addition, both models revealed that male participants produced significantly less feminine completions than females in Female Voice (*p*s < .05), but not significantly more masculine than females in Male Voice (*p*s > .05).

Discussion: Both talker gender and participant gender influenced sentence prediction and production in the predicted directions, but talker gender appeared to have a stronger effect, as participants aligned responses to the talker's stereotypical gender. Regarding participants gender, compared to female participants, males were relatively less sensitive to talker gender, exhibiting less adaptability in producing feminine-oriented responses when hearing voices from the opposite gender, whereas females readily generated masculine-oriented responses to a male voice. Future studies will compare gender biases using pre-trained embeddings and human ratings, as well as explore cross-language and cultural differences.

[1] trained on GoogleNews; https://www.kaggle.com/datasets/leadbest/googlenewsvectorsnegative300
[2] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Table 1. Example of experimental stimuli

Talker gender	Participant gender	Example sentence prompts
Female Voice	Female/Male	My favorite hobby is _____.
Male Voice	Female/Male	My favorite hobby is _____.

Table 2. Mean and standard deviation of Gender Bias for 2 models

Model	Talker gender	Participant gender	Mean	SD
MiniLM	Female Voice	Female	-0.0069	0.0330
		Male	0.0115	0.0232
	Male Voice	Female	-0.0013	0.0288
		Male	0.0123	0.0214
Word2Vec	Female Voice	Female	-0.0168	0.0413
		Male	0.0164	0.0359
	Male Voice	Female	-0.0072	0.0381
		Male	0.0228	0.0389

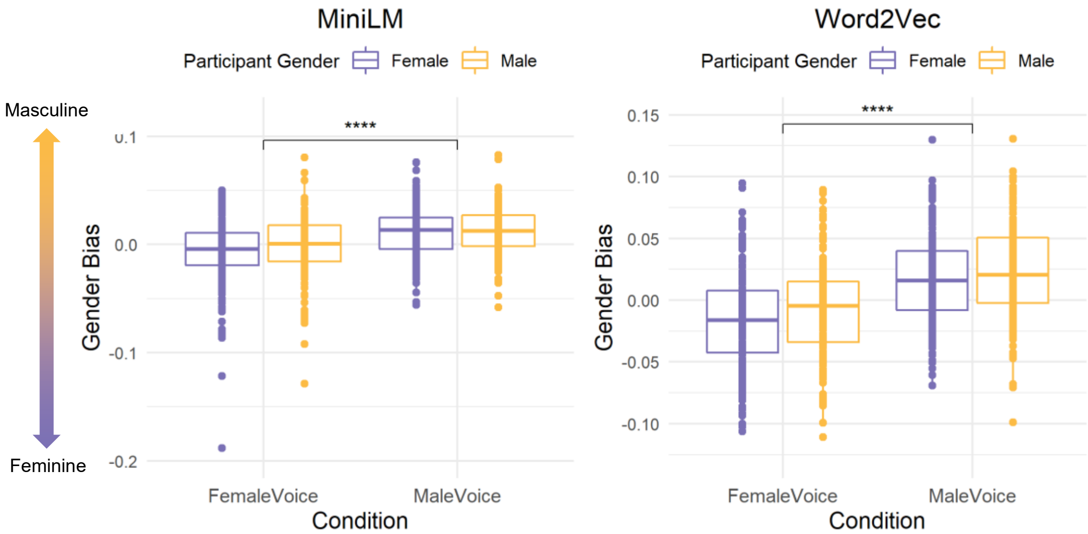


Figure 1. Gender bias from MiniLM and Word2Vec.

References

[1] Feng, Y., Shrestha, S., Xiang, M., Yao, Y. (2025). Integrating talker and message in language processing: the influence of talker gender on sentence prediction. *38th Annual Conference* on *Human Sentence Processing*, College Park, USA.

[2] Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience Letters*, *339*(3), 191-194.

[3] Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580-591.

[4] Nadeem, M., Bethke, A., & Reddy, S. (2020). StereoSet: Measuring stereotypical bias in pretrained language models. *arXiv preprint arXiv:2004.09456*.

[5] Mikolov, T. (2013). Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781, 3781.*

[6] Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems*, 33, 5776-5788.

Immediate sensitivity to thematic role constraints in a lexical decision task

Kate Stone[1], Milena Rabovsky[2], Henning Holle[1]

[1]University of Hull, [2]University of Potsdam

k.stone@hull.ac.uk

The N400 ERP component is usually large at context-incongruent words. However, evidence shows that the N400 is small at a context-incongruent verb in sentences where stereotypical thematic roles have been reversed, e.g. "the waitress that the customer served" (it is unusual for a customer to serve a waitress)[1:6]. There are several accounts of the small N400; we focus on two: The Sentence Gestalt N400 account (SG)[7] proposes that the small N400 in the role-reversed sentence results from uncertainty created by conflict between the literal sentence and knowledge about the usual thematic roles of customers and waitresses in the event "serve"; sensitivity to role information is therefore immediate. In contrast, the Slow Prediction hypothesis (SP)[8] proposes that verb prediction is not immediately constrained by thematic roles, which are assigned later. According to the SP account, the N400 at "served" in the role-reversed sentence is small because "served" is a plausible prediction based on the lexical association of waitress and customer, regardless of their thematic roles. Both accounts thus make similar predictions about the N400 via different mechanisms, making it difficult to distinguish between them using the N400 alone. As an alternative way to test for the presence of role constraints on early verb activations, we used a speeded lexical decision task in which participants responded yes/no to whether the target verb was a real English word.

Immediately after presentation of the two nouns (Ex.1), the SP and SG accounts make different predictions about response times at the verb "served" in role-reversed versus role-canonical sentences: The SP account should predict similar response times in both sentence types because, all else being equal, "served" should be just as activated if role constraints have not been applied. For the same reason, a verb that is consistent with customer-as-agent, such as "tipped", should elicit response times just as fast as "served", all else being equal. In contrast, the SG account predicts that sensitivity to role information already at "served" will trigger conflict in the reversed condition between the literal sentence and the expectation for waitress-as-agent, slowing response times relative to the canonical condition. Moreover, it would similarly predict a reversed vs. canonical slow-down for "tipped" and slower times for "tipped" overall versus "served". The "tipped" condition additionally acts as a cloze-matched control comparison (see Ex.1).

Methods. 100 online participants. 60 critical English sentences strongly predictive of the verb in condition (a) (associated with stronger N400 "illusion"[6]). 2×2 design: Role order (canonical/reversed) × target verb consistency with pre-target expectation (consistent/inconsistent). Sentences presented word-by-word to mimic ERP experiments; lexical decision at verb (Fig.1). 60 fillers: Same structure but weakly constraining; target pseudo-verbs.

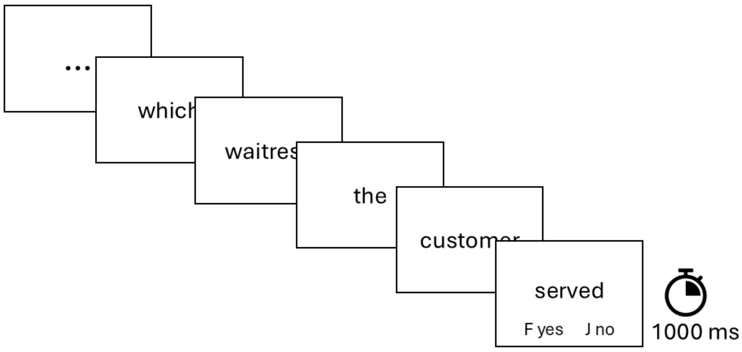
Results. See Fig.2 for full results. Bayesian linear mixed effects models including word length, log cloze probability, and word frequency as covariates revealed main effects of role order and verb consistency, but no interaction. Nested comparisons revealed a reversal effect within both consistent and inconsistent conditions. Allowing 200 ms for motor planning, mean response times were consistent with N400 time window processing.

Conclusions. The findings support the SG account where interpretation is immediately influenced by thematic roles, at least in strongly constraining sentences. Assuming that the differences in response times reflect the same processes underlying N400 amplitude, the findings may also explain why small by-participant effects are observed in N400 role reversal experiments when an N400 effect at the group level is not seen[6].

Example 1. Contexts were considered to have a strong, pre-target expectation for the verb in (a) if there was a moderate- to high-cloze verb given in the canonical condition of the cloze test that was also given in the reversed condition. The pre-target expectation-inconsistent verb was the most frequent, plausible verb completion in the reversed condition of the cloze test, or was selected by the researchers based on cosine similarity with the context if no appropriate verb was given.

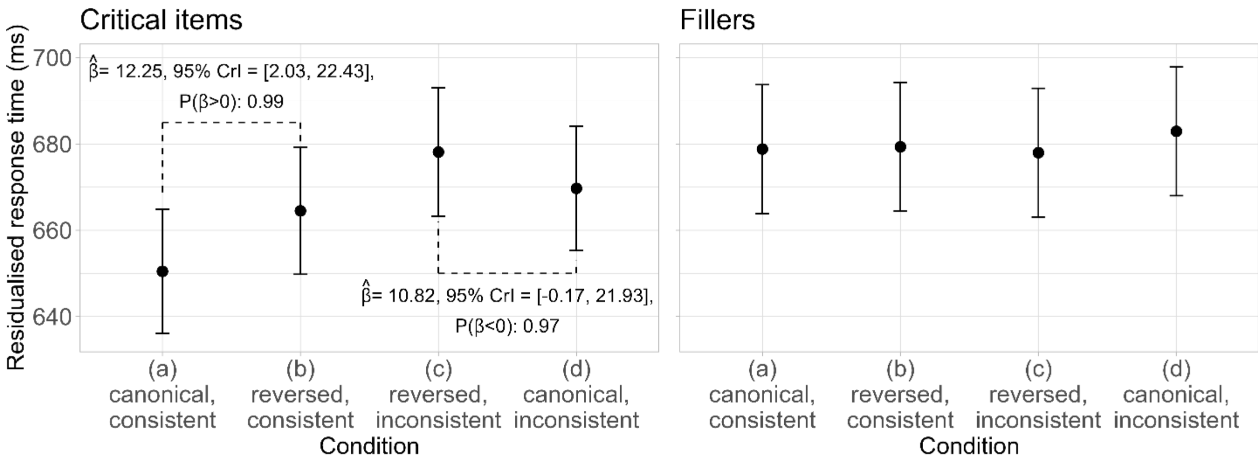
	Role order based on verb	Verb consistency with pre-target expectation	*Yesterday in the restaurant we saw…*	Cloze
(a)	Canonical	Consistent	…which customer the waitress served	0.55
(b)	Reversed	Consistent	…which waitress the customer served	0.14
(c)	Reversed	Inconsistent	…which customer the waitress tipped	0.00
(d)	Canonical	Inconsistent	…which waitress the customer tipped	0.06
Filler: The accountant checked which invoice the client / which client the invoice paied				-

Figure 1. Experimental paradigm. Participants were instructed to respond as quickly and accurately as possible as to whether the last letter string was a real English word or not within a 1000 ms response deadline. Feedback was given during practice trials ("correct/incorrect" or "too slow!").



References. [1] Kuperberg et al., 2003, *Cog Brain Res*; [2] Kim & Osterhout, 2005, *JML*; [3] Hoeks et al., 2004; [4] Chow et al., 2018, *Lang Cog Neurosci*; [5] Nakamura et al., 2024, *Cog Sci*; [6] Stone & Rabovksy (2025) *J Cog Neurosci*; [7] Rabovsky et al. (2018) *Nat Hum Behav*; [8] Liao et al. (2022) *JML*

Figure 2. Results of the nested comparisons overlaid on residualised response times. Not shown: The model of the 2×2 interaction revealed a main effect of consistency associated with slower responses for verbs that were inconsistent with the pre-target expected verb versus those that were expectation-consistent, $\hat{\beta} = 17.93,\ 95\%\ CrI = [6.65, 29.22], P(\beta > 0) = 0.99$, and a main effect of role order with a slow-down for verbs in the reversed versus canonical condition, $\hat{\beta} = 11.58,\ 95\%\ CrI = [5.56, 17.79], P(\beta > 0) = 0.99$. The model was not consistent with an interaction, $\hat{\beta} = -1.93,\ 95\%\ CrI = [-18.50, 14.78], P(\beta < 0) = 0.59$. The fillers were not analysed.



When “Mayor apologized citizens” becomes acceptable: ERP investigations on the transitive use of intransitive verbs in Mandarin

Chia-Hsuan Liao
National Tsing Hua University, Taiwan
chiahsuanliao@mx.nthu.edu.tw

Background: On social media, an increasing number of Mandarin intransitive verbs are being used transitively (e.g., "Mayor apologized citizens"). Linguists have defined these verbs as light verbs, proposing that they are underlyingly intransitive and become transitive through a syntactic operation [1, 2]. However, from a processing perspective, it remains unclear whether comprehenders mentally represent light verbs as transitive or continue to treat them as fundamentally intransitive. To address this gap, we conducted two ERP experiments comparing the processing profiles of light verbs with those of unaccusative verbs [3]—a class of intransitive verbs that can never be used transitively (e.g., *"Fear disappeared participants"). If light verbs have shifted toward a transitive representation, their processing profiles should differ from those of unaccusative verbs. Prior work has shown that violations of verb subcategorization elicit a P600 effect [4]. Accordingly, we expect a larger P600 with unaccusative verbs in transitive contexts, and examine whether light verbs elicit similar effects under the same conditions.

Method: Both experiments used a 2×2 design, crossing Verb Type (Light, Unaccusative) with Transitivity (Transitive, Intransitive). In Exp.1, each sentence began with an NP_1-Adverb-Verb-NP_2 sequence. Transitivity was indicated by punctuation: a comma after NP_2 in the Transitive conditions, and a comma after the verb in the Intransitive conditions, signaling that NP_2 was the subject of the following clause. To control for potential clause-boundary effects induced by punctuation, Exp.2 manipulated transitivity via lexical category: Nouns served as direct objects in the Transitive conditions, while temporal or frequency adjuncts followed the verb in the Intransitive conditions. Although lexical category differed across the post-verb targets, an interaction between the two factors should still be detectable if present. Participants (Exp.1: N = 36; Exp.2: N = 24) read sentences presented via RSVP (30 per condition, 800 ms/word) and performed a binary acceptability judgment after each trial. ERP responses were time-locked to verb onset and continued through the post-verb region.

Results: Behaviorally, intransitive verbs were judged equally acceptable across verb types. However, participants tended to reject transitive forms of unaccusative verbs but accept those of light verbs (Fig.1). The ANOVA analyses of the ERP results revealed a consistent P600 main effect of Transitivity at the post-verb region in both experiments, with no interaction with Verb Type (Figs. 2-3). Exp.1 additionally showed a LAN effect at the verb, possibly driven by design. Since all no-comma sentences were followed by direct objects, participants likely associated comma absence with transitivity. In Exp.2, a P200 main effect of Transitivity at the target may suggest early sensitivity to structural well-formedness during processing.

Discussion: Although participants behaviorally accepted Light verbs in transitive contexts, ERP results suggest their processing was not qualitatively different from that of Unaccusative verbs. This implies that light verbs may still be mentally represented as intransitive. The mismatch between behavioral and neural responses points to an early stage of syntactic change, where grammatical representations are slow to adapt. Future work could examine how individual differences, such as linguistic experiences and cognitive flexibility, influence the restructuring of argument structure.

Figure 1: Acceptability rate of each condition in Experiment 1 (Left) and Experiment 2 (Right).

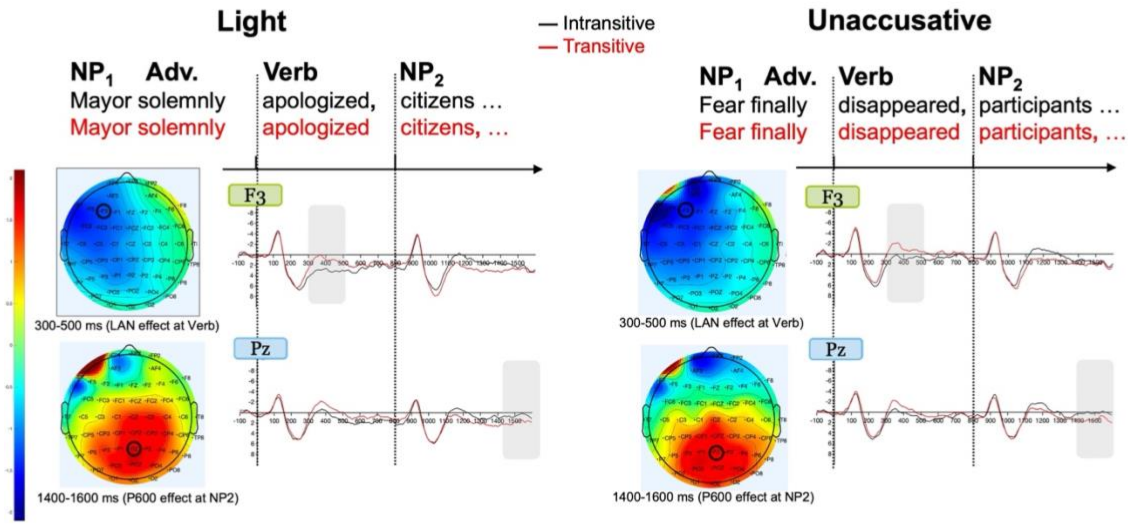


Figure 2: Grand average ERPs and their topographic distributions to the transitivity effect in Light verbs (Left) and Unaccusative verbs (Right) in Experiment 1.

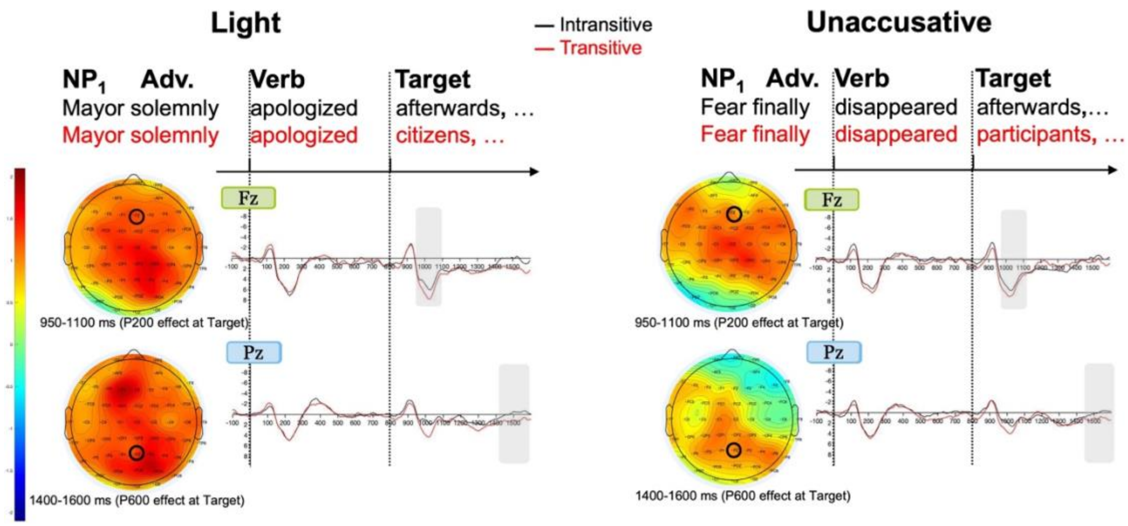


Figure 3: Grand average ERPs and their topographic distributions to the transitivity effect in Light verbs (Left) and Unaccusative verbs (Right) in Experiment 2.

References

[1] Tsai W.-T. (2017). Transitivization, Applicative Construction and Light Verb Analysis. *Contemporary Research in Modern Chinese 19,* 1-13.

[2] Lin, T.-H. J. (2001). *Light verb syntax and the theory of phrase structure.* Doctoral Dissertation. University of California, Irvine.

[3] Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society,* 157–1890. Berkeley, CA: Berkeley Linguistics Society, University of California, Berkeley.

[4] Kielar, A., Meltzer-Asscher, A., & Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia, 50*(14), 3320-3337.

Task-dependent neural modulation during sentential meaning computation

Yao-Ying Lai[1], Maria M. Piñango[2], Hiromu Sakai[3], Michiru Makuuchi[4]
[1] National Chengchi University, Taiwan; [2] Yale University, USA; [3] Waseda University, Japan;
[4] National Rehabilitation Center for Persons with Disabilities, Japan
yaoyinglai@gmail.com

Background: Neurolinguistic studies on sentence comprehension often employ post-trial tasks, yet their impact on neural activation remains unclear. Beyond the core semantic mechanisms, neural responses may vary by task due to top-down *semantic control*, which highlights certain conceptual aspects based on task demands [1]. For instance, explicit vs. implicit semantic tasks differentially engaged combinatorial vs. lexical-level processing in N-N composition [2]. The current study examines how task type affects cortical activation during the processing of sentences with compositionally underspecified iteration (e.g., "*The patient coughed for 2 minutes/2 months.*") vs. transparent counterparts (e.g., "…*napped for 2 minutes.*"). Using an iterativity judgment (**ITER**) task (determining how many times the event occurred), Lai et al. (2023) [3] found left inferior frontal gyrus (L.IFG) and parietal activation for the former, reflecting contextual meaning enrichment [4] and event numerosity computation [5]. Similar L.IFG activation was reported by Piñango et al. (2024) [6], whose post-trial task queried sentence content irrelevant to iterativity. Nuanced discrepancies between studies on semantic computation have been observed as well, hypothesized to result from task-dependent modulation. Accordingly, we address *how the nature of post-trial tasks modulates neural activation during the processing of semantically-underspecified sentences.* We first adopted Lai et al.'s stimuli but employed naturalness-rating (**NAT**) as the post-trial task. Compared to iterativity judgments that explicitly query the underspecified iteration, naturalness ratings are implicit about the target meaning while still requiring full comprehension. Combining Lai et al.'s data with ITER and our novel data with NAT, we performed a joint analysis to examine task effects.
Method: 32 sets of 4 conditions: *NoIter*, *PuncIter*, *DurIter*, and *GapIter* (Table 1) were used in both NAT and ITER fMRI experiments, with the key contrast being NoIter with transparent meaning vs. {PuncIter, DurIter, GapIter} with underspecified iteration. Each experiment included a distinct set of 40 fillers additionally. All sentences were normed in a prior questionnaire (n=22) to ensure naturalness and interpretations. Thirty right-handed native Japanese speakers were recruited for NAT and twenty for ITER (aged 18~35). They read sentences and performed either naturalness ratings (*1: very unnatural ~ 4: very natural*), or iterativity judgments for the denoted event (*once, 2~10 times, 11~100 times, > 101 times*). Except for the post-trial task, the NAT and ITER experiments were identical in procedure and paradigms (TR = 1 sec, TE = 30 ms). With outliers removed, data of 25 participants for NAT and 18 for ITER were analyzed. The fMRI analysis targeted the sentence-reading window, excluding the task period. The joint analysis identified 12 ROIs commonly activated for the *semantically-underspecified (PuncIter, DurIter, GapIter) > transparent (NoIter)* conditions across experiments. Regional activation (beta values) of the ROIs were compared to assess task differences.
Findings: For all critical contrasts against the NoIter baseline, the same sentences elicited stronger activation followed by the ITER task than NAT per ROI (Fig. 1). While both experiments engaged frontal regions supporting the contextualization of underspecified meaning, explicitly probing the target meaning (ITER) directed attention to the specific iterativity computation (compared to naturalness evaluation), enhancing neural activation. Results suggest that the nature of the post-trial tasks influences cognitive allocation during comprehension, modulating neural responses beyond the core combinatorial meaning computation.

Table 1: Sample stimuli of the 4 critical conditions shared by the two experiments (NAT & ITER)

Condition	Sample sentence
(A) NoIter Durative verb + Short interval	kaeru-ga 20 pun-kan ike-de oyoi-da-to kodomo-wa haha-ni shabet-ta. frog-NOM 20 minute for pond-LOC swim-PAST-COMP child-TOP mother-DAT talk-PAST 'The child told the mother that the frog **swam for 20 minutes** in the pond.'
(B) PuncIter Punctual verb +Short interval	kaeru-ga 20 pun-kan ike-de haneta-da-to kodomo-wa haha-ni shabet-ta. frog-NOM 20-minute for pond-LOC jump-PAST-COMP child-TOP mother-DAT talk-PAST 'The child told the mother that the frog **jumped for 20 minutes** in the pond.'
(C) DurIter Durative verb + Long interval	kaeru-ga 2 kagetsu-kan ike-de oyoi-da-to kodomo-wa haha-ni shabet-ta. frog-NOM 2-month for pond-LOC swim-PAST-COMP child-TOP mother-DAT talk-PAST 'The child told the mother that the frog **swam for 2 months** in the pond.'
(D) GapIter Punctual verb + Long interval	kaeru-ga 2 kagetsu-kan ike-de haneta-da-to kodomo-wa haha-ni shabet-ta. frog-NOM 2-month for pond-LOC jump-PAST-COMP child-TOP mother-DAT talk-PAST 'The child the mother that the frog **jumped for 2 months** in the pond.'

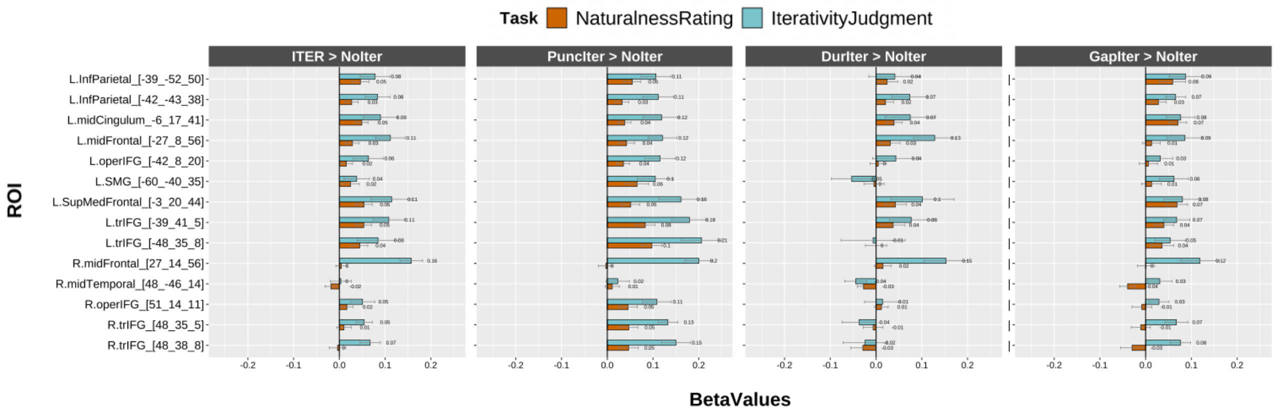


Figure 1. Imaging results of task comparisons showing the regional activation level per critical contrast

Abbreviations: L = left hemisphere; R = right hemisphere; IFG = inferior frontal gyrus; tr = pars triangularis; oper = pars opercularis; mid = middle; SupMed = superior medial; Inf = inferior.
On the left-most panel, ITER = {PuncIter, DurIter, GapIter} with underspecified iterative meaning combined

References

[1] Jackson, R. L. (2021). The neural correlates of semantic control revisited. *NeuroImage, 224*, Article 117444.

[2] Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage*, *53*(2), 638-646.

[3] Lai, Y. Y., Sakai, H., & Makuuchi, M. (2023). Neural underpinnings of processing combinatorial unstated meaning and the influence of individual cognitive style. *Cerebral Cortex*, *33*(18), 10013-10027.

[4] Deo, A. S., & Piñango, M. M. (2011, September). Quantification and context in measure adverbs. In *Semantics and Linguistic Theory* (pp. 295-312).

[5] Dehaene S, Piazza M, Pinel P, Cohen L. Three parietal circuits for number processing. *Cognitive Neuropsychology*. 2003:20(3–6):487–506.

[6] Piñango, M. M., Lai, Y. Y., Deo, A., Foster-Hanson, E., Lacadie, C., & Constable, T. (2024). Comprehension of English *for*-adverbials: The Nature of Lexical Meanings and the Neurocognitive Architecture of Language. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12746

Automatic processing of relational structure in language: A frequency-tagging EEG study of Chinese compounds

Fengyun Hou[1] (fengyun.hou@bristol.ac.uk), Alexander Anderson[1], Nina Kazanina[2]

1 School of Psychological Science, University of Bristol

2 Department of Basic Neuroscience, University of Geneva

Background: Language is a primary means of exchanging information about the world, and much of the exchange concerns relations between entities in the world. On a par with lexical means for expressing relations, such as prepositions (e.g., lamp <u>on</u> the bench, lesson <u>after</u> the break), relations can also be expressed morphologically either using overt affixation (possession: Ivy<u>'s</u> cat) or null morphemes (raincoat 'a coat <u>for</u> rain', snowball 'ball <u>made-of</u> snow').

Previous studies argued that relational structure is accessed routinely, as shown by facilitatory priming effects between compounds sharing the same semantic relation, e.g. *raincoat* [*coat <u>for</u> rain*] facilitates *rainboots* [*boots <u>for</u> rain*] relative to *rainwater* [*water <u>originating from</u> rain*] [1,2]. However, this finding was confounded by the fact that compounds sharing semantic relation frequently also contained semantically related second constituents (*coat & boots* are semantically closer than *coats & rain*), hence the facilitation couldn't be definitively attributed to the shared semantic relation. More definitive demonstration of automatic extraction of relational structure is needed; we address this question via a frequency-tagging EEG study of Chinese compounds.

Design and Method: Twenty native Chinese speakers viewed 60 second-long sequences of Chinese compounds presented at a regular rate of 166.7ms/compound (i.e. frequency-tagged @6Hz). Each sequence consisted of four base stimuli alternating with a fifth oddball (@1.2 Hz). Base compounds all featured the FUNCTION relation between constituents within each XY compound interpretable as 'Y for X; e.g. 酒瓶 (/jiǔpíng/, wine bottle – 'bottle for wine'), the second constituent 瓶 (/píng/, bottle) specified the function of the first constituent 酒 (/jiǔ/, wine). Oddball stimuli varied across five conditions:

a) POSSESSION ('Y of X'): 蛋 (/dàn/, egg) + 壳 (/ké/, shell) → 蛋壳 (/dànké/, eggshell, 'shell of egg');
b) COORDINATION ('X and Y'): 花 (/huā/, flower) + 草 (/cǎo/, grass) → 花草 (/huācǎo/, 'flowers and grass');
c) ACTION ('X act on Y'): 弹 (/tán/, play) + 琴 (/qín/, instrument) → 弹琴 (/tánqín/, 'play[action] instrument [object]');
d) PSEUDO-COMPOUND ('PC'): 头 (/tóu/, "head") + 柴 (/chái/, "firewood") → 头柴 (/tóuchái/, illicit compound);
e) FUNCTION ('Y for X'): 药 (/yào/, medicine) + 箱 (/xiāng/, box) → 药箱 (/yàoxiāng/, medicine box, 'box for medicine').

Results and Discussion: If semantic relations are extracted automatically, we expected that a spectral peak would appear at the rate of oddballs (i.e. 1.2 Hz and its harmonics) in conditions (a-d) in which the semantic relation within the oddball stimuli was different from that in the base stimuli. This prediction was confirmed by the results of non-parametric cluster-based permutation tests: the power at the oddball rate 1.2 Hz and its harmonics was significantly higher than in the neighbouring bins in conditions (a-d), but not in condition (e) in which base and oddball stimuli shared the same relation. These results demonstrate that participants automatically extracted the detailed semantic relations between constituent morphemes in compounds, providing the first neural demonstration of extraction of latent relational structures in language. We place these results into a larger theoretical framework that considers the role of relations across cognitive domains, including vision and memory [3].

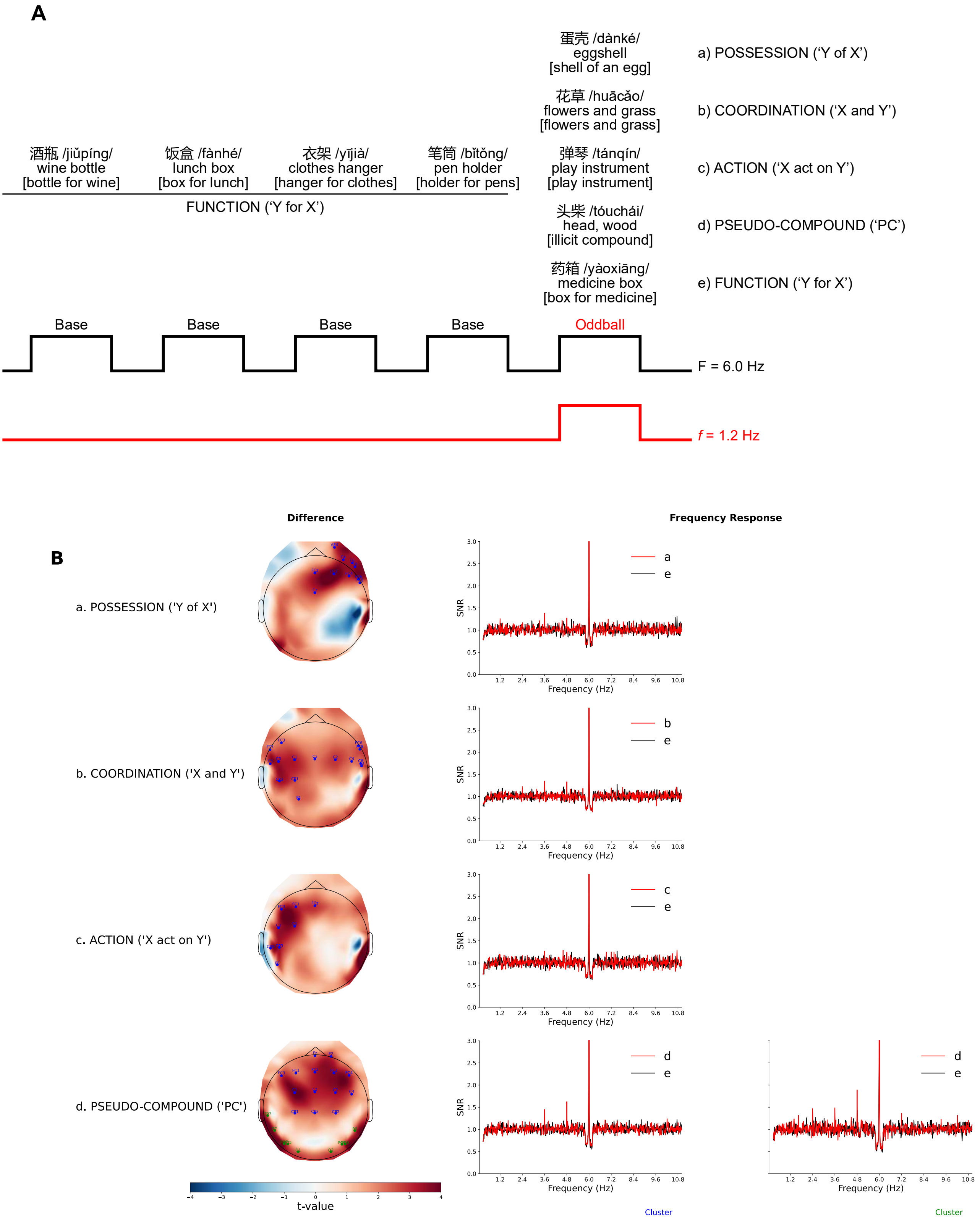


Figure 1: **(A)** Schematic illustration of the experimental paradigm. **(B)** Topographical maps showing significant differences across conditions (a–d) where the semantic relations within oddball stimuli differed from those in the base stimuli. These differences are reflected in the SNR of *f*+ (i.e., the sum of 1.2 Hz and its harmonics) within regions of interest identified through cluster-based permutation analysis. Frequency response plots depict the mean SNR across electrodes within each significant cluster.

Reference

[1] Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language, 60*(1), 20–35. https://doi.org/10.1016/j.jml.2008.07.003

[2] Raffray, C. N., Pickering, M. J., & Branigan, H. P. (2007). Priming the interpretation of noun–noun combinations. *Journal of Memory and Language, 57*(3), 380–395. https://doi.org/10.1016/j.jml.2007.06.009

[3] Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences, 25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Influence of the syntactic function on the production of negated sentences in German

Fenja Plate-Güneş[1], Jana-Elina Jordan[1], Markus Bader[1], Sascha Bargmann[1], Gert Webelhuth[1]
[1] Goethe University Frankfurt
plate-guenes@lingua.uni-frankfurt.de

Background: While the comprehension of negation has been widely studied, much less is known about how negated sentences are produced. This is surprising considering that negation is often associated with word-order alternations, a central topic in production research. In German, for instance, there are two ways to express an existential quantifier within the scope of negation: The indefinite article *ein* ('a') can either follow the negation particle *nicht* ('not') and contract to *kein* ('no'), resulting in surface scope, or it can precede the negation, yielding inverse scope. Studies on sentence comprehension have shown a preference for surface scope over inverse scope. The former is semantically transparent, whereas the latter is considered more processing intensive [1]. Nevertheless, several recent studies (e.g. [2], [3]) have shown that inverse scope is accessible in German.

Hypotheses: We investigate whether the syntactic function of the existential NP influences the choice between *ein-nicht* and *kein*. Based on evidence that subjects are planned early and sentence production is incremental [4], we derive the following predictions: (1) Subjects should precede negation more often than objects. (ii) Subjects with an agent role (unergative verb) are subjects at the surface and underlyingly, whereas subjects with a patient/theme role (unaccusative and passive verbs) have properties of underlying objects [4]. Consequently, we expect sentences with unergative verbs to elicit more recalls with *ein* preceding negation compared to unaccusative and passive verbs. (iii) For sentences with an indefinite object, we expect less inverse scope (*ein nicht*) for PP than for accusative objects because PP objects can appear after the negation without contraction.

Method: This experiment employs a variant of the production-from-memory paradigm. It follows a 1x5 factorial design, with syntactic function of the indefinite NP as the independent variable at 5 levels (see example Table 1): subject of unergative verb, subject of unaccusative verb, subject of passive verb, accusative object, prepositional object. Participants (n = 27 so far) read out a context sentence and a negated main clause with a sentence initial indefinite. Participants then read the context again, followed by a prompt like "Es heißt, dass..." ('It is said that...'), and transformed the main clause to an embedded clause. This procedure allows us to determine under which conditions they produce sentences with either *kein* or *ein-nicht*.

Results: We tested the hypotheses derived above in a preliminary analysis of the data from 15 participants with the `lme4` package [5] (see Table 2). (i) *ein-nicht* is produced significantly more often when the indefinite is the subject of the sentence rather than the object. (ii) There were no significant differences between the three types of subjects. (iii) *ein-nicht* is produced significantly more often for accusative objects than for PP objects. PP objects were often produced after the negation without contraction (see Figure 1a).

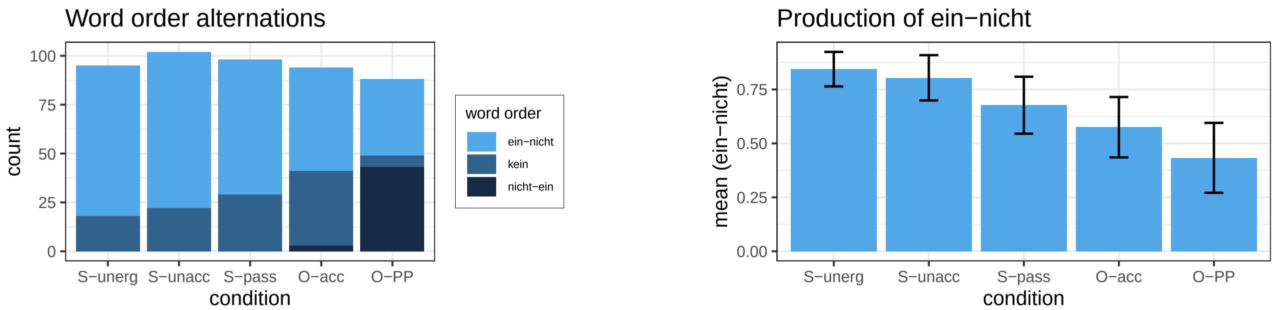
Discussion: The results show that the syntactic function of the indefinite influences the production of negation in German. During sentence planning, negation appears to be processed after the subject, leading to more inverse scope (*ein-nicht*) for subjects than for objects. The lack of differences between the three types of subjects indicates that the surface status of the subject is crucial in this regard. The results for PP objects suggest that *kein* contraction is avoided if possible. This in turn may be one of the reasons why *ein-nicht* is produced despite inverse scope.

Example item

Table 1: A stimulus item in all five versions

Context	In der Galerie um die Ecke wurde vergangene Woche eine neue Ausstellung eröffnet. ’In the gallery around the corner, a new exhibition opened last week.’
Target sentence	
S-V-unergative	Ein lokaler Künstler hat zur Überraschung aller nicht mitgewirkt. A local artist has to surprise of-all not participated ’To everyone’s surprise, a local artist has not participated.’
S-V-unaccusative	Ein lokaler Künstler ist zur Überraschung aller nicht erschienen. A local artist is to surprise of-all not appeared ’To everyone’s surprise, a local artist has not appeared.’
S-V-passive	Ein lokaler Künstler wurde zur Überraschung aller nicht eingeladen. A local artist was to surprise of-all not invited ’To everyone’s surprise, a local artist was not invited.’
O-accusative	Einen lokalen Künstler hat man zur Überraschung aller nicht eingeladen. A local artist has one to surprise of-all not invited ’To everyone’s surprise, one didn’t invite a local artist.’
O-PP	Nach einem lokalen Künstler hat man zur Überraschung aller nicht gefragt. After a local artist has one to surprise of-all not asked ’To everyone’s surprise, one didn’t ask for a local artist.’
Prompt	Es heißt, dass (man) zur Überraschung aller… ’It is said that to everyone’s surprise...’

Results



(a) Produced word orders

(b) Produced inverse scope word order

Figure 1: Preliminary results of word order and inverse scope

Table 2: Mixed-effects model for production of inverse scope (see Figure 1b).

Formula: $inverseScope \sim Condition + (1|Participant) + (1|SentenceNr)$

Contrast	Estimate	SE	z value	p value
Subjects versus objects	1.7497	0.2828	6.186	6.17e-10 ***
S-unergative vs S-unaccusative	0.3770	0.4472	0.843	0.39920
S-unaccusative vs S-passive	0.6498	0.4156	1.564	0.11793
O-accusative vs O-PP	0.9326	0.3985	2.341	0.01926 *

References

[1] Wurmbrand, S. (2018). The cost of raising quantifiers. *Glossa: a journal of general linguistics*, *3*(1), 19.

[2] Radó, J., & Bott, O. (2018). What do speaker judgments tell us about theories of quantifier scope in german? *Glossa: a journal of general linguistics*, *3*(1), 91.

[3] Fanselow, G., Zimmermann, M., & Philipp, M. (2022). Accessing the availability of inverse scope in german in the covered box paradigm. *Glossa: a journal of general linguistics*, *7*(1).

[4] Momma, S., & Ferreira, V. S. (2019). Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, *114*, 101228. https://doi.org/https://doi.org/10.1016/j.cogpsych.2019.101228

[5] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Language in the Dyad: Linking Linguistic and Neural Alignment.

Kristof Strijkers[1] & Emilia Kerr[1]
[1] Aix-Marseille University, CNRS & LPL, Aix-en-Provence, France.
Kristof.Strijkers@univ-amu.fr

Language is inherently interactive, yet much research in linguistics, psychology, and neuroscience focuses on individuals, neglecting dialogue as the primary context for communication [1]. This talk presents two studies that start addressing this limitation and investigate language and the brain in their most functional context, the dyad.

A central focus is *alignment*, observed both in linguistic behavior and neural activity. Linguistic alignment concerns convergence in language use [2], while neural alignment refers to correlated brain activity across interlocutors [3]. Whether both phenomena are linked remains unresolved, as linguistic alignment is typically studied for specific levels of representation (e.g., lexico-semantics or syntax), whereas neural alignment tends to focus on broader communicative dynamics. Here we explore whether neural alignment might serve as a biological marker of linguistic alignment, and whether prediction could mechanistically link the two.

Using a novel paradigm combining interactive language games with EEG hyperscanning (i.e., simultaneous EEG recordings in two or more participants; [4]), pairs of participants engaged in a semantic association task involving animal and tool words (***Figure 1***). Interlocutor A named a picture (e.g., *"dog"*), and Interlocutor B had to reply with a categorically related word (e.g., *"cat"*), following either predictive or neutral sentence cues. Prediction facilitated speech production in both interlocutors but was twice as large for the interlocutor replying with an associate to predictable content (Interlocutor B) compared to the interlocutor naming the actual predictable content (Interlocutor A). Moreover, when doing the task individually, this incremental prediction effect disappeared highlighting that it was driven by the interactive context (***Figure 2***). We then ran this paradigm with EEG hyper-scanning and observed neural alignment between the interlocutors in the theta-band which was specific to interactions about the animal and tool words (***Figure 3***).

These findings reveal that (a) linguistic and predictive processing differ fundamentally in dyads versus isolation, hereby emphasizing the necessity to explore the mechanisms of language in social interaction, and (b) neural alignment can reflect specific lexico-semantic representations, suggesting a possible direct link of alignment in the brain and in behaviour. This may allow us to start expanding dialogue models with neurobiological principles.

Figures

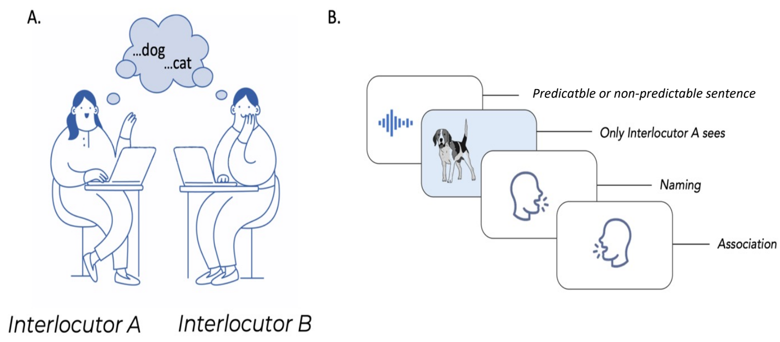


Figure 1. Design and trial sequence. (A) Simple schematic of set-up. (B) Example trial sequence of the semantic association game preceded by a predictable or non-predictable context

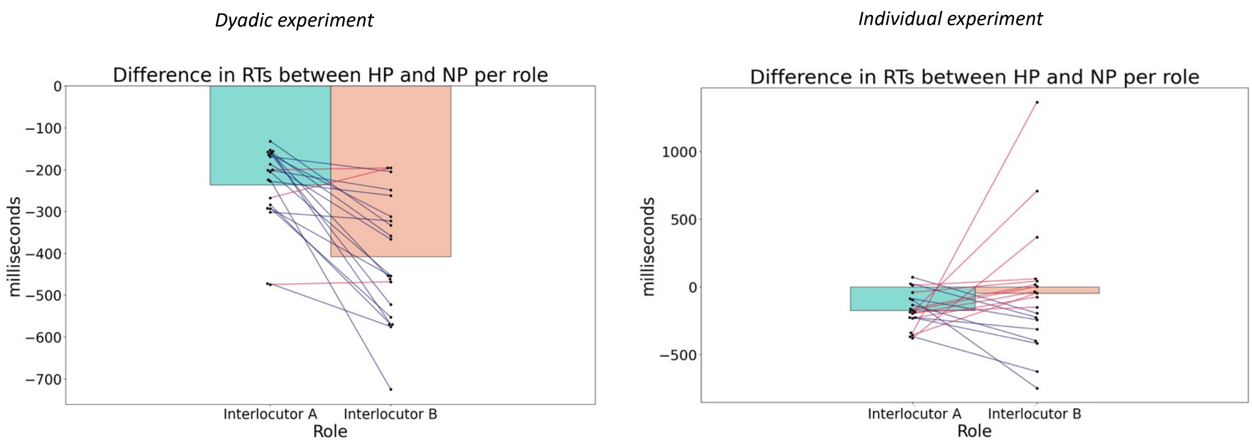


Figure 2. Prediction effect divided by role (Interlocutor A who names the picture in green and Interlocutor B who replies with a categorically related word in orange) by subtracting the reaction times (RTs) of the Highly Predictable (HP) sentence context from the Non-Predictable (NP) sentence context. The left panel displays the prediction effects for the dyadic experiment and the right panel the individual control experiment. Blue lines correspond to interlocutors an incremental prediction effect (larger for Interlocutor B) and red lines correspond to the absence of a prediction effect.

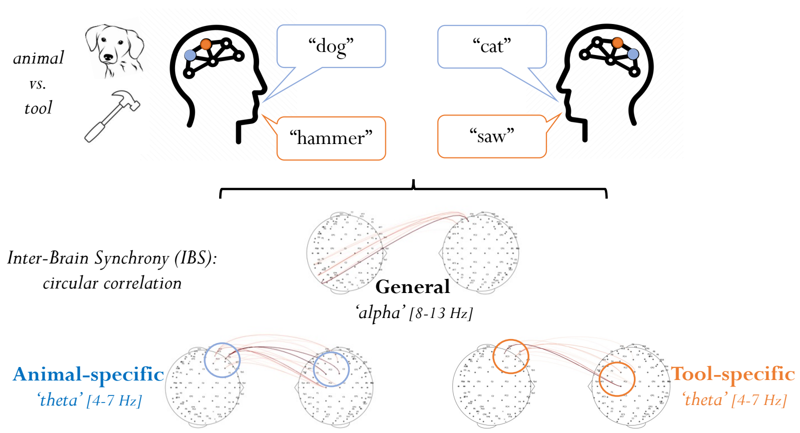


Figure 3. Neural alignment (as measured with inter-brain synchrony of EEG) between the interlocutors performing the semantic association game of animal and tool words. Results show meaning-specific neural alignment in the theta-band for animal (blue) and tool (orange) words, which is qualitatively distinct from neural alignment in the alpha-band to the task in general (black).

References

[1] Kuhlen, A. K., & Abdel Rahman, R. (2023). Beyond speaking: neurocognitive perspectives on language production in social interaction. *Philosophical Transactions of the Royal Society B*, *378*(1875), 20210483.

[2] Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

[3] Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114-121.

[4] Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PloS one*, *5*(8), e12166.

Linguistic and Non-Linguistic Cues during Colour Discrimination May Function as Differential Cues for Expectations: An Approach Using ERP and Oscillatory Analyses

Owen Kapelle(1, 2), Monique Flecken(1), Conrado Bosmann(3), Paul Boersma(1)

1. University of Amsterdam, Amsterdam Center for Language and Communication

2. University of Amsterdam, Data Science Center

3. University of Amsterdam, Faculty of Science, Swammerdam Institute for Life Sciences

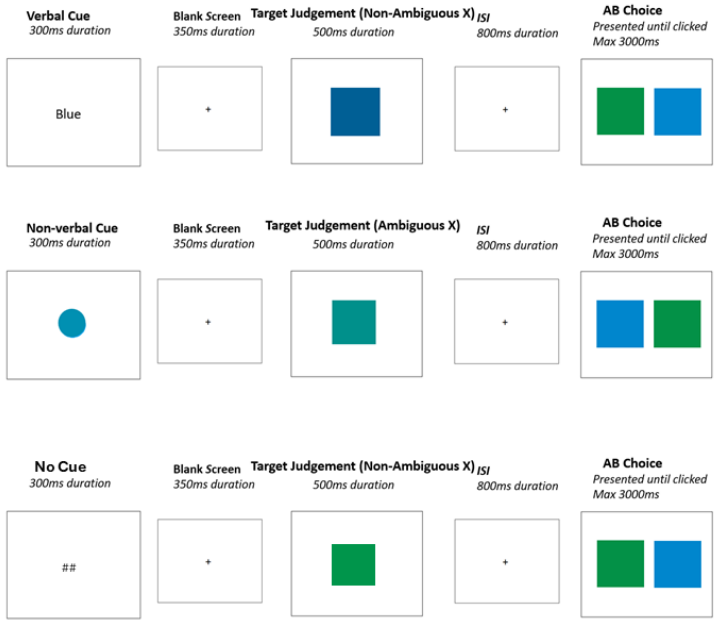
Correspondence: Owen Kapelle, o.e.j.kapelle@uva.nl

A classic way to explore the influence of language on perception focuses on the relationship between language and colour perception. In the current study, we investigated this interaction within a predictive processing framework to explore whether linguistic and/or non-linguistic visual cues act as functionally different types of prior expectations, that could lead to differential cueing effects, because they activate different representations of colour. In previous research using a linguistic cueing paradigm, auditory linguistic cues (words) prompted quicker and more accurate responses than auditory non-linguistic cues (sounds) (Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013), particularly with ambiguous or 'hard to see' stimuli (Ostarek & Huettig, 2017).

We employ a visual 'linguistic cueing' paradigm, in which we contrast the differential effect of linguistic cues as opposed to non-linguistic cues compared to uncued trials (figure 1). We expect linguistic cues to activate categorical perceptual representations, facilitating the activation of category-diagnostic features of perceptual stimuli (a 'prototypical' blue), while non-linguistic cues activate representations of specific members of a category (baby or navy blue) but do not promote categorical processing.

We recorded Electroencephalography (EEG) in 47 participants recruited from the Dutch and English speaking student population in Amsterdam, using a 64-electrode set-up. Preprocessing was carried out with the python software MNE (Gramfort et al., 2013) for resampling, to apply filters, to average the data over the mastoids, and removal of noise and artefacts. We have recently finished pre-processing all participants' data. We aim to analyse event related potentials (ERPs) such as the N400, associated with category typicality (Wang et al., 2016), as well as gamma band oscillations to assess the prediction error related to different cue conditions, and lastly alpha-beta bands to assess top-down influences during perceptual prediction (Michalareas et al., 2016). Our analysis will focus on occipital electrodes around the visual areas for low-level colour perception, as well as the prefrontal cortex as it is associated with higher cognitive functions.

Figure 1. Overview task



Sources

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron, 89(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. Journal of Experimental Psychology: General, 141(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. PNAS Proceedings of the National Academy of Sciences of the United States of America, 110(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. Journal of Experimental Psychology: Human Perception and Performance, 43(3), 499–508. https://doi.org/10.1037/xhp000031

Wang, X., Tao, Y., Tempel, T., Xu, Y., Li, S., Tian, Y., & Li, H. (2016). Categorization Method Affects the Typicality Effect: ERP Evidence from a Category-Inference Task. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00184

Sound-Symbolism Effects in Novel Word Generation

Anna Cameron, Sandie Cleland, Agnieszka E. Konopka
University of Aberdeen
agnieszka.konopka@abdn.ac.uk

Background: Evidence of sound symbolism, i.e., a relationship between word forms and word meanings, has been documented in numerous languages. Cross-linguistically, many basic vocabulary items are associated with certain speech sounds (e.g., “small” with /ɪ/ and /i/ [1]), and participants can perform at above chance accuracy in categorizing words from languages they do not know [2]. For example, participants are more likely to associate the label “bouba” with round shapes and the label “kiki” with spikey shapes (the “bouba/kiki” effect [3]). However, few studies have investigated whether sound-symbolism effects occur in tasks that do not include obvious categories, a limited choice of words, or existing languages. The current study took a new approach: we asked participants to generate novel words in response to prompt words that, unbeknownst to them, varied on three implied dimensions (gender: feminine vs. masculine; size: big vs. small; shape: round vs. spikey) that are associated with different distributions of sounds in existing words, and we tested for evidence of sound-symbolism in their unconstrained productions.

Method: 50 native speakers of English were instructed to generate 60 novel pronounceable words. They received 10 prompt words per category (masculine/feminine, big/small, round/spikey words; see examples in Table 1) and were asked to come up with a word in a new language to be used by a character in a game (“What is the word for this in the main character’s language?”). Any responses resembling existing words (e.g., *gigantoss*, *bulkissimo*) or with ambiguous pronunciations were rejected. The remaining 2050 novel words were scored based on standard English pronunciation rules for the presence of “feminine” (/f/, /ɪ/, /i/, /u/) vs. “masculine” (/g/, /r/, /k/) phonemes, “big” (/ɑ/, /ɒ/, /b/, /g/, /r/,) vs. “small” (/ɪ/, /i/) phonemes, and “round” (/b/, /m/, /oʊ/, /u/) vs. “spikey” (/aɪ/, /ɪ/, /k, /t/) phonemes.

Results: Three generalised linear mixed-effects models compared the likelihood of participants generating novel words with category-congruent phonemes in the three categories of prompt words. Each analysis showed an interaction between phoneme category and prompt word category (feminine/masculine: z=4.78; big/small: z=8.01; round/spikey: z=4.86, all ps<.001), as participants produced more novel words containing the expected phonemes after category-specific prompt words (Figure 1). Because several phonemes were associated with 1+ categories (e.g., /ɪ/ is a “feminine”, “small” and “spikey” phoneme), a complementary analysis was conducted to compare generation of novel words with the scored phonemes after all category-congruent (i.e., matching) and all category-incongruent (i.e., mismatching) prompt words, and showed converging results (main effect of phoneme-prompt word match, z=3.35, p<.001; Figure 2).

Discussion: Sound-symbolism effects were remarkably robust: participants’ productions showed a strong and consistent effect of all three implied categories on sound selection in novel words. It remains to be determined whether participants were guided by their implicit (or explicit) knowledge of how sound patterns in English relate to meaning or whether cross-linguistically pervasive sound-meaning regularities drove their selection. In either case, the findings are particularly interesting given that participants were not explicitly informed that the words they were responding to were drawn from a pool of specific categories and thus suggest that sound-symbolism effects are not restricted to relatively artificial forced-choice tasks.

Table 1. Examples of prompt words per category (prompt words in each category did not include a higher number of category-congruent phonemes than prompt words from other categories).

<u>Masculine/Feminine Category</u>		<u>Big/Small Category</u>		<u>Round/Spikey Category</u>	
Masculine words	**Feminine words**	**Big words**	**Small words**	**Round words**	**Spikey words**
Destruction	Adorable	Castle	Cottage	Ball	Cactus
Handsome	Beautiful	Country	Little	Donut	Shrapnel
Hero	Cherish	Giant	Miniature	Globe	Porcupine

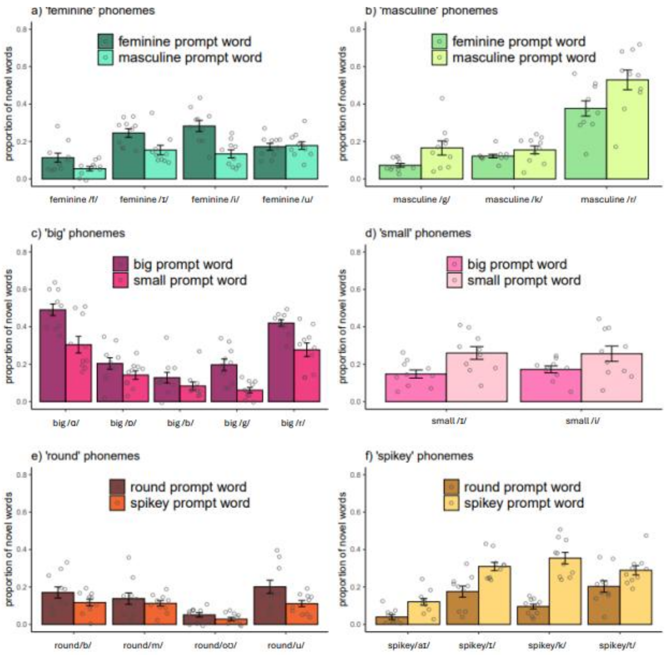


Figure 1. Proportion of novel words (by-item means) with (a-b) "feminine" vs. "masculine" phonemes, (c-d) "big" vs. "small" phonemes, and (e-f) "round" vs. "spikey" phonemes generated in response to category-specific prompt words.

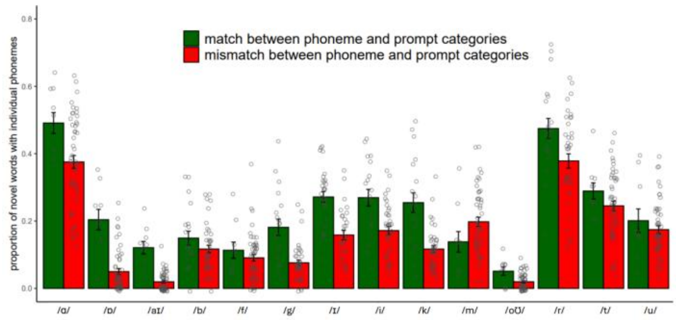


Figure 2. Proportions of novel words (by-item means) generated with the 14 scored phonemes in response to all category-congruent (i.e., matching) prompt words and all category-incongruent (i.e., mismatching) prompt words.

References

[1] Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *PNAS, 113*(39), 10818-10823.

[2] D'anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing meaning from word sounds of unfamiliar languages: A cross-cultural sound symbolism study. *Frontiers in Psychology, 10,* 593).

[3] Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200390.

Mozart's Concert in da Hood:
A VWP Eye-tracking study on the effects of music as formality-context on online comprehension of register nuances.

Angelina Ioannidou-Tsiomou[1], Katja Maquate[1], Sarah Creel[2], Pia Knoeferle[1,3,4]
[1] Humboldt-Universität zu Berlin; [2] University of California, San Diego;
[3] Einstein Center for Neurosciences Berlin; [4] Berlin School of Mind and Brain
angeliki.ioannidou-tsiomou.1@hu-berlin.de

Background: Music as a priming stimulus for spoken language comprehension represents the core object of investigation in this study, with the focus lying on its potential ability to convey socially situated, formality-register-related information. Using a Visual-World-Paradigm Eye-Tracking experiment with young adult comprehenders (18-31years, L1-German, N=32) the study aims to bridge two research gaps, on register [1] and on music [2], as factors potentially influencing language comprehension. The study is motivated by psycholinguistic findings, showing that comprehenders utilize different types and sources of information during processing, such as the speaker's voice [3], activating relevant social stereotypes [4]. Given the parallels between human voice and music, the latter is investigated as a factor possibly influencing language comprehension.

Method: The present study builds on the work of Patarroyo, Maquate, Ito, and Knoeferle [5], and investigates the possible ability of music, its genre, and/or its performing instruments to function as formality-register context, and prime spoken language comprehension. Participant's eye-movements and fixations on a formal or an informal picture (see Figure 1) are recorded while they listen to a melody – originating from a formal or informal genre and being performed by formal or informal instruments – and to a spoken sentence in German (see Table 1). The genre and the performing instruments could match or mismatch the formality of the linguistic stimulus. Stimuli were pretested for their perceived formality. A post-eye-tracking task was implemented to gather additional information on the melodies' formality as perceived by the eye-tracking participants and on participants' musical background.

Results: Pilot results (N=8) yielded a significant effect of the instrument-formality on visual scene exploration (see Figure 2). Even though an effect of the instrument-formality aligned with the third alternative hypothesis (see Table 2), the direction of the effect was not expected. Participants seemed to fixate a referent more when listening to music performed by instruments that mismatched its formality, rather than by ones that matched it. No significant effects on language comprehension were found.

Discussion: The significant effect of instrument-formality on visual scene exploration might suggest that music can convey socially situated information related to formality. Nevertheless, language remains the much more dominant modality. The absence of effects during language comprehension could be attributed to various factors, such as the complexity of the musical stimuli – varying in rhythm, tempo, and frequency –,the low number of participants (N=8), or the differences in participants' musical backgrounds. Alternatively, it could be ascribed to the lack of real-time relation between formality-related aspects of music and language comprehension, hinting at a genuine null effect.

Table 1: Illustration of the experimental conditions.

Condition	Context Music's Genre	Context Music's Performance
High Register Target Sentence: I will right away tie my shoes. (↑formal: Schuhe)		
Full Match	Mozart's Serenade No. 13 in G, K 525	Formal Set of Instruments
Performance Mismatch	Mozart's Serenade No. 13 in G, K 525	Informal Set of Instruments
Genre Mismatch	Lady Gaga's "Poker Face"	Formal Set of Instruments
Full Mismatch	Lady Gaga's "Poker Face"	Informal Set of Instruments
Low Register Target Sentence: I will right away tie my shoes. (↓informal: Latschen)		
Full Match	Lady Gaga's "Poker Face"	Informal Set of Instruments
Performance Mismatch	Lady Gaga's "Poker Face"	Formal Set of Instruments
Genre Mismatch	Mozart's Serenade No. 13 in G, K 525	Informal Set of Instruments
Full Mismatch	Mozart's Serenade No. 13 in G, K 525	Formal Set of Instruments

Table 2: Groups of performing instruments.

Voice	formal instruments	informal instruments
Soprano G-Clef	Violin	Overdrive E-Guitar
Alto/Mezzo Soprano G-Clef	Piano G-Clef	Distorted E-Guitar
Tenor F-Clef	Violoncello	E-Bass
Bass F-Clef	Piano F-Clef	Slap Bass

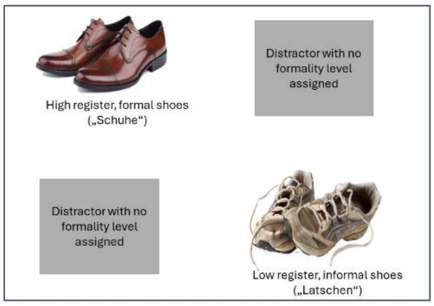


Figure 1: Display example

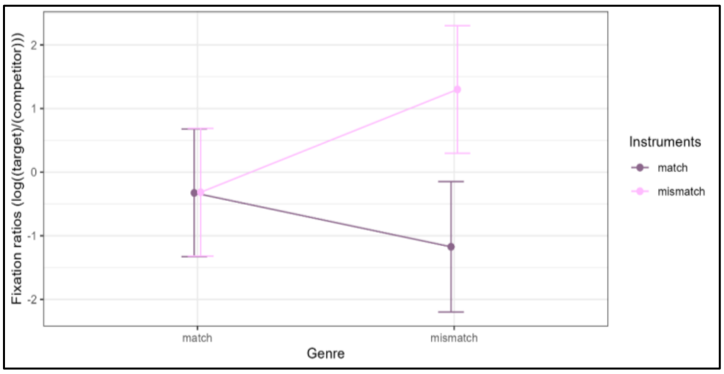


Figure 2: Effects of genre-formality and instrument-formality during visual scene exploration. Error bars = 95% CIs.

Table 3: Hypotheses for the visual exploration interest-period. (AH=alternative hypothesis)

AH1	AH2	AH3
Register related effect of music as a whole	Register related effect of the music-genre	Register related effect of the instruments
More fixations on referents that match the formality of the music, when it is fully formal or fully informal, than on referents that mismatch it.	More fixations on referents that match the formality of the genre, regardless of the instruments-formality, than on referents that mismatch it.	More fixations on referents that match the formality of the instruments, regardless of the genre-formality, than on referents that mismatch it.

References

[1] Agha, A. (1999). Register. *Journal of Linguistic Anthropology, 9*(1/2), 216-219.

[2] Koelsch, S. (2011). Toward a neural basis of music perception–a review and updated model. *Frontier in Psychology, 2,* 110.

[3] Borovsky, A., & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*(5), 1600.

[4] Van den Brink, D., Van Berkum, J., Bastiaansen, M., Tesink, C., Kos, M., Buitelaar, J., & Hagoort, P. (2012). Empathy matters: Erp evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience, 7*(2), 173-183.

[5] Patarroyo, A., Maquate, K., Ito, A., & Knoeferle, P. (2022). Investigating the real-time effect of register-situation formality congruence versus verb-argument semantic fit during spoken language comprehension. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44).

T(w)o Gender(s) or not t(w)o Gender(s) – Gender Assignment in German in English-German and French-German Bilinguals

Regina Hert[1], Barbara Köpke[1], E Jamieson[2], Vera Heyer[3], Monika S. Schmid[2] & Holger Hopp[3]
[1]Université Toulouse II, [2]University of York, [3]TU Braunschweig

Background: Lexical gender assignment is typically taken to be arbitrary, non-transparent and item-specific (thus frequency-reliant), still there are phonological regularities which are probabilistically linked with some genders (e.g., German: [1]; French: [2]). **Lexical gender assignment** therefore involves both **item-based** (bottom-up) and **rule-based** (top-down) processes. To gain insights into which processes are at play in gender assignment in **L1** and **L2** speakers of **German**, we tested different late bilinguals varying in whether their contact language includes a **competing gender system** (French) or not (English).

Method: While data collection is still ongoing, we present data of 63 L2 speakers of German with L1 French or English living in Germany and 68 L2 speakers of French and English with L1 German living in France or the UK. We used a timed lexical gender assignment task of 120 real words and recorded accuracy and RT. We manipulated (a) phonological cue (regularity) – whether words ended on schwa (e.g. *Kufe* 'skid') or began with a consonant cluster (e.g. *Spind* 'locker') – (b) transparency – whether the word matched or mismatched with the phonological cue (schwa → feminine, consonant cluster → masculine) – and (c) word frequency.

Hypotheses: We expect (H1) an interaction of transparency (match vs. mismatch) with L1 German, indicating that participants with German as the L1 use transparency to a lesser extent than participants with German as L2. Further, we hypothesize (H2) an interaction of frequency with L1 German, showing that the effect of frequency will be smaller in participants with German as their L1 than in those with German as their L2. Finally, we assume that bilingual groups will differ in their use of phonological cues and hypothesize that (H3) phonological cue modulates gender assignment in interaction with transparency and contact language (English or French): gender assignment accuracy will be higher when contact language is English than when it is French.

Results: We ran generalized mixed effects models on accuracy. We found the interaction of transparency with L1 German to be marginally significant. Post-hoc comparisons showed that only participants whose L1 is not German were affected by transparency performing better in match than in mismatch conditions. We found a main effect of frequency and of L1 German. Higher frequency led to higher accuracy, and German as L1 led to higher accuracy as well. The interaction of frequency and L1 German was not significant. There was also a main effect of cue and the interaction of cue with transparency was significant showing that gender assignment was facilitated by the schwa cue specifically in match items. The effect of contact language was not significant.

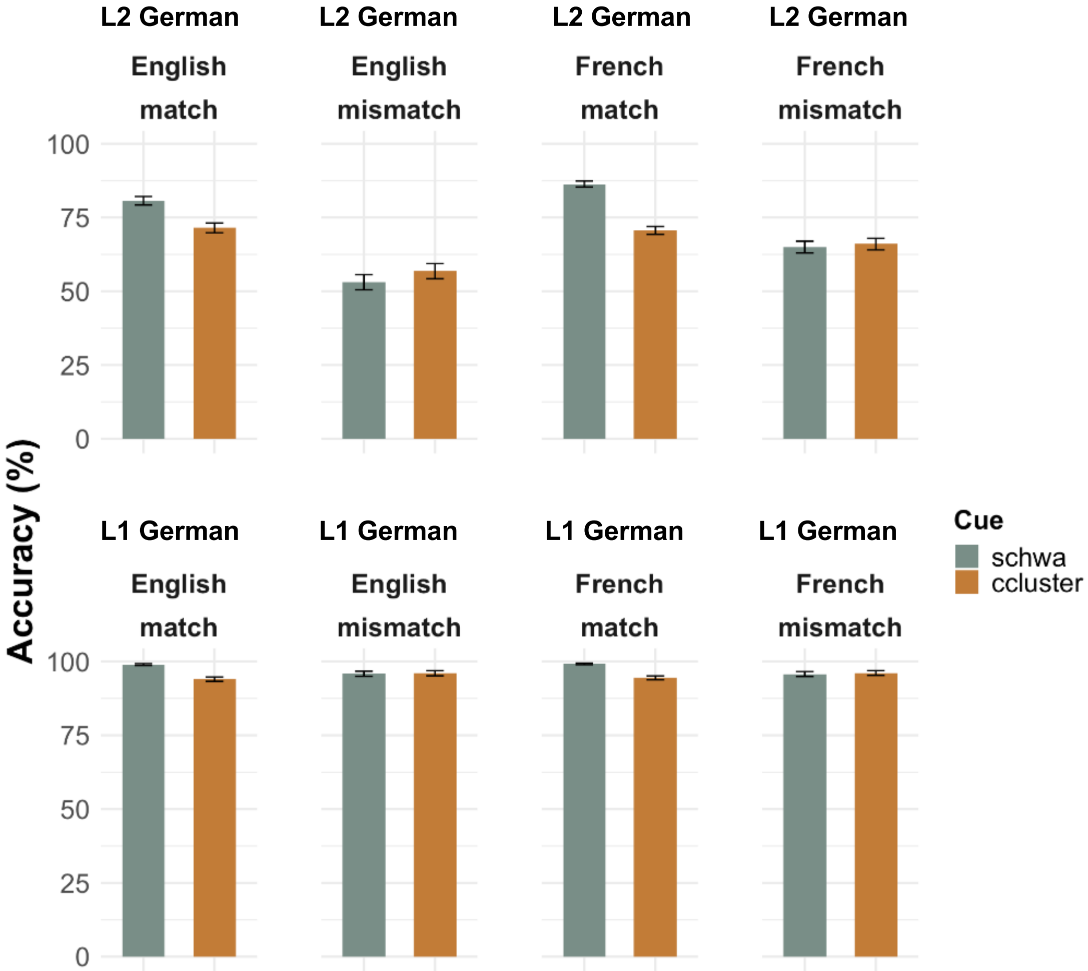
Discussion: These findings suggest that transparency (H1) only influences L2 learners of German, while frequency (H2) affects gender processing in all. This indicates that language development and processing in later ages of acquisition (in adulthood) are driven by both top-down (rules) and bottom-up (frequency) processes whereas when German was acquired earlier (from birth) only bottom-up processes play a role. Cue (H3) affected gender assignment but did not reveal any differences between English and French groups suggesting that these processes may not be vulnerable to cross-linguistic influence.

References

[1] Köpcke, K. & Zubin, D. (1996). Prinzipien für die Genuszuweisung im Deutschen. In E. Lang & G. Zifonun (Ed.), *Deutsch – Typologisch* (pp. 473-491). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110622522-021

[2] Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, *16*(1), 69–92. doi:10.1017/S0959269506002304

Fig. 1: Accuracy across groups, cue and transparency



Dynamic language transfer in bilingualism: How L1 Vietnamese L2 English speakers process filler-gap dependencies in English

Linh Pham & Zuzanna Fuchs
University of Southern California
llpham@usc.edu

Background: Research shows that bilinguals's linguistic systems are not separate but interconnected (e.g. [5]), interacting with each other during language processing. Empirical studies provides support for this interaction at different levels of linguistic representation – ex., phonological (e.g.[3]) or lexical (e.g. [1]). However, evidence is mixed for interaction of more abstract structural rules like island constraints. **This study** extends ongoing debates on language transfer in real-time processing by investigating whether the absence of the abstract relative clause (RC) island constraint in a speaker's L1 affects the processing of island structures in L2, while controlling for language immersion in the L2 to maximize potential L1 influence. *Island constraints* in processing block the active gap-filling strategy (AFS), which normally favors resolving dependencies early (e.g. placing the gap at (a) in (1) instead of waiting for the actual gap at (b), by preventing parsers from positing gaps within certain syntactic boundaries (e.g. RC island in (2)) [6]. However, some languages like Vietnamese allow violations of this constraint (as in (3)) ([8]).
Method: Following [7], we conducted self-paced reading studies using a 2×2 design crossing island (island vs. non-island) and plausibility (plausible vs. implausible) (Tab.1(English) and ex(4) (Vietnamese) for stimuli).
Experiment 1 (n = 60, Vietnamese native speakers) examines AFS in Vietnamese to test whether RC constraints are violated during processing, as previously suggested by offline judgments, thereby validating the design of Experiment 2 (Fig.1). A main effect of plausibility at both spillover regions (p<.0001) suggests that parsers try to posit the gap within the RC island. In other words, no RC island constraint was found.
Experiment 2 compares the processing of English RC island structures in three groups: (1) English native control group (n=38), (2) L1 Spanish-L2 English L2 control group (n=39) (Spanish has similar RC constraint as English), and (3) L1 Vietnamese-L2 English (n=37). (See Fig.2). Both L2 groups were recruited in their home country.
English group: At "regularly", we observed (1) a plausibility × island interaction (ß = -0.13, SE = 0.05, df = 828.6, t = -2.5, p = .012), showing a slowdown for implausible, non-island sentences only, and (2) a main island effect (ß = 0.06, SE = 0.03, df = 828.8, t = 2.09, p = .037). At "about", a main island effect was also found (ß = -0.11, SE = 0.03, df = 839.4, t = -4.28, p < .0001). These effects indicate sensitivity to the island constraint. **Spanish group:** We found a main effect of plausibility at 'regularly' (ß=-0.0.66, SE=0.28, df=853.3, t=-2.4, p=0.02); a nested analysis suggests that only the implausible conditions in non-island sentences are slowed down (p=0.009). At 'about', there is a main effect of island (ß=-0.14, SE=0.02,df=879.8, t=-6.1, p<.0001) with the same patterns as in the English group. **Vietnamese group:** At 'regularly', there is no significant effects, while at 'about', we found a main effect of plausibility (ß=-0.22, SE=0.03,df=795.6, t=-6.9, p<.0001) suggesting that both implausible conditions are slowed down.
Discussion: We found evidence of island sensitivity in the English and Spanish groups, but not in the Vietnamese group. This suggests that the absence of island constraints in L1 influences L2 processing of island structures. Our findings support current accounts of language transfer during processing [4], and contribute to models of bilingual language processing.

(1) Who did Jane see (a)…… Max with (b)_______?

(2) *What did Mary meet [the man who bought _____]?

Non-island/plausible	The book that the author wrote regularly about was named for an explorer.
Non-island/ implausible	The city that the author wrote regularly about was named for an explorer.
Island/ plausible	The book that the author who wrote regularly saw was named for an explorer.
Island/implausible	The city that the author who wrote regularly saw was named for an explorer.

Table 1. *Example of stimuli for experiment 2*

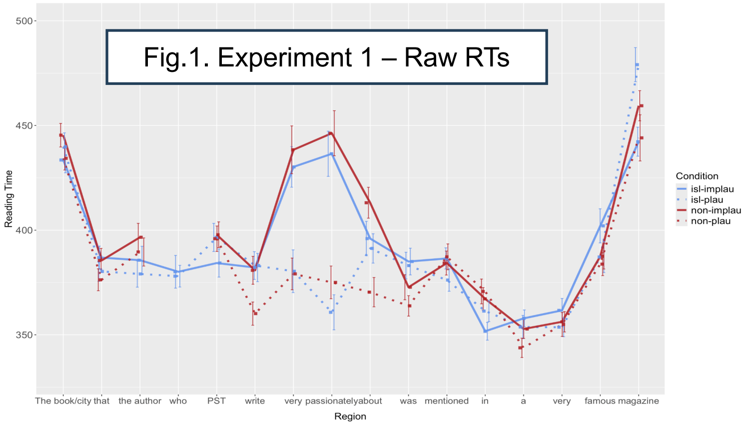


Fig.1. Experiment 1 – Raw RTs

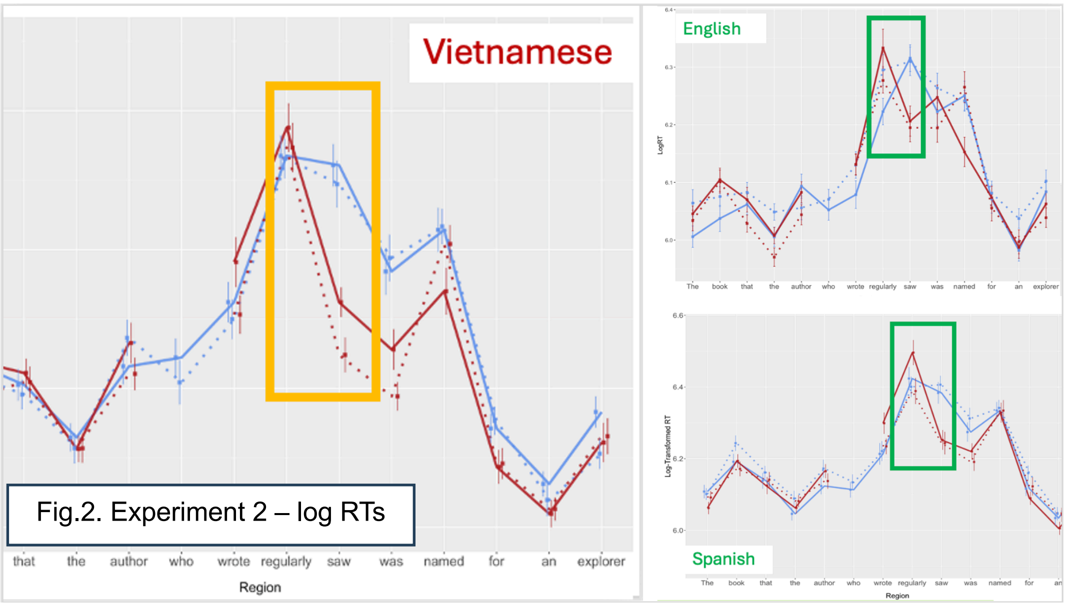


Fig.2. Experiment 2 – log RTs

References

[1] Assche, E. V., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in psychology*.|| **[2]** Clifton, C. Jr., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing* .|| **[3]** Kim, J. E., Cho, Y., Cho, Y., Hong, Y., Kim, S., & Nam, H. (2020). The effects of L1–L2 phonological mappings on L2 phonological sensitivity: Evidence from self-paced listening. *Studies in Second Language Acquisition*.|| **[4]** Kroll, J. F., & Dussias, P. E. (2024). Beyond Transfer: Language Processing in Bilinguals Is Shaped by Competition and Regulation. *Language Teaching Research Quarterly*. || **[5]** Odlin, T. (2003). Cross-linguistic influence. *The handbook of second language acquisition*.|| **[6]** Ross, J. R. (1967). *Constraints on variables in syntax.* PhD dissertation, MIT. || **[7]** Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*. || **[8]** Tsai, C. Y. E. (2009). *Wh-dependency in Vietnamese and the syntax of wh-in-situ*.

(3) Cô gái [$_{RC}$ mà An không biết [ai đã bắt cóc___]] vừa trở về.
The girl [$_{RC}$ that An does not know [who PST kidnap___]] just return.
'The girl that An doesn't know who kidnapped has just returned.'

(4)

(a) Non-island / plausible

Quyển sách mà ông tác giả viết một cách nhiệt tình về được đề cử
The book that the author write passionately about PASS nominate
giải thưởng.
prize
'The book that the author wrote passionately about was nominated for the prize.'

(b) Non-island / implausible

Thành phố mà ông tác giả viết một cách nhiệt tình về được đề cử
The city that the author write passionately about PASS nominate
giải thưởng.
prize
'The city that the author wrote passionately about was nominated for the prize.'

(c) Island / plausible

Quyển sách mà ông tác giả mà viết một cách nhiệt tình nhắc tới được
The book that the author who write passionately mention PASS
đề cử giải thưởng.
nominate prize
'The book that the author who wrote passionately mentioned was nominated for the prize.'

(d) Island / implausible

Thành phố mà ông tác giả mà viết một cách nhiệt tình nhắc tới được
The city that the author who write passionately mention PASS
đề cử giải thưởng.
nominate prize
'The city that the author who wrote passionately mentioned was nominated for the prize.'

Exposure Frequency and Native Language Interference in Early Second Language Auditory Word Learning

Jens Schmidtke
Leipzig University

Background: A central question in second language acquisition concerns the characteristics of phonolexical representations in the mental lexicon [2]. This study examines the representation of novel word forms during the initial stages of acquiring a new language. While earlier research on phonolexical processing has primarily focused on learners at intermediate or advanced proficiency levels, this work adopts a developmental approach to explore how such representations form and evolve. Two competing views offer contrasting predictions about early-stage processing: one holds that lexical access is incremental from the outset [4], while the other proposes that learners begin with holistic representations that only later become segmentally detailed when more similar sounding words are added to the mental lexicon [1].

Method: Fifty-two native German speakers (mean age = 26) participated in the experiment. Participants were introduced to 32 novel words for common objects taken from [3]. The new labels had no phonological overlap with the standard German terms for these objects. Half of the items were classified as high-frequency (HF; 12 exposures) and half as low-frequency (LF; 5 exposures). Sixteen words included a phonological onset competitor (POC), divided into pairs based on similarity: four pairs were more phonetically similar (e.g., gesher-geshem) and four less similar (e.g., makor–makel). These pairs were balanced across participants between HF and LF status. Another set of 16 words was associated with German POCs (e.g., keter–Kette); these objects were presented during training but not referenced directly. Targets and competitors shared 3.1 and 2.9 phonemes in the within and between-language conditions, respectively. Stimuli were generated using speech synthesis with a Hebrew native speaker model. During training, participants performed a cross-situational word learning task that involved presentation of two objects and one auditory label. POC pairs were taught in separate blocks, each comprising 136 trials. After each block and at the conclusion of the experiment, four-alternative forced-choice (4AFC) tests were administered. In the final session, POCs were presented together to assess competition effects.

Results: Participants acquired the word–object associations rapidly. Final 4AFC accuracy was high (M = 92.1%, SD = 7.5), with higher performance for HF (93.7%) than LF (90.2%) items. Accuracy decreased when a within-language competitor appeared (95.7% vs. 80.7%), independent of frequency. Preliminary analyses of the eye-tracking data indicated fewer fixations on LF targets (Fig. 1) and increased attention to within-language competitors compared to controls (Fig.2), effects which emerged post word offset. Competition effects predominantly arose from the eight most phonetically similar pairs, while less similar POCs did not differ from control trials in fixation rates. There was minimal evidence that German POCs influenced processing as competitors.

Discussion: Relative to previous studies involving experienced second language speakers [5], the observed eye-movement patterns suggest that post-lexical access processes related to uncertainty may be more prominent than incremental processing. This interpretation is supported by the delayed emergence of exposure frequency effects. The reduced accuracy in trials with within-language competitors and pronounced competition effects among similar-sounding words but not those that only share the onset indicate that early L2 word forms may lack full specification, with their representations not yet distinct from phonologically similar alternatives.

References

[1] Charles-Luce J, Luce PA. Similarity neighbourhoods of words in young children's lexicons. Journal of Child Language. 1990;17(1):205–15. doi:10.1017/S0305000900013180

[2] Darcy, I., Llompart, M., Hayes-Harb, R., Mora, J. C., Adrian, M., Cook, S., & Ernestus, M. (2024). Phonological processing and the L2 mental lexicon Looking back and moving forward. Studies in Second Language Acquisition, 1–27. https://doi.org/10.1017/S0272263124000482

[3] Duñabeitia, J.A., Crepaldi, D., Meyer, A.S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages.*Quarterly Journal of Experimental Psychology*, 71(4), 808 - 816.

[4] Magnuson J.S., Tanenhaus M.K., Aslin R.N., Dahan D. The Time Course of Spoken Word Learning and Recognition: Studies With Artificial Lexicons. J Exp Psychol Gen. 2003;132(2):202–27.

[5] Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. Journal of Memory and Language, 50(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

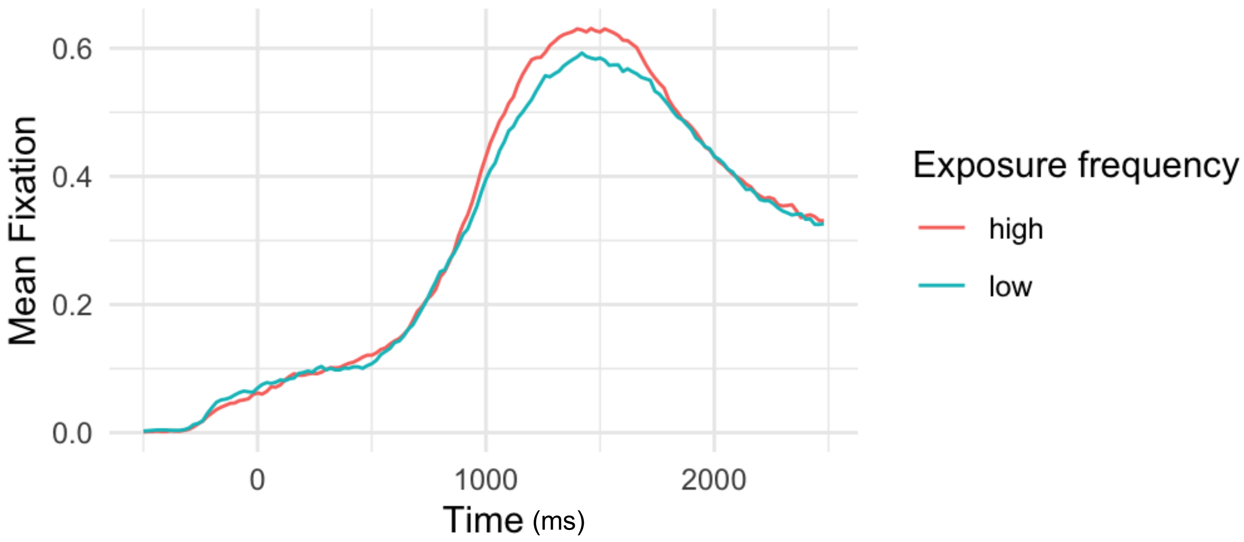


Figure 1. Mean fixation proportion of target images. High frequency targets were seen 12 times during training and low frequency targets 5 times. 0 marks the word onset and the mean word offset was at 553 ms.

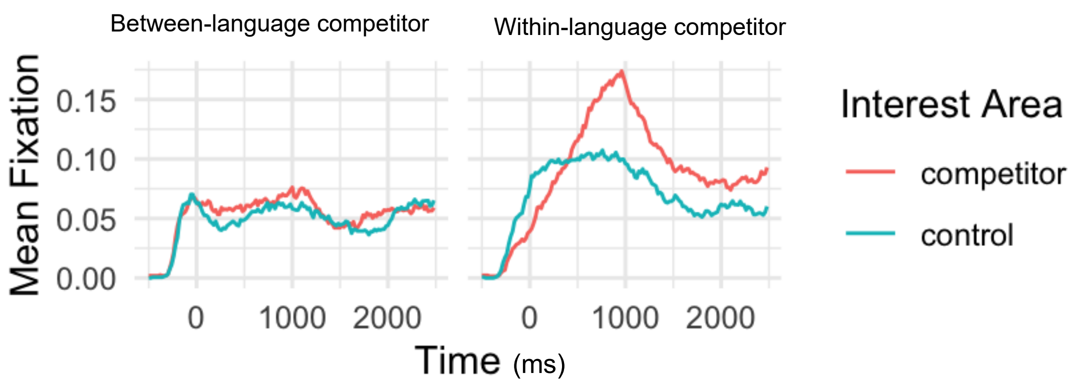


Figure 2. Mean fixation proportion of competitor and control images over time. Control and competitor images were the same, the difference being that they were paired once with a phonologically unrelated target and once with a target that shared the same onset. 0 marks the word onset and the mean word offset was at 553 ms.

Female butchers meet male babysitters: a multilingual maze on the effects of gender stereotypes and grammatical gender during sentence processing

Zunino, G.M.[1,2]; Tomé, C.[3].; Gagliardi, G.[4]; Freitag, R.[5]; Stetie, N.[1,2]; Tzinavos, S[3].; Miola, E.[4]

[1]Universidad de Buenos Aires, [2]CONICET, [3]Universidad de Salamanca, [4]Università di Bologna, [5]Universidade Federal de Sergipe

gmzunino@conicet.gov.ar

Psycholinguistic studies show evidence of the complex relationship between language, gender and thought in two directions: while some research shows that linguistic forms can influence cognition [9], others point out that gender stereotypes can condition language comprehension [7]. Some investigations postulate that the generic masculine tends to be interpreted consistently with an exclusive reference to men, specially for languages with grammatical gender [3, 4, 8, 12]. Furthermore, several studies analyze the influence of gender stereotypes during language comprehension and most agree on a significant incongruence effect [4, 5]: it is more costly when morphological information does not coincide with stereotypical information. The evidence of most of these studies, however, is the result of explorations in a single language, mainly English, which has a morphological gender marking system concentrated on pronouns, not inflections.

Studying how grammatical and stereotypical gender interact during processing across different languages can provide critical insights into variation in language processing. Particularly, Romance languages are paradigmatic for studying this phenomenon, and show ongoing debates regarding the so-called (in)congruence effect [2, 5]. No previous cross-linguistic work with the same experimental design that considers grammatical and stereotype interaction on the (in)congruence effect during sentence processing was done. Furthermore, all previous studies have considered just European varieties of Spanish and Portuguese, reinforcing the Global North bias in psycholinguistic research [1].

Within the framework of the MultiLingualGender - Horizon MCSA project, we proposed a grammatical maze study in three Romance languages (Italian, Peninsular and Rioplatense Spanish and Brazilian Portuguese) to study potential interaction effects between gender stereotypes and grammatical gender during sentence comprehension. We predict that 1) both male and female stereotypical bias will generate an obstacle during processing in the incongruent grammatical gender condition (e.g., *camionera*, truck driver-F; *niñero*, babysitter-M); 2) we expect different processing costs for the pronouns modulated by Language, while it is expected that the contrastive vs. anaphoric interpretation generates differences in pronominal processing, also considering generic or specific interpretations of the masculine morphological marking. Testing these hypotheses in languages with grammatical gender could provide deeper insights into the relationship between gender, language and mental representations.

In a 2x2x3 design we manipulated Stereotipical Bias, that is, the stereotypicality of role names (male, e.g., *camionero*, truck driver-M; female, e.g., *niñera*, babysitter-F;); Gender morphology (masculine, e.g., *los camioneros*, truck drivers-M; feminine, e.g., *las camioneras*, truck drivers-F); and Language (Italian, Portuguese, Spanish).

We designed 44 sentences organized in 4 counterbalanced lists within a grammatical maze paradigm (Figure 1), based on two normative tasks to select suitable role names [11].

Data was collected online through PCIbex [7] in four countries: Argentina, Brazil, Italy and Spain (details in page 3). For examples of the sentences used see Table 1.

Preliminary analysis shows a specific incongruence effect for male role names with feminine grammatical gender, but not vice versa, showing that the acceptability and possibility of representing women in roles typically associated with men might be more difficult than the inverse situation. Furthermore, we found diatopic differences in the interpretation of the pronoun between both varieties of Spanish, which poses interesting challenges in the framework of the psycholinguistic effects of linguistic variation. Even within romance languages, all grammatical gendered languages, there are differences among them that might be interpreted as the result of an interaction between linguistic variation with socio-cultural factors.

Figure 1. Grammatical maze example in Spanish

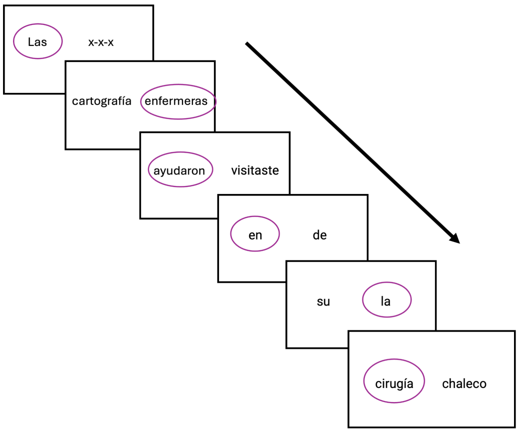


Table 1. Example of female Stereotypical bias items in the Spanish, Italian and Portuguese

	Incongruence Stereotypical bias/Morphological gender	Congruence Stereotypical bias/Morphological gender
Spanish (Argentina and Spain)	*Los enfermeros* ayudaron en la cirugía; *ellos* prepararon los instrumentos necesarios para la operación.	*Las enfermeras* ayudaron en la cirugía; *ellas* prepararon los instrumentos necesarios para la operación.
Italian	*Gli infermieri* controllano i turni di guardia del giorno; *tutti* trovano degli errori madornali.	*Le infermiere* controllano i turni di guardia del giorno; *tutte* trovano degli errori madornali.
Brazilian Portuguese	*Os enfermeiros* verificaram a escala de plantão do dia; *eles* corrigiram vários erros.	As enfermeiras verificaram a escala de plantão do dia; *elas* corrigiram vários erros.

References

1. Berghoff, R., & Bylund, E. (2025). Diversity in research on the psychology of language: A large-scale examination of sampling bias. *Cognition*, *256*, 106043. https://doi.org/10.1016/j.cognition.2024.106043

2. Bürki, A., van den Hoven, E., Schiller, N. & Dimitrov, N. (2023). Cross-linguistic differences in gender congruency effects: Evidence from meta-analyses. *Journal of Memory and Language, 131*. 104428 https://doi.org/10.1016/j.jml.2023.104428
3. Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, 28(2), 277-293. http://doi.org/10.1017/S0142716407070142
4. Carreiras, M., Garnham, A., Oakhill, J., Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, 49(3), 639-663. http://dx.doi.org/10.1080/713755647
5. Casado, A., Sá-Leite, A., Pesciarelli, F., & Paolieri, D. (2023). Exploring the nature of the gender-congruency effect: implicit gender activation and social bias. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1160836
6. Duffy, S.A., & Keir, J.A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32(4), 551-559. https://doi.org/10.3758/BF03195846
7. Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, *4*(10), 1021-1028. https://doi.org/10.1038/s41562-020-0918-6
8. Pinheiro, B. F. M., & Freitag, R. M. K. (2020). Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência. *Linguística*, 16(1), 85-107. http://doi.org/10.31513/linguistica.2020.v16n1a31637
9. Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology*, *6*, 1631. https://doi.org/10.3389/fpsyg.2015.01631
10. Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832
11. Zunino, G. M., Gagliardi, G., Stetie, N. A., & Miola, E. (in press). Female truck drivers and male babysitters? Interactions between gender stereotypes and grammatical gender in Spanish and Italian: a psycholinguistic proposal. *Linguistik Online*, in press.
12. Zunino, G.M., & Stetie, N.A. (2022). Binary or non-binary? Gender morphology in Spanish: Differences dependent on the task. *ALFA: Revista de Lingüística*, 66, e14546. https://doi.org/10.1590/1981-5794-e14546

Particularities for this multilingual study

For each language, the selection of role names that were used to design the sentences was made based on the previous normative in each linguistic community [11].

The construction and adequate adaptation of the experimental materials to the three languages and their varieties was a great challenge. Although the priority was to maintain the naturalness of the items for each linguistic community (this is why for some items, we decided not to offer exact translations but rather ecological adaptations for each language), certain basic restrictions were imposed in all languages: (a) total length of each sentence between 13 and 15 words, with an average of 13.5; (b) number of words 4 to 6, with an average of 4.5, between the NP and the personal pronoun.

In particular, one key element should be highlighted: since we required that the pronoun that begins the second clause of each sentence have gender marking, in Italian, instead of a personal pronoun such as “ellas/ellos” (Spanish) or “elas/eles” (Portuguese), the pronoun “tutte/tutti” was used, to avoid specific interpretation that are generated in Italian but not in Spanish or Portuguese.

For the two varieties of Spanish (Argentina and Spain), a meticulous lexical adaptation was carried out to guarantee the naturalness of the items in each linguistic community.

Composition-sensitive predictions: Incremental Processing of Experiential Perfects

Nino Grillo[1], Buhan Guo[1], Keir Moulton[2], Shayne Sloggett[1]

[1]University of York, [2]University of Toronto

nino.grillo@york.ac.uk

Background: Much of the work on memory in sentence processing—whether focused on retrieval, maintenance, or prediction—has emphasized dependencies that are resolved by accessing or anticipating specific lexical items or morpho-syntactic features. These include well-studied phenomena like subject-verb agreement [1], filler-gap dependencies [2, 3, 4], and lexical predictability effects [5, 6, 7], where the relevant information can be directly linked to a stored representation. However, not all grammatical constraints operate over discrete lexical features. Some require the online composition of complex meanings from multiple linguistic elements—meanings that are not recoverable from any single word alone (1). These cases raise important questions about how compositional interpretations are built and maintained during real-time comprehension.

We explore this issue through the lens of *Experiential Perfect* constructions in English, a use of the present perfect aspect often cued by the polarity-sensitive item *ever* (e.g., *John hasn't ever eaten seal*; [8]). These constructions are only felicitous when the verb phrase denotes a repeatable event, i.e. an *event kind* [9, 10, 11]. As seen in (1), *kill flies* is acceptable, while *kill the fly/John* is not, since the latter describes a specific, one-time event. Crucially, repeatability is not a lexical or morpho-syntactic property of either the verb or the object DP alone but emerges as an interaction from the composition of verb semantics and object definiteness: *kill* allows repeatable readings with indefinites but not definites; *see* permits both [12].

(1) a. Have you ever killed a fly? / b. *Have you ever killed the fly/John? / c. Have you ever seen John?

Because *ever* can trigger the experiential reading early on, this raises the question of whether comprehenders generate expectations about upcoming verb phrases based on the requirement for repeatable events, and when and how this compositional information is computed during processing. Specifically, we ask whether comprehenders (i) predict the semantic compatibility of upcoming material with the experiential frame, and (ii) compute repeatability incrementally as the verb and object are encountered. This study sheds light on how real-time comprehension integrates compositional semantic constraints with mechanisms of predictive processing.

Method: Participants read sentences in a Grammatical-Maze task (n=133). Stimuli followed a 2×2 within-subject design crossing VERB TYPE (Consumption vs. Repeatable) and DETERMINER TYPE (Definite vs. Indefinite), as shown in Table 1. If the repeatability constraint is computed incrementally, definites should be harder to process than indefinites under consumption verbs in unambiguous experiential contexts.

Results: As shown in Figures 1 and 2 respectively, we found a significant interaction of VERB TYPE * DETERMINER TYPE at both the Determiner (β=0.04, p=.01) and the spillover noun (β=0.05, $p<.01$). While reading times were generally longer for Definites than Indefinites–likely due to definite DPs being unsupported out context in this study–this effect was crucially larger under Consumption verbs than Repeatable verbs as predicted.

Discussion: These results suggest that comprehenders maintain and integrate abstract semantic constraints—like event repeatability—during incremental processing. Such constraints are not computed via retrieval of lexical or morpho-syntactic features alone, but must be evaluated against verb phrase semantics, supporting models that allow for composition-sensitive prediction in comprehension. Ongoing work tests experientials over longer distances, via perceptual reports (2), and ties these results to semantic processing models [13].

V-Type	D-Type	Example
Consumption	Def	Maria knew that Sam hasn't ever eaten the cookie although he wanted to
Consumption	Ind	Maria knew that Sam hasn't ever eaten a cookie although he wanted to
Repeatable	Def	Maria knew that Sam hasn't ever touched the cookie although he wanted to
Repeatable	Ind	Maria knew that Sam hasn't ever touched a cookie although he wanted to

Table 1: Experimental design and example stimuli

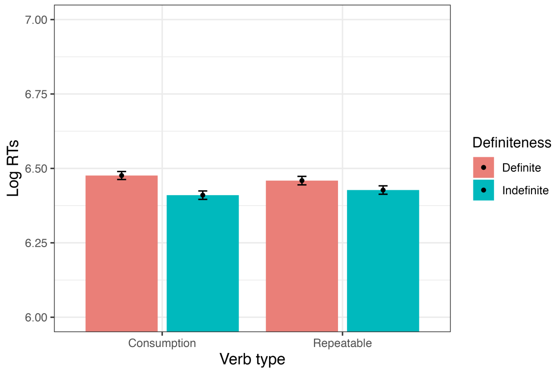


Figure 1: Log RTs at the determiner (*the/a*); Error bars indicate standard errors.

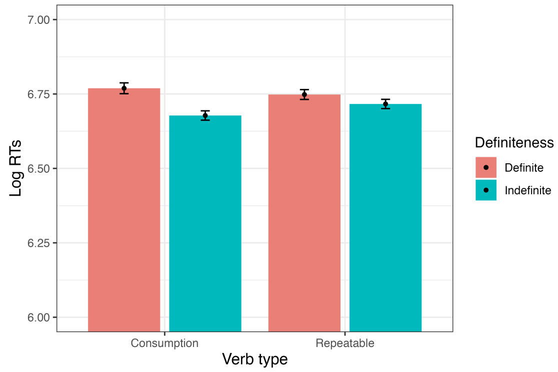


Figure 2: Log RTs at the noun (e.g. *cookie*); Error bars indicate standard errors.

(2) Mary hasn't ever seen Sam watch John kill a fly / *the fly/ *Bill.
(Note: when combined with direct percetion of an event, experientials only allow perception of a repeatable event [14], the ungrammatical example can only be rescued in a situation involving e.g. a video recording of the event of Sam killing the fly / Bill).

References

[1] Wagers, Lau, & Phillips. (2009). Agreement attraction in comprehension. *JML*, *61*.

[2] Lewis & Vasishth. (2005). An activation-based model of sentence processing. *Cognitive Science*, *29*.

[3] Wagers & Phillips. (2009). Multiple dependencies & role of the grammar in real-time comp. *J.of Ling.*, *45*.

[4] Ness & Meltzer-Asscher. (2017). Wm in the processing of long-distance dependencies. *JPR*, *46*.

[5] Altmann & Kamide. (1999). Incremental interpretation at verbs. *Cognition*, *73*(3), 247–264.

[6] Lau, Holcomb, & Kuperberg. (2013). Dissociating N400 effects of prediction/association. *J.C.Neuro.*, *25*.

[7] Pickering & Gambi. (2018). Predicting while comprehending language. *Psychological bulletin*, *144*.

[8] Comrie, B. (1976). Aspect: An introduction to the study of verbal aspect and related problems. *CUP*.

[9] Inoue, K. (1975). *Studies in the perfect* [Doctoral dissertation, University of Michigan].

[10] Dahl, O. (1985). *Tense and aspect systems*. Blackwell Publishers.

[11] Gehrke. (2019). Event kinds. In Truswell (Ed.), *The oxford handbook on event structure*. OUP.

[12] Grillo, Gil, Marsden, Radkevic, Sloggett, Tsoulas, & Yeo. (2022). Properties of L1 experientials modulate the acquisition of *Ever* and *Any* in Chinese and Korean L2 speakers of English. *HSP 35*.

[13] Brasoveanu & Dotlačil. (2020). *Computational cognitive modeling and linguistic theory*. Springer Nature.

[14] Grillo & Moulton. (2016). Event kinds and the pseudo relative. *Proceedings of NELS 46*, *2*, 11–20.

Dissociating Speaker-Specific Effects on Referential Precedent Interpretation

Andrés Contreras[1, 2], Anita Tobar-Henríquez[3], Ernesto Guerra[3], Edmundo Kronmüller[2]
[1] Faculty of Medicine, University of Chile; [2] School of Psychology, Catholic University of Chile; [3] Center for Advanced Studies in Education, University of Chile
andrescontreras@ug.uchile.cl

Background: Conversations often rely on referential precedents to reduce ambiguity, but how these work, especially the roles of speaker-specific cues and contextual inference, remains debated [1,2]. The two-stage extended model suggests both automatic cue-driven and inferential processes are involved [1]. This study aims to separate these processes, which past research has typically conflated.

Method: Participants played a referential game in the context of a visual-world eye-tracking paradigm. We created twenty-four item sets of homophone pairs of six displays of three images of everyday objects each, and participants heard an instruction to click one by a co-present confederate (whom the participants believed was a naive participant), or a pre-recorded speaker. In the test trial, two of the three images represented a homophone pair (e.g, a candle and a sail, both *vela* in Spanish), one target and the other competitor, always referred to by the co-present speaker. We used the remaining displays to create the four experimental conditions: Same Speaker Precedent (SS/P), where the co-present speaker gave the precedent of the target object; Different Speaker Private Precedent (DS/PP), where the pre-recorded speaker gave the precedent by headphones only heard by the participant; Different Speaker Shared Precedent (DS/SP), where the pre-recorded speaker gave the precedent by loudspeakers heard by both participant and confederate; and No Precedent (NP), where there was no precedent given. We hypothesize early differences between the DS/SP and SS/P conditions (the latter favoring the target image, perceptual cue effect), and a later difference between the DS/SP and the DS/PP conditions (the former favoring the target image, common ground effect).

Results: There is a reliable effect of precedent in target selection. Participants selected the target more often when there was a precedent (67.9%), compared to when there was not (50.5%). This difference is statistically significant (β=0.634; Z=4.906; p<0.001). Participants were faster to make their choice when there was a precedent (M=1612; SD=878; Md=1345) compared to when there wasn't (M=1907; SD=943; Md=1417) (β=-0.062; t=-2.490; p=0.0166). Eye movement data shows a preference for the target image in all precedent conditions, but there was no effect of common ground when keeping constant perceptual cues. To test our hypothesis, we conducted three cluster permutation analyses, one for each contrast. First, we test the precedent effect by comparing the NP condition against the mean of the other three conditions where a precedent was present. There is a reliable cluster that starts at 500 ms after noun onset and lasts until the end of the window by subjects (p < 0.001) and by items (p < 0.01). Regarding the effect of perceptual cues when keeping constant common ground, we compared the DS/SP against the SS/P condition, finding a reliable cluster starting at 1220 ms. and ending at 1340 ms. in the analysis by subject (p < 0.001) and by items (p < 0.01), which was opposite as hypothesized. No significant cluster was found when looking at the effect of common ground when keeping constant perceptual cues.

Discussion: We found only evidence concerning the effect of precedent, and found no evidence of an effect of common ground. We believe participants relied heavily on precedents, rather than on common ground. We propose a follow-up experiment, where we can more heavily influence participants to rely on common-ground, and where issues regarding frequency of use in homophone pairs can be addressed.

References

[1] Kronmüller, E., and Guerra, E. (2020). Processing Speaker-Specific Information in Two Stages During the Interpretation of Referential Precedents. *Front. Psychol., 11*:552368. doi: 10.3389/fpsyg.2020.552368

[2] Kronmüller, E., and Barr, D. J. (2015). Referential precedents in spoken language comprehension: a review and meta-analysis. *J. Mem. Lang. 83*, 1–19. doi:10.1016/j.jml.2015.03.008

Figures

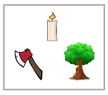
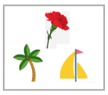
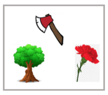
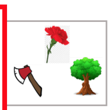
	Grounding Phase			Test Phase		
Visual Context / **Condition**						
Same Speaker Precedent	**La vela (The candle)**	**El hacha (The axe)**	**La flor (The flower)**	**La flor (The flower)**	**La vela (The candle)**	**El hacha (The axe)**
Same Speaker No Precedent	**El árbol (The tree)**	**El hacha (The axe)**	**La flor (The flower)**	**La flor (The flower)**	**La vela (The candle)**	**El hacha (The axe)**
Different Speaker Private Precedent	La vela (The candle)	El hacha (The axe)	El clavel (The carnation)	**La flor (The flower)**	**La vela (The candle)**	**El hacha (The axe)**
Different Speaker Shared Precedent	*La vela (The candle)*	*El hacha (The axe)*	*El clavel (The carnation)*	**La flor (The flower)**	**La vela (The candle)**	**El hacha (The axe)**

Figure 1. Implementation of an item in each experimental condition. A schematic representation of a six-trial sequence in which a Test trial is embedded (highlighted in red). The sequence is divided into two phases: Grounding (first three trials) and the Test (subsequent three trials). All conditions are implemented in the grounding phase, whereas the testing phase remains the same. In bold are instructions given by the co-present speaker. In the DS/PP (regular font) condition, instructions in the grounding phase are played through earphones, heard only by the participant, and in the DS/SP condition (italics), they are played through loudspeakers, so both the confederate and participant can hear them.

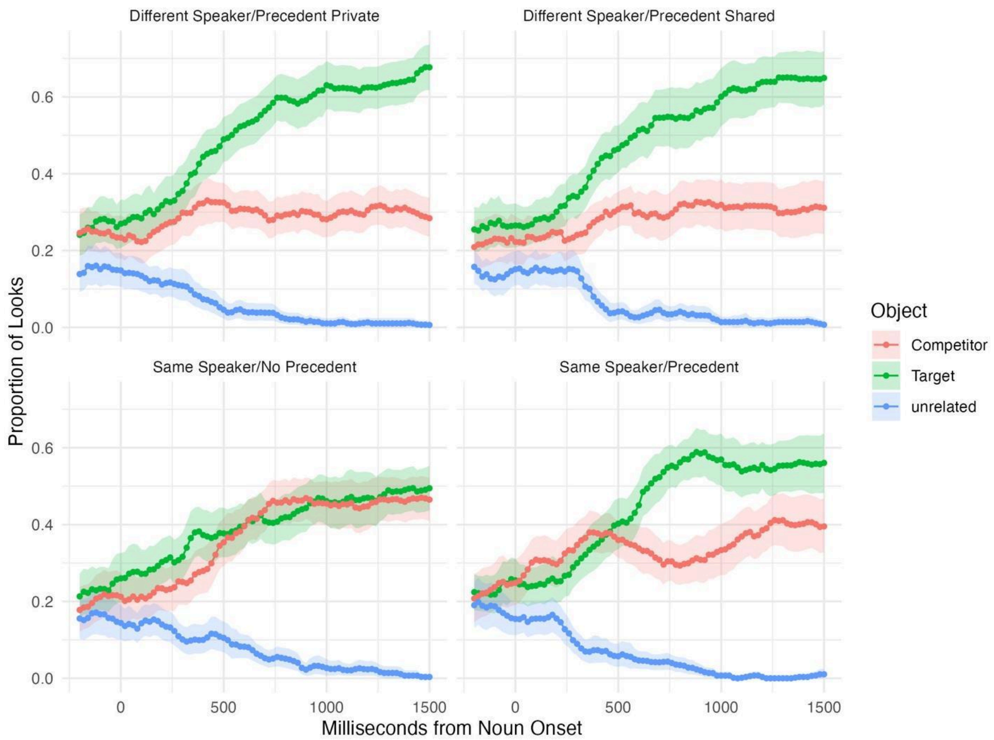


Figure 2. Proportion of looks to each object per condition from onset of noun. Confidence intervals 95% calculated by bootstrap subjects (n=10000)

Item sets used in the experiments: Chilean Spanish

The experiments presented in this abstract were done in the context of Chilean Spanish, so the homophone pairs used correspond to nouns referring to everyday objects in Chile. They were the following (in italics, the translation to Spanish):

1. cat - car jack: *gata*
2. doll - wrist: *muñeca*
3. llama - flame: *llama*
4. sail - candle: *vela*
5. cup - glass: *copa*
6. siren - mermaid: *sirena*
7. dumbbell - weight: *pesa*
8. leaf - paper sheet: *hoja*
9. playing card - letter: *carta*
10. tail - glue: *cola*
11. pomegranate - hand grenade: *granada*
12. comet - kite: *cometa*
13. basketball hoop - earring: *aro*
14. bank - bench: *banco*
15. drum kit - battery: *batería*
16. button - button: *botón*
17. TV screen - lampshade: *pantalla*
18. sneaker - plug-in: *zapatilla*
19. key - faucet: *llave*
20. floor - stool: *piso*
21. photo camera - bicycle inner tube: *cámara*
22. computer keyboard - piano keyboard: *teclado*
23. bomb - firehouse: *bomba*
24. foot sole - plant: *planta*

Network Properties of Chinese Characters and their Effect on Processing

Yixia Wang, Peter Hendrix, Emmanuel Keuleers
Department of Cognitive Science and Artificial Intelligence, Tilburg University
y.wang_1@tilburguniversity.edu

Background: In recent years, methods from network analysis have been used to reveal insights about the structure of the lexicon [1], cross-linguistically [2], and developmentally [3]. While these studies have focused on alphabetic languages, little is known about lexical networks for writing systems such as Chinese, which uses sequences of hierarchically organized characters instead of linear strings of letters to represent words. Knowing how such networks are organized can be an essential component in understanding the mental representation and processing of more complex script systems.

While for alphabetic scripts, the similarity between words can be easily determined using existing string similarity measures, the hierarchical nature of Chinese characters makes determining similarity non-trivial. Recently, [4] have proposed using ideographic description sequences (IDS) as a linearization of the hierarchical structure of Chinese characters to which string similarity methods can be applied. In this paper, we use IDS-based networks to analyze and compare the structural properties of the Chinese orthographic and phonological lexicon. In addition, we investigate how centrality measures from these networks predict lexical decision times for Chinese characters.

Method: To construct the networks, we represented character orthography using IDS [4], which describes characters as sequences of stroke patterns and spatial relations, and phonology using pinyin, a romanized transcription of character pronunciation. We built an orthographic network by connecting characters that had a one-component edit distance between their IDS representations and a phonological network, connecting nodes that had a one-letter edit distance in pinyin. The orthographic network included 7,825 characters; the corresponding phonological network covered 1,370 unique syllables (with tones). This nearly 6:1 ratio between characters and their pronunciations is indicative of the large amount of homophony in Chinese.

For network measures, we computed the *degree* for nodes in both the phonological and orthographic networks and two additional centrality measures for nodes in the orthographic network—*clustering coefficient* and *betweenness centrality*. These measures were then used to perform a correlation analysis. Next, we used linear mixed effects models to evaluate the extent to which these measures contribute to predicting lexical decision times for characters from the Simplified Chinese Lexicon Project [5], which collected latencies and accuracies for **all** 8,105 characters in the *List of Commonly Used Standard Chinese Characters*.

Results: 1) The phonological network consisted of a monolithic, densely connected component, illustrative of the compressed phonological space in Mandarin, where a small inventory of syllables leads to a high level of overlap in pronunciation. **2)** The orthographic network, in contrast, was sparse and fragmented. Unexpectedly, the use of ideographic description sequences showed that the structural composition of the Chinese character network is similar to what has been found for alphabetic languages. In contrast, using existing approaches (for instance, connecting nodes that share a radical or a stroke) would have lead to a fragmented network composed of small, disconnected clusters, offering little insight in the lexical network of written Chinese beyond local orthographic neighbourhoods. **3)** The orthographic and phonological metrics were not correlated, further confirming their structural differences, and contrasting with alphabetic languages, where the relationship between spelling and sound is governed by the grapheme-phoneme correspondence (GPC) rules, to varying extents. **4)** The centrality measures computed on the orthographic network significantly predicted character recognition, after controlling for word frequency as Zipf value [6] and visual complexity (number of strokes), highlighting how a character's position within the network influences how quickly it is recognized.

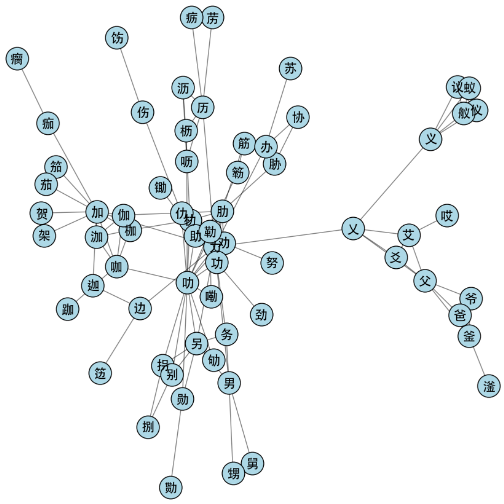


Figure 1: The second largest connected component in the orthographic network with 67 nodes and 135 edges.

	Zipf value	NStrokes	degree	clustering coefficient	betweenness centrality	degree-py
RT	-0.353	0.257	-0.093	0.103	-0.207	0.003
Zipf value		-0.298	0.136	-0.150	0.273	-0.001
NStrokes			-0.352	0.127	-0.462	-0.010
degree				0.242	0.550	-0.002
clustering coefficient					-0.293	-0.045
betweenness centrality						0.032

Table 1: A Spearman’s rank correlation matrix.

Variable	Estimate	Std. Error	*t*	*df*	*p*
(Intercept)	-1.921	0.231	-8.332	2765.406	<.001***
sqrt NStrokes	0.111	0.005	23.800	2828.177	<.001***
Zipf value	-0.086	0.002	-46.965	2872.992	<.001***
clustering coefficient	0.032	0.007	4.618	2832.538	<.001***
log degree	-0.001	0.003	-0.460	2831.969	.645
transformed betweenness centrality	-0.383	0.231	-1.654	2737.974	.098
sqrt degree-py	0.001	0.003	0.498	2822.587	.619
(Intercept)	-1.550	0.023	-67.504	28.397	<.001***
RC1 (transformed betweenness centrality .97)	-0.023	0.002	-10.482	2729.040	<.001***
RC2 (clustering .97)	0.021	0.002	10.089	2833.929	<.001***
RC3 (sqrt degree-py 1.00)	0.001	0.002	0.516	2822.390	.606
RC4 (Zipf value .99)	-0.112	0.002	-53.623	2875.024	<.001***
RC5 (sqrt NStrokes .96)	0.072	0.002	33.954	2828.186	<.001***
RC6 (log degree .95)	-0.014	0.002	-6.685	2835.495	<.001***

Table 2: Comparison of model results for fixed effects factors before and after applying PCA.

References

1 Siew, C. S. (2018). The orthographic similarity structure of English words: Insights from network science. *Applied Network Science*, *3*(1), 13.

2 Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679–685.

3 Siew, C. S., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General*, *149*(12), 2376.

4 Wang, Y., & Keuleers, E. (2024). Simplified Chinese character distance based on ideographic description sequences. *Proceedings of the Second Workshop on Computation and Written Language (CAWL)@ LREC-COLING 2024*, 59–66.

5 Wang, Y., Wang, Y., Chen, Q., & Keuleers, E. (in press). Simplified Chinese Lexicon Project: A lexical decision database with 8105 characters and 4864 pseudocharacters.

6 Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-uk: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology*, *67*(6), 1176–1190.

Chinese script and orthographic structure of characters We here give a brief introduction to the aspects of the Chinese writing system relevant to our network construction. Chinese script is **morphosyllabic**: each character typically corresponds to a morpheme with a specific pronunciation, almost always a single syllable. Like alphabetic writing systems, Chinese writing consists of a linear sequence of characters. However, unlike alphabetic languages where letters are typically building blocks of words, Chinese characters are hierarchically organized structures and almost every character is also a word, with the majority of words consisting of two-characters. In addition, unlike alphabetic scripts where readers only need to recognize a small set of letters (e.g., 26 in English), Chinese readers need to recognize several thousand characters (around 8000). These characters have a wide range of visual complexity, from visually simple (e.g., 一 *yi1* 'one', contains 1 stroke) to very complex (e.g., 龖 *da2* 'depicting the majestic soaring of a dragon', contains over 30 strokes) or more [4]. Despite this variation, each character is written within a uniform square block, creating a "standardized" spatial arrangement across the script.
IDS is a system to describe each character using stroke patterns and their spatial relations (e.g., side-by-side, top-bottom). It works recursively, allowing for systematic descriptions and thus comparisons of characters of different visual complexity. For example [4]:

- 作 is represented as ⿰亻乍,
- 网 as ⿵冂⿰メメ,
- 鬱 as ⿳⿲木⿱メ⿻丿㇏木冖⿰⿱⿶凵⿻メ⿳丶⿰丶丶丶⿺㇄丿彡.

Sentence Processing and Memory: Immediate Recall of Information from Adjectives with Different Syntactic Status in Czech Adult Speakers

Eva Pospíšilová[1], Jan Chromý[1]
[1]Faculty of Arts, Charles University
eva.pospisilova@ff.cuni.cz

Background: Earlier studies have observed systematic differences in speakers' ability to recall information from written sentences [1]. Both studies on English [2] and Czech speakers [3, 4] have shown that people tend to recall the subject and object of a sentence more easily, whereas information conveyed through verbs, attributes or adverbial adjuncts specifying place or time is more difficult to remember. These findings suggest that during sentence processing, people selectively direct their attention to information they have learned to be important, while selectively ignoring peripheral details. Our study builds on Czech research [3, 4] and examines recall differences for adjectives modifying either direct objects or adverbial adjuncts. Since direct objects are generally recalled better than adjuncts, we expect that the same effects will extend to the modifiers of these sentence elements, meaning that adjectives modifying the direct object will be recalled better.

Method: Sentence stimuli were adapted and modified from Chromý and Tomaschek [4]. Each sentence included a temporal adverbial adjunct, a locative adverbial adjunct modified by an adjective, a subject expressed by a full name, a verb, and a direct object modified by an adjective (see Table 1). Before the experiment began, the pre-selected stimuli were rated for acceptability by native speakers of Czech. Based on these ratings, a final set of 24 target sentences was selected. Two open-ended questions were then created for each of these sentences: one targeting the information conveyed by the adjective modifying the direct object, and one probing the information from the adjective modifying the adverbial adjunct. These 24 sentences with questions were accompanied by a total of 92 filler sentences with questions targeting different syntactic information (e.g., subject, object, locative adjunct). A total of two experiments were conducted using the self-paced reading method followed by open-ended questions. Each participant saw each sentence once, questions were assigned to the target sentences via the Latin square design. In Experiment 1 (N=149), sentences were shown word-by-word; in Experiment 2 (N=154), they appeared as a whole. The accuracy of recall in the responses was subsequently analyzed and compared.

Results: Figure 1 shows the differences in recall accuracy between adjectives modifying direct objects and adjectives modifying adverbial adjuncts from the Experiment 1. We implemented a generalized linear mixed model, which revealed a significantly higher percentage of incorrect responses for adjectives modifying the adverbial adjunct ($\beta = 2.72$, $SE = 0.32$, $z = 8.49$, $p < 2e-16$). Similar findings were observed in Experiment 2 as well, with generalized linear mixed model also revealing significant effect of the type of adjectives on the recall accuracy ($\beta = 1.15$, $SE = 0.29$, $z = 3.91$, $p < 0.001$). As shown in Figure 2, in the Experiment 2 - where participants saw the sentence as a whole - there were overall fewer errors in recalling adjectives linked to adverbials.

Discussion: Our findings show poorer recall for adjectives linked to adverbial adjuncts and better recall for adjectives modifying direct objects, indicating that recall effects extend beyond heads to dependent elements. Higher error rates in the word-by-word condition suggest that limited re-reading impairs memory for peripheral information. This suggestion could be further explored using online methods, such as eye-tracking.

Key words: sentence processing, information recall, self-paced reading, adjectives

sentence	question	targeted information
V devět hodin v útulném pokoji četl Josef Novotný novou knihu. *At nine o'clock in the cozy room Josef Novotný read a new book.*	Jaká to byla kniha? *What kind of book was it?*	adj_obj
V devět hodin v útulném pokoji četl Josef Novotný novou knihu. *At nine o'clock in the cozy room Josef Novotný read a new book.*	Jaký to byl pokoj? *What kind of room was it?*	adj_adv

Table 1: *Experimental items example: adj_obj = combination of a sentence with a question targeting an adjective modifying direct object, adj_adv = combination of a sentence with a question targeting an adjective modifying adverbial adjunct*

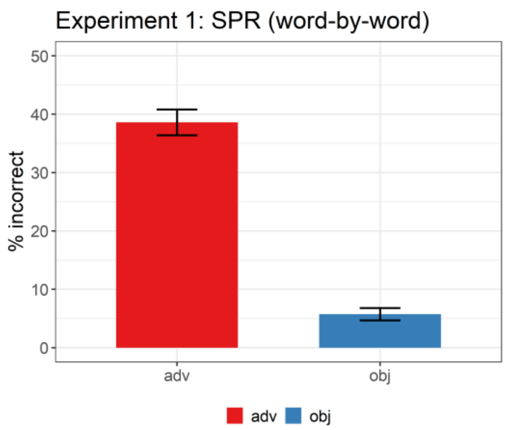


Figure 1: *% of incorrect answers to questions targeting adjectives modifying adverbial adjuncts (adv) and questions targeting adjectives that modify direct objects (obj) from the Experiment 1*

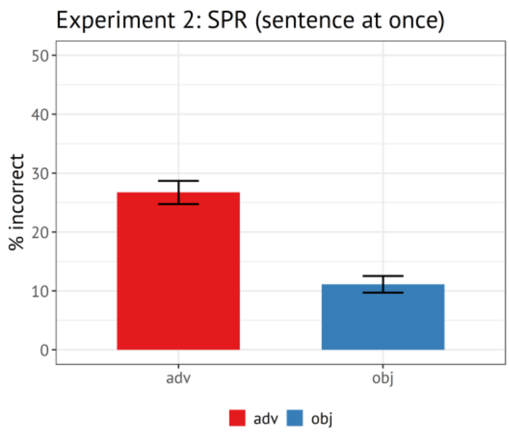


Figure 2: Figure 1: *% of incorrect answers to questions targeting adjectives modifying adverbial adjuncts (adv) and questions targeting adjectives that modify direct objects (obj) from the Experiment 2*

References

[1] Clark, H. H. (1966). The prediction of recall patterns in simple active sentences. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 99-106.

[2] James, C. T., Thompson, J. G., & Baldwin, J. M. (1973). The reconstructive process in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 51-63.

[3] Chromý, J., & Vojvodić, S. (2024). When and where did it happen? Systematic differences in recall of core and optional sentence information. *Quarterly Journal of Experimental Psychology*, *77*(1), 111-132.

[4] Chromý, J., & Tomaschek, F. (2024). Learning or boredom? Task adaptation effects in sentence processing experiments. *Open Mind*, *8*, 1447-1468.

[5] Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, 56(7), 485–495.

Predicting ellipsis usage with a game-theoretic model informed by production data

Robin Lemke[1]
[1] Saarland University
robin.lemke@uni-saarland.de

Game-theoretic approaches have become increasingly popular in pragmatics, but most evidence comes from controlled studies [1] and it is unclear whether it scales up to more complex situations. I explore this with at the case of fragments (1a), which speakers can often use instead of a complete sentence (1b). I hypothesize that their choice depends on a trade-off between the lower production cost of the shorter fragment and the risk of misunderstanding: Fragments are ambiguous, so the listener might infer (1c) instead of (1a). This is modeled as a signaling game [1, 2]: The speaker sends an utterance $u \in U$ to communicate their message $m \in M$ and the listener has to infer which m the speaker had in mind (See Eq. 1, 2). Shorter utterances have a lower cost c. The $m \in M$ can differ in prior probability Pr(M) and the listener assigns the most likely m_{max} to a fragment u_{frag}, so the speaker chooses u_{frag} if the lower c outweighs the risk of u_{frag} not being interpreted as m_{max}.

Experiments: M, U, Pr(m) and $[[u]]_m$ were estimated from crowd-sourced production data by [3] and the model predictions evaluated with 3 web-based experiments (60 subjects each) (Fig. 2). The participant takes the speaker role, while the listener role is simulated. In each trial (n = 15), the participant reads a story, 3 messages that could be communicated and 6 utterances. 3 of the utterances are sentences equivalent to the messages. Among the fragments, one is ambiguous between two of the sentences, one unambiguously refers to the third sentence. An explicit cost structure is implemented by a system of coins: sentences (130) are more expensive than fragments (40) and successful communication rewarded with a payoff (100).The 3 experimental conditions differ in the message that the speaker is asked to communicate: (i) the *most likely* message given the ambiguous fragment, (ii) a *less likely* one, (iii) the message corresponding to the *unambiguous* fragment. My account predicts a gradual increase of fragment ratio as a function of the fragment's p(m|u).

Results: The data for **Exp. 1** (See Fig. 3) were analyzed with LMMs [4], confirming that fragment ratio increases with fragment p(m|u) ($\chi^2 = 16.01, p < 0.001$). Since this does not hold within the subset of ambiguous conditions ($\chi^2 = 16.01, p > 0.5$), the effect could result from subjects avoiding any ambiguity instead of game-theoretic reasoning. **Exp. 2** addressed this by replacing the low p(m|u) condition by an unambiguous one, making the experiment fully unambiguous. Fragment ratio higher is higher than in exp. 1, ($|z| = 4.14, p < 0.001$), but there is also a p(m|u) main effect ($|z| = 8.92, p < 0.001$). **Exp. 3** used a new low p(m|u) condition (not based on the production data), which addresses the possibility that the messages produced by subjects were not unlikely enough. Unlike in exp. 1 and 2, the p(m|u) effect was also significant for the ambiguous conditions only ($|z| = 5.82, p < 0.001$). **Exp. 4** aimed at replicating this with a free production task, where subjects entered their answer into a text field (each word cost 20 coins). The study found no effect of p(m|u) on fragment ratio, but utterances communicating predictable messages were longer (in words) ($|z| = 2.57, p < 0.05$).

Discussion The utterance selection studies confirm the game-theoretic prediction that fragments are used to communicate predictable meanings: Exp. 2 shows the effect cannot be explained by ambiguity avoidance and exp. 3 replicates it even within the ambiguous conditions. This extends the evidence for game-theoretic reasoning from tightly controlled experiments to more diverse and realistic situations. Exp. 4 suggests that subjects do not only use fragments, but also other means of reducing utterance length to communicate efficiently.

(1) Passenger approaches conductor at the train station next to a waiting train.

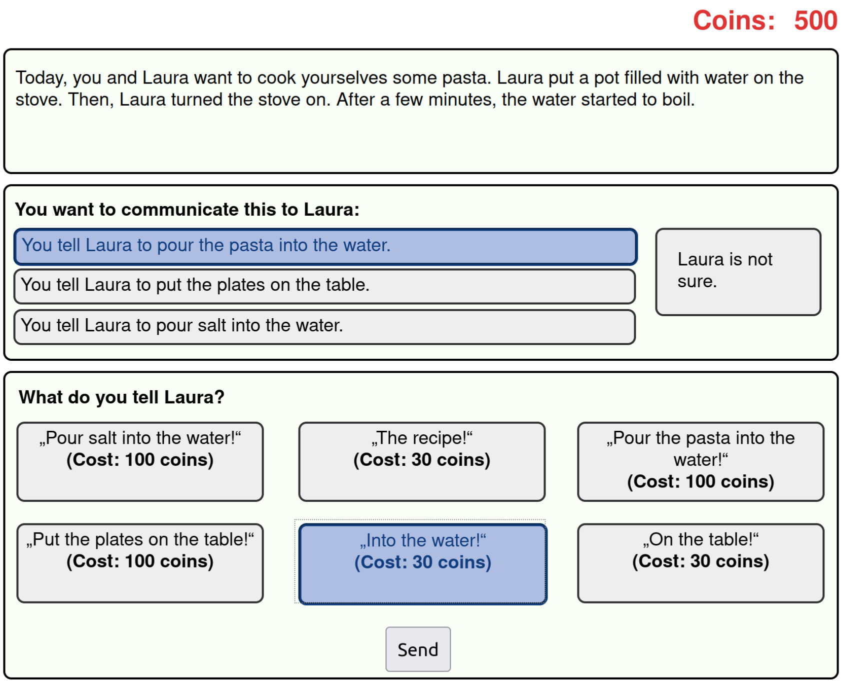
a. To Paris? (Fragment)

b. Does this one go to Paris? (Sentence)

c. Have you ever been to Paris? (Alternative interpretation)

$$L_0(m, u) = \frac{Pr(m) \times [[u]]_m}{\sum_{m'} Pr(m') \times [[u]]_{m'}} \quad (1) \qquad EU_{S1}(u, m) = L_0(m, u) \times payoff - cost \quad (2)$$

The listener reweighs Pr(M) among the messages from which the fragment could have been derived by grammatical omission ($[[u]]_m = 1$). The speaker maximizes their Expected Utility, which depends on the listener behavior, the payoff, and the utterance cost. Since cost and payoff are kept constant within each experiment, I use p(m|u) as a predictor in my analyses.



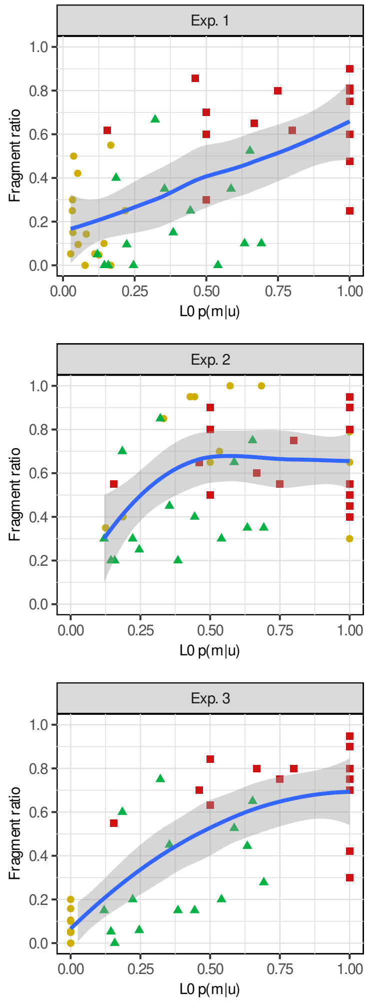


Figure 1: Screenshot of the experiment, which was conducted in German, translated to English for convenience.

Figure 2: Ratio of fragments (errors excluded) by item and condition across the three utterance selection experiments. The low p(m|u) condition is displayed in yellow ●, the high p(m|u) condition in green ▲ and the unambiguous condition in red ■.

References

[1] Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

[2] Franke, M. (2009). *Signal to act: Game theory in pragmatics* (Doctoral dissertation). Universiteit van Amsterdam.

[3] Lemke, R. (2021). *Experimental investigations on the syntax and usage of fragments*. Language Science Press.

[4] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

A large-scale investigation of pronoun interpretation biases in LLMs

Valerio Pepe[1], Joshua K. Hartshorne[2]
[1] Harvard University; [2] Massachusetts General Hospital, Institute of Health Professions
jhartshorne@mghihp.edu

Humans have strong intuitions about pronoun interpretation even in the absence of much information [1–4]:

1. Alfred frightened Bart, because he… [he = Alfred]
2. Alfred liked Bart, because he… [he = Bart]
3. Because Alfred frightened Bart, he… [he = Bart]
4. Because Alfred liked Bart, he… [he = Alfred]

Are these pronoun biases a phenomenon in their own right (e.g., perhaps they exist to speed real-time comprehension), or do they fall out of general pronoun processing mechanisms? A partial answer could come from Large Language Models (LLMs), which perform quite well on difficult pronoun interpretation tasks [5, 6] but are generally not trained for pronoun biases.

A number of studies have investigated pronoun biases in LLMs, with variable results [7–11]. Some focused on the 305 verbs reported in [12], though others tested only a few dozen verbs. Because there is a long history in pronoun bias research of reporting results that did not generalize to additional stimuli [13], we return to this question with a much larger data set. We also test newer LLMs; most published studies use GPT-2 [which performs poorly on pronouns; 14] or similar.

We created a large dataset by combining two public datasets of human judgments [2, 13], which in total contain human pronoun bias results for 1,484 **implicit causality** sentences [similar to (1-2)] involving 1,212 distinct verbs, and 501 **implicit consequentiality** sentences [similar to (3-4)] involving 501 distinct verbs. We determined the noise ceiling (the best correlation between model and human data we could achieve, given statistical noise in the human data) by comparing two synthetic human datasets obtained by resampling the data with replacement. The noise ceiling was $r = .91$ for implicit causality and $r = .66$ for implicit consequentiality (far fewer participants had provided judgments for the implicit consequentiality dataset).

We tested two transformer-based LLMs. **T5 11B** requires fine-tuning to the task [for details, see 15], whereas **OLMo-2-1124-13B-Instruct** can do zero-shot, in-context learning. Both perform quite well on difficult pronoun interpretation tasks, though T5's accuracy is unsurprisingly higher [6, 16].

Results are shown in Fig. 1. Both models did reasonably well on implicit causality (T5: $r = .770, p < .001$; OLMo: $r = .773, p < .001$) if well below the noise ceiling ($.91$). Interestingly, both models (esp. T5) were more certain than humans (Fig. 1, top), though this might reflect inference parameters. T5 performed at the noise ceiling for implicit consequentiality ($r = .68, p < .001$), but OLMo failed entirely ($r = .02, p = .71$). Follow-up analyses showed that OLMo was not merely failing to notice the differences between the two sentence types.

These findings show that very good performance on difficult pronoun interpretation tasks does not necessarily result in human-like pronoun biases. Of course, the two may be linked in humans; it may just be that LLMs process language differently. Alternatively, it may indicate that pronoun biases are an over-and-above adaptation utilized to speed sentence comprehension in humans. Either way, these findings suggest caution in using (current) LLMs as replacements for human subjects.

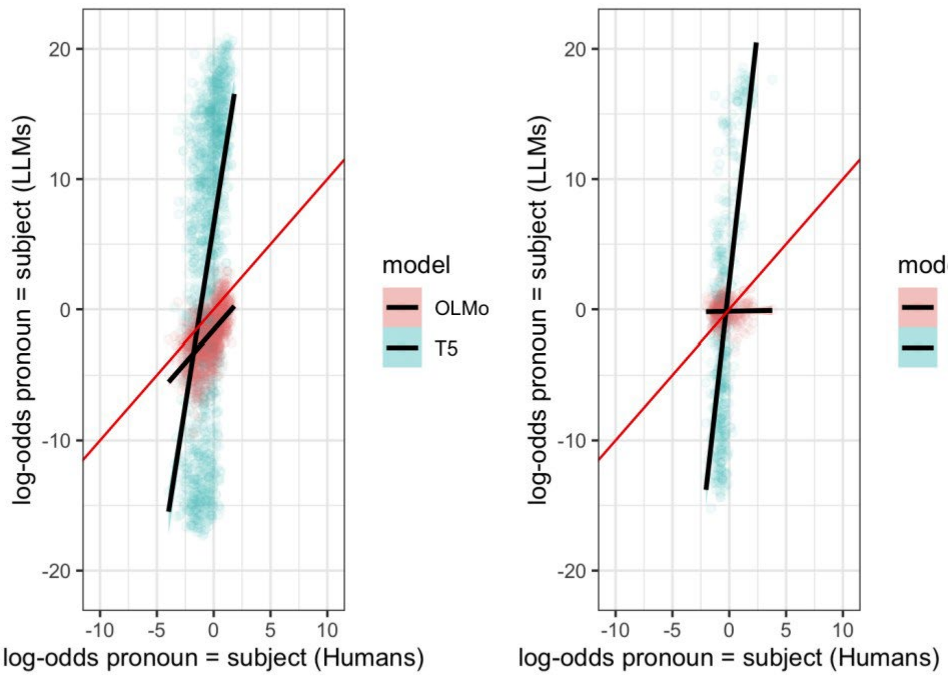


Figure 1: Log-odds of choosing subject for humans (x-axis) and models (y-axis), with implicit causality on the top and implicit consequentiality on the bottom. Positive numbers indicate a subject bias; negative numbers, and object bias. Each stimulus is plotted individually. Separate linear fits are shown for each model, with the 95% confidence interval indicated if not visible. Red line depicts what would be a perfect match.

[1] Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic inquiry, 5*(3), 459–464.

[2] Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, cognition and neuroscience, 30*(6), 716–734.

[3] Stewart, A., Pickering, M., & Sanford, A. (1998). Implicit consequentiality. *Proceedings of The 20th Annual Conference of the Cognitive Science Society.*

[4] Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes, 15*(3), 225–262.

[5] Kocijan, V., Davis, E., Lukasiewicz, T., Marcus, G., & Morgenstern, L. (2023). The defeat of the winograd schema challenge. *Artificial Intelligence, 325,* 103971.

[6] Allen Institute for Artificial Intelligence. (2025). Winogrande leaderboard - public submissions [Accessed: 2025-04-30].

[7] Kankowski, F., Solstad, T., Zarriess, S., & Bott, O. (2025). Implicit causality-biases in humans and llms as a tool for benchmarking llm discourse capabilities. *arXiv preprint arXiv:2501.12980.*

[8] Lam, S.-Y., Zeng, Q., Zhang, K., You, C., & Voigt, R. (2023). Large language models are partially primed in pronoun interpretation. *Findings of the Association for Computational Linguistics: ACL 2023*, 9493–9506.

[9] Kementchedjhieva, Y., Anderson, M., & Søgaard, A. (2021). John praised mary because he? implicit causality bias and its interaction with explicit cues in lms. *arXiv preprint arXiv:2106.01060.*

[10] Upadhye, S., Bergen, L., & Kehler, A. (2020). Predicting reference: What do language models learn about discourse models? *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 977–982.

[11] Davis, F., & van Schijndel, M. (2020). Discourse structure interacts with reference but not syntax in neural language models. *arXiv preprint arXiv:2010.04887.*

[12] Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in english: A corpus of 300 verbs. *Behavior Research Methods, 43*, 124–135.

[13] Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes, 28*(10), 1474–1508.

[14] Sakaguchi, K., Bras, R. L., Bhagavatula, C., & Choi, Y. (2021). Winogrande: An adversarial winograd schema challenge at scale. *Communications of the ACM, 64*(9), 99–106.

[15] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research, 21*(140), 1–67.

[16] Groeneveld, D., Beltagy, I., Walsh, P., Bhagia, A., Kinney, R., Tafjord, O., Jha, A. H., Ivison, H., Magnusson, I., Wang, Y., et al. (2024). Olmo: Accelerating the science of language models. *arXiv preprint arXiv:2402.00838.*

Listeners *without* the pin-pen merger find 'pin' and 'pen' ambiguous: Evidence for a parallel activation account of dialect processing

Ebony D. Pearson, Van Rynald T. Liceralde, Duane G. Watson
Vanderbilt University
ebony.d.pearson@vanderbilt.edu | van.liceralde@vanderbilt.edu | duane.g.watson@vanderbilt.edu

This project compares two accounts of how listeners in multidialectal environments process words that are potentially ambiguous across dialects: (1) perceptual adaptation and (2) parallel activation. The ***perceptual adaptation*** account stipulates that listeners adjust category boundaries based on their experience, whereas the ***parallel activation*** account stipulates that listeners co-activate multiple phonological representations from various dialects during word recognition, much like bilinguals do. These accounts offer contrasting predictions in the processing of phonological mergers, such as the pin-pen merger in which the vowels /ɪ/ ("pin") and /ɛ/ ("pen") are pronounced identically before nasal consonants. This merger is common in Southern U.S. English and some varieties of African American English, while other dialects of English maintain the distinction. Both accounts predict that listeners who have the merger (“merged” listeners) would hear words like “pin” and “pen” as ambiguous and difficult to process in isolation, but the accounts diverge for listeners who don’t have the merger (“non-merged” listeners). The adaptation account predicts that nonmerged listeners would hear “pin” and “pen” as distinct and easy to process in isolation because of their experience maintaining the distinction. In contrast, the parallel activation account predicts that nonmerged listeners would encounter some difficulty because they still co-activate phonological representations consistent with the merger, where these words are ambiguous. We tested these predictions using a novel visual search task.

In the visual search task, participants (n = 98; Merged = 28, Nonmerged = 70) heard a target word spoken by a non-merged speaker and selected one of four images. In control trials, the four images depicted phonologically distinct words. In critical trials, the four images consisted of (1) the target, (2) a same-category cohort (e.g., *pen* as a cohort for *pin*), (3) a competitor item with the distinct vowel /æ/ (e.g., *pan*), and (4) an unrelated distractor. Critical trials either had minimal pairs (e.g., pin-pen) or nonminimal pairs (e.g., dinner vs. denim) as target and cohort (**Table 1**). There were 45 total trials, 30 control trials and 15 critical trials (7 minimal, 8 nonminimal).

Mixed-effects modeling revealed that average accuracy was lowest for minimal-pair trials, followed by nonminimal-pair trials, and highest for control trials across merged and non-merged listeners (all $p < .01$), where merged listeners found minimal-pair trials significantly harder than non-merged listeners ($\beta = 1.25$, $z = 4.11$, $p < .001$) (**Figure 1**). Within merged listeners, accuracy was lowest for minimal trials, moderate for nonminimal, and highest for control trials, with all contrasts significant (all $p < .01$). A similar pattern was found in reaction times, where (log) RTs were significantly slower for both minimal- ($\beta = 0.106$, $t = 3.48$, $p = .001$) and nonminimal-pair ($\beta = 0.125$, $t = 4.32$, $p < .001$) trials compared to control trials across participants (**Figure 2**).

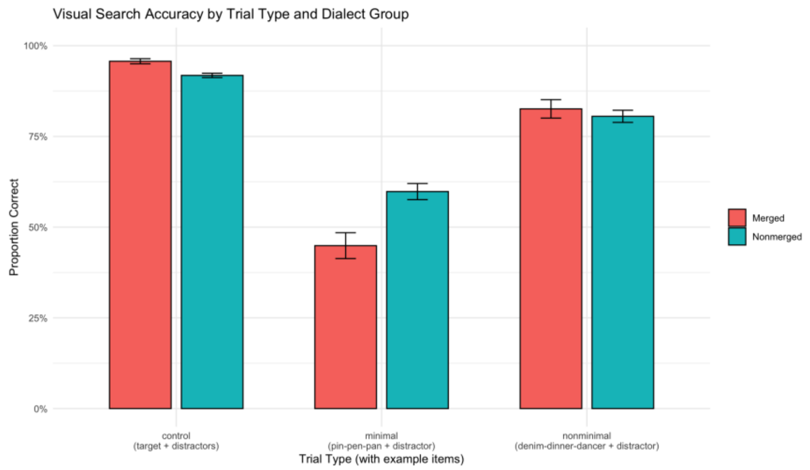
Together, these results support the idea that listeners, particularly those from non-merged dialect backgrounds, activate multiple phonological alternatives even when the speaker clearly distinguishes them. This supports a bilingual-like parallel activation account of dialect processing over a perceptual adaptation account, which predicted that non-merged listeners would find critical trials as easy as control trials.

Table 1.

Trial Type	Target	Cohort	/æ/ Competitor	Distractor
Minimal Pair	PIN	PEN	PAN	GUITAR
Nonminimal Pair	DENIM	DINNER	DANCER	BIKE
***Control**	BUCKET	SUIT	MUSTACHE	KIWI

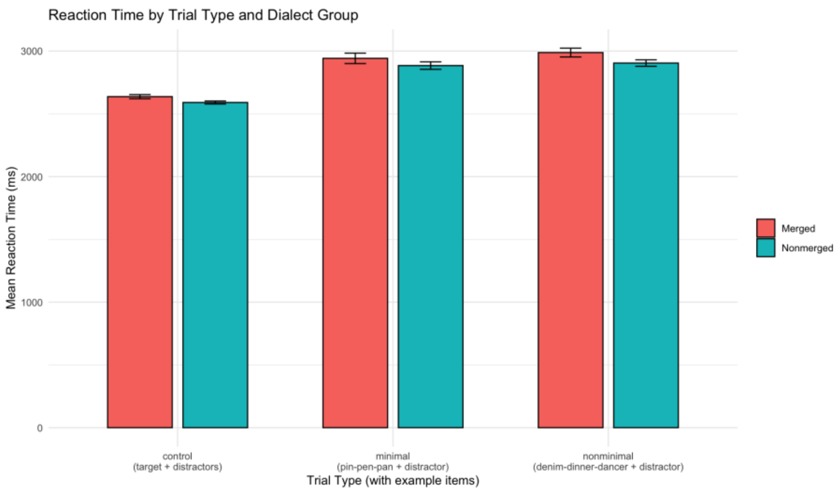
Example trial items for each condition in the visual search task. Each trial includes a target image, a same-category cohort (potential phonological competitor within the merger), an /æ/ vowel competitor (e.g., PAN), and an unrelated distractor. Control trials include only unrelated distractors.

Figure 1. Visual Search Task Accuracy by Target Type and Dialect Group



Proportion of correct responses in the visual search task by trial type (control, minimal, nonminimal) and dialect group (Merged vs. Nonmerged). Nonmerged listeners were significantly more accurate than merged listeners on minimal trials (p < .001). Within merged listeners, accuracy was significantly lower on minimal than nonminimal and control trials (p < .01), suggesting greater lexical competition during phonological ambiguity.

Figure 2. Visual Search Task Reaction Time by Target Type and Dialect Group



Mean reaction times in milliseconds by target type and dialect group. Both minimal and nonminimal trials elicited significantly longer reaction times than control trials (p < .001), indicating increased processing cost. No significant RT difference was found between minimal and nonminimal trials (p = .996), and no interaction with dialect group was observed.

The Role of Emotional Valence of Head-NP in the Korean Relative Clause Attachment

Sumin Jo, Yunju Nam, Jaewon Choi
Hanyang University
lovesuminn@hanyang.ac.kr

Background: This study investigates how Korean native speakers resolve relative clause (RC) attachment ambiguities and examines the role of the emotional valence of head nouns in this process during both real-time and offline comprehension. Previous research has shown that RC attachment preferences in ambiguous noun phrases (e.g., NP1 of NP2) vary across languages and depend on task types. In head-final languages such as Korean, offline tasks typically favor high attachment (HA), while online tasks show greater flexibility or a low attachment (LA) tendency [1–3]. Additionally, recent findings in Spanish suggest that emotional-semantic properties of nouns can influence attachment decisions early during parsing [4]. Building on this background, we used a self-paced reading (SPR) task and an offline judgment task to (i) reassess the RC attachment strategy in Korean and (ii) investigate whether the emotional valence of NPs modulates attachment preference during real-time processing.

Material and Procedures: Two experiments were conducted with 21 native Korean speakers. Sentences followed the structure: $\text{Subject}_{\text{-main}}$ + [$\text{Subject}_{\text{-rel}}$ + $\text{Verb}_{\text{-rel}}$] + $\text{NP1}_{\text{-possessive marker}}$ + $\text{NP2}_{\text{-accusative marker}}$ + $\text{Verb}_{\text{-main}}$ (see Table 1).

Experiment 1 (SPR + Attachment judgment task): Participants read 60 ambiguous (Neutral-Neutral) or disambiguated sentences, semantically biased toward either HA or LA. Reading times were measured at NP1 and NP2, analyzed using linear mixed-effects regression models (LMERs). We predicted slower reading times under LA conditions if HA was preferred.

Experiment 2 (SPR + Attachment judgment task): Participants judged attachment preferences for 100 sentences systematically varying in emotional valence (positive vs. negative) and emotional word position (NP1 vs. NP2). Emotional valence was based on words selected from the ANEW and validated in a prior SAM (Self-Assessment Manikin) task [5]. Generalized linear mixed-effects models (GLMMs) analyzed the data, comparing to the neutral-neutral baseline.

Results: Experiment 1. Contrary to predictions, reading times increased significantly at NP2 in HA conditions ($\beta = 33.02$ ms, $t(1158) = 3.74$, $p < .001$), suggesting that initial parsing favors LA. HA attachment appears to arise through later reanalysis when initial integration with NP2 fails, but the comprehension task instead showed a 53% preference for NP2 (HA), although the difference was not statistically significant ($z = -1.60$, $p = .109$).

Experiment 2: Emotional valence effects were robust: negative valence at NP2 significantly increased HA choices compared to neutral conditions ($p < .001$). Furthermore, SPR reading times at NP2 were significantly longer in negative emotional NP2 conditions (N-Neg), suggesting that emotional negativity prompts re-integration and promotes HA.

Conclusion: The findings suggest that in Korean RC processing, initial parsing favors LA, but HA preferences emerge through later semantic integration, not during the earliest stages of parsing Emotional valence—especially negativity—plays a critical role in biasing attachment decisions, indicating that parsing is sensitive to emotional-semantic information from an early stage. These results highlight that syntactic processing interacts dynamically with semantic-affective factors. Future research employing eye-tracking or ERP methods could further disentangle early and late processing stages beyond what SPR can reveal.

References

[1] Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30*(1), 73–105.

[2] Kamide, Y., & Mitchell, D. C. (1997). Relative clause attachment: Nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research, 26*(2), 247–254.

[3] Lee, D., & Kweon, S. O. (2004). A sentence processing study of relative clauses in Korean with two attachment sites. *Studies in Linguistics*, *20*(1), 123–145.

[4] Fraga, I., Piñeiro, A., Acuña-Fariña, J. C., Redondo, J., & García-Orza, J. (2012). Emotional nouns affect attachment decisions in sentence completion tasks. *The Quarterly Journal of Experimental Psychology, 65*(9), 1740–1759.

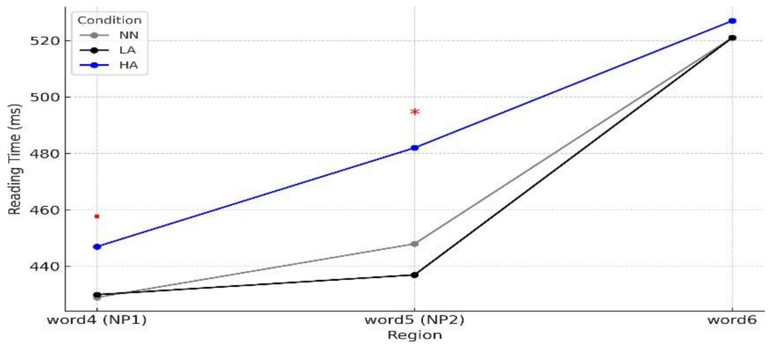
[5] Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Technical Report No. C-1). The Center for Research in Psychophysiology, University of Florida.

Table 1. Examples of Stimuli Sentences Across Conditions (Experiment 1 & 2)

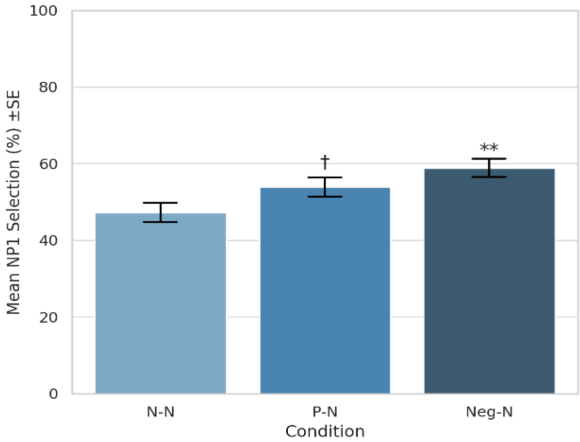
	word1	word2	word3	word4(NP1)	word5(NP2)	word6
Ambiguity RC (NN)	나는 I-top.	주희가 Juhee-nom.	설명한 explain-rel.	기계의 machine-poss.	부품을 part-acc.	찾았다. find-past-decl.
	I found the part of the machine that Juhee explained.					
NP1 Attachment (LA)	나는 I-top.	윤지가 Yoonji-nom.	개업한 open-rel.	식당의 restaurant-poss.	직원을 employee-acc.	칭찬했다. praise-past-decl.
	I praised the employee of the restaurant that Yoonji opened.					
NP2 Attachment (HA)	나는 I-top.	지윤이가 Jiyeon-nom	주문한 order-rel	요리사의 chef-poss.	음식을 food-acc.	음미했다. restore-past-decl.
	I savored the food of the chef that Jiyeon ordered.					

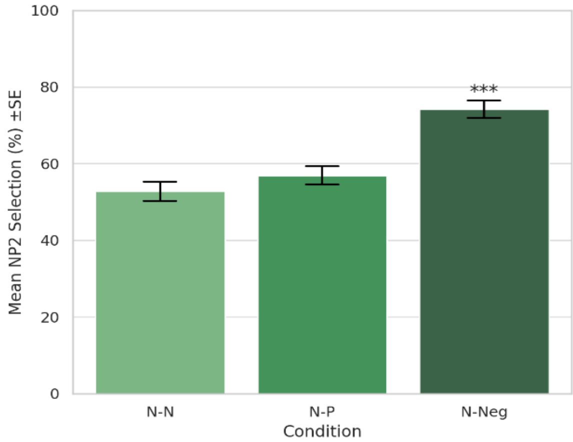
Mean reading times for Experiment 1 across regions word4 (NP1), word5 (NP2), and word6. A significant increase was observed at NP2 in the HA condition ($p < .001$).

Condition	word4	word5	word6
NN	429	448	521
LA	430	437	521
HA	447.	482*	527



Mean NP1 and NP2 Selection Rates by Condition with Standard Errors Bars.





Revisiting Uniform Information Density and *that*-reduction in English Complement Clauses

Hailin Hao, Elsi Kaiser
University of Southern California
hailinha@usc.edu

Language production is highly flexible, and speakers often have multiple options to express the same meaning. For instance, in English complement clauses (CCs), the complementizer *that* is optional (**Ex. 1**). According to the Uniform Information Density (UID) hypothesis, speakers exploit this variability to maintain a steady rate of information transmission. Supporting this, [1] found that *that* is more likely to be omitted in English CCs when the **information density** of the CC, -logP(CC|context) is low. However, [1] quantified P(CC|context) using matrix verbs' subcategorization probabilities estimated from corpus frequencies. This static measure does not fully capture dynamic predictive processes and may conflate predictability with verb-specific preferences in *that*-mentioning. Our study improves prior work by (i) using machine learning methods and contextual word embeddings [2] to more robustly estimate CC structural predictability, and by (ii) analyzing a modern large-scale conversational corpus, CANDOR [3] (8 million words with 1,456 unique speakers). After parsing the transcripts using spaCy [4], we extracted 51,276 CCs with 50 unique matrix verbs (post-cleaning). This broader lexical and speaker coverage enables us to test the generalizability of UID effects in syntactic reduction.

CC Structural Predictability We trained neural networks to predict P(CC | context) using all 236,504 occurrences of the 50 matrix verbs in the CANDOR corpus. Seven relevant linguistic features were identified, and feature selection was conducted using incremental methods and lasso regression [5]. We also trained a model based on *contextual word embeddings* of the matrix verbs—numeric representations of word meaning informed by its surrouding context (in this case, the pre-CC context)—extracted from GPT-2 [2] and reduced to 50 dimensions via PCA [6]. Results (**Table 1**) show that no linguistic feature beyond verb subcategorization probability improved prediction performance in terms of F1 and log loss. Notably, the embedding-based model outperformed all others. Based on this, we proceeded to test how well information density derived from **(i)** verb subcategorization probabilities (as in [1]) and **(ii)** from contextual embeddings predicts *that*-mentioning.

Modeling *that*-reduction: We fitted generalized mixed-effects models predicting the presence of *that* (Eq. 1), using CC information density (verb-based; embedding-based) as predictors, and control variables informed by availability-based [6] and grammaticalization accounts [7], syntactic priming [8], disfluencies, and a speaker random intercept. **Results**: In both models, higher CC information density increased *that*-mentioning (p<.001, Fig.1-2). In addition, the model with verb-based information density yielded better AIC/BIC (**Table 2**). However, as noted above, verb-based measures may encode lexical variation. Therefore, we fitted two new models with a **verb random intercept**. Strikingly, now effects of verb-based information density become non-significant, but effects of embedding-based information density remain significant. Model comparisons (**Table 2**) show that patterns of *that*-mentioning are best captured by the model with embedding-based information density and verb random intercepts. This suggests (i) *that*-mentioning is guided by the identity of matrix verbs and (ii) embedding-based information density better predicts *that*-mentioning compared to verb-based ones.

Discussion: As predicted by UID, higher CC information density leads to that being mentioned more. We also show that verbs' subcategorization probabilities (e.g., [1]) encode substantial verb-specific variation. Crucially, more holistic measures of information density from word embeddings predict *that*-mentioning beyond lexical biases. This work also highlights the value of machine learning and NLP for psycholinguistics.

Ex. 1 *I understand (that) you've got all kinds of online events.* (from CANDOR corpus)

Eq. 1 that_presence ~ information_density + verb_freq + CC_nsubj_freq + coref + CC_nsubj_form + CC_word_count + CC_nsubj_deps_before + distance_type + verb_position + priming + filler + repetition + verb_nsubj_form + double_that + (1|speaker)

Table 1: Summary of model performance in predicting CCs. Models were optimized for log loss (lower is better), with higher F1 indicating better performance. AIC and BIC comparisons (not shown here) are also consistent with log loss and F1 score.

Feature set	Log Loss	F1 score
Baseline Model	0.643	0
Subcategorization Probability	0.49	0.66
Subcategorization Probability, Verb Frequency	0.49	0.66
Subcategorization Probability, Factivity	0.49	0.66
Subcategorization Probability, Tense	0.49	0.65
Subcategorization Probability, Position	0.49	0.66
Subcategorization Probability, Subject Form	0.49	0.65
Subcategorization Probability, Subject Frequency	0.48	0.65
Lasso-selected features: Subcategorization Probability, Subject Form, Factivity, and Verb Frequency	0.48	0.65
GPT-2 Contextual Embeddings	**0.44***	**0.69***

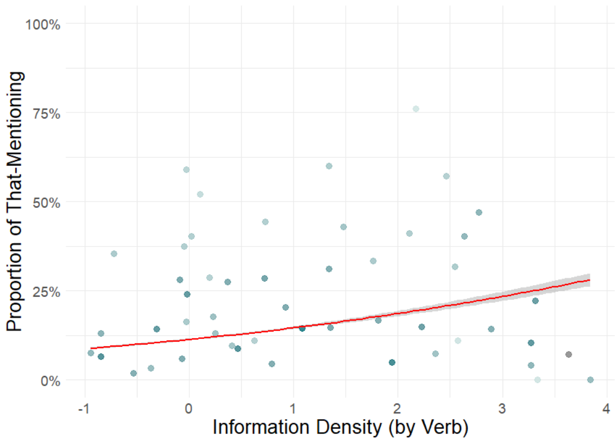


Fig. 1. Effects of **verb-based** information density (dots represent each verb, with color darker shades representing higher frequency)

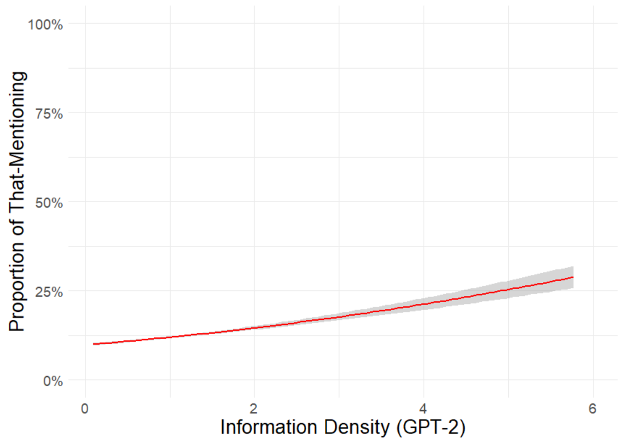


Fig. 2. Effects of **embedding-based** information density

Table 2: Model AIC and BIC (lower the better)

Models	AIC	BIC
Verb-based information density, without verb random intercept	33344	33521
Embedding-based information density, without verb random intercept	33736	33912
Verb-based information density, with verb random intercept	32302	32488
Embedding-based information density, with verb random intercept	**32277***	**32462***

References: [1] Jaeger (2010) Cogn. Psychol. [2] Radford et al. (2019) OpenAI. [3] Reece et al. (2023) Sci. Adv. [4] Honnibal et al. (2020). [5] Tibshirani (1999) J. R. Stat. [6] Ferreira & Dell (2000) Cogn. Psychol. [7] Roland, Elman, & Ferreira (2005) Cognition. [8] Bock (1986) Cogn. Psychol.

Effects of written but unpronounced morphemes on auditory word recognition

Hannah Krueger, Samantha Wray
Dartmouth College
hannah.e.krueger.26@dartmouth.edu, samantha.c.wray@dartmouth.edu

Background: Auditorily-presented morphologically related words prime one another as morphologically complex words are decomposed during lexical access [1, among many others]. Priming has been widely attested for word stems, but less so for affixes [2, 3]. Furthermore, in French, certain inflectional suffixes, specifically 3rd person plural -ent, and feminine noun/adjective suffix -e are partially or sometimes entirely unpronounced. Therefore, there is a potential that these are perceived by literate French speakers as 'silent morphemes' having the same function as pronounced inflectional morphemes. Previous evidence in French shows facilitated response to targets that have these affixes, but the prime-target pairs share stems [4]. Assuming a model where inflectional morphemes prime, we predict then that pronounced inflectional morphemes in French may also prime their unpronounced allomorphs, supporting a morphological decomposition in the auditory domain which is influenced by orthography.

Method: N=50 (age M=34.8 SD=10) L1 French speakers performed a primed auditory lexical decision task via Gorilla [5] . Thirty-six prime-target pairs (in addition to distractors and nonwords) made up two main conditions across three word classes (verbs, nouns, and adjectives) : (1) related prime-target pairs for which the plural *-ent* or feminine *-e* suffixes were indicated in the pronunciation of the prime but not the target, despite being written in both (e.g. charpentièr-e/fiancé-e) and (2) unrelated prime-target pairs had the suffix in the target only, lacking a morphological relationship (e.g. puissance/logéant-e).

Results: Reaction times to individual prime-target pairs were analyzed using t-tests for each word class. Reaction time was chosen based on the wealth of research which shows that affected reaction times to a target based on its prime is evidence for a relationship between the prime and target. Results showed that there was a significant difference between related and unrelated conditions, but only for two out of three word classes: verbs with the feminine third plural suffix *-ent* ($t(724) = -2.11$, $p = 0.03$), and nouns with the feminine singular suffix *-e* ($t(731) = 2.39$, $p = .02$). Prime-target pairs which were adjectives sharing the feminine suffix *-e* did not show significant differences compared to their unrelated counterparts ($t(704) = 1.01$, $p = 0.3$). However, the differences diverged in direction: verbs with shared affixes were facilitated, and nouns with shared affixes were inhibited, indicating potential phonological interference for the latter [6].

Discussion: From this work, we learn about the interaction of literacy and language by investigating if the spoken forms of words activate the written forms even when those forms may deviate. However, the fragility of inflectional affix priming is also confirmed, as only one word class exhibited priming for pairs sharing affixes and another exhibited inhibition.

References

[1] Emmorey, K. D. (1989). Auditory morphological priming in the lexicon. *Language and cognitive processes*, *4*(2), 73-92.

[2] Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. The Quarterly Journal of Experimental Psychology Section A, 38(3), 351-365.

[3] Goodwin Davies, A., & Embick, D. (2020). The representation of plural inflectional affixes in English: Evidence from priming in an auditory lexical decision task. *Language, Cognition and Neuroscience*, *35*(3), 393-401.

[4] Meunier, F., Seigneuric, A., & Spinelli, E. (2005). The Morpheme Gender Effect: An Evidence for Decomposition. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 27, No. 27).

[5] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*, 388-407.

[6] Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of experimental psychology: Learning, memory, and cognition*, *18*(6), 1239.

The Influence of Conceptual and Syntactic Interaction on Syntactic Structure Selection in Chinese Speakers' Language Production

Xiao Ke 1, Silke Brandt 2, Katherine Messenger 1
1 Department of Psychology, Lancaster University, Lancaster, UK
2 Department of Linguistics and English Language, Lancaster University, Lancaster, UK
x.ke1@lancaster.ac.uk

Abstract: This study tested whether the conceptual structure of causal and transfer events influences syntactic choice in Mandarin sentence production, contrasting the autonomous (e.g., Bock & Loebell, 1990) and functionalist (e.g., Pickering & Branigan, 1998) accounts.

Two primed recall experiments were conducted with native Chinese speakers (N = 60 per experiment). In each trial, participants read a target sentence, then a prime, repeated the prime, and recalled the target.

Experiment 1 examined whether "bǎ (把)" constructions (*S+ba+O+V*) and resultatives in the primes influenced the recall of resultative-free simple targets (*S+V+O*). Results showed that "bǎ" primes significantly increased "bǎ" recall (p = .001), while the presence of resultatives in the primes had no effect on "bǎ" recall (p = .322). In contrast, both syntactic structure and resultatives significantly affected resultative recall (p = .010, p = .008), with a significant interaction between them (p = .037). The conceptual effect of resultatives was stronger in the SVO condition than in the "bǎ" condition, indicating both syntactic and conceptual priming.

Experiment 2 tested the recall of double object (DO) dative targets ($S+V+g\check{e}i+O_1+O_2$) after six types of primes: baseline SVO (*S+V+O*), DO datives, prepositional object (PO) datives ($S+V+O_2+g\check{e}i+O_1$), PO locatives ($S+V+O_2+d\grave{a}o+O_1$), "bǎ" datives ($S+ba+O_2+V+g\check{e}i+O_1$), and "bǎ" locatives ($S+ba+O_2+V+d\grave{a}o+O_1$), with 36 primes per condition. Results showed that syntactic structure significantly influenced PO recall (p < .001), while conceptual structure did not (p = .247). For locative recall, conceptual structure significantly influenced recall (p = .002), but syntactic structure did not (p = .821). No interaction was observed (p = .193). For "bǎ" recall, syntactic priming was again significant (p < .001); conceptual structure showed a marginal effect (p = .052), and a significant syntactic × conceptual interaction emerged (p = .014).

Together, these findings support a functionalist account: syntactic and conceptual information both shape sentence production in Mandarin Chinese.

References

[1] Bock, J. K., & Loebell, H. (1990). Framing sentences. Cognition, 35, 1–39.
[2] Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntacticpriming in language production . Journal of Memory and Language, 39(4), 633-651.

Processing (dis)continuous explicit and implicit discourse relations in European Portuguese

Paula Luegi[1,3], Isabel Falé[2,3], Jéssica Gomes[1,3], Amália Mendes[1,3]
[1]University of Lisbon, FLUL; [2]Universidade Aberta; [3]CLUL
paularibeiro@edu.ulisboa.pt

Background: Discourse relations (DRs), which can convey different meanings (e.g., Cause or Contrastive DRs), might be expressed explicitly, through the presence of a discourse marker, or implicitly, where such a marker is absent. However, the naturalness of a connective's implicitness as well as its processing costs might depend on various factors. For instance, the 'continuity' hypothesis [1] proposes that comprehenders expect events in discourse to progress in a linear temporal order, with continuity characterized across multiple dimensions (time, space, action progression, and perspective) [2, 3]. In fact, several studies have shown that, overall, comprehenders benefit from explicit connective information during sentence comprehension [e.g., 4], particularly when the connective conveys discontinuous relations. Corpus-based studies have shown that discontinuous relations, such as Contrastive DRs, are explicitly expressed more frequently than continuous relations, such as Cause [5, 6], and experimental studies suggest that the processing of implicit-discontinuous DRs is harder than implicit-continuous DRs [7]. In the present study, we investigated the effects of (dis)continuity of explicit and implicit DRs of three types in European Portuguese: (i) a causative reading of *de facto* 'indeed', where the second argument provides a subjective cause for the first argument (continuous); (ii) a confirmation reading of *de facto* 'indeed', with the first argument conveying the perspective of an external source and the second the speaker's perspective (discontinuous), and (iii) a contrastive reading of *na verdade* 'in fact' which combines both perspective shift and negative polarity (discontinuous). Hypotheses: H1 – implicit DRs are harder to process than explicit DRs; H2 – discontinuous DRs are harder to process than continuous DRs; H3 – implicit-discontinuous DRs are harder to process than implicit-continuous DRs; H4 – Contrastive DRs, due to their negative polarity, are harder to process than Cause and Confirmation DRs.
Method: two self-paced reading experiments: **Experiment 1** (N=100), testing implicit and explicit Cause and Confirmation DRs (32 items/64 fillers); **Experiment 2** (N=120), testing implicit and explicit Cause, Confirmation, and Contrastive DRs (36 items/72 fillers). The participants read segmented sentences, such as (1a), (2a), or (3a), either with the connective (Explicit) or without it (Implicit), followed by sentence (b), and then answered a 'yes/no' comprehension questions, such as (4). [Analyses: LMMs with maximal structure (simplified when needed). Implicitness of the connective: coded as Explicit (-0.5) and Implicit (+0.5). DR type: coded as Cause (-0.5) and Confirmation (+0.5), in Exp1; coded with treatment contrast (Cause as the baseline), in Exp2.]
Results and discussion: The results (Figure 1 and Table 1) of both experiments show that: (i) Explicit DRs are consistently easier to process, leading to lower RTs, confirming H1; (ii) Cause DRs (continuous) are easier to process, yielding lower RTs and confirming H2; (iii) implicit-continuous DRs (Cause) are easier to process, showing lower RTs, than implicit-discontinuous DRs (Confirmation and Contrastive), supporting H3; (iv) however, our results do not confirm H4, as Contrastive DRs show RTs comparable to Confirmation DRs, differing only from Cause DRs. This lack of effect when two concurrent discontinuity factors (perspective shift and negative polarity) were present seems to suggest that the presence of a single discontinuity factor is sufficient to influence DRs' processing, indicating that discontinuity may not be a gradient.

Stimuli		
Cause (Sentence 1)	(1a)	*O Pedro / deve estar / a desenvolver-se / de forma saudável. / (De facto,)…/*
		'Pedro must be developing healthily. (CONNECTIVE)…'
Confirmation (Sentence 1)	(2a)	*A Sofia / esperava / que o Pedro / estivesse mais alto. / (E de facto,)…/*
		'Sofia expected Pedro to be taller. (CONNECTIVE)…'
Contrastive (Sentence 1)	(3a)	*A Sofia / pensava / que o Pedro / estava da mesma altura. / (Na verdade,)…/*
		'Sofia thought Pedro was the same height. (CONNECTIVE)…'
(Sentence 2)	(b)	*…o rapaz cresceu* [Critical] */ muito nos últimos meses* [Post-Critical] */ deste verão.* [Wrap-up]
		'the boy has grown a lot in the last months of this summer.'
Question	(4)	*O Pedro engordou no final deste verão?*
		'Did Pedro gain weight at the end of this summer?'

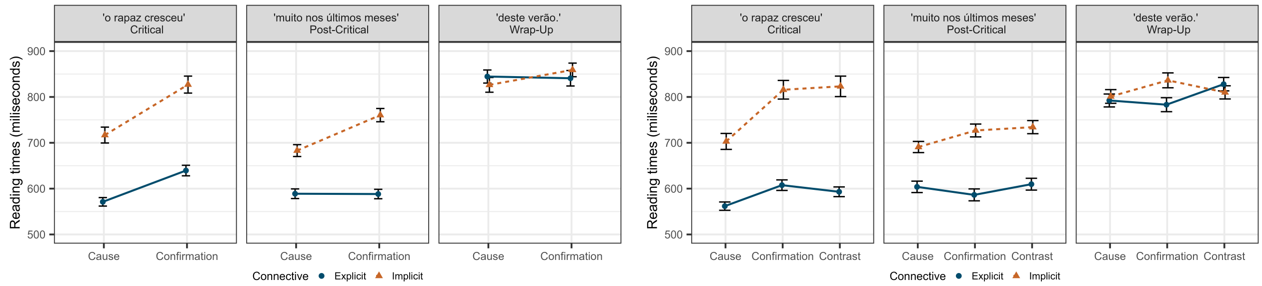


Figure 1: Reading times in Experiment 1 (left panel) and Experiment 2 (right panel). [Filled line: Explicit connective; Dashed line: Implicit connective]

		Critical		Post-Critical		Wrap-Up	
		Est (SE)	t-value	Est (SE)	t-value	Est (SE)	t-value
Experiment 1	(Intercept)	6.419 (0.025)	260.777	6.386 (0.025)	253.469	6.632 (0.032)	207.079
	DR	0.116 (0.022)	**5.204**	0.051 (0.015)	**3.478**	0.017 (0.018)	0.952
	Connect	0.174 (0.027)	**6.410**	0.204 (0.017)	**11.869**	0.001 (0.014)	0.071
	DR:Connect	0.028 (0.035)	0.803	0.087 (0.025)	**3.456**	0.054 (0.030)	1.827
Experiment 2	(Intercept)	6.270 (0.022)	289.000	6.300 (0.029)	215.000	6.580 (0.033)	197.000
	DRConf	0.066 (0.021)	**3.210**	-0.022 (0.019)	-1.180	-0.020 (0.019)	-1.020
	DRContr	0.048 (0.023)	**2.090**	0.008 (0.019)	0.394	0.041 (0.021)	**1.970**
	Impl	0.156 (0.029)	**5.360**	0.154 (0.021)	**7.330**	0.011 (0.019)	0.590
	DRConf:Impl	0.069 (0.028)	**2.470**	0.060 (0.025)	**2.410**	0.062 (0.025)	**2.480**
	DRContr:Impl	0.083 (0.030)	**2.760**	0.037 (0.025)	1.470	-0.028 (0.025)	-1.140

Table 1: LMM results of log(RTs) of Experiment 1 and Experiment 2 (bold represent p-values lower than 0.05).

References

[1] Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory and Cognition*, *25*, 227–236.

[2] Segal, E. M., Duchan, J. F., & Scott, P. J. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, *14*(1), 27–54.

[3] Das, D., & Egg, M. (2023). Continuity in discourse relations. *Functions of Language*, *30*(1), 41–66.

[4] Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, (29), 37–60.

[5] Asr, F. T., & Demberg, V. (2012, December). Implicitness of discourse relations. In M. Kay & C. Boitet (Eds.), *Proceedings of COLING 2012* (pp. 2669–2684).

[6] Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113–131.

[7] Zufferey, S., & Gygax, P. M. (2016). The role of perspective shifts for processing and translating discourse relations. *Discourse Processes*, *53*(7), 532–555.

Learning Articles in an Artificial Minilanguage: Error-based Learning or Propose-but-verify?

Filip Smolík[1], Maroš Filip[1]
[1] Institute of Psychology, Czech Academy of Sciences, Prague
smolik@praha.psu.cas.cz

Error-based learning has been proposed as an important mechanism in language learning [1, 2]. According to this mechanism, learning relies on predictive processing of upcoming stimuli and the prediction errors are particularly informative learning instances. The mechanism might explain important phenomena for which there were no generally accepted explanations, such as retreat from overgeneralization. However, the exact conditions under which error-based learning occurs, as well as the parameters of the learning process, need to be investigated empirically. In some contexts, incorrect hypotheses in learning, followed by feedback, resulted in worse learning than correct hypotheses [3]. We executed two experiments testing if adults show better learning after errors or not. The experiments mimicked morphosyntactic learning of the relation between a determiner and word ending.

Two experiments presented adult participants (N=120 and 80, respectively, members of the student pool) with 540 learning items consisting of three repetitions of 180 nonwords, each paired with one of two articles (ut, hep). Nonwords had 3 possible endings, -a, -o, or -u, and the words differed in the preference for the articles. The majority (90%) of -a nonwords were preceded by one article (counterbalanced across versions), the majority (90%) of -u nonwords with the opposing article, while the -o words were split half to half between articles. Each of the 3 times participants saw a stimulus nonword, they had to guess the article (Exp. 1) or the ending (Exp. 2). The whole set of nonwords was presented before any of the stimuli was presented the second or third time. Experiments were presented remotely using PsychoPy, and results were analyzed using binomial mixed models with random intercepts for subjects and items.

In both experiments, the success rate increased during the experiment, although the increases were small, and the rate was only slightly above chance level (Figure 2). For -a endings, the rate increased significantly faster than for the remaining endings. In all three nonword groups (-a, -u, -o), success rate increased with second and third presentations, revealing learning of both the predominant pattern, and the lexically idiosyncratic exceptions for the -a and -u endings. Critically, though, the success rate in the second and third round was higher for words that had correct responses (articles or endings) in the previous, first or second round (Figure 3).

Error on a word presentation did not result in increased learning, rather the opposite. The pattern observed in the experiment appears to contradict the principles of error-based learning but it is similar to the pattern observed in lexical learning studies and characterized as propose-but-verify [3]. Presumably, this mechanism results from participants forming hypotheses about a word, and only keeping track of whether this hypothesis is confirmed. While the present experiments did not involve learning of word meaning, it appears that they activated a similar mode of learning used in lexical acquisition, and the propose-but-verify principle. Lack of error-based learning effects indicates that the possibility of making an error is not sufficient to induce error-based learning.

Figure 1: trial structure in Exp. 1 (left) and 2 (right). In Exp. 2, choice is made on the third screen.

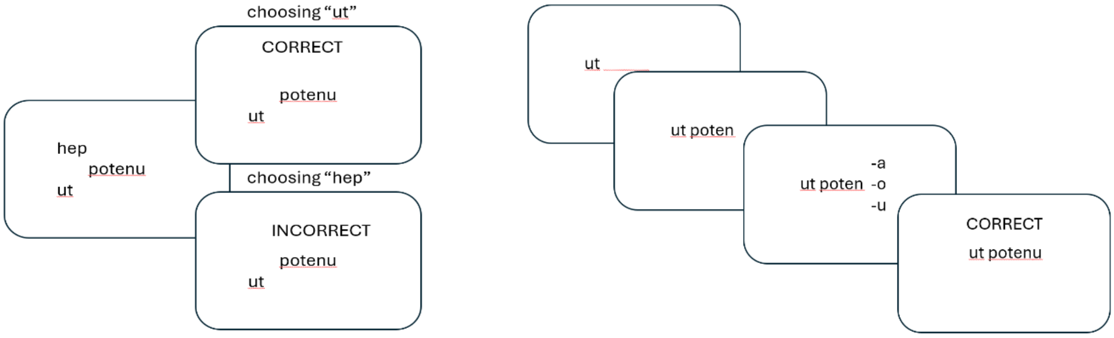


Figure 2: Experiment 1 (left) and 2 (right) accuracy for word classes

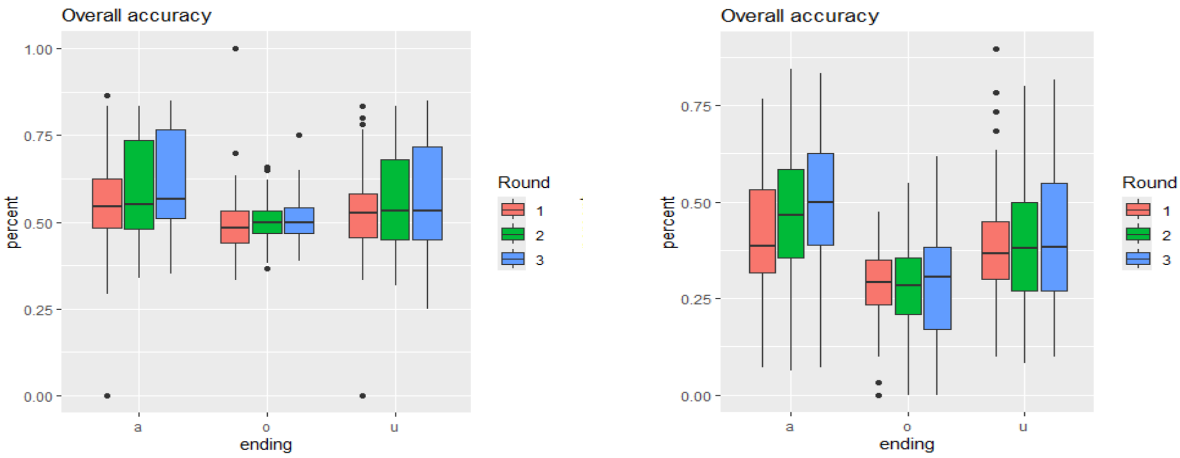
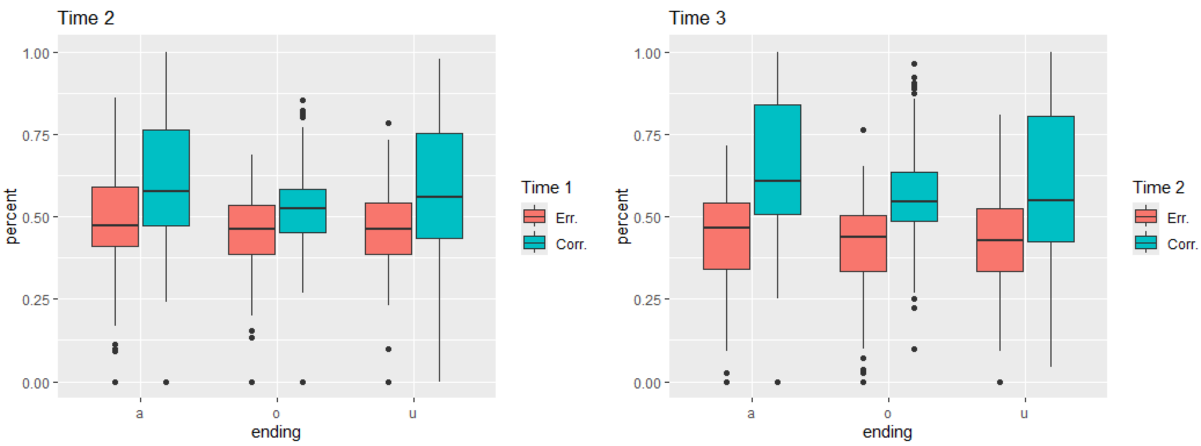


Figure 3: Experiment 1 accuracy depending on the previous round success/error for a given word



References

[1] Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, *89*(4), 760-793.
[2] Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. MIT Press.
[3] Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126-156.

Communicative Goals Influence Conceptual but Not Structural Alignment

When speaking, people often unconsciously copy each other's speech patterns [1]. For example, English speakers were more likely to say a passive sentence (e.g., *'The boy is awakened by a noisy alarm*') after hearing another passive sentence (e.g., *'The car's windshield was struck by a brick')* than an active sentence (e.g., *'A brick struck the car's windshield'*) [2]. This tendency – termed "alignment" in psycholinguistics – is thought to be influenced by communicative factors [3], such as the presence of a conversation partner. However, while some studies using transitive events (such as a boy pushing a girl) found stronger alignment when a conversation partner was present rather than absent [4], other studies using dative events (such as a ballerina giving a soldier an apple) found no such difference [5]. One explanation for this discrepancy is that transitive and dative sentences trigger different types of alignment, which respond differently to communicative factors. Specifically, while both sentence types trigger structural alignment (e.g., choice of sentence structure), transitive sentences in active-passive alternation may additionally trigger conceptual alignment (e.g., emphasising an important element). Structural and conceptual alignment often coincide and bias speakers toward the same choice (e.g., passives), hence distinguishing between them has been a challenge for the field.

Using unique structures in Mandarin (e.g., '*shi...de'* cleft focus structure, topicalisation, left-dislocation, and patient-emphasising BA-structure vs. canonical SVO structure), our study was able to disentangle the two alignment types. In three experiments, we tested 151 Mandarin speakers in semi-naturalistic conversations. Each time, a real participant and a trained "confederate" (pretending to be another participant while reading scripted lines) took turns describing pictures of transitive events and checking if their own picture matched the other's description. We manipulated communicative goals: In Experiments 1 and 2, participants and the confederate described their pictures without any prior questions from the other, whereas in Experiment 3, they asked each other a pre-scripted question about the picture before the other gave descriptions. The question-and-answer component of this manipulation was intended to make the task more communicative, addressing concerns that the null effect of communicative factors on structural alignment observed in [5] may have resulted from the mere presence of a conversation partner not being sufficiently social.

We found that structural and conceptual alignment occurred independently (both showed significant effects in Experiments 1 and 2). Importantly, only conceptual alignment was influenced by communicative goals, as shown by persistent structural alignment and diminished conceptual alignment in Experiment 3. These results suggest that conversation partners align with each other's sentences across the board, but align with each other's perspectives (e.g., emphasising a particular event role) only when they do not already have an explicit communicative intent. Our study demonstrates a novel relationship between alignment and communicative factors in speaking, helping to reconcile mixed findings from previous research and shedding new light on the mechanisms of human communication.

References:

[1] Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*, e282. doi:10.1017/S0140525X16002028

[2] Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437-458. https://doi.org/10.1016/j.cognition.2006.07.003

[3] Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(2), 212-225. doi:10.1017/S0140525X04450055

[4] Schoot, L., Hagoort, P., & Segaert, K. (2019). Stronger syntactic alignment in the presence of an interlocutor. *Frontiers in Psychology*, *10*, 685. doi:10.3389/fpsyg.2019.00685

[5] Ivanova, I., Horton, W. S., Swets, B., Kleinman, D., & Ferreira, V. S. (2020). Structural alignment in dialogue and monologue (and what attention may have to do with it). *Journal of Memory and Language*, *110*, 104052. doi:10.1016/j.jml.2019.104052

Task effects in *good-enough* parsing in Turkish: Human and LM comprehension of thematic roles

Onur Keleş and Nazik Dinçtopal Deniz
Bogazici University
onur.keles1@bogazici.edu.tr

Introduction. The *good-enough parsing* model maintains that language users occasionally rely on syntactically shallow, semantically plausible interpretations and revise them only when necessary [1, 2]. This often occurs in role-reversal contexts leading to implausible events [3] such as 'the dog was bitten by the man' misinterpreted as 'the dog bit the man.' While studies in English show that task demands can modulate if these reversals are detected online [4,5], evidence from Turkish is limited. To address this, we conducted two self-paced reading (SPR) experiments with comprehension tasks that differ in their requirements for processing depth: a Semantic Plausibility Judgment (PJ) task, with relatively less cognitive demands, and an Agent–Patient Detection (APD) task, a cognitively more challenging task due to its thematic role assignment and event structure processing requirements. We examined both human behavior and surprisal estimates from GPT-2 and LLaMA-3 models trained on Turkish [6] and asked if (i) task demands as well as word order, structure, and animacy factors affect the extent of *good-enough* interpretations, and (ii) if language models (LMs) approximate human language processing. **Methodology.** Turkish speakers (N = 52 in the PJ task and N = 72 in the APD task) participated in the experiments. The experimental sentences as in (1,2) were taken from [3] and translated into Turkish. They were manipulated for structure (active, passive) and word order (canonical, reversed). Half of the sentences had animate arguments whose thematic roles, given real-world plausibility, were reversible but *Biased* as in (1) and the other half had one animate and one inanimate argument whose thematic roles were *Irreversible* as in (2). The participants either judged sentence plausibility on a binary scale (PJ task) or identified the agent/patient in the sentence (APD task). Each participant saw a set of 21 experimental sentences in each set, counterbalanced for word order and structure, along with 21 symmetrical controls (where the reversal yielded equally plausible events). Accuracy, decision times (DTs), and word reading times (RTs) were recorded. **Results.** Accuracy and logRT/DT data were analyzed using (generalized) mixed-effects models, with fixed effects for Structure [Active, Passive], Word Order [Canonical, Reversed], and Set [Biased, Irreversible]. Surprisal estimates from Turkish-adapted LMs (GPT-2-Base, GPT-2-Large, LLaMA-3-8B; see Table 1 for details) were added to the base models to compare their behavior to human RTs. *Accuracy data* are summarized in Table 2. In the PJ task, there was an interaction between Set and Word Order (OR = 24.39, p < .001): the participants were more accurate for canonical orders (1a,b; M = 97.5) than reversed orders (1c,d; M = 78.5%) in the biased set (OR = 0.05, p < .001). In the irreversible set, they were more accurate in the reversed (2c,d) than the canonical order (2a,b; OR = 2.87, p = .004). In the APD task, the accuracy was overall higher in the irreversible set (M = 95%) than the biased set (M = 94%; OR = 1.58, p = 0.007). *Sentence-final DTs* are illustrated in Figure 1. In the PJ task, there was an interaction between Set and Word Order; β = –0.22, p < .001. The DTs were slower for reversed sentences than canonical ones in the biased set (β = 0.13, p < .004) and longer in the irreversible set; β = –0.13, p < .001. In the APD task, there was an interaction between Set, Structure, and Word Order; β = –0.12, p = .033. DTs were overall slower for reversed orders in both sets (p's < .001) but they were reduced for passive sentences in the reversed order conditions for the irreversible set; β = –0.09, p = .019. Figure 2 shows *Word RTs*. For the PJ task, RTs for the verb were slower for the reversed orders than canonical orders only in the irreversible set; β = 0.09, p = .034. They were slower in the biased set for passive sentences; β = 0.10, p = .039. For the APD task, both sets had shorter RTs for the reversed orders than canonical orders (p's < .022). The *LM analyses* showed that GPT-2-Large surprisal predicted human reading times in both tasks (PJ: β = 0.10, $p \le$ .009; APD: β = 0.02, $p \le$ .010); GPT-2-Base predicted RTs only in the PJ task (β = 0.07, p = .006), and LLaMA-3 predicted RTs only in the APD task (β = 0.01, p = .031). **Discussion.** The results show that task type influences the extent of good-enough parsing, supporting [4]. When participants were required to assign thematic roles to arguments (the APD task), they employed syntactically detailed parses (higher accuracies and lack of word order effects). When the task did not require to do so (the PJ task), they showed good-enough parsing (reduced accuracies and lack of word order effects on the verb RTs for the biased but implausible sentences). The observation of faster DTs for reversed sentences in the irreversible set compared to the biased set shows use of animacy cue may also be task-dependent and more essential for plausibility judgments. The size of the LMs appear to affect the model's similarity to the human data, with more medium-size model (GPT-2-Large) matching human behavior in both tasks. The larger model (LLaMA-3) was more sensitive to syntax and aligned with human reading only in the APD task and the smaller model (GPT-2-Base) was more heuristic and aligned with human behavior only in the PJ task.

1) Reversible but Biased Set

Active, canonical

a. Köpek-ler adam-ı **ısır-dı** sanır-ım
dog-PL man-ACC bite-PST think-1SG
'I think the dogs bit the man.'

Passive, canonical

b. Adam-lar köpek tarafından **ısır-ıl-dı** sanır-ım
man-PL dog by bite-PASS-PST think-1SG
'I think the men were bitten by the dog.'

Active, reversed

c. Adam-lar köpeğ-i **ısır-dı** sanır-ım
man-PL dog-ACC bite-PST think-1SG
'I think the men bit the dog.'

Passive, reversed

d. Köpek-ler adam tarafından **ısır-ıl-dı** sanır-ım
dog-PL man by bite-PASS-PST think-1SG
'I think the dogs were bitten by the man.'

2) Irreversible Set

Active, canonical

a. Şefler-ler önlüğ-ü **giy-di** sanır-ım
chef-PL apron-ACC wear-PST think-1SG
'I think the chefs wore the apron.'

Passive, canonical

b. Önlük-ler şef tarafından **giy-il-di** sanır-ım
apron-PL chef by wear-PASS-PST think-1SG
'I think the aprons were worn by the chef.'

Active, reversed

c. Önlük-ler şef-i **giy-di** sanır-ım
apron-PL chef-ACC wear-PST think-1SG
'I think the aprons wore the chef.'

Passive, reversed

d. Şef-ler önlük tarafından **giy-il-di** sanır-ım
chef-PL apron by wear-PASS-PST think-1SG
'I think the chefs were worn by the apron.'

Table 1. Specifications of the Turkish LMs Used in the Study

Model Name	Parameters	Layers	Attention Heads
GPT-2-Base Turkish	124m	12	12
GPT-2-Large Turkish	774m	36	20
LLaMA-3-8B Turkish	8m	32	32

Table 2. Speakers' Accuracy on the APD and PJ Tasks

Set	Word Order	Structure	APD	PL
Biased	Nonreversed	Active	93%	97%
Biased	Nonreversed	Passive	97%	98%
Biased	Reversed	Active	92%	74%
Biased	Reversed	Passive	94%	83%
Irreversible	Nonreversed	Active	97%	96%
Irreversible	Nonreversed	Passive	96%	93%
Irreversible	Reversed	Active	95%	97%
Irreversible	Reversed	Passive	95%	99%

References

[1] Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. Current Directions in Psychological Science, 11(1), 11–15.

[2] Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. Cognitive Psychology, 42(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

[3] Fernanda F. 2000. Syntactic Vulnerability and Thematic Misinterpretation.

[4] Ehrenhofer, L., Lau, E., & Phillips, C. (2022). A possible cure for 'N400 blindness' to role reversal anomalies in sentence comprehension.

[5] Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. Journal of Memory and Language, 52(2), 205-225. https://doi.org/10.1016/j.jml.2004.11.001

[6] Kesgin, H. T., Yuce, M. K., Dogan, E., Uzun, M. E., Uz, A., İnce, E., Erdem, Y., Shbib, O., Zeer, A., & Amasyali, M. F. (2024). Optimizing large language models for Turkish: New methodologies in corpus selection and training. In 2024 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1–6). IEEE.

Figure 1. Sentence-final DTs in the SPR Tasks, by Set, Structure, and Word Order

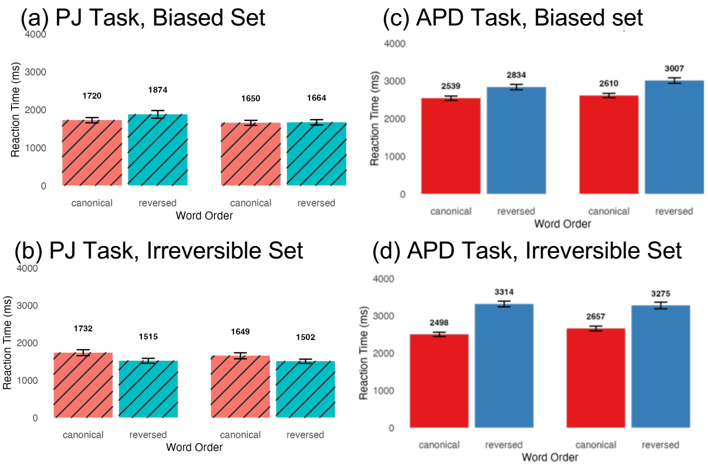
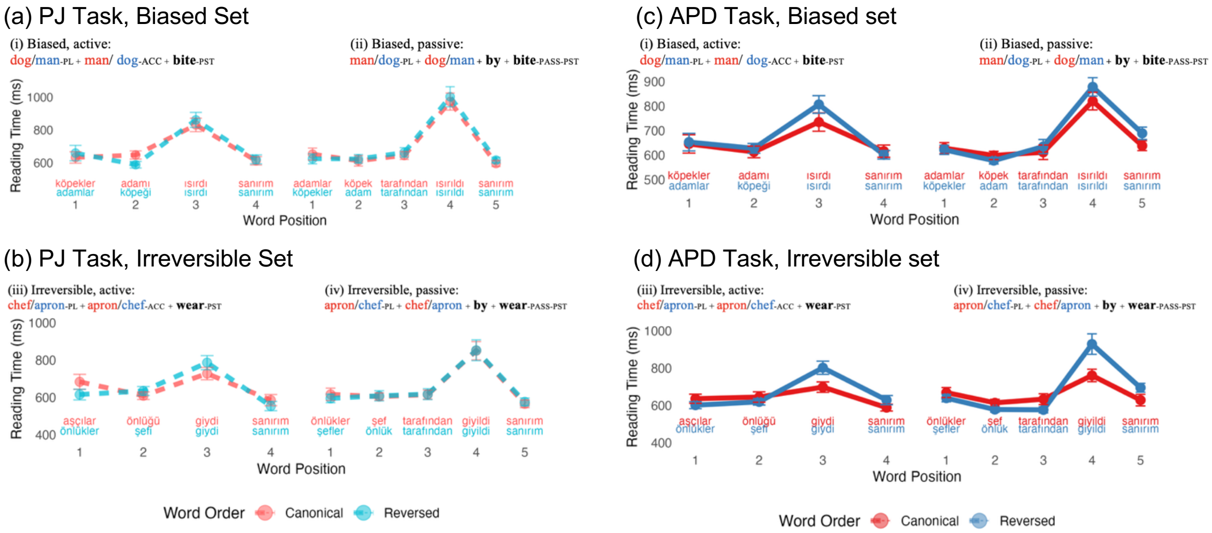


Figure 2. Word-by-word RT in the SPR Tasks by Set, Structure, and Word Order.



How *good-enough* is L2 sentence comprehension? Processing of thematic role (reversal)s in L2 Turkish

Onur Keleş[1], Nazik Dinçtopal Deniz[1], and Zeynep İrem Bayraklı Keleş[2,3]
[1]Bogazici University, [2]Yıldız Technical University, [3]İstinye University
onur.keles1@bogazici.edu.tr

Introduction. Language comprehension occasionally employs heuristic strategies that prioritize semantic plausibility and real-world knowledge over detailed syntactic analysis, leading to "good enough" (GE) interpretations [1]. This can result in misinterpretations, particularly with noncanonical structures like passives [2]. For second language (L2) processing, the Shallow Structure Hypothesis (SSH) posits that L2 speakers employ such "good-enough" parsing, especially for syntactically complex constructions, more than L1 speakers as their computations lack syntactic detail [e.g., 3]. However, other studies have shown that advanced L2 learners can engage in deep syntactic processing under certain conditions such as when they have achieved high proficiency in the target language, and when processing requires structural expectations [e.g., 4].This study investigates how native and L2 Turkish learners process active and passive sentences with varying plausibility, aiming to shed light on the interplay between heuristic and syntactic processing in L2 comprehension. **Methodology.** A self-paced reading (SPR) experiment was conducted with native (N = 52) and L2 Turkish speakers (N = 20) with upper intermediate proficiency. Sentences were taken from [2] and translated/adapted into Turkish. They manipulated syntactic structure (active vs. passive) and word order (canonical vs. reversed) as in (1) and (2). Half of the sentences had animate arguments which were reversible but *Biased* as in (1) and the other half had one animate one inanimate argument which were *Irreversible* as in (2). Participants read 42 experimental sentences across counterbalanced lists, along with 21 symmetrical controls (where the reversal yielded equally plausible events). Sentences were presented word-by-word. Each participant saw a set of 21 experimental sentences in each set, counterbalanced for word order and structure, along with 21 symmetrical controls (where the reversal yielded equally plausible events). After each sentence, they judged if the described event was plausible on a binary scale. Plausibility judgment accuracy, end-of-sentence decision times, and word-by-word reading times were recorded. **Results.** Accuracy and log RT/DT data were analyzed using (generalized) mixed-effects models, with fixed effects for Structure [Active, Passive], Word Order [Canonical, Reversed], Set [Biased, Irreversible], and Group [L1, L2], and with random effects for participants and items. We examined the RTs for the second region (where morphosyntactic cues for argumenthood become available) and the verb region. Accuracy data are summarized in Table 1. There was an interaction among Set, Word Order, and Group; OR = 0.03, p < .001. The L1 group was more accurate for canonical orders (1a,b; M = 97.5) than reversed orders (1c,d; M = 78.5%) in the biased set; OR = 0.05, p < .001. In the irreversible set, they were more accurate in the reversed orders (2c,d) than the canonical ones (2a,b; OR = 2.87, p = .004). However, reversed orders (i.e., implausible events) decreased L2 group's accuracies in both sets (*OR's* = 0.16-0.20, *p's* < .001). DTs are illustrated in Figure 1. For the L1 group, the DTs were slower for reversed sentences than canonical in the biased set (β = 0.13, p < .004) and longer in the irreversible set; β = –0.13, p < .001. As for the L2 readers, reversed word orders increased DTs in both sets; *β's* = 0.18-0.12, *p's* < .018. Word RTs are shown in Figure 2. L1 RTs for the verb were slower for the reversed orders than canonical orders only in the irreversible set; β = 0.09, p = .034. Verb RTs were also slower in the biased set for the passive sentences; β = 0.10, p = .039. As for the second regions, an interaction between structure and word order was observed only in the biased set: the region two in the passive and reversed sentences had longer RTs; β = 0.09, p = .029. As for the L2 group, verb RTs were significantly increased by reversed orders in both sets; *β's* = 0.32, *p's* = .002. The second region RTs among L2 readers increased in reversed orders only in the irreversible sentences; *β's* = 0.14, *p's* = .045. **Discussion.** The results showed that L1 participants used both active/passive and animacy information in their parsing decisions and showed *good-enough* parsing behavior only in structurally more complex (i.e., passive) constructions and when role reversals were plausible with the two animate arguments (i.e., in the biased condition). The animacy cue in the irreversible set (animate agent, inanimate patient) prevented *good-enough* parsing. The L2 data did not appear to benefit from the animacy cue to the same extent as the L1 speakers as evidenced by their lower accuracies for the reversed/implausible sentences regardless of structure and set. As such, we conclude that L2 speakers use morphosyntactic cues in their initial parsing decisions but their lower accuracies and longer end-of-sentence DTs for reversed/implausible sentences indicate that they have difficulty in integrating thematic roles sentence-finally, leading to greater reliance on heuristics [5] regardless of syntactic complexity. This may also be due to their proficiency or monitoring of their judgments more than L1 participants, as well as limited number of participants (data collection is in progress).

(1) Example Biased Set

Active, canonical

a. Köpek-ler adam-ı **ısır-dı** sanır-ım
dog-PL man-ACC bite-PST think-1SG
'I think the dogs bit the man.'

Passive, canonical

b. Adam-lar köpek tarafından **ısır-ıl-dı** sanır-ım
man-PL dog by bite-PASS-PST think-1SG
'I think the men were bitten by the dog.'

Active, reversed

c. Adam-lar köpeğ-i **ısır-dı** sanır-ım
man-PL dog-ACC bite-PST think-1SG
'I think the men bit the dog.'

Passive, reversed

d. Köpek-ler adam tarafından **ısır-ıl-dı** sanır-ım
dog-PL man by bite-PASS-PST think-1SG
'I think the dogs were bitten by the man.'

(2) Example Irreversible Set

Active, canonical

a. Şefler-ler önlüğ-ü **giy-di** sanır-ım
chef-PL apron-ACC wear-PST think-1SG
'I think the chefs wore the apron.'

Passive, canonical

b. Önlük-ler şef tarafından **giy-il-di** sanır-ım
apron-PL chef by wear-PASS-PST think-1SG
'I think the aprons were worn by the chef.'

Active, reversed

c. Önlük-ler şef-i **giy-di** sanır-ım
apron-PL chef-ACC wear-PST think-1SG
'I think the aprons wore the chef.'

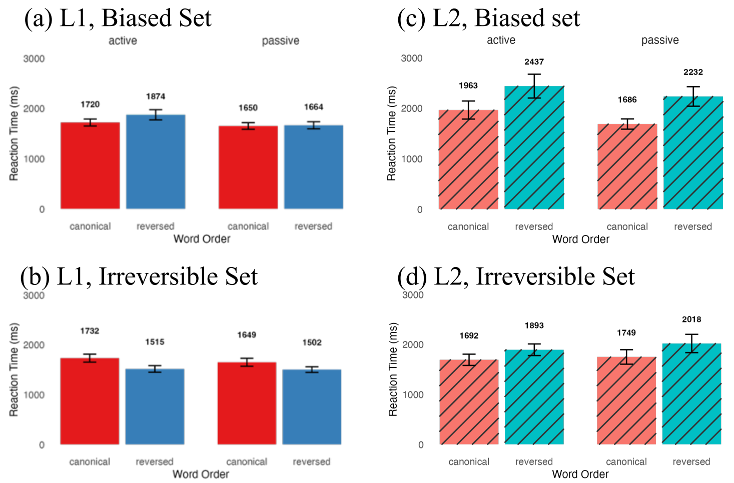
Passive, reversed

d. Şef-ler önlük tarafından **giy-il-di** sanır-ım
chef-PL apron by wear-PASS-PST think-1SG
'I think the chefs were worn by the apron.'

Table 1. L1 and L2 Speakers' Accuracy

Set	Word Order	Structure	L1 Acc.	L2 Acc.
Biased	Nonreversed	Active	97%	92%
Biased	Nonreversed	Passive	98%	91%
Biased	Reversed	Active	74%	71%
Biased	Reversed	Passive	83%	73%
Irreversible	Nonreversed	Active	96%	89%
Irreversible	Nonreversed	Passive	93%	89%
Irreversible	Reversed	Active	97%	66%
Irreversible	Reversed	Passive	99%	71%

Figure 1. End-of-sentence DTs in the SPR Task by Set, Structure, and Word Order



References

[1] Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. Current Directions in Psychological Science, 11(1), 11–15. https://doi.org/10.1111/1467-8721.00158
[2] Fernanda F. 2000. Syntactic Vulnerability and Thematic Misinterpretation.
[3] Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. Applied Psycholinguistics, 27(1), 3–42. https://doi.org/10.1017/S0142716406060024
[4] Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. Applied Psycholinguistics, 33(2), 419–456. https://doi.org/10.1017/S0142716411000427
[5] Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Bilingualism: Language and Cognition, 16(2), 367–382. https://doi.org/10.1017/S1366728912000302

Figure 2. Word-by-word RT in the SPR Task by Set, Structure, and Word Order.