Linguistic Evidence 2024

22 – 23 February, Potsdam

SFB1287 'Limits of Variability in Language'

University of Potsdam

Venue: Campus Griebnitzsee, Building 6, Lecture Hall H02

Programme

Thursday, 22 February

8:15 – 9:00

Registration

9:00 – 9:30

Welcome

9:30 – 10:35

Keynote talk: Possible and impossible determiners 🖼

Luisa Marti (Queen Mary University of London)

10:35 – 11:00

Coffee break

11:00 – 11:30

An experimental investigation of island effects in Romanian and the internal consistency of data sets from Romance languages 🖼

Gert-Jan Schoenmakers (Utrecht University) and Irina Stoica (University of Bucharest)

11:30 – 12:00

When children are more pragmatic than adults: Norwegian children's comprehension of contextualized absolute adjectives 🖼

Camilo Rodriguez Ronderos (University of Oslo), Emma Mathisen (University of Oslo), Ira Noveck (CNRS) and Ingrid Lossius Falkum (University of Oslo)

12:00 – 12:30

Microvariation and syntactic change: Numeral classifiers in Yucatec Maya 🖼

Barbara Blaha Pfeiler (CEPHSIS Merida) and Stavros Skopeteas (University of Göttingen)

12:30 – 13:45
Lunch

13:45 – 15:30
Poster session *(see below for details)*

15:30 – 16:00
Do syntactic and semantic similarity lead to interference effects? 🖳
Pia Schoknecht (University of Potsdam) and Shravan Vasishth (University of Potsdam)

16:00 – 16:30
On the similarity of causatives and modals: Targeting abstract modal representations 🖳
Angelica Hill (University of Massachusetts Amherst)

16:30 – 17:00
A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900 🖳
Sophia Voigtmann (Saarland University)

Friday, 23 February

9:00 – 10:05
Keynote talk: Interactive adaptation to others' speech: Evidence from patients with neurological conditions 🖳
Wolfram Ziegler (LMU München)

10:05 – 10:30
Coffee break

10:30 – 11:00
The Functional Relation of Prediction and Integration in Discourse Processing: Evidence from Reading Times and ERPs 🖳
Mathias Barthel (Leibniz-Institute for the German Language), Rosario Tomasello (FU Berlin) and Mingya Liu (HU Berlin)

11:00 – 11:30
A crosslinguistic survey of the at-issue status of ideophones: Experimental evidence from German, Akan, English and Vietnamese 🖳
Kathryn Barnes (Goethe University), Cornelia Ebert (Goethe University), Theresa Stender (Goethe University), Kurt Erbach (Goethe University), Yvonne Portele (Goethe University), Reginald Duah (HU Berlin), Prince Asiedu (University of Ghana), Josiah Nii Ashie Neequaye (University of Ghana) and Mavis Boateng Asamoah (University of Ghana)

11:30 – 12:00

A New Technique to Detect Lexical Content in Gaps and Its Application to the Question of Reconstruction in Wh-Movement 🖽

Benjamin Bruening (University of Delaware) and Rebecca Tollan (University of Delaware)

12:00 – 13:15

Lunch

13:15 – 13:45

Acceptability judgment experiments on PPs with gaps in 🖽

James Griffiths (University of Tübingen) and Craig Sailor (Trinity College Dublin)

13:45 – 14:15

Experimental findings for a cross-modal account of dynamic binding in gesture-speech interaction 🖽

Cornelia Ebert (Goethe University), Kurt Erbach (Goethe University) and Magnus Poppe (Goethe University)

14:15 – 14:45

Deontic priority – converging evidence for a universal in modal semantics 🖽

Wataru Uegaki (The University of Edinburgh), Anne Mucha (The University of Edinburgh), Nathaniel Imel (University of California, Irvine) and Shane Steinert-Threlkeld (University of Washington)

14:45 – 15:15

Coffee break

15:15 – 15:45

A Bounded Rationality Account of Constituent Order in SOV languages 🖽

Sidharth Ranjan (University of Stuttgart) and Titus von der Malsburg (University of Stuttgart)

15:45 – 16:15

Auditory phonotactic wellformedness intuitions depend on the nativeness of a speaker's pronunciation 🖽

Leonardo Piot (University of Potsdam & University of Paris-Cité), Thierry Nazzi (Université Paris-Cité) and Natalie Boll-Avetisyan (University of Potsdam)

16:15 – 16:45

Constraints on word exchanges during noisy-channel inference 🖽

Markus Bader (Goethe University) and Michael Meng (Merseburg University of Applied Sciences)

16:45 – 17:00
Closing remarks

Poster presentations

1. **Adjuncts and complements beyond English: testing the applicability of the do so test in Croatian** 🖼
   Ana Werkmann Horvat (University of Osijek), Matea Birtić (Institute for Croatian language and linguistics) and Martina Gračanin-Yuksek (Middle East Technical University)

2. **An Argument for Symmetric Coordination: A Replication Study** 🖼
   Adam Przepiórkowski (University of Warsaw & ICS Polish Academy of Sciences), Magdalena Borysiak (University of Warsaw) and Adam Głowacki (University of Warsaw)

3. **Punctuation Modulates the Valence of Referents in Exclamative Clauses in CMC** 🖼
   Kalle Glauch (Ruhr-Universität Bochum)

4. **The processing of quotation marks in German: Evidence from eye-tracking** 🖼
   Natascha Raue (University of Kassel), Holden Härtl (University of Kassel) and Álvaro Cortés Rodríguez (University of Potsdam)

5. **Korean causative change of state predicates and non-culminating readings** 🖼
   Paola Fritz-Huechante (HU Berlin) and Elisabeth Verhoeven (HU Berlin)

6. **Ellipsis (not) as deaccentuation: evidence from Icelandic** 🖼
   Güliz Güneş (University of Tübingen) and Nicole Dehé (University of Konstanz)

7. **L1-Acquisition of Deontic and Epistemic Meanings of Czech muset [must]** 🖼
   Edita Schejbalová (Charles University) and Radek Šimík (Charles University)

8. **Ja or Jaaaaa? The Influence of Iconically Lengthened Response Particles on the Scalar Interpretation of Adjectives** 🖼
   Lennart Fritzsche (Goethe University)

9. **(Un)marked indefinites in Russian and Bulgarian: An experimental investigation** 🖼
   Luca Molinari (University of Warsaw & Ca' Foscari University of Venice) and Daria Seres (University of Graz)

10. **Experimenting with principle C in German ATB movement** 🖼
    Timea Szarvas (University of Potsdam)

11. **Acquisition of auxiliary selection in French and Italian** 🖼
Balthazar Lauzon (Ulster University), Raffaella Folli (Ulster University), Juliana Gerard (Ulster University) and Christina Sevdali (Ulster University)

12. **Production and comprehension of case marking after local two-case prepositions in German-speaking preschoolers** 🖼
Tanja Diederich (FU Berlin) and Flavia Adani (FU Berlin)

13. **Agreement attraction effects depend on the goal of processing** 🖼
Anna Laurinavichyute (University of Potsdam), Himanshu Yadav (Indian Institute of Technology Kanpur), Titus von der Malsburg (University of Stuttgart) and Shravan Vasishth (University of Potsdam)

14. **'You don't hang a Frida Kahlo next to a Jackson Pollock.' The effect of referential features and gender congruency on the comprehension of unfamiliar artist-for-work metonymies in German** 🖼
Franziska Kretzschmar (Leibniz Institute for the German Language), Sandra Hansen (Leibniz Institute for the German Language), Anna Volodina (Leibniz Institute for the German Language) and Christian Lang (Leibniz Institute for the German Language)

15. **Prosodic Prominence and Negation** 🖼
Frank Kügler (Goethe University) and Markus Bader (Goethe University)

16. **Is there a conjunctive default in the interpretation of disjunction? A nonce word approach** 🖼
Adina Camelia Bleotu (University of Bucharest), Andreea Nicolae (ZAS Berlin), Gabriela Bilbiie (University of Bucharest), Mara Panaitescu (University of Bucharest), Anton Benz (ZAS Berlin) and Lyn Tieu (University of Toronto)

17. **It's not just all in the head: Towards a processing model of German adjective-noun-noun constructions and the bracketing paradox** 🖼
Anna Prysłopska (University of Stuttgart) and Titus von der Malsburg (University of Stuttgart)

18. **Experiments on anaphora resolution of generic masculine nouns in German** 🖼
Philipp Rauth (Saarland University), Robin Lemke (Saarland University) and Lisa Schäfer (Saarland University)

19. **The psycholinguistics of historical language change – Structural priming and the dative alternation in Middle English** 🖼
Gunnar Jacob (University of Mannheim) and Mirja Maier (University of Mannheim)

Keynote Talk

Possible and impossible determiners

Luisa Martí (Queen Mary University of London)

In this talk I propose that the form-meaning mapping in the category D(eterminer) is regulated by a principle of contiguity that applies to a hierarchy of determiner meanings. The meanings in question are those described in the literature for demonstratives, anaphoric and non-anaphoric definites and indefinites. One of the case studies discussed in the talk is based on my fieldwork on Atara Imere, a Polynesian language of Vanuatu.

An experimental investigation of island effects in Romanian and the internal consistency of data sets from Romance languages

Gert-Jan Schoenmakers and Irina Stoica

Syntactic island phenomena have long been a central topic in the linguistic literature, in part because of apparent differences between languages. Romance languages, for instance, ostensibly allow extraction from subject and interrogative islands (Rizzi 1982, Torrego 1984). The past decade has seen a strong proliferation of experimental investigations of syntactic islands in different languages, which have provided the discussion with a richer empirical basis. However, the new data suggest that there is still much variation between the Romance languages: while extraction from Spanish subjects (López-Sancio & Laka 2019) and embedded questions (Pañeda & Kush 2021) is seemingly unproblematic, significant subject and interrogative island effects were found in Italian (Sprouse et al. 2016), subject island effects in French (Abeillé et al. 2020), and interrogative island effects in Brazilian Portuguese (Almeida 2014) (we restrict ourselves to *wh*-dependencies here). Moreover, the reported effect sizes are vastly different.

The present paper has two main goals. First, Romanian is claimed to display island sensitivity (Dobrovie-Sorin 1994), but we could not find any experimental data corroborating this claim. Our first goal is thus to add data from Romanian to the discussion. Second, we aim to chart the variation patterns in our novel data, as well as in data sets from five prominent studies on syntactic islands in the Romance languages. We established the internal consistency of eighteen data sets using Generalizability Theory.

An experimental investigation of four different island effects in Romanian

We conducted two judgment experiments in Qualtrics, one testing adjunct and complex NP islands, the other interrogative and subject islands. The stimuli were adapted from the Italian stimuli from Sprouse et al. (2016), which were designed according to the factorial design that isolates the factors *distance* (short/long-distance movement) and *construction* (island/non-island), see (1) for a sample item.

(1) a. **Cine** ___ a auzit [că a copt Maria un tort]?   [non-island/short]
'Who heard that Maria baked a cake?
b. **Cine** ___ a auzit [zvonul că a copt Maria un tort]?   [island/short]
'Who heard the rumor that Maria baked a cake?'
c. **Ce** a auzit îngrijitorul [că a copt Maria ___]?   [non-island/long]
'What did the janitor hear that Maria baked?
d. **Ce** a auzit îngrijitorul [zvonul că a copt Maria ___]?   [island/long]
'What did the janitor hear the rumor that Maria baked?'

Procedure Participants were asked to rate how natural the target sentences would sound when uttered by a native speaker of Romanian, on a 7-point scale. Eight items per island type were distributed over four lists for both experiments (eight lists in total). 32 (un)grammatical filler items were added to each list and each list started with three practice items. Participants were randomly assigned to a list, in which items were presented in one of two distinct orders.

Participants 211 native speakers of Romanian without language deficits completed the questionnaire. Data from seven participants were removed prior to statistical analysis because their average score for ungrammatical fillers was higher than 4.5. We entered data from 103 participants in the adjunct/CNP-island task (M_{age}: 20.6, SD: 3.1, 18–46) and 101 participants in the subject/*whether*-island task (M_{age}: 21.7, SD: 5.1, 18–47) into statistical analysis.

Analysis We performed a series of linear mixed effect models on the *z*-transformed data using R (v. 4.2.3) and the *lme4* package (Bates et al. 2015). We entered the variables *distance* (short vs. long movement) and *construction* (island vs. non-island) as fixed effects into the models, using deviation contrasts (-0.5; 0.5). The random structure included by-item and by-participant intercepts; the random slopes and their interactions were included only if they improved the model fit (assessed by LRTs).

Results We find significant interaction (island) effects in each of the island types (adjuncts: $\beta = 1.10$, SE = 0.30, $t = 3.71$, $p = .006$; CNP: $\beta = 1.39$, SE = 0.09, $t = 15.47$, $p < .001$; interrogative: $\beta = 1.82$, SE = 0.08, $t = 21.86$, $p < .001$; subject: $\beta = 0.41$, SE = 0.09, $t = 4.56$, $p < .001$). The data are visualized in Figure 1. However, the patterns in Figure 1 show that closer scrutiny of the data is warranted. That is, long distance movement from non-islands increases ratings in the adjunct and *wh*-islands. This finding raises the question whether and to what extent the effect is reliable and replicable in future experiments.
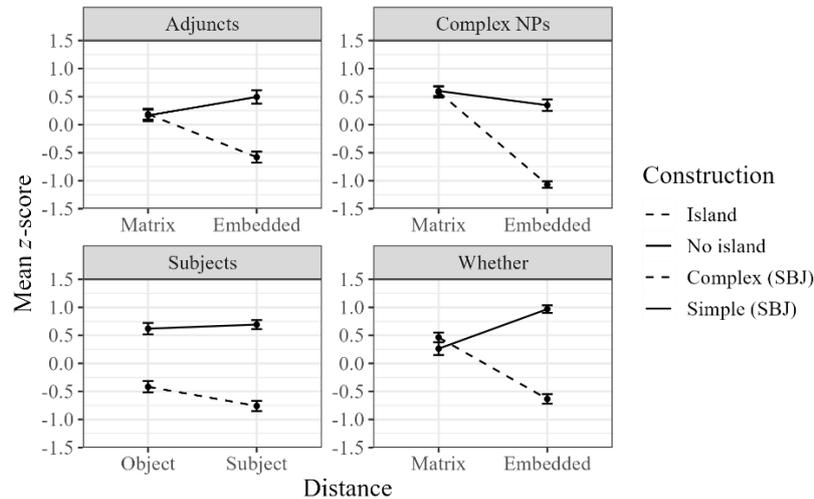
Figure 1: Romanian wh-dependencies, interaction plots for each island type (error bars represent 95% CIs)

The internal consistency of data

As a next step, we established the internal consistency of each of our data sets and, for the purposes of comparison, we computed the internal consistency of the data sets from Almeida (2014), Sprouse et al. (2016), Abeillé et al. (2020), Pañeda & Kush (2021), and Rodríguez & Goodall (2023) as well, making use of Generalizability Theory (see Schoenmakers & van Hout 2024). The outcome G-coefficient (ρ^2) represents the relative contribution of a true score component and an error component, like the ICC, but unlike the ICC it can handle nested structures and takes into account different sources of variation (such as participants or items). If the residual error component, which consists of the error attributable to the interaction of the sources of variation and random or unidentified effects (Shavelson & Webb 1991), is relatively large, the internal consistency of the data set is low and the judgment patterns are inconsistent.

We observe that in ten out of eighteen data sets (55.6%) participants did not perform the task in an adequately similar or consistent manner ($\rho^2 < .75$). This shows that island phenomena in Romance are prone to high degrees of (random) variation and that ratings cannot always be taken at face value. However, based on a series of decision studies, we show that the internal consistency can be improved by using larger numbers of items in the materials. We thus recommend experimental researchers to use at least 24 experimental items in a 2×2 design. More generally, we argue that these analyses, which are not common in the field of experimental syntax, can help understand the variability in judgment data.

References

◆ **Abeillé, A., B. Hemforth, E. Winckel, & E. Gibson (2020)**. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition 204*. Article 104293. ◆ **Almeida, D. (2014)**. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista de ABRALIN 13*(2), 55–93. ◆ **Bates, D., M. Mächler, B. Bolker, & S. Walker (2015)**. Fitting linear mixed effects models using lme4. *Journal of Statistical Software 67*(1). 1–48. ◆ **Dobrovie-Sorin, C. (1994)**. *The syntax of Romanian: Comparative studies in Romance*. Mouton de Gruyter. ◆ **López-Sancio, S. & I. Laka (2019)**. *Dependency type modulates island effects: Evidence from Spanish*. Presentation at the 93rd Annual Meeting of the Linguistic Society of America. ◆ **Pañeda, C. & D. Kush (2021)**. Spanish embedded question island effects revisited: an experimental study. *Linguistics 60*(2), 463–504. ◆ **Rizzi, L. (1982)**. *Issues in Italian syntax*. Foris. ◆ **Rodríguez, A. & G. Goodall (2023)**. *Clitic Left Dislocation in Spanish: Island sensitivity without gaps*. Presentation at the 36th Annual Conference on Human Sentence Processing. ◆ **Schoenmakers, G. & R. van Hout (2024)**. *Consistency and variability in acceptability judgments from naïve native speakers*. Unpublished manuscript. ◆ **Shavelson, R. & N. Webb (1991)**. *Generalizability Theory: A primer*. SAGE ◆ **Sprouse, J., I. Caponigro, C. Greco, & C. Cecchetto (2016)**. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory 34*(1). 307–344. ◆ **Torrego, E. (1984)**. On inversion in Spanish and some of its effects. *Linguistic Inquiry 15*(1), 103–129.

When children are more pragmatic than adults: Norwegian children's comprehension of contextualized absolute adjectives

Camilo R. Ronderos[1], Emma Mathisen[1,2], Ira Noveck[2] & Ingrid Lossius Falkum[1]
[1]University of Oslo, [2]CNRS, Paris
camilorr@uio.no

Adjectives like *straight* (known as **maximum standard absolute gradable adjectives**, henceforth **absolute adjectives** for short) can have both a precise ("literal") (*perfectly straight*) and an imprecise ("non-literal") (*straight enough*) interpretation. Their precise interpretation has been claimed to be part of the adjective's semantic meaning (Kennedy, 2007; Syrett et al., 2010; Aparicio et al., 2015; i.a.). The imprecise interpretation is seen as a pragmatic phenomenon that arises after setting a contextual threshold for tolerance (see Lasersohn, 1999; Leffel et al., 2016). How and when do we learn to set this threshold in order to decide whether people are speaking precisely or imprecisely?

It could be the case that as a form of pragmatic reasoning, imprecise interpretations develop over time, with young children first showing a preference for literal interpretations, similarly to what has been found for other pragmatic phenomena (e.g., Noveck, 2001).

Alternatively, it could be that children show the opposite trajectory. Firstly, it is reasonable to assume that there is a prevalence of contexts that favor imprecise interpretations (how many things around a child are perfectly straight?). Secondly, Lee & Kurumada (2020) showed that second language adult learners can reason about the intentions of a speaker in order to learn the precise meaning of a novel absolute adjective despite receiving only imprecise input. It could therefore be that children first develop an imprecise interpretation of *straight*, and only with time (i.e., after multiple exposures to imprecise input) are able to bootstrap the precise, semantic meaning of the absolute adjective.

The current study tests these two hypotheses using a paradigm based on the landmark study by Syrett et al. (2010), who investigated children's understanding of gradable adjectives. They asked participants (3, 4, 5-year-olds and adult controls) to select *the straight wire* when presented with two different objects. If neither of the objects could be described as a referent of the utterance, they could choose neither. When children had to select between two objects that did not have the maximal degree of the relevant property (i.e., neither of the wires was straight), they preferentially selected *the straighter wire*, that is, they preferred an imprecise interpretation. Adults overwhelmingly selected neither of the pictures, consistent with a precise, semantic understanding of the adjective *straight*. This study was limited to a single item (*full jar*, in Experiment 1, and *straight wire* in Experiment 2), tested only 10 participants per age bracket, and did not discuss a possible developmental trajectory. Importantly, the imprecise picture used represented a large deviation from the standard (i.e., the wire was bent). It is uncertain how these results translate to the interpretation of more common, conventional uses of imprecision (e.g., an only slightly bent line being referred to as *straight*).

We therefore set out to replicate and extend the findings of Syrett et al. (2010). In a pre-registered experiment, we tested 100 native speakers of Norwegian ages 3-8 along with 33 adults. In the experiment, participants saw 12 critical items (using 6 different absolute adjectives) and 12 fillers on a tablet screen. In each trial, participants heard an instruction to select a picture. They then saw a set of three pictures including an incorrect referent, a target referent (depicting either a precise, imprecise or incorrect control picture in each condition) and a red X, signifying that neither of the two pictures was the correct referent (See Figure 1). Participants were told in advance that they should select the red X if neither of the two images complied with the spoken instruction. The Experiment had CONDITION (three levels: precise, imprecise, control), AGE (continuous predictor measured in days) and their interaction as fixed effects. The dependent variable was whether participants selected the target picture (coded as 1) or the red X (coded as 0). Instances of selecting the distractor image (less than 2% of the data) were discarded.

We fitted a mixed-effects, 'maximal' logistic regression model to the data, which included random intercepts and slopes for both factors and their interaction by items and random intercepts and a slope term for the factor CONDITION by subjects. The results suggest a different developmental trajectory for imprecise relative to precise interpretations: There was an interaction between CONDITION (precise vs. imprecise levels) and AGE (z-value= 2.4, $p<0.05$), suggesting that, with age, children were less likely to tolerate imprecise interpretations and increasingly preferred precise ones (see Figure 2). A post-hoc test (with matched number of participants per group) comparing adults to the

youngest 33 children showed that adults were significantly less likely to tolerate imprecision ($p < 0.05$), whereas there was no significant difference in their acceptance for precise referents.

Overall, our findings suggest that children behave more pragmatically than adults when understanding imprecision, and that only with age do they become less tolerant of imprecise interpretations of absolute adjectives. This could be indicative of differences in the lexical representation of absolute adjectives between children and adults: It could be the case that children's lexical entry is underspecified for precision (i.e., that their representation of absolute adjectives does not contain the requirement that the adjective's argument possess the maximal quantity of the relevant property). Only with repeated exposure to different usages of the adjectives can they bootstrap the precise meaning, in line with the suggestion of Lee and Kurumada (2022).

Another alternative is that children do have a precise representation of absolute adjectives but are simply more tolerant of imprecision, as suggested by Syrett et al. (2010). We are currently conducting further studies investigating the contextualized interpretation of absolute adjectives in order to test between these possible alternatives.

Figure 1: The three possible visual displays shown to participants.
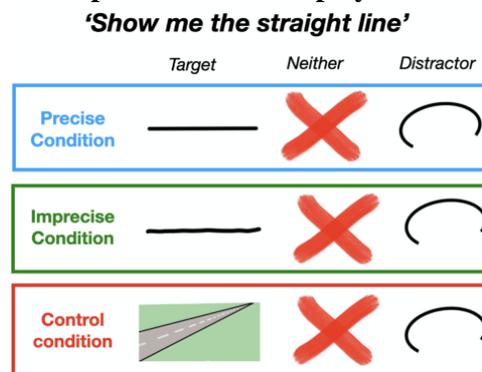
Figure 2: Results of the picture selection task.
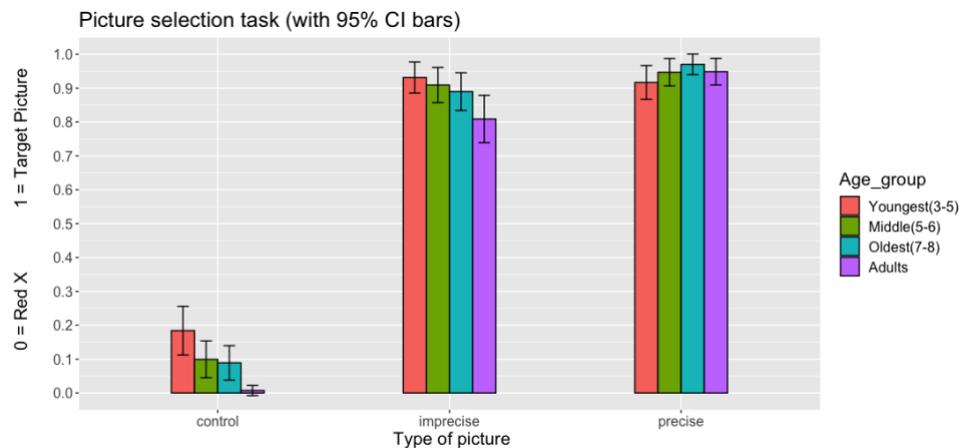
References

Aparicio, H., Xiang, M., & Kennedy, C. (2016). Processing gradable adjectives in context: A visual world study. In *Semantics and Linguistic Theory* (Vol. 25, pp. 413-432).

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, *30*(1), 1-45.

Lasersohn, P. (1999). Pragmatic halos. *Language*, 522-551.

Lee, C., & Kurumada, C. (2021). Learning Maximum Absolute Meaning Through Reasoning About Speaker Intentions. Language Learning, 71(2), 326-368.

Leffel, T., Xiang, M., & Kennedy, C. (2016). Imprecision is pragmatic: Evidence from referential processing. In *Semantics and linguistic theory* (Vol. 26, pp. 836-854).

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165-188.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of semantics*, *27*(1), 1-35.

Microvariation and syntactic change: Numeral classifiers in Yucatec Maya
Barbara Blaha Pfeiler and Stavros Skopeteas

SUMMARY: The methodological goal is to draw inferences from the variation between speakers from different locations and age groups to the underlying processes of syntactic change. In particular, we examine the use of numeral classifiers in the *Atlas of Yucatec Maya* and test whether the current variation in the generalized use of an abstract classifier results from the grammaticalization of this element or from the reanalysis of mensural classifiers to nouns.

BASIC FACTS: Yucatec Maya has an inventory of classifiers (cf. Miram 1983), including sortal classifiers (= types of entities, e.g., 'thing', 'person', 'long object', etc.) and mensural classifiers (= quantities, e.g., 'slice', 'ball', 'drop'). Both types of classifier can be selected directly via the numerals, which cannot appear without a classifier; see (1). Sortal classifiers as in (1a-b) are needed by the *numerals*; they assume the *cardinality function* that is not encoded by the numerals in these languages (see Bale et al. 2019 and Little et al. 2022 on Ch'ol; based on Krifka 1995). Mensural classifiers express measures, but differ from ordinary nouns expressing measure in that they may directly complement the numerals (without the mediation of a classifier).

(1) a. *jun**(-*p'éel*) *naj*   (one-CL.UNIT house) 'one house'

b. *jun**(-*ts'íit*) *kib*   (one-CL.LONG candle) 'one candle'

c. *jun**(-*wóol*) *sakam*   (one-CL.BALL dough) 'one ball of dough'

Numeral classifiers in Yucatec Maya are subject to two processes of language change that create variation in the contemporary language. The first process is the use of a general classifier -*p'éel* 'CL.UNIT' instead of specific sortal classifiers; see (2a). The second process is the use of the general classifier along with mensural classifiers; see (2b).

(2) a. *jun-p'éel kib*   (one-CL.UNIT candle) 'one candle' (cf. 1b)

b. *jun-p'éel wóol sakam*   (one-CL.UNIT BALL dough) 'one ball of dough' (cf. 1c)

HYPOTHESES: The phenomena in (2a)-(2b) may be due to the following hypotheses (that do not exclude each other): A. The change lies in the classifier -*p'éel*, that develops to a general marker of the *Cardinality Function*; B. The change lies in the mensural classifiers, that are reanalyzed as nouns expressing measure, which gives rise to the use of the classifier in (2b).

If the current developments are reducible to a change in the use of the classifier -*p'éel* (hypothesis A), then the constructions in (2a-b) are expected to appear in the same groups of the contemporary population (certain regions, age groups, etc.). If mensural classifiers are reanalyzed as nouns, then the use of the construction in (2b) is predicted to depend on classifier, and the occurrence of the construction in (2b) will not necessarily appear in the same groups of speakers.

METHOD: We examine data from 176 native speakers (birth years range between 1906 and 1989) from a sample of locations containing all regions in which Yucatec Maya is currently (this data is part of the Atlas of Yucatec Maya); see locations in Fig. 1. The data were collected through translation of simple expressions (illustrative prompts: 'one candle', 'two portions of beans', etc.) from Spanish to Yucatec Maya. We examine 24 prompts that elicited numeric expressions (this is only a subset of the data provided by the Atlas of Yucatec Maya. The data are analyzed with generalized additive models (GAM, Wood 2017).

RESULTS: The results of this study provide evidence for both hypotheses:

- The use of the general classifier instead of sortal classifiers (as in 2a) is a significant predictor of the use of the general classifier in mensural constructions (as in 2b).
- In mensural expressions, the choice of the construction with the general classifier (as in 2b) varies between types of mensural classifier (expressing 'part', 'sum' or 'portion').
- The geographical dispersion of the general classifier differs: compare the dispersion of the use of the general classifier in sortal expressions in Figure 2 with the dispersion in measure expressions in Figure 3. This pattern suggests that these developments (partly) take place in different parts of the population.

Fig 1. Sortal Classifiers: Probability of generalized classifier (brown), as in (2a), viz. specific classifiers (green), as in (1b)	Fig 2. Space coefficients of the GAM on sortal classifiers (2a vs. 1b)	Fig 2. Space coefficients of the GAM on mensural classifiers (2b vs. 1c)
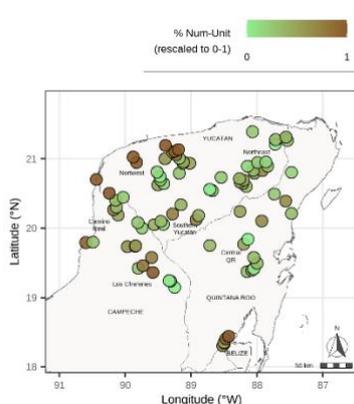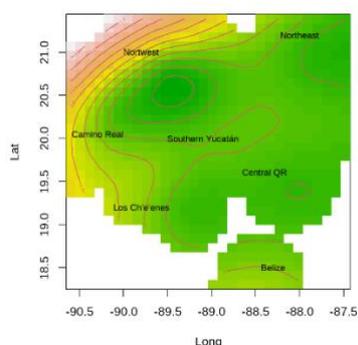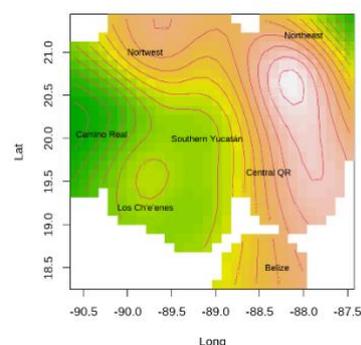
CONCLUSIONS: Our findings show that the two processes at issue are partially related: the general classifier develops towards a general marker of the *Cardinality Function*, mediating between numerals and nouns in all classifier expressions. A second source of variation lies with the mensural classifiers: some of them are reanalyzed as nouns expressing measure. Crucially, the current trend towards the generalization of a single classifier is cumulated to this variation, i.e., it does not reducible to the lexical developments of mensural classifiers.

References

Bale, A., J. Coon & N. Arcos. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: a journal of general linguistics* 4(1). 77.1‑30;

Krifka, M. 1995. Common nouns: a contrastive analysis of English and Chinese. In G. Carlson & F.J. Pelletier (eds.), *The generic book*, 398‑411. Chicago: University of Chicago Press.

Little, C.-R., M. Moroney & J. Royer. 2022. Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: a journal of general linguistics* 7(1). 1‑35.

Wood, S. N. 2017. *Generalized Additive Models*. Chapman and Hall/CRC.

Do syntactic and semantic similarity lead to interference effects?
Pia Schoknecht & Shravan Vasishth, University of Potsdam

When dependent elements are not close together within a sentence, comprehenders need to rely on their memory to integrate them. But memory processes can go awry and cause processing difficulty. Cue-based retrieval, a theory on memory processes underlying sentence processing, postulates that dependency completion is a cue-based process [1]. For example, in (1) the retrieval cues {+subject} and {+animate} can be generated at the position of the verb *moaned* to retrieve its subject *the lady*. If the retrieval cues match with not only the features of the subject, i.e., the retrieval target, but also with features of other nouns in the sentence, then cue-based retrieval theory predicts processing difficulty at the verb. This difficulty to identify the correct dependent is called similarity-based interference. In (1), a noun intervening between subject and verb was manipulated to function as a distractor, i.e., induce interference. The animacy of the distractor was manipulated to induce semantic interference: The animate distractor *man* is assumed to cause high semantic interference, while the inanimate distractor *seat* is assumed to cause low semantic interference. Syntactic interference was manipulated via the syntactic status of the distractor. In (1a) it has subject marking and in (1b) it is inside a prepositional phrase, i.e., not a subject.

(1)   The pilot remembered that the lady$_{+anim.,+subj.}$ who . . .
  a.   said that the man$_{+anim.,+subj.}$ / seat$_{+anim.,+subj.}$ was smelly
  b.   who was sitting near the smelly man$_{+anim.,-subj.}$ / seat$_{+anim.,-subj.}$
  . . . yesterday afternoon <u>moaned</u> about a refund.

Previous findings suggest that syntactic and semantic cues are used during sentence processing [2, 3], but it is not obvious which cues are relevant for a particular retrieval [4, 5]. A large body of reading studies has investigated retrieval interference [for a review, see 6], but event-related potential (ERP) studies on the matter are sparse. ERP results on interference have been inconsistent across studies.

We present large-sample self-paced reading (N=774) and ERP experiments (N=103) using a well-established design [2, 3, see (1)] which investigated interference due to syntactic cues and semantic cues. A Bayes factor analysis showed that both the self-paced reading times as well as ERPs provided clear evidence for a semantic interference effect. In the self-paced reading experiment, semantically matching distractors caused a slow-down in reading time which persisted throughout following regions, including the critical verb region (see Figure 1). ERPs elicited by the verb in the standard spatio-temporal N400 window were more negative when the distractor matched the semantic cue vs. when it did not (see Figure 2). In contrast, in both experiments, Bayes factor analyses showed evidence against i) interference due to syntactic cues (see Figure 1, 3) and ii) the interaction of syntactic and semantic interference.

These findings contradict the prediction of cue-based retrieval theory [7], which assumes that both syntactic and semantic cues play an equal role in retrieval. We show through computational modeling that cue-based retrieval also shows no syntactic interference, if the parser is assumed to use hierarchical syntactic information. A modified cue-based retrieval model, which uses a hierarchically informed syntactic cue {+subject-in-same-clause} instead of the usually assumed {+subject} cue, predicts effects which are close to the observed patterns in our data. Importantly, our data suggest that the parser uses syntactic and semantic cues independently of each other. In sum, this large-sample study suggests that both syntactic and semantic cues drive retrieval at the verb, but that syntactic cues may include hierarchical syntactic information.

[1.] Lewis, R. L. & Vasishth, S. *Cognitive Science: A Multidisciplinary Journal* **29,** 375–419 (2005).
[2.] Mertzen, D. et al. *Glossa Psycholinguistics* **2** (2023). [3.] Van Dyke, J. A. *Journal of Experimental Psychology: Learning, Memory & Cognition* **33,** 407–430 (2007). [4.] Dillon, B. et al. *Journal of Memory and Language* **69,** 85–103 (2013). [5.] Sturt, P. *Journal of Memory and Language* **48,** 542–562 (2003). [6.] Jäger, L. A. et al. *Journal of Memory and Language* **94,** 316–339 (2017). [7.] Engelmann, F. et al. *Cognitive Science* **43,** e12800 (2019).

Figure 1: Self-paced reading times across the whole sentence; separately for high (A) and low (B) syntactic interference due to differing sentence structure.
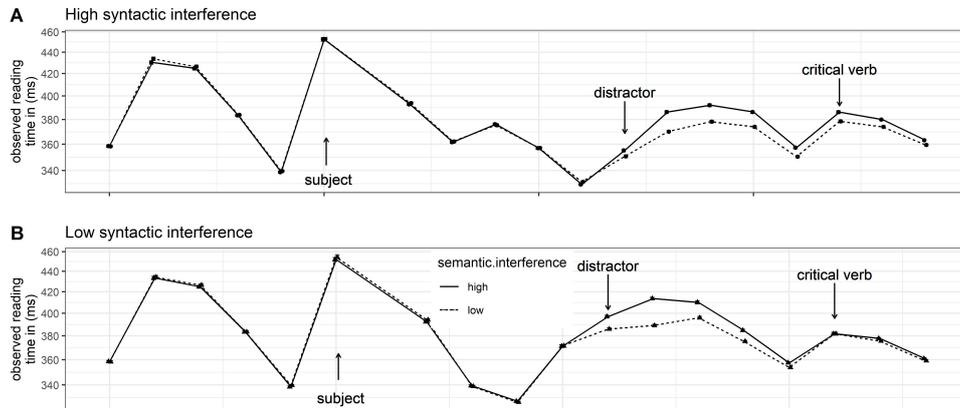
Figure 2: Grand average event-related potentials for semantic interference at electrode Cz.
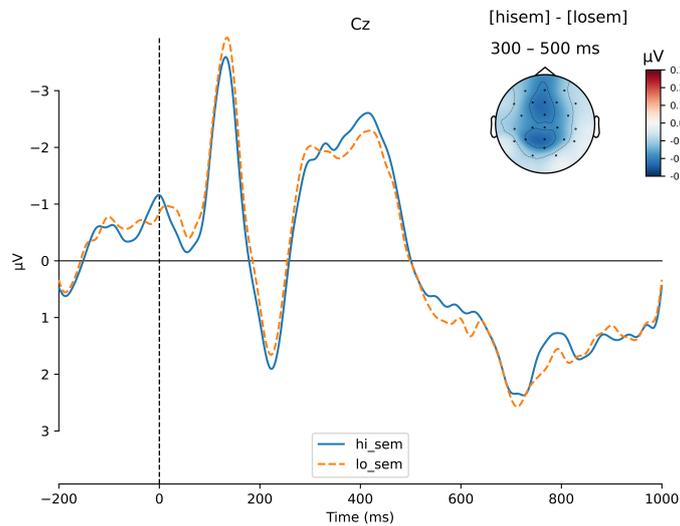
Figure 3: Grand average event-related potentials for syntactic interference at electrode Cz.
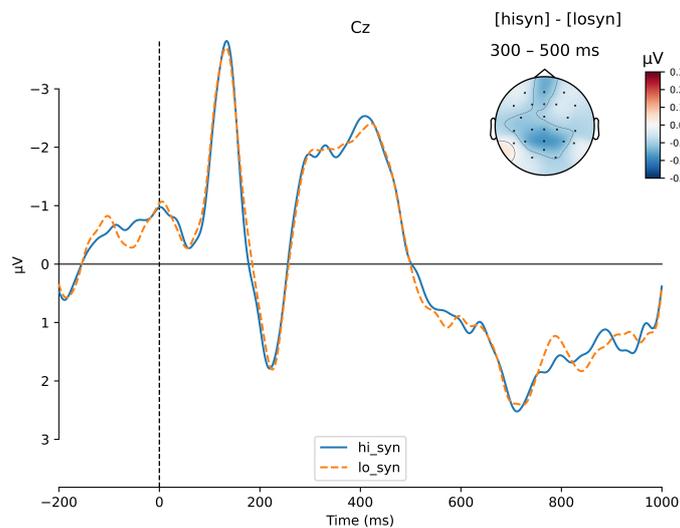
On the similarity of causatives and modals: Targeting abstract modal representations
Angelica Hill

University of Massachusetts Amherst

Motivation. It's debated whether causatives contain a modal component [6, 7, 10, 5] or not [3, 8]. Any semantic theory must capture the fact that obligation and permission interpretations are associated with causatives shown here with a cancellation test, "*Giulia made Giorgio go to the store, ?but he didn't have to.*" To account for this [10, 5] provide modal analyses of causatives, in which *made* universally quantifies over possible situations, entailing necessity, which is also entailed by strong deontic modals, such as *had to*. If causatives and modals share the same core meaning, and priming effects can target shared meaning, then **it's possible the shared modal necessity conveyed by causative *made* and deontic modal *had to* can be targeted with priming**. Here we tested the hypothesis that strong causatives and strong deontic modals in English share a necessity entailment using semantic priming in a **sentence recall task** [2, 4, 9]. Our results show that ***had to* can be primed by *made* despite their structural differences, suggesting that causatives and deontic modals indeed share the same core meaning and that priming can arise from the shared meaning.**

Background. There are differences between causatives such as *made* and deontic modals such as *had to*: *made* yields actuality entailments, *had to* doesn't; *made* takes a small clause as argument, *had to* doesn't. Yet modal meanings arise with causative sentences, shown in the example above. Similar cancellation tests have shown the inability to cancel deontic meaning in causative sentences cross-linguistically, such as in Serbian, Japanese, Yu'pik, and Italian [6, 7]. Causatives and modals are semantically similar in that they can vary in strength. For causatives: *cause* vs. *allow*, for modals: *necessity* vs. *possibility*, and they have similar entailment patterns between strength: *cause* entails *allow*, and *necessity* entails *possibility*. For [10, 5] the modal interpretation is available with causatives because they convey modal necessity in the case of *cause* and possibility in the case of *allow*, entailing the respective obligation and permission (see [7] for a causal model account based on presupposition). These similarities between *made* and *had to* motivate the hypothesis that both expressions share the same core meaning (quantification over possible situations). We use priming to test this hypothesis.

Design. To observe whether people could be primed to produce a *had to*-sentence after uttering a *made*-sentence, we tested sentence production in a priming experiment using the sentence recall task. In each trial participants (n=48) were asked to read aloud and memorize two sentences, one of which they were asked to recall later given a cued prompt, an uninflected verb that appeared in the to-be-recalled sentence (Fig. 1). Sentences appeared as either *had to*-sentences (modal), *made*-sentences (causative), or in the simple past (control). For critical trials (n=24), target sentences were in the simple past, prime sentences were either a *made*-sentence (n=12) or in the simple past (n=12), and people were asked to recall the target (Fig. 1). Of the 72 total trials, 36 of them contained a *had to*-sentence, meaning people were biased to insert *had to* during recall. If *made* indeed primes *had to*, we should observe that the rate of ***had to* production should be higher when the prime sentence contains *made*, compared to when the prime is a control sentence** where no causative or modal expressions are involved.

Results. During recall people inserted *had to* after uttering a *made*-sentence 47.4% of the time, and after uttering a control sentence 41.3% of the time (Fig. 2). A maximally-structured mixed effects logistic regression model [1] suggests that **the rate of *had to*-insertion after uttering a *made*-sentence is significantly higher than the rate of *had to*-insertion after uttering a control sentence (p = 0.037)**. This replicates a prior, similarly designed, pilot experiment with less statistical power (p = 0.08).

Conclusion. Our results show that **priming can be used to target high-level representations used when processing *made* and *had to***, two syntactically dissimilar constructions, and support semantic accounts in which the two share semantic meaning. Future work will test possible priming effects between causatives and deontic modals of different strengths. For example, we predict a weak causative such as *allow* should prime the insertion of a weak deontic modal. Our results show that abstract semantic representation can be primed using the standard psycholinguistic method and contribute to the methodological and theoretical progress of understanding semantic processing and its effect on language production.

[1] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. [2] Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places?. *Cognition*, *90*(1), 29-49. [3] Copley, B., & Harley, H. (2015). A force-theoretic framework for event structure. *Linguistics and Philosophy*, *38*, 103-158. [4] 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. [5] Hill, A. (2024). A counterfactual analysis of causatives. [Ms in preparation]. [6] Ilič, T. (2013). *Modality and causation in Serbian dative anticausatives: A crosslinguistic perspective* (Doctoral dissertation). [7] Ilič, T. (2014). Modality and causation: Two sides of the same coin. In B. Copley & F. Martin (Eds.), *Causation in Grammatical Structures* (72-86). Oxford University Press. [8] Martin, F. (2018). Time in probabilistic causation: Direct vs. indirect uses of lexical causative verbs. In *Proceedings of Sinn und Bedeutung* (Vol. 22, No. 2, pp. 107-124). [9] Momma, S. (2022). Producing filler-gap dependencies: Structural priming evidence for two distinct combinatorial processes in production. *Journal of Memory and Language*, *126*, 104349. [10] Privoznov, D. (2023). On Actuality Entailments, Causation, and Telicity in Balkar. *Languages*, *8*(3), 178.
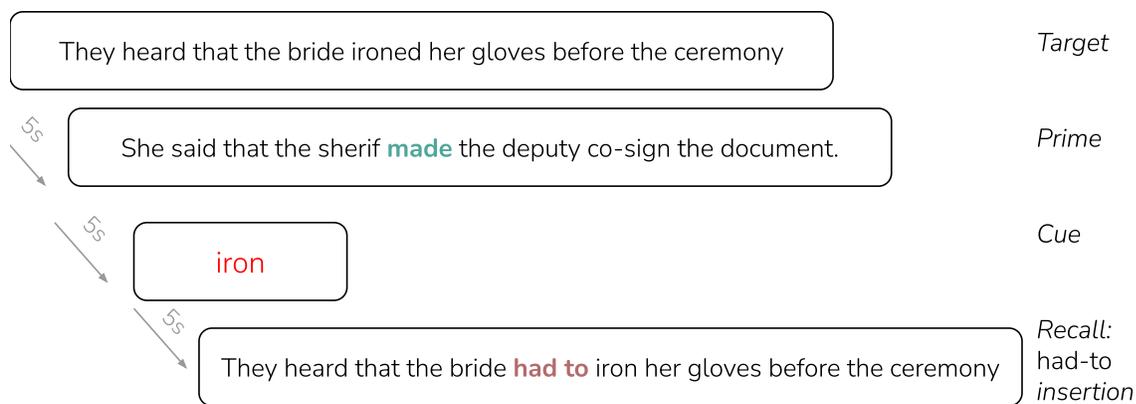
Figure 1. Example of critical trial
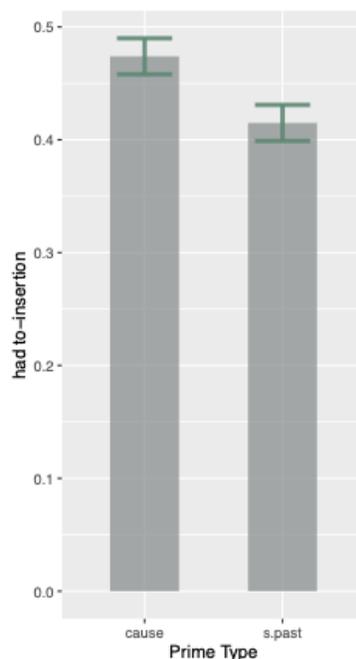
Figure 2. *Had-to* insertion for causative and control conditions

A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900

Sophia Voigtmann
Saarland University

Although in modern German, it is highly marked to place an NP in the postfield,[1] the phenomenon is not as rare as expected in early New High German (1650-1900) data (ex. A).

A. …weil er […] von den meisten Medicis [genennet wird]$_{RSB}$ **ein Schmid aller Kranckheiten**.
    ... as he ... by the most doctors called is a forger of.all sicknesses.
    "...as he is called a forger of sicknesses by most physicians." (Abel 1699, 225)

However, studies concerned with NP-extraposition in diachrony treat the placement of NP behind the so-called right sentence bracket (Wöllstein 2014) as a marginal phenomenon that can nearly exclusively be explained by the length of the NP (Ebert 1980, Sapp 2014) or pragmatic factors like givenness (Light 2011).

Although it is not mentioned as such in these studies, both explanations can be linked to processing difficulties which are resolved by extraposition. Processing difficulties can be rather objectively investigated using Information Density (ID), namely Surprisal (Shannon 1948). Levy and Jaeger (2007; 1) define information as the "amount of information per unit comprising the utterance". It is calculated as the likelihood with which a word occurs in a context (P(word) = -log2(word|context)). More expected combinations of words result in lower surprisal values and, thus, in lower perceiving difficulties (Hale 2001, Levy 2008). Furthermore, low surprisal values reduce processing difficulties (Levy & Jaeger 2007, Hale 2001, Levy 2008). Surprisal values can generally be calculated on any linguistic unit. For this study, we concentrate on the lexical predictability of a word in its context and claim that the surprisal value of NPs is also relevant for their placement in the postfield. Therefore, we propose the following hypothesis: NPs with high surprisal values are more likely to be extraposed.

To investigate this claim, we built a corpus of medical and theological texts from 1650 to 1900 taken from the Deutsches Textarchiv (DTA, BBAW 2019). We manually extracted extraposed and embedded NP and the sentence brackets using WebAnno (Eckart de Castilho et al. 2016). Then, we calculated a 2-Skip-Bigram-Language Model (Guthrie et al. 2016) to gain surprisal values for every word in the context. However, the mere word surprisal value is no sufficient approximation for our study because we aim to quantify the surprisal value on a whole constituent. Therefore, we choose to calculate the mean Skipgram surprisal on lemmata for every annotated NP, which is calculated as the sum of all surprisal values in the annotated constituent divided by the number of words, i.e. the average surprisal. Furthermore, we determined the length of the NP, the text genre (medical vs. theological), and the Orality Score (Ortmann & Dipper 2022) since extraposition is claimed to be more likely in conceptionally oral texts (Koch & Oesterreicher 2007), the time of publication, the period, and the constituent function of every extraposed NP. To determine the most influential factor for extraposition, logistic regression was performed with R (The R Core Team 2022).

As a result (Table 1), we find that extraposition is indeed linked to high surprisal values (z=-2.44, p<.05 *) and that length is not significant (z=-0.48, p<0,63), in contrast to the aforementioned literature. However, we also see a significant interaction between the Orality Score, i.e. the closeness to orality and the period in which a text was published (z=-2.68, p<.001**). That suggests a difference between medical and theological texts and a change over

---

[1] The postfield is the position behind the right sentence bracket (RSB) and the RSB is the position late in the clause where verbal material, which is distributed over two positions in the clause in German, occurs (Wöllstein 2014).

time. The latter is furthermore supported by a slightly significant result for the interaction between length and period (z=-1.75, p<.1) which shows that length becomes tendentially more influential in younger periods of German.

Following Speyer (2015: 499), we suggest that there are more processing capacities available behind the right sentence bracket because the main verb is eventually processed at this point. Due to the characteristic of German, the verbal material of a clause is distributed over two positions, the so-called left and right sentence brackets.

Until the right sentence bracket is processed, where in the case of extraposition only the lexical verb could be fully processed, there is great uncertainty for the recipient about the material that still must appear in the sentence (Gibson 1998, Levy 2008). Uncertainty about the subsequent material consumes cognitive capacity (Levy 2008). Consequently, this capacity is no longer available to process lexically challenging constituents, i.e., constituents with high surprisal values. Thus, if a constituent with a high surprisal value were to be processed at a point where there were still strong uncertainties about the verb and the arguments required by the valence, this could lead to a general overload of cognitive capacity, causing communication to fail. This is prevented by extraposing NPs with high surprisal values. Since after the right sentence bracket, the verb valence is also processed, the recipients know which arguments are still required and do not have to spend any more capacity on predicting them. The capacity thus gained can be used for lexical processing. Extraposition thus ensures successful communication. In our corpus, the effect is even more pronounced than the influence of length.

We further showed that extraposition occurs significantly more frequently in (conceptually oral) medical texts than in theological texts. This indicates a difference between the two genres that can also be explained by processing. Since medical research developed rapidly between 1650 and 1900 (Eckart 2021), numerous new terms and new verbs became necessary - in contrast to theological texts. With new verbs, the valence is naturally also less known, so it is advantageous to present these verbs earlier, which is achieved by extraposition. In addition, the more complex content also requires more cognitive capacity to process. In the case of theological texts, on the other hand, the content is more familiar and undergoes fewer changes. In their case, embedding may consequently provide advantages for processing because the constituents before the right sentence bracket may in turn make the material in the right sentence bracket more expectable (Levy 2008).

Our study thus contributes to exemplifying the complex relationship between processing and word order, using the extraposition of nominal phrases as an example. It also focuses on an area as well as on a German language period which has so far rather rarely been in the spotlight of research. It also shows that both early New High German and scientific texts should receive more attention in research.

	Estimate	Std.-Error	z-values	p-value	
Intercept	2,77	1,64	1,69	0,09	.
Skipgram$_{mean}$	-0,92	0,37	-2,44	0,01	*
Length$_{log}$	-0,33	0,69	-0,48	0,63	
Orality Score	3,45	1,39	2,47	0,01	*
Genre	-2,36	0,92	-2,58	0,009	**
Period	0,38	0,26	1,45	0,15	
Length$_{log}$:Orality Score	-2,39	1,35	-1,77	0,08	.
Length$_{log}$:Genre	1,81	0,97	1,87	0,06	.
Length$_{log}$:Zeitstufe	-0,46	0,26	-1,75	0,08	.
Orality Score:Genre	2,50	0,83	3,00	0,003	**
Orality Score:Period	-0,63	0,23	-2,68	0,007	**

Table 1 Logistic regression of all NPs.

Example taken from:
Abel, H. (1699). *Wohlerfahrner Leib-Medicus Der Studenten*. Leipzig: Groschuff.

References:

BBAW (2019). Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften; http://www.deutschestextarchiv.de/. [last accessed: 2023-01-19]

Ebert, R. P. (1980). Social and stylistic variation in Early New High German word order: The sentence frame (›Satzrahmen‹). 102. Jahresband, 357–398.

Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A. and Biemann, C. (2016): A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In *Proceedings of the LT4DH workshop at COLING 2016*, Osaka, Japan.

Eckart, W. U. (2021): *Geschichte, Theorie und Ethik der Medizin*, Springer eBook Collection, 9. edition 2021. edn, Springer, Berlin, Heidelberg.

Guthrie, D., B. Allison, W. Liu, L. Guthrie, and Y. Wilks (2006). A closer look at skip-gram modeling. *Proceedings of the Fifth International Conference on Language Resources and Evaluation.*

Hale, J. 2001. A probabilistic Early parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics*.

Koch, P. & Oesterreicher, W. 2007. Schriftlichkeit und kommunikative Distanz. *ZGL 35*, 346-375.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition 106(3)*, 1126–1177.

Levy, R. & Jaeger, F. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems 19*. 849-856.

Light, C. (2011). The information structure of subject extraposition in early new high German. In S. Müller (Ed.), *Proceedings of the HPSG 2011 Conference*

Ortmann, K. and S. Dipper (2022a). Coast (conceptual orality analysis and scoring tool). https://github.com/rubcompling/COASTcoast-conceptual-orality-analysis-and-scoring-tool [last accessed: 2023-01-18]

R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Sapp, C. D. (2014). Extraposition in middle and new high German. *The Journal of Comparative Germanic Linguistics 17(2)*, 129–156.

Speyer, A. (2015a). Auch früher wollte man informieren – Zum Einfluss derInformationsstruktur auf die Syntax in der Geschichte des Deutschen. *Zeitschrift für germanistische Linguistik 43(3),* 485–515.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27(3)*, 379 – 423

Wöllstein, A. (2014). *Topologisches Satzmodell* (2 ed.). Heidelberg: Winter.

Keynote Talk

Interactive adaptation to others' speech: Evidence from patients with neurological conditions

Wolfram Ziegler

EKN – Clinical Neuropsychology Research Group
Institute for Phonetics and Speech Processing, Ludwig Maximilian University Munich

Abstract

One of the remarkable characteristics of spoken language is that it is constantly undergoing change. The speakers of a community continually influence each other in the phonetic nuances of their speech, leading to the emergence and drift of accents and continuous sound change. The plasticity of sound patterns is driven by processes of covert mutual adaptation among those who interact in conversations. It thereby reflects a constant interplay of perceptual and motor processes of spoken language within and across social or regional groups and a malleability of the sensorimotor goals of speech across lifetime. Existing models of speech motor control largely neglect this plasticity by focusing on internal sensorimotor processes that aim to protect the alleged invariance of phonological goals against variability and change in the speaker's motor system. Sensorimotor and neuroanatomic mechanisms of adaptation to others' speech are still underrepresented in research.

I will present three studies investigating the propensity of speakers with dysfunctions in different regions of the brain to align with or adapt to others' speech. With the non-clinical audience of this conference in mind, I will focus more on the experimental paradigms than on neuroanatomic and clinical implications.

The Functional Relation of Prediction and Integration in Discourse Processing: Evidence from Reading Times and ERPs

Mathias Barthel (IDS Mannheim)**, Rosario Tomasello** (FU Berlin)**, & Mingya Liu** (HU Berlin)

The mechanisms of prediction and their effects on language processing have recently been a major focus of psycholinguistic research (e.g. Pickering & Garrod, 2013; Heilbron et al., 2022), investigating prediction on a number of linguistic levels ranging from lexical semantics to discourse level pragmatics (e.g. Nieuwland & Van Berkum, 2006). In many of the experimental studies to date, the mental processes of prediction have been investigated via indirect observations, as the critical measures were mostly taken *after* the critical language input had been presented. Especially in EEG studies, ERPs observed after the critical input have been compared between more vs. less predictable conditions (see Kutas & Federmeier, 2011). If the differences between conditions observed after word onset are caused by predictive processing, i.e., by processes that are executed *before* the presentation of the critical linguistic material, then (i) the effects of these processes should also be observable in advance of the more vs. less predictable words, and (ii) the effects before and after the critical words should be found to be related.

In the present study we investigate the processes of discourse level prediction and their relation to language input processing (Barthel et al., 2022; Barthel et al., 2024). We visually presented short discourses in German including conditional sentences containing either the conditional connective *if* or *only if.* Within the presented discourses, the conditional sentences with these different connectives allowed for more or less predictable discourse continuations. Consider the following example:

Sentence 1:   *Leon besuchte seine Eltern und dachte sich:*
              (Leon visited his parents and thought:)
Sentence 2:   ***Wenn / Nur wenn** die Blumenstrause hubsch sind, bringe ich einen mit.*
              (**If / Only if** the bouquets are pretty, I will take some with me.)
Sentence 3:   *Wie sich zeigte, waren die Blumenstrause nicht hubsch.*
              (As became apparent, the bouquets were not pretty.)
Sentence 4:   *Von denen brachte er **einen / keinen** mit und ging weiter.*
              (Of those he took **one / none** and went on.)

S1 set the scenario context. The conditional sentence S2 contained either *if* or *only if*. After S3, which, in critical trials, negated the antecedent of the conditional in S2, *only if* discourses allowed for a strong prediction of a negated conditional consequent in S4, while bare *if* discourses did not allow for a strongly constrained prediction (Herburger, 2015, 2019). S4 finally either negated the consequent of the conditional in S2, containing the critical quantifier *none*, or confirmed it, containing the quantifier *one*. We thus tested a 2 × 2 design, with two levels of conditional (*if* vs. *only if*) and two levels of discourse continuation, disclosed at and by the critical quantifier (*one* vs. *none*).

In Experiment 1 we gained first indirect evidence for the differences in predictability of the discourse conclusion presented in S4 in a self-paced reading study presenting 108 discourses like the exemplified one to 29 participants (Barthel et al., 2022, Exp. 2). We found negated quantifiers to be read significantly faster in discourses with *only if* conditionals than in discourses with bare *if* conditionals (**Fig. 1**).

In order to gain more direct evidence for the effects being due to predictive processing, the target processes need to be observed in situ, i.e. before the critical discourse continuation is presented (Altmann & Kamide, 1999; Pulvemüller & Grisoni, 2020). Measuring participants' EEG signal, and changing the presentation procedure to even-paced visual presentation, we tested 144 items in 38 subjects in Experiment 2 (Barthel et al., 2024). Analyzing subjects' brain responses across trials before the critical quantifier, we observed a significantly increased Prediction Potential (Grisoni et al., 2017), a slowly building negative brain wave before the critical input, in *only if* scenarios as compared to bare *if* scenarios, indicating that subjects built stronger expectations about the upcoming discourse continuation in *only if* scenarios as compared to *if* scenarios (**Fig. 2A**). This finding supports previous linguistic analyses on the semantics of the two conditional connectives. Additionally, in response to the presentation of the critical quantifier, negative quantifiers (*none*) led to significantly decreased

P300 responses in *only if* scenarios as compared to *if* scenarios (**Fig. 2B**). These results match the previous effect observed in reading times in Exp. 1, giving reason to assume that discourse continuations containing negative quantifiers were easier to be integrated into the discourse representation after they were made predictable in *only if* scenarios as compared to bare *if* scenarios.

Notably, in the constraining discourse contexts containing *only if*, where strong Prediction Potentials were observed, the size of the word-induced P300 component in response to both expected and unexpected discourse continuations was found to be predictable by the size of the Prediction Potential before the critical word (**Fig. 3**). The greater the Prediction Potential before the onset of the critical word, the greater the word-induced P300 component in response to unexpected, positive quantifiers, but the smaller the P300 in response to expected, negative quantifiers. In other words, the stronger the expectations generated by participants in the constraining context condition (*only if*), the greater the word-induced processing effort for the integration of the new information in cases where the input was unexpected (*one*), and the smaller the processing effort for word-induced discourse updating when the input matched the expectations (*none*).

This is the first work observing the Prediction Potential for predictions on the discourse level, i.e. triggered by predictions across sentences. We find that the observed Prediction Potential and the word-induced P300 are functionally related. The correlations of prediction effort or commitment before the discourse continuation, as indicated by the Prediction Potential, and the processing effort for integration of the presented discourse continuation, as indicated by the P300, are taken as evidence for a direct link between pre-activation of expected discourse continuations and reduced (or increased) costs of input processing. Our results thus demonstrate that the mental processes of discourse understanding are functionally interconnected with processes of discourse prediction.

Figure 1. Reading times in Experiment 1.

Figure 2. Prediction Potentials (panel A) and word-induced ERPs (panel B) in Experiment 2.

Figure 3. Correlations of Prediction Potential and P300 in *only if* trials in Exp. 2.
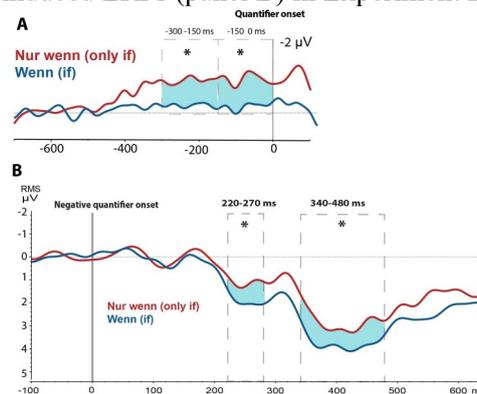
References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs. *Cognition*,

Barthel, M., Tomasello, R., & Liu, M. (2022). Online comprehension of conditionals in context. *Ling Van.*

Barthel, M., Tomasello, R., & Liu, M. (2024). Conditionals in context. *Cognition*.

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *J Neurosci.*

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *PNAS*.

Herburger, E. (2015). Only if: If only we understood it. *Sinn und Bedeutung.*

Herburger, E. (2019). Bare conditionals in the red. *Ling Phil.*

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting. Ann Rev Psych.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love. *J Cogn Neurosci.*

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Beh Brain Sci.*

Pulvermüller, F., & Grisoni, L. (2020). Semantic Prediction in Brain and Mind. *TICS.*

A crosslinguistic survey of the at-issue status of ideophones: Experimental evidence from German, Akan, English and Vietnamese

Kathryn Barnes[1], *Cornelia Ebert*[1], *Kurt Erbach*[1], *Mavis Boateng Asamoah*[2], *Prince Asiedu*[3], *Josiah Nii Ashie Neequaye*[2], *Reginald Duah*[2,3], *Kim Tien Nyugen*[1] *Theresa Stender*[1] *and Yvonne Portele*[1]

1 - Goethe University, Frankfurt am Main, 2 - Humboldt University Berlin, 3 - University of Ghana

Ideophones, such as the German *plitsch-platsch* (splish-splash) or Akan *HWIMM* (moving quickly making a noise of a draft) give a vivid representation of an event and access different modes of perception (hear, see, smell, . . .) (Doke 1935, Dingemanse 2012). We present a crosslinguistic experiment design that tests for the (not-)at-issueness (cf. Potts 2005) of ideophones. We provide initial empirical evidence that in prototypical ideophone languages, such as Akan, ideophones are by default more at-issue than in non-prototypical ideophone languages, such as German and are planning to conduct further research into another prototypical ideophonic language, Vietnamese, and non-prototypical ideophonic language, English, in order to establish whether this is a categorical distinction between ideophonic and non-ideophonic languages.

While there has been much descriptive research on ideophones in individual languages, there is much less crosslinguistic and empirical research into the phenomena. Barnes et al. (2022) was the first experimental work on the at-issue status of ideophones, showing that adverbial ideophones in German are less at-issue than standard adverbials in German. This will therefore be the first crosslinguistic study on the at-issue status of ideophones, with data already having been collected from German and Akan.

We tested the at-issue status of ideophones in Akan (Kwa) (cf. Agyekum 2008) by building on an experimental study of Barnes et al. (2022). We compared Akan adverbial ideophones to ordinary adverbials in matching and mismatching contexts. The experiment has a 2×2-design with the two-level factors CATEGORY (Adverbial vs. Ideophone) and MATCH (Match vs. Mismatch) in a Latin square design. (1) exemplifies the variation of CATEGORY. Whether the target sentence instantiates Match or Mismatch varies depending on its preceding context: e.g., a boy sees a car either going very fast (Match) or slowly (Mismatch) and tells his father: TARGET SENTENCE. Participants rated how well the target sentence matched the context on a scale from 1 = "no match at all" to 5 = "perfect". We presume the mismatch effect (= the decrease in ratings for mismatching contexts as opposed to matching ones) to be stronger if induced by at-issue information, when compared to information that is not at-issue (Barnes et al. 2022)).

In the matching condition, both Akan adverbials ($\bar{x} = 4.35, s = 1.21$) and ideophones ($\bar{x} = 4.51, s = 1.03$) were rated similarly high in quality of match, and in the mismatching condition adverbials ($\bar{x} = 1.99, s = 1.53$) and ideophones ($\bar{x} = 2.13, s = 1.57$) were rated similarly low (see Figure 1a). An ANOVA test was conducted in R on the linear model in (2), and no effect of Category was found (F(1,360.52) 0.0065, p = 0.9358). Moreover, a null model is a better fit in having a lower Bayesian information criterion (BIC=1379.2) than the model in (2) (BIC=1389.8). The experiment thus found no noticeable difference between adverbials and ideophones in Akan. This is in contrast to the results in Barnes et al. (2022), where the mismatch effect was significantly larger for adverbials than for ideophones (see Figure 1b), which we assume indicates that ideophones in Akan are more at-issue than those in German.

These results point to a categorical difference between ideophonic and non-ideophonic languages. In prototypical ideophone languages, such as Akan, ideophones form a large lexical class covering a range of semantic categories and occur frequently in every day speech. In non-prototypical ideophone languages, such as German, however, this lexical class is much smaller and ideophones are rarer in everyday speech. We suggest that the frequency and conventionalization of ideophones in Akan can account for the fact that the ideophones in the language are more at-issue. As German is not a

prototypical ideophone language, ideophones are far less frequently used and much less integrated and conventionalized than in Akan (Dingemanse 2012) and are as such less at-issue. We are currently preparing to replicate this experiment in two further languages, Vietnamese, as an ideophonic language, which uses ideophones similarly to Akan and English as a non-ideophonic language, which uses ideophones similarly to German. This will allow us to establish if there is truly a distinction between ideophonic and non-ideophonic languages and the results of these studies should be ready to present at Linguistic Evidence 2024. This investigation of ideophones across ideophonic and non-ideophonic languages is crucial to better understand the pragmatic behaviour of ideophones from a crosslinguistic perspective.

Agyekum, Kofi (2008). 'The language of Akan ideophones'. *Journal of West African Languages* 35 (1-2).

Barnes, Kathryn Rose, Cornelia Ebert, Robin Hörnig & Theresa Stender (Apr. 2022). 'The at-issue status of ideophones in German: An experimental approach'. *Glossa: a journal of general linguistics* 7 (1): 1–39. DOI: `10.16995/glossa.5827`.

Dingemanse, Mark (2012). 'Advances in the Cross-Linguistic Study of Ideophones'. *Language and Linguistics Compass* 10 (6): 654–672. DOI: `10.1002/lnc3.361`.

Doke, Clement Martyn (1935). *Bantu linguistic terminology*. London: Longmans, Green.

Potts, Christopher (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press. DOI: `10.1093/acprof:oso/9780199273829.001.0001`.

(1) *Kaa bi   pa-a   me ho HWIMM / ho   ntem*
     Car INDEF pass.PST 1SG self IDEO   / quickly
     'A car passed by me HWIMM / quickly.'

(2) lmer(Response Context*Category+(1|Participant)+(1|Item), data=akan, REML=FALSE) where Context is match/mismatch and Category is adverbial/ideophone
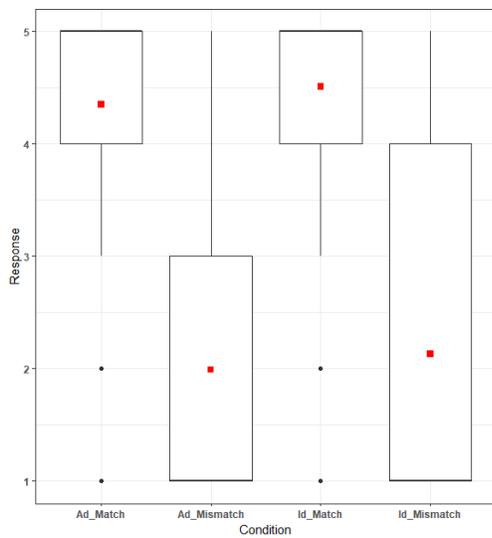
(a) Akan results
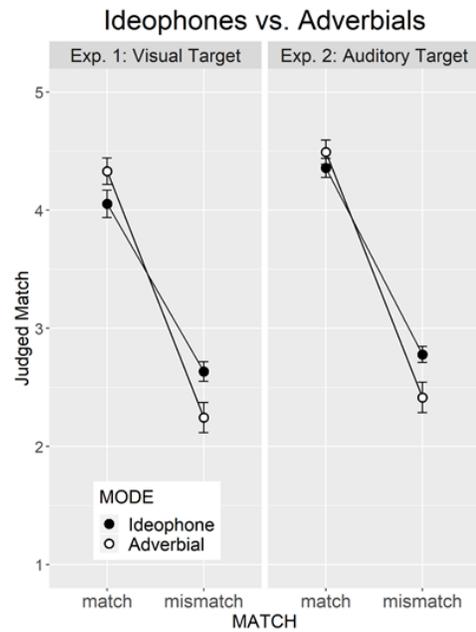
(b) German results (Barnes et al. 2022)

Figure 1: Figure 1. Responses to target sentences where an Adverbial or Ideophone matches or mismatches the preceding context (5 = "perfect", 1 = "no match at all").

A New Technique to Detect Lexical Content in Gaps and Its Application to the Question of Reconstruction in Wh-Movement
Benjamin Bruening and Rebecca Tollan (University of Delaware)

This paper has two goals: (1) To introduce a new experimental technique that, like cross-modal priming, can detect lexical reactivation at gap sites; (2) to apply this technique to the question of whether there is reconstruction of complements and adjuncts of wh-moved NPs.

New technique. Cross-modal lexical priming (Swinney et al. 1979) has proven very useful in showing reactivation of filler material in filler-gap dependencies. However, it is very difficult and time-consuming to program the experiments, and experimental participants have to be run in-person in a laboratory. We sought to come up with an alternative that would be simple and quick to program, and can be run on-line. We have participants read a sentence, and then present them with two words. Their task is to say which of the two words appeared LAST in the sentence. Applied to filler-gap dependencies, the idea is that, if there is reactivation of lexical material in a gap site, we will see interference from a lexical item that is part of a filler, if the gap occurs after the other word choice. Simple example: *Which nurse did the doctor say that the patient was unhappy with? NURSE PATIENT* The word *patient* came last, but if *nurse* is reactivated after *with*, then we may see interference in the form of longer reaction times and lower accuracy compared to a baseline.

Application: The literature on reconstruction in wh-movement currently debates whether complements of moved Ns reconstruct along with the head N. The debate centers around wh-movement examples like *Which corner of John's room was he sitting in?* The point of disagreement is whether the pronoun *he* can refer to *John*. If it cannot, there is presumably reconstruction: *Which corner of John's room was he sitting in [~~which corner of John's room~~]?* If *he* is covalued with *John* in this representation, it is a Condition C violation, since the pronoun c-commands the name in the reconstructed position. The theoretical literature disagrees on the facts: some, like van Riemsdijk and Williams (1981), Fox (1999), Takahashi and Hulsey (2009), claim that the pronoun in questions like this cannot refer to an NP in the complement of the moved N, while others, like Bianchi (1994) and Kuno (2004), say that it can. More recently, experimental work has addressed this same question, with mixed results. Leddon and Lidz (2006), Adger et al. (2017), and Bruening and Al Khalaf (2019) find that experimental participants do permit covaluation, while Stockwell et al. (2021, 2022) claim that they do not.

We enter this debate from a different angle. Rather than looking at Binding Condition C, we ask whether we can find evidence of reactivation of lexical material at the gap site. Unlike previous work, which only compares complements of N to adjuncts of N, we use a three-way comparison: the head N itself, a complement to N, and an adjunct to N:

WH-HEAD: The senators couldn't agree on which discussion question the important subcommittee on foreign relations should issue a statement about__ first.
WH-COMP: The senators couldn't agree on which discussion of a question the important subcommittee on foreign relations should issue a statement about__ first.
WH-ADJ: The senators couldn't agree on which discussion after a question the important subcommittee on foreign relations should issue a statement about__ first.

The lexical item at issue here is *question*. In the HEAD condition, it is part of a compound head noun. In the COMP condition, it is the complement of the head N. In the ADJ condition, it is an

adjunct to the head N. We contrast wh-movement items with baseline items (BASE) that have a similar structure involving mostly the same lexical items, but the N *question* is now not part of a filler-gap dependency:

BASE: The senators couldn't agree on a discussion {**question** HEAD/of a **question** COMP/after a **question**ADJ} when the important subcommittee on foreign relations issued a statement about cooperation.

We expect reactivation to manifest as a longer reaction time to "Which word came last in the sentence?" (*question*, *relations*) for the WH condition as compared to the BASE condition. All researchers agree that the head N reconstructs, so we minimally expect reactivation in the HEAD condition. This contrasts with adjuncts, which everyone agrees do not reconstruct; we therefore expect no reactivation effect for ADJ. How COMP will pattern is the open question.

Experiment. We constructed 12 sets of items like those above and used them as the critical items in a 2 x 3 design (within subjs, within items), crossing Position (HEAD, COMP, ADJ) and Movement (BASE, WH). In the WH items, there were approx. 11 words between the filler and the gap. The second word choice for the task was always part of the embedded subject, which comes between the filler and the gap in the WH items. Presentation order of the two response nouns was counterbalanced across items. The items were distributed among 6 lists, according to Latin Square. We coupled these with 12 filler items, and 6 randomly distributed comprehension questions (participants who answered < half correctly were not included in the analysis). Data from 198 participants recruited through Amazon Mechanical Turk entered the analysis.

Results. The mean response times by condition are shown in Fig. 1. We fitted a 2x3 LMEM to the (log-transformed) RTs. The variable Position was coded with centered Helmert contrasts, comparing HEAD to COMP+ADJ pooled, and COMP and ADJ directly (Movement was sum coded). We found a main effect of Movement: RTs for WH conditions were slower overall than BASE conditions ($t = -2.08$, $p = .037$). Most importantly, we find an interaction of Movement with HEAD vs. COMP+ADJ ($t = -2.34$, $p = .018$), but no interaction of Movement with COMP vs ADJ ($p = .73$). This indicates that reactivation effects for heads are different from complements and adjuncts, but the latter two do *not* differ. There were no significant effects for accuracy (all $ps > .1$). These results indicate that only the head noun is reactivated at the gap site; neither complements nor adjuncts to the moved N are reactivated, since neither gave rise to an interference effect. The experiment also demonstrates the viability of the new technique.
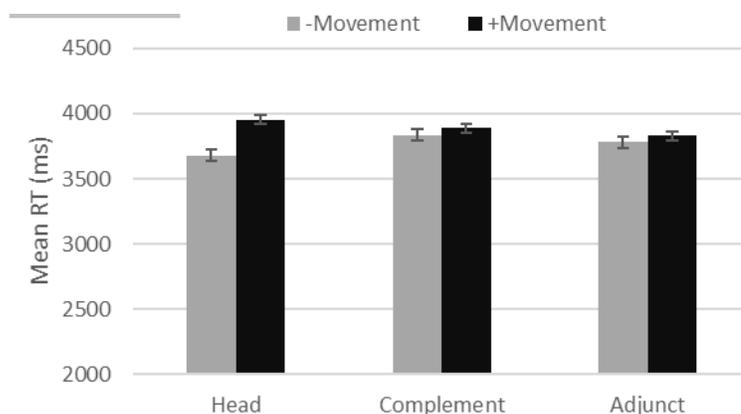
Fig 1. Mean RTs by condition. Error bars indicate +/- 1 S.E.

Acceptability judgment experiments on PPs with gaps in

James Griffiths
University of Tübingen
james.griffiths@uni-tuebingen.de

Craig Sailor
Trinity College Dublin
craig.sailor@tcd.ie

Synopsis. Omission of the object of the preposition in configurations such as (1) is permitted in British (BrE) but not North American (NAE) English. These are *Prepositional Object Gap* (POG) configurations. Griffiths & Sailor (G&S, 2015) conducted the first systematic generative study of POGs and propose that POGs are traces of A-movement (2) (see also G&S 2017, S&G 2019). Stockwell & Schütze (S&S 2019) critique G&S's analysis and voice scepticism about the basic empirical facts as reported by G&S. They present an alternative empirical picture and suggest that POGs would most profitably be analyzed as akin to mainland Germanic R pronouns (albeit, null ones) and/or French "orphan" prepositions.

> (1) He was carrying a box with cups in (it).
>
> (2) He was carrying [a box]$_1$ with cups in t_1.

The differing empirical pictures sketched by G&S and S&S are both based on acceptability judgment data collected informally from a handful of (different) BrE speakers. Although employing this informal data collection method is typically unproblematic for generative syntax research (Sprouse & Almeida 2012, 2017, 2018, Sprouse et al. 2013), we contend that it is problematic on this occasion, for reasons related to small effect sizes, the potential for regiolectal variation on POGs within BrE, and the likelihood of high idiolectal variation (which is common for omission phenomena; see e.g., Thoms 2019 for VP ellipsis). In an effort to create a reliable empirical foundation on which competing analyses of POG configurations can be assessed, we conducted five large-scale acceptability judgment studies using formal experimental methods (e.g., large sample sizes, modelling idiolectal variation as a random effect). Each experiment addresses an issue related to POGs, most often a point of empirical contention between G&S and S&S. Our results support G&S on some points of debate and support S&S on others. We conclude that, overall, a revised version of G&S's analysis currently seems best-placed to capture the properties associated with POGs. In this abstract, we summarize only experiment 1 (287 BrE speakers, 218 NAE speakers), and briefly describe the findings of experiments 2 to 5. The other four experiments will be discussed during the talk itself.

Background to experiment 1. G&S and S&S agree that POGs are licensed only with locative spatial prepositions. But which locative spatial Ps permit POGs? This question cannot be answered by obtaining judgments from only BrE speakers, as all English varieties permit a wide variety of phrases to be omitted after locative spatial Ps (3-4). Therefore, one must collect judgments from both BrE and NAE speakers. If the BrE speakers accept the configuration but the NAE speakers reject it, then one has discovered a genuine POG configuration. G&S applied this methodology to Svenonius' (2010) taxonomy of locative spatial prepositions; see (5) for their list of POG-licensing Ps. S&S report different BrE judgments to G&S: for S&S, POGs are fully acceptable only with *in* and *on*. S&S's judgments come from speakers of Standard Southern British English, whereas G&S's come from other regions of England (London, Cambridge, Bristol, East Midlands). One possible explanation for the different judgments could therefore be that there is regiolectal variation within BrE regarding which Ps license POGs. The purpose of Exp1 was therefore to determine which Ps are POG-licensors.

> (3) That film was just a remake with the plot taken away ([$_{PathP}$ from it]).
>
> (4) Nils looked over the snow drift. The frozen fjord beyond ([$_{DP}$ it]) was dotted with seals.
>
> (5) **projective:** above, behind, below, beyond
> **extended:** across, along, around, over
> **bounded:** between
> **particle:** down, in, on, up
>
> (6) Mine's the mug with the coaster **under** (it). [extended]
>
> (7) I think this crowd has some undercover police officers **among** (it). [bounded]
>
> (8) Look – this table here has stools **beneath** (it). Let's sit here! [projective]
>
> (9) Mine's the house with a bus-stop in front **of** (it). [non-locative]

Procedure. The five experiments (each 1-7 Likert scale) were conducted between April 2020 and April 2023. Participants were sourced via *Prolific* (online, unsupervised). BrE and NAE speakers completed 2 different versions of each experiment, where stimuli differed across these versions only regarding salient

natiolectal variants, for instance in spelling conventions (e.g., *colour* vs. *color*) and lexis (*flat* vs. *apartment*). Each experiment contained the *standard fillers* for English from Gerbrich et al. (2019). Comparison across the BrE and NAE groups (henceforth COHORT) was made possible by z-scoring a participant's raw ratings for test items over her ratings for the standard fillers. We fit linear mixed effect models (LMMs) using R's *lmer*. When required, we conducted post-hoc Tukey-adjusted t-tests of the best fitting model's estimated marginal means (using R's *emmeans*). Regarding experiment 1, it had a 2 x 2 x 6 design. The
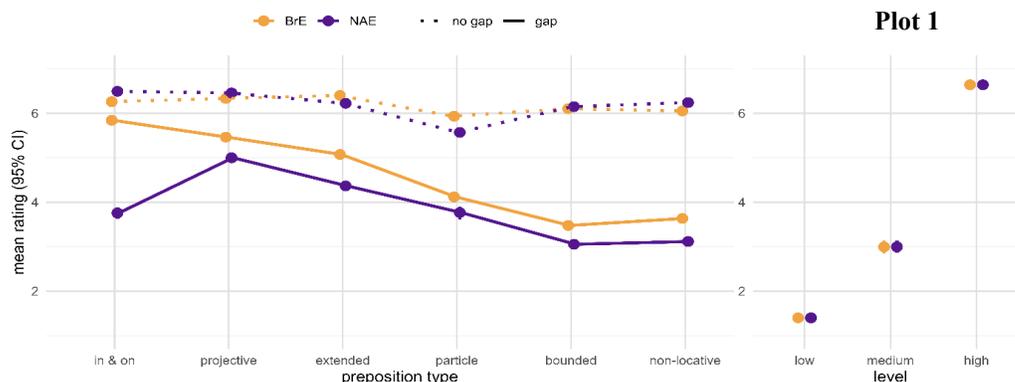
Plot 1

factor PREPTYPE had 6 levels: each of the classes listed in (6), *in&on*, and *non-locative*. The factor GAP (*yes*, *no*) compared P-object omission and retention. See (7-9) for example test items. Plot 1 shows the mean ratings for test items (left) and standard fillers (right).

Because the best fitting LMM of the z-scored results (10) returned significant interactions, we conducted post-hoc comparisons. **Results:** As expected, in the *no-gap* condition, differences across COHORT (BrE vs. NAE) and PREPTYPE are insignificant. In the *gap* condition, a small (~0.3 on the 1-7 scale) but statistically significant difference between the BrE

> (10) score ~ cohort * gap * preptype
> + (cohort | item) + (1 | subject)

and NAE groups was observed at each level of PREPTYPE aside from *in&on*, where the difference is highly significant (t = 13.55, p > 0.01). BrE speakers treat gap configurations with *in* and *on* as on par with those with projective Ps (t = 1.86, p = 0.42), whereas NAE speakers treat *in* and *on* like just another particle (*in&on* vs. *particle*: t = 0.09, p = 0.99). Other tests revealed no regiolectal variation within BrE. **Discussion:** S&S are correct that POGs are only licensed with *in* and *on*. P-omission with other Ps looks like *ground omission*: the acceptability cline aligns with Svenonius's (2010) claim that that ground omission is most acceptable with projective Ps, less acceptable with extended Ps, and unacceptable – relatively so, it transpires – with bounded and non-locative Ps. G&S's claim that the Ps in (5) are POG-licensors likely arises from a misinterpretation of the upward shift of the acceptability cline on Plot 1 for BrE speakers. That the 'worst' cases of P-omission receive 'medium' scores in absolute terms (i.e., when compared to the standard fillers) is unsurprising: these sentences involve only omission of a contextually-recoverable pronoun, not the garbled syntax the 'low' fillers (e.g., *Historians wondering what cause is disappear civilization.*).

G&S's A-movement analysis predicts that POG configurations show freezing effects, and that POGs can be licensed only in superstructures containing possessive lexemes such as *have* and *with*. The results of exp2, exp4, and exp5 bear these predictions out. Conversely, S&S suggest that figure extraction causes unacceptability in POG configurations. Exp3 shows that such extraction causes mild (and statistically insignificant) degradation in acceptability; a result that conflicting with findings reported very recently in Stockwell et al. 2023 and which potentially undermines the analysis offered therein. In the talk, we outline the analytical possibilities going forward.

Sel. Refs. Griffiths, J. & C. Sailor. 2015. Prepositional object gaps in British English. *Linguistics in the Netherlands* 32, 63-74. • Stockwell, R. & C. Schütze. 2019. Objectless locative prepositions in British English. *Proceedings of the Linguistic Society of America* 4 (1): 48:1–15. • Stockwell, R., Himmelreich, A. & C. Schütze. 2023. An extraction restriction with complement-less prepositions in British English but not dialectal German. To appear in *Proceedings of 35th Comparative Germanic Syntax Workshop*. Lang. Sci. Press • *Proceedings of the Linguistic Society of America* 4 (1): 48:1–15. • Svenonius, P. 2010. Spatial P in English. In G. Cinque & L. Rizzi (eds.), *Mapping spatial PPs: The cartography of syntactic structures*, 127–160. OUP.

Experimental findings for a cross-modal account of dynamic binding in gesture-speech interaction

Cornelia Ebert, Goethe-Universität Frankfurt, ebert@lingua.uni-frankfurt.de
Kurt Erbach, Goethe-Universität Frankfurt, Universität des Saarlandes, erbach@hhu.de
Magnus Poppe, Goethe-Universität Frankfurt, magnus.poppe@web.de

We report experimental results of two experiments on pronoun and presupposition binding across modalities. We show that (1) ordinary pronouns (in the spoken/written domain) can be dynamically bound to gesturally introduced discourse referents and (2) that presuppositions induced by presupposition triggers in the spoken/written domain (as e.g. *again* or *too*) can be bound and satisfied by propositions that have been introduced in the gestural domain.

Background. Ebert, Ebert & Hörnig (2020) (based on Ebert & Ebert 2014) suggest a formal framework for gesture semantics where certain iconic and pointing co-speech gestures introduce discourse referents that can serve as antecedents in anaphoric reference. Crucially, this necessitates a unidimensional dynamic system that allows for binding effects across dimensions and, in this case, modalities. Based on the dynamic system of Anderbois et al. (2015) that can handle binding effects across dimensions (with appositives introducing non-at-issue material), Ebert, Ebert & Hörnig (2020) suggest that gestures behave and can be handled on a par with appositives since both contribute propositional non-at-issue information by default. Furthermore and crucially, pointing gestures and iconic gestures introduce discourse referents for rigid designators as their core 'lexical' meaning, i.e. when a pointing gesture is performed this triggers the introduction of a discourse referent that is identified with the rigid concept of the gesture referent. This discourse referent (DR) can then be anaphorically picked up by a pronoun in later discourse. Importantly, in this dynamic semantic framework it is predicted that gesturally introduced DRs allow for anaphoric binding across dimensions, i.e. gesturally introduced DRs can be referents of speech pronouns.

While the introduction of DRs by gesture has been claimed and implemented in the formal system of Ebert, Ebert & Hörnig (2020), this has not been experimentally demonstrated. Here we show that dynamic binding across dimensions can be made with respect to both pronouns and presupposition triggers. It can be shown that gesture can introduce discourse referents which can be picked up by a speech pronoun later-on (as illustrated in (1)). Furthermore, gestures can introduce propositional content that can serve as presupposition binders for presupposition triggers in speech (see ex. (2)).

In the constructed example (1a), the pointing co-speech gesture in the form of extending an index-finger towards a piece of cake as opposed to other baked goods is assumed to introduce a DR for the gesture concept for the referent of said piece of cake and allows it to be bound to the pronoun "it" in the hypothetical follow-up (1b). If (1a) had included a hand-over-stomach gesture to indicate having eaten (1c) and crucially not introducing a DR, then presumably "it" in (1b) cannot be bound. In our experiment, we add as a control (1d) as a possible follow-up. While it seems unlikely that (1a) would be followed by (1d) where a confirmatory response is given that ignores the pointing gesture, (1d) could presumably follow (1c) where no specific referent is indicated. Similarly with presupposition triggers like *again*, the jogging gesture in (2a) - adding the propositional content that Paul was jogging (when the speaker met him) - is assumed to be an additional propositional information given in the visual modality via gesture that can serve to satisfy the presupposition that is triggered by *again* in (2b), namely that Paul went jogging before. In the absence of such a gesture the presupposition triggered by *again* would not be satisfied, at least under the assumption that people don't commonly meet while jogging and hence such a proposition cannot be accommodated. Conversely, a follow up like (2d) is presumably odd following a jogging gesture (2a) under the assumption that people do not jog in cafes, but following (2c) ought to be fine assuming people often meet in cafés.

(1)   a. Have you <u>eaten</u>[pointing to a piece of cake]?   (2)   a. Yesterday <u>I met Paul</u>[jogging gesture]

      b. It was too sweet for me.   b. He went jogging again today.

c. Have you eaten[placing hand over stomach]?
d. Yeah, a few too many cookies.

c. Yesterday I met Paul[pointing backwards]
d. Was it in the café again?

Experiments. Two experiments were designed in German to test the contrasts demonstrated in (1) and (2). Given the similarity in contrasts, albeit distinct form of gesture and anaphora, the designs were complementary and allowed each to be used as filler for the other. Both experiments had two factors each with two levels, yielding two treatment factor levels (felicitous or infelicitous). Experiment 1 had the levels GESTURE (pointing (1a) or iconic (1c)) and to-be-bound-PRONOUN (present (1b) or absent (1d)), and Experiment 2 the levels: GESTURE (pointing (2c) or iconic (2a)) and to-be-bound-PRESUPPOSITION (present (2b) or absent (2d)). Each participant participated in each of the within subject conditions in (3)-(4). The minimal pairs resembling (1) and (2) were distributed across four groups of participants. We recruited 60 native German speaking participants via Prolific, following the 2x2 repeated-measures design in Brysbaert (2019). In a variation of the covered-box task (cf. Fanslow et al. 2019), the sentence pairs were presented with the context, e.g. (1a) presented in video form, and the follow-up, e.g. (1b), being presented in written form as one choice in a pair of alternatives, the other being 'covered' (lit. "[geschwärzt]" ('redacted')). Participants were instructed that one of the alternatives was a reasonable follow-up to the context and the other wasn't, and they should select whichever they believe to be more reasonable.

(3)   a. GESTURE—pointing + PRONOUN—present   (felicitous, (1a)+(1b))
        b. GESTURE—pointing + PRONOUN—absent   (infelicitous (1a)+(1d))
        c. GESTURE—iconic + PRONOUN—present   (infelicitous (1c)+(1b))
        d. GESTURE—iconic + PRONOUN—absent   (felicitous (1c)+(1d))
(4)   a. GESTURE—iconic + PRESUPPOSITION—present   (felicitous, (1a)+(1b))
        b. GESTURE—iconic + PRESUPPOSITION—absent   (infelicitous (1a)+(1d))
        c. GESTURE—pointing + PRESUPPOSITION—present   (infelicitous (1c)+(1b))
        d. GESTURE—pointing + PRESUPPOSITION—absent   (felicitous (1c)+(1d))

Results. Starting with the pronoun experiment, for items with pointing gestures, follow-ups with pronouns meant to be bound to the gesture DR were largely accepted (3a, n=115), and, surprisingly, those without such a pronoun were accepted nearly as much (3b, n=105). As expected, with iconic gestures, pronouns that could not be bound to a DR were not accepted (3c, n=63) unlike those with other continuations (3d, n=133). In the presupposition experiment, items with iconic gestures plus follow-ups with presuppositions meant to be bound to iconic gestures were largely accepted (4a, n=129), and those with such presuppositions absent less so (4b, n=86). As expected, the same items albeit with pointing gestures plus follow-ups with to-be-bound-presuppositions were generally not accepted (4c, n=72) and those without were accepted (4d, n=138). Responses for each experiment were analyzed with a 2x2 ANOVA with the within-subject factors. A significant interaction of GESTURE+PRONOUN was found $(F(1,716)$ 39.54, $p < 0.001$, η2 0.055) as well as GESTURE+PRESUPPOSITION $(F(1,716)$ 75.32, $p < 0.001$, η2 0.105)—i.e. the null hypothesis of no interaction between gesture and anaphora is unlikely.

Discussion. There are two key contrasts targeted in this study: (i) when pronouns have gesture DR vs. when they have no obvious referent (cf. (1a+1b) vs. (1c+1b) and (ii) when presupposition triggers can be bound to a gesture-introduced proposition vs. when they have no obvious referent (cf. (2a+2b) vs. (2c+2b). In both contrasts the former has been assumed to be felicitous, and the latter not, and the interaction between gesture and binding found in the experiments support these assumptions.

References. Brysbaert, M. 2019 How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, 2(1): 16, pp. 1–38. DOI: https://doi.org/10.5334/joc.72 ● Anderbois, Scott & Brasoveanu, Adrian & Henderson, Robert. 2013. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32(1). 93–138. ● Ebert, Cornelia & Ebert, Christian. 2014. Gestures, demonstratives, and the attributive/referential distinction. *Talk at Semantics and Philosophy in Europe 7, Berlin: ZAS.* ● Ebert, Christian & Ebert, Cornelia & Hörnig, Robin. 2020. Demonstratives as dimension shifters. In Franke, Michael & Kompa, Nikola & Liu, Mingya & Mueller, Jutta L. & Schwab, Juliane (eds.), *Proceedings of Sinn und Bedeutung* 24. 161–178. ● 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).

Deontic priority – converging evidence for a universal in modal semantics

Summary: Theoretical and typological research on modality has yielded various fruitful hypotheses on the range of crosslinguistic variation and possible linguistic universals in the modal domain (Nauze 2008, Rullmann & Matthewson 2018, Steinert-Threlkeld et al. 2022, a.o.). We present novel evidence contributing to this research program. Based on a crosslinguistic fieldwork study, we propose a universal in the domain of negative modality, which we label Deontic Priority (DP): If a language has a lexicalized form for impossibility, it has a lexicalized form for *deontic* impossibility. We explore the idea that this generalization can be captured in terms of a utility bias in an informativeness/complexity trade-off model (see e.g. Imel & Steinert-Threlkeld 2022), supported by a computational modeling study and experimental data.

Crosslinguistic data: In our crosslinguistic study, we adapted Vander Klok's (2021) revised modal questionnaire and added contexts for eliciting negative modality (non-necessity and impossibility) expressions with epistemic, deontic, teleological and (pure) circumstantial flavor. In our sample of 24 languages, we observed that non-necessity is always realized as a combination of morphologically overt negation and a modal marker, while impossibility is lexicalized to some extent in several languages. Among these, we identify two patterns: either a language uses a lexicalized impossibility modal across all flavors, or only in the deontic flavor (which is the more common case in our sample). In (1) and (2) below, Basque exemplifies the first pattern, Hausa the second. In (1) we illustrate lexicalized deontic impossibility in both languages and contrast it with epistemic impossibility (2), where only Basque allows for the use of the impossibility modal.

(1) Deontic impossibility: You are going to visit your friend in the hospital. When you enter into the hospital, you stop at the information desk to inquire what room your friend is in. But the woman at the information desk tells you that you can't visit your friend now because it's already 8pm. She says: "I'm sorry, the hospital regulations say that... **Visitors mustn't stay after 6pm**."

   a. Bisitariak **ezin** dira 6 ostean gelditu.   (Basque)
       Visitors   MOD($\neg\Diamond$) be.3pl.ind 6 after   stay

   b. **Kada** maziyarta su wuce ƙarfe 6 na yamma.   (Hausa)
       MOD($\neg\Diamond$) visitors   3pl.prosp stay   hour 6 pm

(2) Epistemic impossibility: Ben goes swimming every day. Ben is not obliged or required to go swimming; it is just a habit of his. It is now time for Ben to be swimming, so... **Ben can't be at home**.

   a. Benat **ezin** da etxean egon.   (Basque)
       Ben   MOD($\neg\Diamond$) aux home   be

   b. **Ba** zai **yiwu** Ben ya kasance a gida **ba**.   (Hausa)
       NEG 3sg.fut MOD($\Diamond$) Ben 3sg be   at house NEG

   c. (# **Kada** Ben ya kasance a gida.)

Table 1 summarizes the two patterns and lists the languages in our sample that exhibit them:

	Non-necessity	**Impossibility**		
	(any flavor)	epistemic	deontic	other root flavors
Basque, Turkish	✕	✓	✓	✓
Hausa, Hebrew, Thai Hungarian, Russian, (Kîîtharaka)	✕	✕	✓	✕

Table 1: ✓ means the meaning is lexicalized, ✕ means it is not

Deontic Priority: As Table 1 illustrates, in our sample there is no language that lexicalizes impossibility but does not lexicalize deontic impossibility. This observation motivates the Deontic Priority (DP) generalization: If a language lexicalizes any impossibilities, then it lexicalizes deontic impossibility.

Computational modeling: The DP generalization suggests that a theory of modal lexicalization must capture the contrast between flavors (deontic vs. others). In a computational modeling study, we explored how Imel & Steinert-Threlkeld's (2022) existing model in terms of a complexity/informativeness trade-off can be extended to capture the contrast. In particular, we explored if asymmetries in the communicative utility function yield a picture in which bias correlates with the optimality of languages that satisfy the DP generalization. The results in Fig. 1 suggest that optimizing the trade-off between simplicity and informativeness, in the presence of a bias for the deontic flavor, could explain the DP generalization. In particular, given a utility bias for deontic flavor, all languages that are closest to the optimal trade-off between complexity and informativeness (depicted as a black line) are languages that adhere to the DP generalization (depicted in blue triangles, as opposed to red circles

Figure 1: Complexity vs. communicative cost

that represent non-DP languages). This result contributes to a growing body of literature suggesting that augmenting the basic efficient communication analyses with certain biases may be necessary to account for semantic typology in certain domains (e.g. Chen et al. 2022, Zaslavsky et al. 2021).

Experiment: The modeling results raise the issue to what extent the utility bias for the deontic flavor is empirically grounded. We tested the hypothesis that the bias relates to particularly high pressure to communicate successfully in the case of deontic impossibility, since its communicative function (prohibition) is to prevent negative and potentially dangerous situations. In a 2×3 design, we crossed the factors FORCE (levels: impossibility, possibility) and FLAVOR (levels: deontic, circumstantial, epistemic). Modal flavor was disambiguated by means of designated lexical items (*allowed/ not allowed* for deontic, *able/ not able* for circumstantial and *it's possible / not possible* for epistemic possibility and impossibility, respectively). We constructed 18 items of the form in (3) in 6 conditions. The items were distributed across 6 lists in a Latin square

Figure 2: importance ratings

design and intermixed with fillers. The participants (64 English native speakers) were asked to rate how important it is that the content of a modal sentence as conveyed by the speaker was correctly understood by the addressee on a scale from 1 (not important) to 7 (very important).

(3)　Paul says to John: Max is not allowed to play the piano.　　　　　[condition: deontic impossibility]
　　　Question: How important is it that John heard correctly what Paul said?

The results are depicted in Fig. 2. An ordinal logistic mixed effects model fitted to the data reveals a significant effect of FLAVOR: deontic modal sentences were rated as more important than circumstantial (β = -2.88, 95% CI [-3.78, -1.98], p < .001) and epistemic sentences (β = -4.37, 95% CI [-5.21, -3.53], p < .001). This result provides motivation for a utility bias for deontic modality as assumed in our modeling study. Also, we observe a significant main effect of FORCE (β = -1.82, 95% CI [-2.6, -1.04], p < .001). If we focus on deontic vs. epistemic comparison, we observe a FORCE-FLAVOR interaction (β = 1.51, 95% CI [0.57, 2.44], p < .01). We discuss interpretations of this interaction and their typological predictions in the presentation.

Selected references: ● Chen, Futrell & Mahowald (2022). Investigating information-theoretic properties... SIGTYP 4. ● Imel & Steinert-Threlkeld (2022). Modal semantic universals optimize... SALT 32 ● Nauze (2008): *Modality in typological perspective*. UvA diss. ● Vander Klok (2021). Revised modal questionnaire for cross-linguistic use. Open access field linguistics tool.

A Bounded Rationality Account of Constituent Order in SOV languages

Sidharth Ranjan and Titus von der Malsburg

University of Stuttgart

{sidharth.ranjan, titus.von-der-malsburg}@ling.uni-stuttgart.de

Languages with flexible word order provide numerous ways to express an idea (see Example 1 below). Yet, only *one* is ultimately produced, among many available expressions. What drives this choice, and what cognitive mechanisms underlie this decision-making process? The principle of dependency length minimization (DLM) has been very influential in explaining such choices across languages, primarily driven by efficiency considerations stemming from limited memory capacity (Futrell et al., 2015). According to DLM, language processing system strives to maintain close proximity between syntactically related words within a sentence. However, the extent to which DLM is employed in a given language is not well understood. Additionally, the cognitive mechanisms governing the minimization process is unclear. Inspired by recent work by Ranjan and von der Malsburg (2023), we test the hypothesis that placing a short preverbal constituent (possibly the shortest) immediately adjacent to the main verb explains preverbal constituent ordering decisions better than global minimization of dependency length in SOV languages. This "least-effort" strategy concurrently reduces the length of all preverbal dependencies connected to the main-verb without the need to simulate through entire search space of possible constituent orders, thereby facilitating efficient communication. This approach aligns with the concept of bounded rationality in decision-making, favoring *fast-but-frugal* heuristics over exhaustive searches for optimal solutions (Gigerenzer and Goldstein, 1996; Simon, 1982).

(1)
a. C_1 C_2 C_3 C_4 V
b. C_3 C_2 C_1 C_4 V
c. C_4 C_1 C_2 C_3 V
d. C_1 C_2 C_4 C_3 V

Design. Our dataset comprises sentences from all head-final (SOV) languages that are prominently represented in the Universal Dependency Treebank (Zeman et al., 2022, version 2.11), containing at least 2000 sentences belonging to projective dependency trees with a minimum of two preverbal constituents. Our dataset includes *Basque, Hindi, Japanese, Korean, Latin, Persian, and Turkish*. For each reference sentence in the corpus, we created at most 120 counterfactual variants by randomly permuting the preverbal constituents whose head was an immediate child of the root verb in the dependency tree. Conceptually, Example 1 illustrates this generation process for an SOV sentence containing four preverbal constituents C_i. Sentences that originally appeared in the corpus are considered the more preferred syntactic choice compared to those that are counterfactually generated. Next, we examined the distributions of the length of preverbal constituents and the total dependency lengths in both reference and variant sentences. The length of a constituent was calculated by counting the number of words within it, and the dependency length of a sentence by summing the distances (in terms of words) between all head-dependent pairs in a dependency tree. We then deployed measures (based on least-effort strategy) in a logistic regression classifier to distinguish reference sentences from the alternative variants.

Results. Firstly, we examined the length of all preverbal constituents within the corpus reference sentences. The global DLM would predict a gradual decrease in the lengths of preverbal constituents as they approach the main-verb. In contrast, the least-effort strategy would predict optimization primarily focused on the preverbal constituent next to the main-verb. Figure 1 presents our analysis. The plots for Persian and Hindi appropriately depict the expected pattern for the least-effort strategy. The average constituent length remains relatively consistent across positions until the position next to the main verb, where the length suddenly decreases. Turkish and Korean also follow a similar pattern for the most part, but with a deviation observed in sentences with two preverbal constituents, potentially indicating low priority in optimizing dependency length due to the low memory pressure associated with these sentences. The plots for Basque and Japanese exhibit a more gradual decrease in average constituent length as one moves towards the main verb, with the on-average shortest constituent positioned next to the main verb. This observation may indicate certain language-specific properties at play. Interestingly, in Latin, while the shortest preverbal constituent is commonly positioned in the first position, the constituent at the last preverbal position is also short. Further linguistic analysis on Latin text revealed that many sentences begin with shortest constituent (single-word conjunction: "et") indicating stylistic preference in the corpus. Yet, the pressure to minimize the length of the preverbal constituent next to the main verb is consistent with other languages. In summary, our analysis in Figure 1 suggests that naturally occurring corpus sentences across the SOV languages exhibit a prefer-

ence for optimizing the length of the preverbal constituent next to the main verb or at least prioritize it.

Secondly, we compared the average dependency length of corpus reference sentences to four alternative variants with different constituent orders: (i) random order of preverbal constituents, (ii) least-effort order, where the shortest constituent was simply moved next to the main-verb in the random ordering obtained in (i) previously, (iii) ascending order of preverbal constituents leading to maximal dependency length in the sentence, and finally, (iv) descending order of preverbal constituent leading to minimal dependency length in the sentence. Our results in Figure 2 show

Figure 1: Average constituent length of preverbal constituents within corpus sentences with only-2 to only-5 constituents (see the legend of the plot)

that the dependency length of sentences in the corpus generally tracks the dependency length of the least-effort solution across SOV languages, indicating the efficacy of least-effort strategy in explaining the naturally occurring constituent orders in SOV languages. Additionally, the plot also suggests that the tendency for corpus sentences to align with the least-effort solution becomes more pronounced as the number of preverbal constituents – and therefore cognitive load due to vast search space – goes up.

Finally, if speakers employ the hypothesized least-effort strategy, we should be able to predict whether a sentence is a corpus reference sentence or a randomly generated variant by examining the length of the preverbal constituent next to the main-verb (CL Last). Further, these predictions should be *better* than those obtained when total dependency length (Total DL) is used as the predictor. To test these predictions, we deployed logistic regression model to identify corpus reference sentences (amidst variants) using 'Total DL' and 'CL Last' as predictors for each of the investigated SOV languages. Consistent with our

Figure 2: Average total dependency length across various reference-variant sentence types; Grey shaded region indicates the entire spectrum of possible dependency lengths (Minimal to Maximal values) of sentences

hypothesis, the results indicate that 'CL Last' was significantly better ($p < 0.001$ using McNemar's test) at predicting reference sentences than 'Total DL' in terms of classification accuracy (% of correctly predicted reference sentences) for all the languages except for Basque and Japanese. Furthermore, adding 'CL Last' feature over a baseline model containing only 'Total DL' feature, induced a significant increase in the classification accuracy ($p < 0.001$ using McNemar's test) for all SOV languages, including Basque and Japanese.

Conclusion. Overall, our results indicate that speakers of SOV languages minimize dependency length when making constituent ordering decisions. This minimization, however, is achieved through a least-effort strategy that considers only a small fraction of the overall search space of constituent orders, presumably to preserve resources within the constraints of bounded rationality (Gigerenzer and Goldstein, 1996).

References. ● Futrell et al (2015, PNAS) ● Zeman et al (2022, UD@Consortium) ● Simon (1982, MIT Press) ● Ranjan & von der Malsburg (2023, CogSci) ● Gigerenzer & Goldstein (1996, PsycReview)

Auditory phonotactic wellformedness intuitions depend on the nativeness of a speaker's pronunciation
Leonardo Piot, Thierry Nazzi & Natalie Boll-Avetisyan

Adults possess gradient probabilistic knowledge of their language-specific phonotactics (i.e., whether or not and if so with what probability phonemes co-occur in their language(s)), and use this knowledge when judging the acceptability of nonwords: novel sequences of phonemes are more accepted as potential words of their language if those sound sequences are attested or highly frequent in their native language than when they are infrequent or unattested (e.g. Vitevitch & Luce 1999; Bailey & Hahn 2001; Needle, Pierrehumbert & Hay, 2017). While phonotactic wellformedness intuitions have been shown to be robust, and because phonotactics operate on abstract phonological segments, it is unclear whether acoustic properties, such native versus foreign-accented pronunciations, modulates phonotactic wellformedness intuitions. Since pronunciation has been shown to influence the degree of activation of other lexical and phonological processes, such as bilinguals' parallel activation of their lexicons (e.g. Lagrou et al., 2013); or bilinguals' phoneme categorization (Gonzales & Lotto, 2013), it could be posited that it also influences the degree of native-like phonotactic activation. Hence, the present study set out to test whether phonotactic wellformedness intuitions depend on whether speech is produced by a native or non-native speaker. Based on previous findings, we expected to replicate the well-studied effect of gradient phonotactic wellformedness intuitions. Furthermore, we assumed that listeners would overall give lower ratings when listening to a non-native speaker than when listening to a native speaker, but we did not have specific hypotheses about the interaction of pronunciation and phonotactic intuitions. We reasoned that that if pronunciation played a role in the activation of native phonotactic processing, we would find that the association between gradient phonotactics and participants' judgments is stronger when the participants had to judge nonwords pronounced by a speaker of their native language than nonwords pronounced by a speaker of an unfamiliar language.

Using a web-based meta-linguistic wellformedness judgement task, German (N = 23) and French (N = 25) monolingual adults listened to nonwords containing word-initial clusters varying in phonotactic probability in their languages, and were asked to rate how acceptable these nonwords would be as new words of their language. In total, there were 64 unique nonwords that were each pronounced once by a French and once by a German monolingual speaker.

For data analysis, we used linear mixed-effects models (one for each language group, see outputs in Table 1). The results were that, as expected, participants' judgments were significantly predicted by language-specific phonotactic probability: the higher the probability of the word-initial cluster in the nonword, the more acceptable it was rated. Pronunciation, which also predicted participants' judgments (i.e. native pronunciations were rated higher than non-native pronunciations), interacted with phonotactic probability: the association between phonotactic probability and participants' judgments was significantly larger when the stimuli were pronounced by a speaker of the participants' native language compared to when they were pronounced by the speaker of the other language (Figure 1).

These results add to evidence that monolingual adults possess gradient knowledge of the phonotactics of their language, and that they use this knowledge when encountering new words. Interestingly, we found that the nativeness of a speaker's pronunciation modulated the listeners' gradient phonotactic intuitions. Thus, even when having to process the same sequences of phonemes, monolingual listeners appear to apply to a greater extent their gradient phonotactic knowledge when listening to a native pronunciation compared to when they are listening to a foreign-accented pronunciation. This novel finding implies that pronunciation properties interfere with the degree of phonotactic activation. Whether phonotactic information is partially disregarded when processing foreign pronunciation, or whether phonotactic activation is reduced in the context of a generally more demanding phonological processing remains an open question, to be addressed in further studies. Nevertheless, this finding contributes to our understanding of the mechanisms involved when adult speakers process native and foreign-accented speech.

Table 1: Results of the French (left panel) and German (right panel) linear mixed-effect models.

Formula : Rating ~ Language-specific Phonotactic Probability X Pronunication + (1| Participant) + (1| Item)

| Predictors | Model 1 : French Monolinguals | | | Model 2 : German monolinguals | | |
	Estimates	CI	p	Estimates	CI	p
(Intercept)	2.98	2.61 – 3.35	**<0.001**	3.27	2.96 – 3.58	**<0.001**
Phonotac. Prob.	0.39	0.24 – 0.53	**<0.001**	0.31	0.18 – 0.44	**<0.001**
Pronunciation	-1.35	-1.44 – -1.26	**<0.001**	1.00	0.89 – 1.10	**<0.001**
Phonotac x Pronun	-0.29	-0.38 – -0.20	**<0.001**	0.15	0.04 – 0.25	**0.006**
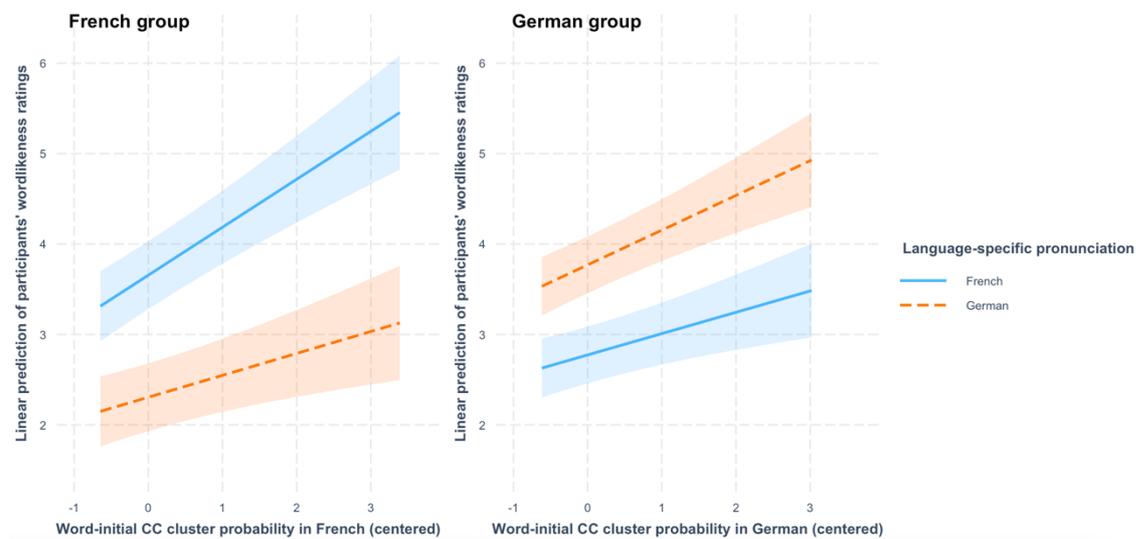
Figure 1: Linear prediction of wordlikeness ratings as a function of phonotactic probability in each pronunciation condition for the French group (left panel) and the German group (right panel). Error bands represent confidence intervals.

References:

Bailey, Todd M., & Ulricke Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods?. *Journal of Memory and Language* 44(4), 568-591.

Gonzales, K., & Lotto, A. J. (2013). A bafri, un pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological science*, 24(11), 2135-2142.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, 16(3), 508-517.

Needle, J. M., Pierrehumbert, J. B., & Hay, J. B. (2022). Phonotactic and morphological effects in the acceptability of pseudowords. *Morphological diversity and linguistic cognition*, 79-112.

Vitevitch, Michael S., & Paul A. Luce. 1999. Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language* 40, 374–408.

Constraints on word exchanges during noisy-channel inference

Markus Bader & Michael Meng
Goethe University Frankfurt & Merseburg University of Applied Sciences

According to the Noisy Channel Model of Gibson, Bergen, and Piantadosi (2013), communication can succeed despite the possibility of corrupted input because comprehenders entertain alternative hypotheses about what the speaker intended. Comprehenders are hypothesized to estimate how likely possible intended sentences s_i are for a given input sentence s_p by computing the conditional probability $P(s_i|s_p)$ on the left hand side of the Bayesian formula below.

(1) $\quad P(s_i|s_p) \propto P(s_i) \times P(s_i \rightarrow s_p)$

The conditional probability $P(s_i|s_p)$ can be estimated from the two probabilities given on the right hand side of the formula: the a-priori probability $P(s_i)$ of producing s_i and the conditional probability $P(s_i \rightarrow s_p)$ that s_i has been corrupted by noise, resulting in a corrupted input string s_p.

With regard to the noise model $P(s_i \rightarrow s_p)$, Gibson et al. (2013) demonstrated that comprehenders consider simple edits (deletions or insertions of a single word), but not more complex edits. In particular, Gibson et al. hypothesized that comprehenders do not consider word exchanges. This hypothesis was based on the finding that implausible passives like *The girl was kicked by the ball* were hardly ever interpreted in a nonliteral way, despite the fact that undoing a noun exchange would restore the sentence to plausibility. Poppels and Levy (2016), however, did find evidence for word exchanges. Implausible double PP sentences such as *The package fell from the floor to the table* were often interpreted nonliterally. Such sentences can be restored to plausibility by undoing a transposition of prepositions.

Poppels and Levy (2016) proposed the Function Word Constraint and the Adjunct Constraint to differentiate between passive sentences (few/no non-literal interpretations) and double PP sentences (many non-literal interpretations), without deciding between the two constraints. Chen, Nathaniel, Ryskin, and Gibson (2023) instead proposed the Intervening Verb Constraint.

(2) a. *Function Word Constraint* (Poppels & Levy, 2016)
         Only function words may be exchanged.
     b. *Adjunct Constraint* (Poppels & Levy, 2016)
         Exchanges can involve elements in adjuncts but not elements in arguments.
     c. *Intervening Verb Constraint* (Chen et al., 2023)
         Two words may be exchanged if no main verb intervenes between them.

We ran three experiments investigating German sentences with non-canonical word order in order to decide between the three constraints in (2). Our experiments used the same experimental task as Gibson et al. (2013): participants were presented a series of sentences and had to answer a yes-no question after each sentence. The question was displayed simultaneously with the sentence, and sentence and question were visible until participants selected either 'yes' or 'no' as answer.

Experiment 1 Experiment 1 varied German main clauses according to three factors: *Word Order* (subject-before object/SO versus object-before subject/OS), *Plausibility* (plausible versus implausible) and *Intervening Verb* (Auxiliary versus Main verb). An example is given in (3) (only implausible sentences are shown, in plausible sentences the two nouns were reversed).

(3) SO: Der   Knochen hat den   Hund gegessen. – Der   Knochen aß den   Hund.
         the$_{NOM}$ bone   has the$_{ACC}$ dog eaten   the$_{NOM}$ bone   ate the$_{ACC}$ dog
     OS: Den   Hund hat der   Knochen gegessen. – Den   Hund aß der   Knochen.
         the$_{ACC}$ dog has the$_{NOM}$ bone   eaten   the$_{ACC}$ dog ate the$_{NOM}$ bone

Results show that implausible OS sentences elicited a high rate of nonliteral interpretations, regardless of whether a main verb or an auxiliary intervened (see Figure 1, left). This finding is consistent with the Function Word Constraint, but not with the Adjunct Constraint and the Intervening Verb Constraint:
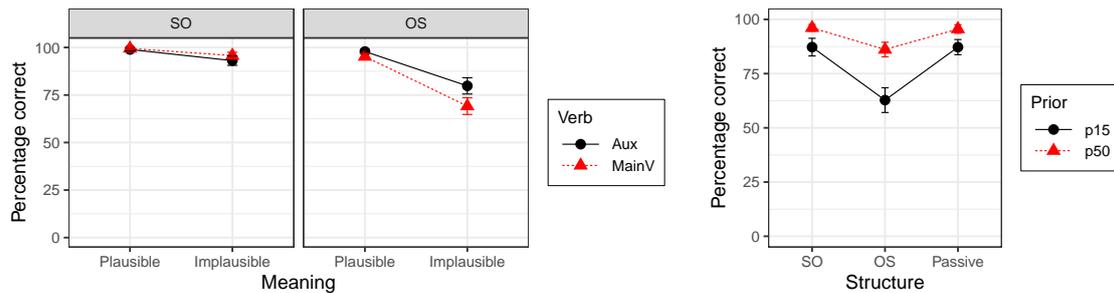
Figure 1: Percentages of correct answers in Exp. 1 and 3. Error bars show 95% confidence intervals.

exchange edits with implausible OS sentences affect arguments and are not blocked by a main verb.

Experiments 2 and 3 Experiments 2 and 3 examined whether nonliteral interpretations of implausible OS sentences are indeed due to exchange edits applied to function words (*der/den*), in accordance with the Function Word Hypothesis, or to exchange edits applied to nouns, transposing *Hund* and *Knochen*. For English, Poppels and Levy (2016) excluded noun exchanges because noun exchanges should result in a high number of nonliteral interpretations for passive sentences, contrary to fact. However, due to the word-order flexibility of German, native speakers of German may be more willing to consider noun exchanges as a possible source of corruption during noisy-channel inference. Therefore, Experiments 2 and 3 included German passive sentences besides SO and OS sentences.

(4)   Passive, implausible: Der   Hund wurde vom   Knochen gegessen.
        the.NOM dog   was   by-the bone   eaten

Results of Experiment 2 (not shown here) confirmed that nonliteral interpretations are rare for German passives, similar to what has been found for English.

In addition to effects of sentence structure (SO vs. OS vs. passive sentences), Experiment 3 also examined whether these effects can be attributed to noisy channel inferencing. If the nonliteral interpretations observed for implausible OS sentences are indeed the result of noisy channel inferences and not just due to perceptual confusion of the two determiners *der* and *den*, the rate of nonliteral interpretions should depend on the a-priori probability $P(s_i)$. To test this prediction, we followed Gibson et al. (2013) and manipulated the rate of implausible sentences within the complete sentence list (15 experimental + 85 filler sentences) presented to participants. In condition p15, the rate of implausible sentences was 15%, whereas it was 50% in condition p50. A higher rate of implausible sentences means that implausible sentences have a higher a-priori probability, which should decrease the number of nonliteral interpretations. Results for Experiments 3 again demonstrate that implausible OS sentences are susceptible to nonliteral interpretations, whereas accuracy for passive sentences was as high as for SO sentences (see Figure 1, right). Moreover, nonliteral interpretations are clearly modulated by the overall rate of implausible sentences and occur less often if the a-priori probability of encountering an implausible sentence is high.

Overall, our findings show that comprehenders consider word exchanges as a potential source of noise and apply exchange edits to restore implausible to plausible sentences. The probability assigned to exchange edits is modulated by word type, in line with the Function Word Constraint, but depends neither on the argument status of the constituents affected by exchange edits nor the intervening verb.

References

Chen, S., Nathaniel, S., Ryskin, R., & Gibson, E. (2023). The effect of context on noisy-channel sentence comprehension. *Cognition*, *238*, 105503. doi: 10.1016/j.cognition.2023.105503

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. doi: 10.1073/pnas.1216438110

Poppels, T., & Levy, R. P. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 378–383). Austin, TX.

Adjuncts and complements beyond English: testing the applicability of the *do so* test in Croatian

Ana Werkmann Horvat, University of Osijek
Matea Birtić, ICL
Martina Gračanin-Yuksek, Middle East Technical University

The distinction between arguments and adjuncts is tacitly present in many grammatical theories and has been supported by psycholinguistic research in various languages (Tutunjian & Boland 2008; Akal 2017; Chromý & Vojvodić 2023). However, drawing a clear line between arguments and adjuncts is difficult even in a single language, let alone universally. One way of distinguishing between arguments and adjuncts is by using the *do so* test (Lakoff & Ross 1976) which relies on the fact that *do so* must replace a verb and its arguments, but need not include adjuncts, and if the replacement results in grammaticality (e.g. *John ate a banana yesterday, while Geraldine did so today*), it indicates that the element that 'survives' the replacement (here *today*) is an adjunct, while ungrammaticality (**John ate a banana, while Geraldine did so an apple*) suggests it is an argument (*a banana*). The *do so* test is often cited in the theoretical syntactic literature as a means for distinguishing between complements and adjuncts (e.g., Zagona 1988; Baxter 1999). There are, however, surprisingly few studies that test the outcomes of the *do so* replacement in an experimental setting (with either offline or online experiments). Therefore, our study explores whether the *do so* replacement test can successfully discriminate between VP-internal arguments (complements) and adjuncts in Croatian.

We conducted two experiments: an offline acceptability judgment task (AJT) and an online self-paced reading (SPR) task. The experiments were designed to determine whether the *do so* replacement test in Croatian is sensitive to (i) the status of the constituent left out of the *do so* substitution (complement vs. adjunct) and (ii) the case borne by this constituent (accusative vs. instrumental). The critical stimuli consisted of 16 sets of 4 similar sentences across 4 possible combinations, depending on the case/type of the noun following *to čini* 'does so': 1) accusative adjunct, 2) instrumental adjunct, 3) accusative complement, 4) instrumental complement. The same verb was used for the two adjunct sentences, while a different verb was used for the two complement sentences. The nouns following the *do so* pro-form (either adjuncts or complements) were the same across the four sentences. For the complement sentences, we used 16 Croatian verbs belonging to a very restricted class of verbs that can take either an accusative or an instrumental theme (complement). In the creation of our stimuli, we assumed that instrumental phrases denoting instruments would behave like adjuncts. The status of instruments as arguments/adjuncts is controversial, with some authors treating them as arguments (e.g., Belaj & Tanacković Faletar 2017), some treating them as adjuncts (e.g., Van Valin & LaPolla 1997), and yet some others as an in-between category (e.g., Rissman et al. 2015). However, the *do so* test typically classifies instruments as adjuncts (e.g., Lakoff & Ross 1976) and since this is the test we used, instrumental phrases denoting instruments were taken as adjuncts. In Table 1, the critical stimuli, which were used in both experiments, are presented.

ToC	Case	Primjer	English translation
Adj	Acc	Sara omata u žicu, a Rita to čini u traku usput psujući.	Sara is wrapping in wire.ACC, and Rita is doing so in tape.ACC while cursing.
Adj	Inst	Sara omata žicom, a Rita to čini trakom usput psujući.	Sara is wrapping with wire.INST, and Rita is doing so with tape.INST while cursing.
Compl.	Acc	Sara trza žicu, a Rita to čini traku usput psujući.	Sara is pulling the wire.ACC, and Rita is doing so the tape.ACC while cursing.
Compl.	Inst	Sara trza žicom, a Rita to čini trakom usput psujući.	Sara is pulling the wire.INST, and Rita is doing so with the tape.INST while cursing.

Table 1. Critical stimuli

Crucially, in the AJT, the accusative complements received lower ratings compared to instrumental complements (t(1,72.9) = -15.55, p <.0001) and accusative adjuncts (t(1,72.9) = -7.54, p <.0001). The SPR task revealed similar patterns, with the accusative complements being read slower than instrumental complements (t(1,112) = 4.14, *p* = .0001) and accusative adjuncts ((t(1,634) = -4.16, *p* <.0001). In Table 2

and Table 3 raw means for acceptability ratings and reaction times are presented. These findings indicate that the *do so* test successfully discriminated between accusative complements and adjuncts. However, instrumental complements showed comparable acceptability and processing speed to adjuncts, suggesting a different behavior than accusative complements.

Type	Case	Mean rating
Adjunct	Acc	2.32
Adjunct	Inst	3.46
Complement	Acc	1.35
Complement	Inst	3.59

Table 2. Raw means for acceptability ratings

Type	Case	Mean RT (2nd noun)	Mean RT (spill-over)
Adjunct	Acc	668.63	586.27
Adjunct	Inst	760.39	570.68
Complement	Acc	613.70	767.92
Complement	Inst	715.94	585.34

Table 3. Raw mean reading times for the critical regions across the two critical regions

The findings support the argument-adjunct distinction in Croatian, as the pro-verb *to činiti* 'to do so' can replace the verb alone only when followed by an accusative adjunct, not an accusative complement. However, instrumental NPs (both those denoting themes and those denoting instruments) displayed different behavior with respect to the *do so* test. This suggests either that the *do so* test only works with accusatives or that instrumental themes in Croatian are not complements of the verb, i.e., they do not occupy the same structural position that accusative themes occupy. We propose two possible future approaches to distinguish between these possibilities: testing accusative and instrumental themes on additional syntactic tests for argument/adjunct distinction and applying the *do so* test to NPs with other cases (e.g., genitive, dative) that could potentially function as complements of the verb, which would provide insights into the behavior of instrumentals. In addition to this, comparing the behavior of instrumental themes with instrumental phrases having temporal interpretations could shed light on the role of the instrumental case itself in relation to the *do so* test.

References
Akal, T. (2017). Wh-argument/adjunct asymmetry in sentence processing. *Dilbilim Araştırmaları Dergisi* 2017/2, 43-71.
Baxter, D. P. (1999). *English Goal Infinitives*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
Belaj, B., & Tanacković Faletar, G. (2017). *Kognitivna gramatika hrvatskoga jezika. Knjiga prva, imenska sintagma i sintaksa padeža.* Zagreb: Disput.
Chromý, J., & Vojvodić, S. (2023). When and where did it happen? Systematic differences in recall of core and optional information. *Quarterly Journal of Experimental Psychology*. 1-22.
Lakoff, G., & Ross. J. R. (1976). Why you can't do so into the sink. In *Syntax and Semantics. Volume 7. Notes from the linguistic underground,* McCawley, James D (ed), 101-131. New York: Academic Press.
Rissman, L., Rawlins, K., & Landau, B. (2016). Using Instruments to Understand Argument Structure: Evidence for Gradient Representation. *Cognition* 142.266–290.
Tutunjian, D., & Boland, J. (2008). Do We Need a Distinction between Arguments and Adjuncts? Evidence from Psycholinguistic Studies of Comprehension. *Language and Linguistics Compass* 2/4, 631-646.
Van Valin, R., & LaPolla, R. (1997). *Syntax: form, meaning, and function.* Cambridge, UK: Cambridge University Press.

Zagona, K. (1988). *Verb Phrase Syntax: A Parametric Study of English and Spanish, volume 13 of Studies in Natural Language and Linguistic Theory*. Dordrecht: Kluwer.

An Argument for Symmetric Coordination: A Replication Study

Adam Przepiórkowski[*,†], Magdalena Borysiak[*], Adam Głowacki[*]

[*]*University of Warsaw,* [†]*ICS Polish Academy of Sciences*

Introduction Since the beginning of the replication crisis in early 2000s (Ioannidis 2005), there has been a growing recognition of the importance of replication studies. Also within linguistics, a steadily increasing number of researchers put emphasis on replicability and repeated testing of claims, rather than just on producing novel claims (Sönning and Werner 2021). The aim of this paper is to report on a replication of a novel result reported in Przepiórkowski and Woźniak 2023 (henceforth: PW23) that the dynamics of conjunct lengths in English coordinations provides an argument for symmetric approaches to coordination (e.g., that utilized in Prague Dependency Treebanks, in Word Grammar or that of Neeleman *et al.* 2023) and against asymmetric approaches (e.g., that used in Universal Dependencies or that of Munn 1993).

PW23's empirical findings are based on PTB$_\&$ (Ficler and Goldberg 2016), a version of the Penn Treebank (PTB; Marcus *et al.* 1993) containing more explicit representations of coordinate structures. PW23 extract 21,825 binary coordinations from PTB$_\&$ and investigate how the proportion of coordinations with the left conjunct shorter than the right conjunct (with respect to all coordinations with unequal lengths of the two conjuncts) changes with the absolute difference between the two lengths, depending on the presence and position of the governor. Their findings are summarized in the plots to the left, which present the result of fitting logistic models to the PTB$_\&$ data.[1]

Whether length is measured in words (see the first column of plots) or in characters (see the second column), when the governor is on the left of the coordinate structure, as in *Bring apples and oranges!* (where *bring* is the governor of the coordinate structure *apples and oranges*), the proportion of coordinations with the left conjunct shorter grows with the absolute difference of lengths (see the first row). In all four plots, the slopes of the curves are significantly positive ($p \ll 0.001$). However, and this is the crucial new observation of PW23, this effect disappears when the governor is on the right, as in *Apples and oranges fell* (see the third row). Here, when the length is counted in words, the slope is insignificantly negative, and when it is counted in characters, it is insignificantly positive.

PW23 argue that, given the principle of Dependency Length Minimization (DLM; Temperley and Gildea 2018), these tendencies are incompatible with asymmetrical representations, which assume that coordinations are headed by initial conjuncts. On such representations, aggregate dependency length is minimized when the first conjunct is shorter *regardless of the presence and position of the governor*, so all six slopes should be significantly positive. On the other hand, symmetrical approaches predict positive slopes when the governor is on the left or absent, but – depending on the particular approach – they predict a flat (conjunction-headed approaches) or negative (multiple-headed approaches) slope when the governor is on the right. (See PW23 for the complete reasoning, which we cannot present here for lack of space.)

---

[1]Because of data scarcity, PW23 grouped all observations of length differences greater than 3 words into only two further buckets (apart from the three buckets for the differences 1, 2, and 3): one for differences from 4 to 6, and the other for differences from 7 to 25 (and similarly for characters) – see the black boxes indicating these 5 buckets at the bottom of each plot. Due to dataset size, such bucketing was not needed in the current replication study.
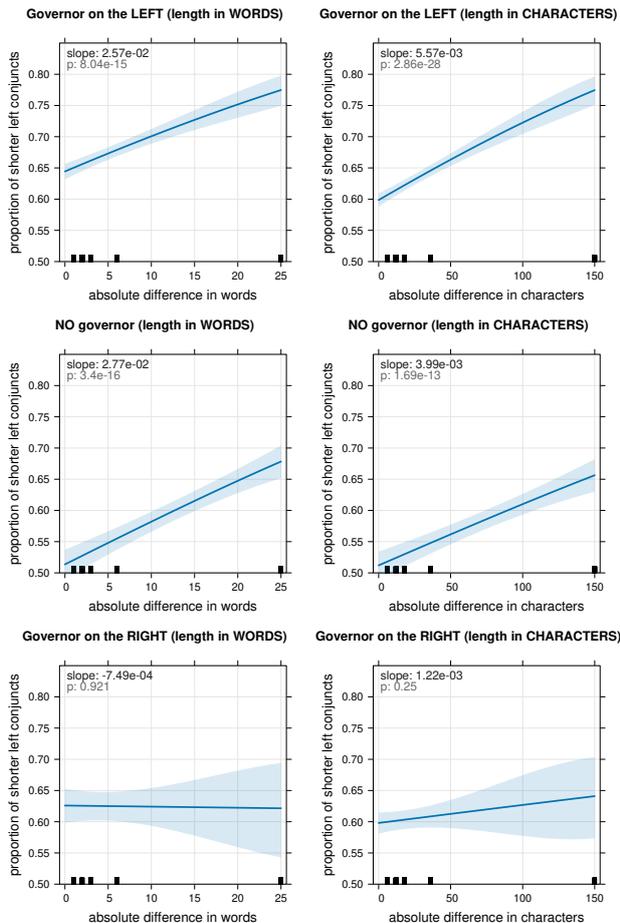
Replication   In this replication study, instead of using a high-quality but small dataset, as in PW23, we used a large but low-quality dataset, namely, large parts of the Corpus of Contemporary American English (COCA; Davies 2023) automatically parsed with Stanza (Qi *et al.* 2020). We parsed texts from 6 genres: newspapers, magazines, academic, fiction, blogs, and other web pages. From dependency trees produced by Stanza, we extracted information about 11,502,053 coordinations.
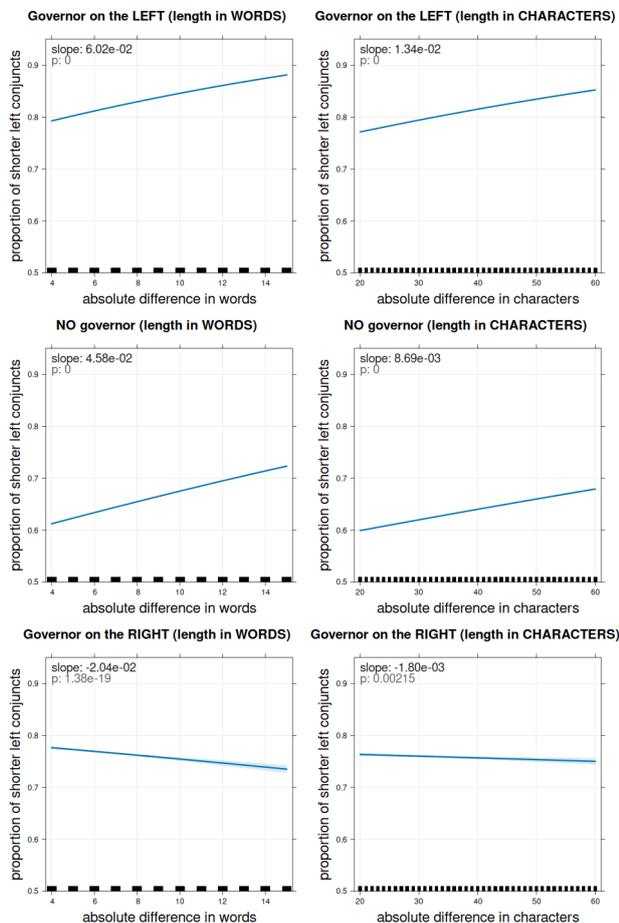
For the statistics based on lengths measured in words, we took into consideration coordinations with absolute length differences between the first and the last conjunct in the range of 4 to 15 words (not shorter, to minimize the effect of fixed binomial expressions); for lengths measured in characters, we inspected differences in the range of 20 to 60 characters. In these ranges, there were between around 1000 and over 40,000 observations for each combination of governor position (left, absent, right) $\times$ particular length difference.

The results are presented in the logistic regression plots to the left. The first two rows, for coordinations with the governor on the left or absent, are analogous to PW23's results. However, while PW23's results did not make it clear whether the actual slopes are negative, zero, or positive when the governor is on the right (see the wide confidence bands in the relevant plots on the previous page), here the slopes are very significantly negative ($p < 0.01$), especially when length is measured in words ($p \ll 0.001$).

Hence, this replication study confirms PW23's empirical findings and their argument against asymmetric approaches to coordination, which predict a positive slope regardless of the presence and position of the governor. However, it also sharpens their results: it favours multi-headed symmetric representations of coordinations, as proposed in Word Grammar (Hudson 1984), as they predict such a negative tendency when the governor is on the right, while conjunction-headed symmetric approaches predict the lack of any positive or negative tendency in such cases. (Again, see PW23 for the reasoning.)

References   • **Davies, M.** (2008–2023). The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.   • **Ficler, J. and Goldberg, Y.** (2016). Coordination annotation extension in the Penn Tree Bank. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 834–842.   • **Hudson, R.** (1984). *Word Grammar*. Blackwell.   • **Ioannidis, J. P. A.** (2005). Why most published research findings are false. *PLoS Medicine*, **2**(8), 0696–0701.   • **Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A.** (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, **19**, 313–330.   • **Munn, A. B.** (1993). *Topics in the Syntax and Semantics of Coordinate Structures*. Ph.D. dissertation, University of Maryland.   • **Neeleman, A., Philip, J., Tanaka, M., and van de Koot, H.** (2023). Subordination and binary branching. *Syntax*, **26**(1), 41–84.   • **Przepiórkowski, A. and Woźniak, M.** (2023). Conjunct lengths in English, Dependency Length Minimization, and dependency structure of coordination. In A. Rogers, J. Boyd-Graber, and N. Okazaki, eds., *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15494–15512. Association for Computational Linguistics.   • **Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D.** (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 101–108.   • **Sönning, L. and Werner, V.** (2021). The replication crisis, scientific revolutions, and linguistics. *Linguistics*, **59**(5), 1179–1206.   • **Temperley, D. and Gildea, D.** (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, **4**, 67–80.

Kalle Glauch, Ruhr-Universität Bochum

Punctuation Modulates the Valence of Referents in Exclamative Clauses in CMC

In computer-mediated-communication (CMC), punctuation marks can convey expressive meanings that reflect the writer's attitude towards specific situations or discourse referents, especially in highly interactive forms of CMC like messaging services (e.g., WhatsApp) and social media (e.g., Twitter) (Gutzmann & Turgay, in press). The current study investigates the influence of punctuation marks on the perceived valence of discourse referents in interrogative-style exclamative clauses (e.g. *What a view!*) in CMC.

Exclamative clauses indicate a heightened emotional state of the speaker (Potts & Schwarz 2008). Regarding the emotional dimension of valence, this may manifest as an unspecific valence-strengthening function, i.e. the embedded referents receive a polarity conforming valence boost, making them seem more positive or negative than in isolation.

While exclamative clauses typically use exclamation marks, corpus data reveals greater variation in punctuation marks, including: exclamation marks (!), multiple exclamation marks (!!!), full stops (.), ellipses signs (…) and null punctuation ( ). Given that punctuation marks have expressive functions to represent various speaker attitudes, they can be hypothesized to also differentially influence the overall valence of discourse referents within exclamative clauses.

To investigate the hypothesis, a valence rating study was conducted using Magpie (https://magpie-experiments.org/). In a 1x5 between-subject design, participants were presented with 120 exclamative clauses manipulated by clause final punctuation mark. The sentences were presented randomly in WhatsApp-like speech bubbles with varying names and timestamps. Participants rated the valence of the embedded referent on a 1-9 scale. 196 native English speakers with a rejection rate of ≤1% were recruited from Prolific. Figure (1) visualizes an experimental item from the participant's point of view.

Figure 1: Experimental item from participant's perspective

Using the mgcv package in R (Wood 2011), a generalized additive mixed model (GAMM) was employed to predict the embedded valence of referents based on their isolated valence ratings (Warriner et al. 2013) and punctuation, with random intercepts for participants and words: gam(embedded_valence ~ s(isolated_valence, by=interaction(punctuation)) + punctuation + s(participant, bs = "re") + s(words, bs = "re"), data = data, method = "REML"). Figure (2) shows the predicted distributions by punctuation

mark. The dashed diagonal line represents the equality line that would be expected if the isolated valence ratings perfectly predicted the embedded valence of discourse referents.
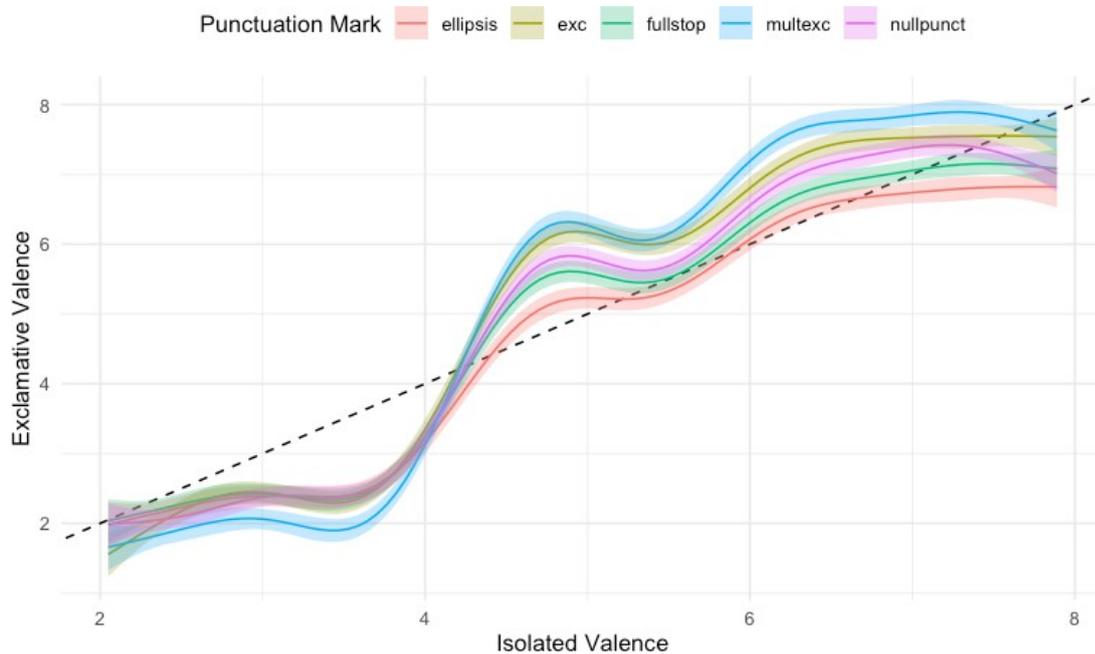
Figure 2: GAMM predicted exclamative valence by punctuation

Difference plots for pairwise comparisons of the conditions were generated using the itsadug R package (van Rij et al. 2022) to identify inherent valence value ranges of significant differences for each contrast. The difference plots suggest following positivity hierarchy of embedded valence ratings based on clause final punctuation for the positive end of the scale: multiple exclamation marks > exclamation marks > null punctuation > full stop > ellipsis sign. On the negative end of the scale, conditions largely overlap, with only multiple exclamation marks showing significant differences from other punctuation marks. While the positivity hierarchy for the positive end of the scale aligns with prior assumptions about specific expressive meanings of the punctuation marks investigated, explaining the minimal differences on the negative end proves challenging.

References:

Gutzmann, D., & Turgay, K. (in press). Expressive Interpunktion!?! Interpunktion zwischen Grammatik(?) und Pragmatik!*. Linguistische Berichte.

Potts, C., & Schwarz, F. (2008). Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora. Manuscript, UMass Amherst.

van Rij J, Wieling M, Baayen R, van Rijn H (2022). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.4.1.

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.

Wood SN (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models." Journal of the Royal Statistical Society, 73(1), 3-36.

The processing of quotation marks in German: Evidence from eye-tracking

Natascha Raue (Universität Kassel)

Holden Härtl (Universität Kassel)

Álvaro Cortés Rodríguez (Universität Potsdam)

1 Introduction

The following study aims to investigate the processing of quotation marks in name-mentioning constructions (NMCs). NMCs typically involve naming predicates such as *nennen* ('call') and *bezeichnen* ('refer to as').

(1)   Dieses Phänomen wird „Sonnenfinsternis" genannt.
        'This phenomenon is called "solar eclipse".'

Quotational constructions of this type are characterized by introducing the name of a lexicalized concept, in this case *solar eclipse* (Härtl 2020, Cortés Rodriguez et al. 2022). The expression introduced in NMCs is often but not preferably enclosed in quotation marks. Recent corpus data revealed that NMCs involving *nennen* do not show a preference for being accompanied by quotes (Raue 2022).

Eye-tracking studies testing the processing of punctuation marks are sparse (see Hill & Murray 2000 for the processing of commas). Further, Yao & Scheepers (2011) present evidence on the reading of written stories in direct and indirect speech reports, using eye-tracking data related to the silent reading process. The linguistic context implied either a fast or slow-speaking quoted speaker, which resulted in longer reading times for (oral/silent) reading direct speech as opposed to indirect speech. In addition, recent research using the self-paced reading paradigm revealed that the processing of irony in NMCs is facilitated if the nominal is enclosed in quotation marks (Schlechtweg & Härtl 2023). Despite this, the nature of the reading process involved in the processing of quotational constructions and, more specifically, of quotations of the type in (1), known as pure quotations, has not been investigated.

2 Research question and hypotheses

The focus of the present study is on German NMCs with and without quotes, as exemplified in (2), where the investigated variables are labelled *Quotes* and *noQuotes*, respectively.

(2)   Kim weiß, dass man dieses Verfahren Neuwahl/„Neuwahl" nennt, und belehrt Anna darüber.
        ‚Kim knows that this procedure is called re-election/ "re-election" and she informs Anna about this.'

The study aims to investigate the processing of written quotation marks during silent reading, using a combination of early and late eye-tracking measures to gain insight into temporal differences in the processing of quotes. The duration of fixations, as well as dwell time, has been shown to be sensitive to linguistic properties, and different types of fixations have been argued to reflect different stages of processing. For instance, the first fixation is associated with the lexical activation process while later fixations serve a discourse integrative process (Holmqvist & Nyström 2011). Differences in eye movements are expected to be manifested as longer fixations in the *Quotes* condition as opposed to the *noQuotes* condition, i.e., the first fixation duration and more subsequent fixations on the quoted item, reflected in fixation durations and a longer dwell time on the Interest Area of the target word.

3 Empirical investigation

All eye-movements were collected using the EyeLink 1000 Plus (SR Research Ltd., Toronto, Ontario, Canada). The eye-tracker was running at 500 Hz sampling rate. Viewing for the participants was binocular but only the dominant eye was tracked. The experimental items were created in German and consisted of 60 filler items and 20 minimal pairs as target items, which were manipulated by varying the presence and absence of quotes around the nominal in NMC constructions. All target nominals were controlled for lexical frequency and word length. Twenty-four native speakers of German (mean age=25) with normal or corrected-to-normal vision participated in the experiment.

4 Results and discussion

After filtering for correct answers to the comprehension question, trials with a high imprecision and offset data were removed, resulting in a total of 435 trials. Linear Mixed Models were created in R (2022), using the respective measurement as continuous dependent variable and the condition as fixed effects. Word length was entered as a covariate and random effects for both participants and items were included in the model. The analysis proved no significant effect for early measures like first fixation duration, second fixation duration and third fixation duration. However, the results revealed significant effects in the dwell time measurement for the target Interest Area for *ConditionName* with and without quotes (Pr(>|t|) = 0.0444*), and *WordLength* between long and short words (Pr(>|t|) = 0.0249*). The *NoQuotes* condition displayed a shorter dwell time (mean = 598ms) as opposed to the *Quotes* condition(mean = 660ms). This eye-tracking evidence reveals that quotation marks reveal processing effects in later windows.

We interpret this effect as reflecting a top-down process typically associated with higher cognitive discourse-based functions (see Rayner & Pollatsek 1989; Orquin & Mueller Loose 2013; Orquin & Holmqvist 2018). This interpretation is compatible with the pragmatic account of pure quotation (Schlechtweg & Härtl 2020). Further, the effect we found for word length are consistent with the well-established word length effect (Just & Carpenter 1980; Rayner 2009; Hautala et al. 2011). The implications of our results will be discussed in light of the processing architecture used for quotation as well as the implementation of pure quotation at the interface between semantics and pragmatics.

References

Carter, Benjamin T. & Steven G. Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155, 49–62.

Cortés Rodríguez, Álvaro, Holden Härtl, Natascha Raue & Kristina Weissbecker. 2022. Referring nouns in name-informing quotation: An empirical investigation. *Linguistic Research* 39(2), 327–354.

Härtl, Holden. 2020. Referring nouns in name-informing quotation: A copula-based approach. In Michael Franke, Nikola Kompa, Liu Mingya, Jutta Mueller, and Juliane Schwab (eds.)*, Proceedings of Sinn und Bedeutung* 24, 291–304. Osnabrück/Berlin: University Osnabrück/Humboldt-Universität Berlin.

Hautala, Jarkko, Jukka Hyönä & Mikko Aro. 2011. Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns, *Vision Research* 51(15), 1719-1727.

Hill, Robin L. & Wayne S. Murray. 2000. Commas and spaces: Effects of punctuation on eye movements and sentence parsing. In A. Kennedy, R. Radach, D. Heller & J. Pynte (eds). *Reading as a perceptual process*, 565-589. Oxford: Elsevier.

Holmqvist, Kenneth & Marcus Nyström. 2011. Eye Tracking. A comprehensive guide to methods and measures. Oxford: Oxford University Press.

Just, Marcel Adam & Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354.

Orquin, Jacob L. & Kenneth Holmqvist. 2018. Threats to the validity of eye-movement research in psychology. Behavior Research Methods 50, 1645–1656.

Orquin, Jacob L., & Simone Mueller Loose. 2013. Attention and choice: A review on eye movements in decision making. *Acta Psychologica* 144, 190–206.

R Core Team. 2022. *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software].

Raue, Natascha. 2022. A corpus-based investigation into the linguistic characteristics of name-mentioning constructions. Paper presented at the *18th Workshop on Syntax, Semantics and Phonology* (WoSSP 2022), Universitat Autònoma de Barcelona.

Rayner, Keith & Alexander Pollatsek. 1989. *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Rayner, Keith. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

Schlechtweg, Marcel & Holden Härtl. 2023. Quotation marks and the processing of irony in English: Evidence from a reading time study. *Linguistics* 61(2), 355-390.

Schlechtweg, Marcel & Holden Härtl. 2020. Do we pronounce quotation? An analysis of name-informing and non-name-informing contexts. *Language and Speech* 63(4), 769-798.

Yao, Bo & Christoph Scheepers. 2011. Contextual modulation of reading rate for direct versus indirect speech quotations. Cognition 121(3), 447–453.

Korean causative change of state predicates and non-culminating readings
Paola Fritz-Huechante & *Elisabeth Verhoeven*
Humboldt-Universität zu Berlin

Different languages have shown that a group of change of state (hereafter CoS) predicates in the perfective is able to produce contrasting readings between: (a) a non-culminating (henceforth NC) reading in which the result state specified by the meaning of the verb's semantic core fails to obtain, and (b) a culminating reading in which the result state obtains (cf. Tatevosov & Ivanov 2009; Demirdache & Martin 2015; Martin & Schäfer 2017, a.o.). Korean causative CoS predicates are also able to produce such readings (cf. Beavers & Lee 2020; Fritz-Huechante, Verhoeven & Rott 2020). A well-studied factor allowing NC readings is the agenthood properties of the external argument (Agent Control Hypothesis –ACH, cf. Demirdache & Martin 2015). The ACH argues that in the presence of an agentive subject, as in (1a), the result state does not need to take place, as seen by the felicitous continuation of the *but*-clause. In contrast, in the presence of an (inanimate) causer as in (1b) the result state cannot be defeated. Hence, the continuation with the *but*-clause generates a contradiction.

(1) a. Yuli-ka   ipwul-ul   mal-ly-ess-ta.   haciman ipwul-i   malu-ci   anh-ass-ta.
       Yuri-NOM blanket-ACC be.dry-CAUS-PST-DECL but   blanket-NOM be.dry-CONN NEG-PST-DECL
       'Yuri dried the blanket, but the blanket was not dry.'
    b. hayspyeth-i ipwul-ul   mal-ly-ess-ta.   #haciman ipwul-i   malu-ci   anh-ass-ta.
       sun-NOM   blanket-ACC be.dry-CAUS-PST-DECL but   blanket-NOM be.dry-CONN NEG-PST-DECL
       'The sun dried the blanket, but the blanket was not dry.'

In this study, we investigate two additional, hitherto less explored factors in their impact on NC readings in Korean: scale structure and causative structure. In terms of scale structure, predicates are classified regarding to the possibility to identify a standard degree of comparison (i.e. the degree –or bound– that manifests the property specified by the semantic core) into: (i) lower-bounded (e.g. *to wet*), (ii) upper-bounded (e.g. *to dry*), (iii) open-bounded (e.g. *lengthen*), and (iv) closed-bounded (e.g. *fill*) predicates (cf. Hay et al. 1999; Kennedy & McNally 2005; Kennedy & Levin 2008). This classification is also pertinent in Korean. Focusing on causative upper-bounded (e.g. *mallita* 'to dry') and lower-bounded (e.g. *ceksita* 'to wet') predicates, we observe that NC readings are (more easily) available with the former but not so with the latter. (1) is an instantiation of an upper-bounded predicate, i.e. in order for (1) to hold true, the blanket has to be maximally dry. A NC reading arises in the presence of an agentive subject as in (1a), yielding an interpretation that Yuri acted upon the blanket to dry it without necessarily causing the crucial CoS (i.e. the blanket being completely dry). In contrast, NC readings are not available with lower-bounded predicates. Lower-bounded predicates hold true at the presence of a minimal amount of change. In (2), a minimum CoS occurs as soon as the subject referent acts upon the object, hence negating the result state generates a contradiction irrespectively of the type of subject.

(2) Yuri-ka   / pipalam-i   ipwul-ul   ceks-y-essta.   #haciman ipwul-i
    Yuri-NOM   / rainstorm-NOM blanket-ACC be.wet-CAUS-PST-DECL but   blanket-NOM
    cec-ci   anh-ass-ta.
    be.wet-CONN NEG-PST-DECL
    'Yuri / the rainstorm wetted the blanket, but the blanket was not wet.'

Regarding causative structures, Korean possesses: (a) morphological causatives (1) formed by attaching the causative morpheme *-i* (or its allomorphs) to the stem of the stative verb (e.g. *mal-li-ta* 'be.dry-CAUS-DECL'), and (b) periphrastic causatives (3) built by adding the light verb *hata* 'do' to the stative verb stem (e.g. *malu-key ha-ta* 'be.dry-ADVR do-DECL') (cf. Lee 2007). The morphosyntactic features of the causative predicates allow NC readings, in that: (a) morphological causatives built on upper-bounded verbs are able to produce such readings with an agentive subject (1a), and (b) periphrastic causatives produce such readings when built on either a lower- or upper-bounded verb and with an agent. (b) is possible due to the morphosyntactic composition (*-key hata*) of the predicate yielding two possible readings: an activity reading (an action performed by the subject without affecting the object) or an action that causes a change in the object (cf. Fritz-Huechante, Verhoeven & Rott 2020; Choe 2022).

(3) Yuri-ka   ipwul-ul   malu-key   / cec-key   hay-ss-ta.   haciman ipwul-i
    Yuri-NOM blanket-ACC be.dry-ADVR / be.wet-ADVR do-PST-DECL but   blanket-NOM
    malu-ci   / cec-ci   anh-ass-ta.
    be.dry-CONN / be.wet-CONN NEG-PST-DECL
    'Yuri made the blanket dry / wet, but the blanket was not dry / wet.'

In a 1–5 Likert scale acceptability study (1 = very bad, 5 = very good), we tested the participants' acceptability ratings w.r.t. cancelling the result state by means of a *but*-clause as in sentences (1) – (3). Two experiments were built according to the verbs' availability to form both morphological and periphrastic causative structures testing the factors: (a) subject type (agent vs. causer) and (b) scale structure (lower vs. upper). 8 target items (4 lower-bounded and 4 upper-bounded) were selected plus 24 fillers per experiment. Sentences were presented online on IBEX. Expectations were: (a) the factor subject type has an impact on the acceptability of a sentence to the extent that NC readings are more easily achieved in the presence of an agent in contrast to causers, and (b) an interaction between scale structure and causative structure to the extent that NC readings are more easily available for upper-bounded predicates in a morphological causative construction, whereas NC readings are available for both lower- and upper-bounded predicates in a periphrastic construction. 32 Korean native speakers participated in the experiments (N = 16 per list). One participant was excluded due to missing data points (N = 31, 15 female, 15 male, 1 no gender. Age: 22–42. M = 29,90). Fig. 1 for morphological causatives shows an interaction of the factors subject type and scale structure to the extent that the acceptability of NC readings is higher in the condition with upper-bounded predicates constructed with an agentive subject. In the case of periphrastic causatives, Fig. 2 shows that this interaction is not present, i.e. both lower- and upper-bounded predicates behave similarly to the extent that the acceptability of NC readings is higher in the presence of an agentive subject (lower-bounded mean 3.11, upper-bounded mean 3.18) than with a causer. The data was fitted with a linear mixed effects model, based on a maximal random effects model (cf. Barr et al. 2013). The random effects structure contained intercepts for items (different verbs) and participants. Results showed a significant effect of subject type ($p<.001$), scale structure ($p=.02$), a significant interaction of causative structure^scale structure ($p<.001$), and a marginal significant interaction of causative structure^scale structure^subject type ($p=.07$).
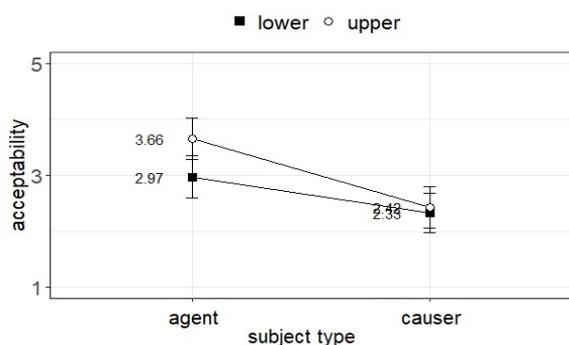
Fig. 1: Acceptability NC readings morphological causatives (95% C.I.)
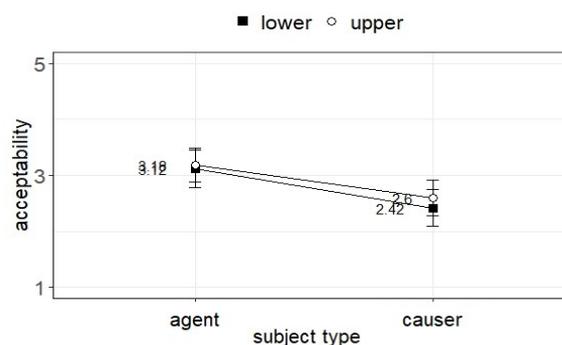
Fig. 2: Acceptability NC readings periphrastic causatives (95% C.I.)

This study experimentally confirms the availability of NC readings in Korean CoS predicates not only w.r.t. the agenthood of the subject (ACH), but also w.r.t. less-studied factors such as scale structure and the morphosyntactic structure of the causative predicates, contributing to elucidate the mechanisms that allow for NC readings in languages with a transparent event structure.

References: Barr, D. J., R. Levy, C. Scheepers & H. J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. Beavers, J. & J. Lee. 2020. Intentionality, scalar change, and non-culmination in Korean caused change-of-state predicates. *Linguistics* 58(5). 1233–1283. Choe, J. S. 2022. Structure and meaning of Korean causatives: Evidence from structural priming. *Discourse and Cognition* 29(2). 195–214. Demirdache, H. & F. Martin. 2015. Agent control over non-culminating events. In E. Barrajón, J. L. Cifuentes & S. Rodríguez (eds.), *Verb classes and aspect*, 185–217. Amsterdam: John Benjamins. Fritz-Huechante, P., E. Verhoeven & J. Rott. 2020. Agentivity and non-culminating causation in the psych domain: Cross-linguistic evidence from Spanish and Korean. *Glossa* 5(1). 1–35. Hay, J., C. Kennedy & B. Levin. 1999. Scale structure underlies telicity in degree achievements. In T. Matthews & D. Strolovitch (eds.), *Proceedings of SALT*, 127–144. Ithaca, NY: CLC Publications. Kennedy, C. & B. Levin. 2008. Measure of change: The adjectival core of degree achievements. In L. McNally & C. Kennedy (eds.), *Adjectives and adverbs: Syntax, semantics and discourse*, 156–182. Oxford: Oxford University Press. Kennedy, C. & L. McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381. Lee, H. C. 2007. The interclausal syntactic and semantic relations of the periphrastic causative. *The Linguistic Association of Korea Journal* 15(4). 163–178. Martin, F. & F. Schäfer. 2017. Sublexical modality in defeasible causative verbs. In A. Arregui, M. Rivero & A. Salanova (eds.), *Modality across syntactic categories*, 87–108. Oxford: Oxford University Press. Tatevosov, S. & M. Ivanov. 2009. Event structure of non-culminating accomplishments. In L. Hogeweg, H. de Hoop & A. L. Malchukov (eds.), *Crosslinguistic semantics of tense, aspect, and modality*, 83–130. Amsterdam: John Benjamins.

Ellipsis (not) as deaccentuation: evidence from Icelandic

Güliz Güneş & Nicole Dehé

University of Tübingen & University of Konstanz

Ellipsis as radical deaccentuation: Tancredi's (1992) theory of ellipsis as Radical Deaccentuation (RD) proposes that material can only be elided if it can be subject to givenness-driven deaccentuation, i.e., reduction or non-realisation of pitch accents on *given* (Schwarzschild 1999) constituents (Ladd 1978). The RD theory has two conceivable forms, strong and weak. The strong form claims that ellipsis applies to actual phonological content in PF, solely targeting already-deaccented material with "distinguished low-flat intonation" (e.g., Chomsky & Lasnik 1993:564). The weaker form treats 'deaccentuation' as a cover term for any phonological effect related to prosodic marking of givenness (e.g., pitch compression, diminished peaks, faster speech rate) and merely states an indirect and rather abstract link between the prosodic marking of givenness and ellipsis (e.g., Ott & Struckmeier 2016).

RD accounts (strong or weak) are based on impressionistic evidence, lacking a systematic inspection of the actual prosodic realisation of the area that can be elided. In this talk, via a production experiment, we test whether the strong or the weaker version of the RD theory is better suited to the data.

The Icelandic puzzle: At first glance, Icelandic represents a challenge to the strong form of RD, as it permits (clausal) ellipsis (Wood et al. 2019) yet does not display canonical deaccentuation of the type observed in other Intonational languages (Nolan & Jónsdóttir 2001, and Dehé 2009). The issue is that: the presence of clausal ellipsis and the absence of givenness-related deaccentuation decouples ellipsis from deaccentuation. However, the literature is not certain about whether Icelandic is really a "non-deaccenting" language. Dehé (2009) finds evidence supporting and contradicting the assumption that Icelandic content words are not deaccented when they are information structurally given.

The study: We conducted a prosodic production experiment to determine if Icelandic indeed undermines the strong RD theory as a "non-deaccenting" language. We compared clauses without any given information (e.g., the italicised clause in (1); the *all-new* condition) to those that are fully given in the preceding discourse and are potential candidates of clausal ellipsis (e.g., the italicised clause in (2); the *given* condition).

(1)   Ég velti því fyrir mér **hver** *eldaði kartöflu fyrir Sigríði fyrir miðnætti*.   [all-new]
        'I wonder   **who** cooked a potato for Sigríður before midnight.'

(2)   Einhver eldaði kartöflu fyrir Sigríði fyrir miðnætti,
        en ég veit ekki **hver** *eldaði kartöflu fyrir Sigríði fyrir miðnætti*.   [given]
        'Someone cooked a potato for Sigríður before midnight,
        but I don't know **who** cooked a potato for Sigríður before midnight.'

We also controlled the given condition such that given material was the potential target for either sluicing or fragment answers (as the representatives of two possible types of clausal ellipsis in Icelandic). All 4 conditions had 4 lexicalisations, with each target sentence repeated 3 times. In addition to 48 target utterances, experiment included 50 fillers. We recorded nine native speakers of Icelandic. In total, 384 target sentences were analyzed (for mean F0, duration and register of the accented syllables). The hypothesis testing was carried out via fitting linear mixed-effect models to the data (*lmer* package in R).

The results show that, although Icelandic does not exhibit canonical deaccentuation (partially confirming previous literature), it exhibits givenness-related prosodic attenuation both in post-sluice remnant and in post-fragment remnant conditions: pitch accents on the lexical item in the given area *are* realized,

yet they are also diminished, exhibiting register compression, lower mean overall F0 (t = 15.07, p < 0.001), and shorter duration (t = 7.86, p < 0.001) when compared to their non-given counterparts. See Fig.1 for a comparison of F0s of an utterance in *all-new/neutral* (black) versus *given/post-new* (red) conditions.
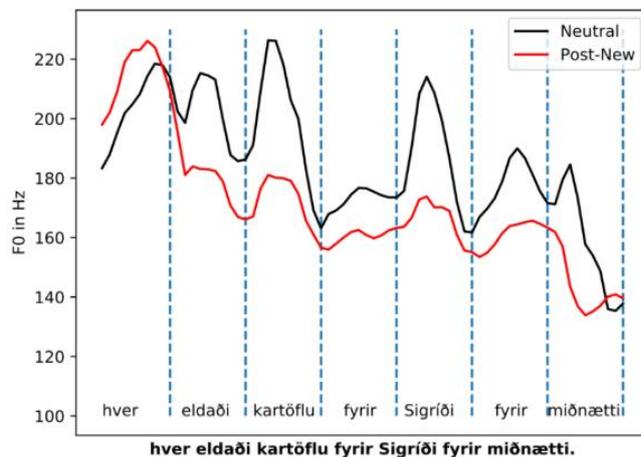
Figure 1. Mean F0 (over normalized time) of the embedded clause in (1-2), averaged across 9 speakers and 3 repetitions (n=27)

Discussion: Regarding the idea that deaccentuation is a pre-condition on ellipsis, our findings directly undermine the strong RD theory indirectly favours weak RD accounts. Comparing Icelandic to other intonation languages and discussing the prosodic behaviour of givenness in pre-focal area, we claim that weak RD theory, unlike its strong version, can cover a wider empirical domain and can be integrated into different syntactic approaches to ellipsis (in particular both to move-and-delete accounts, and in-situ deletion accounts).

Beyond ellipsis, we will discuss how our results support the view that givenness-related deaccentuation is not a defining property of Intonational languages (Kügler & Calhoun 2020). We will also discuss the repercussions of our results in relation to the role of prosodic attenuation (and the lack of deaccentuation) in the **prosodic grammar of Icelandic**. We discuss why Icelandic, which does not have lexically contrastive pitch accents, behaves more like Swedish and Japanese, which do, in terms of lack of complete deaccentuation and pitch suppression. Swedish and Japanese prosodic grammars, unlike Icelandic, have been argued to lead to avoidance of deaccentuation to preserve pitch accents on the items that are information structurally given. We consider two possible explanations, (i) historical effects, and (ii) eurythmic effects. As for (i), we discuss the possibility of the presence of tonal pitch accents as a defining feature of the prosodic grammar of Old Norse, and that, unlike Norwegian and Swedish, Icelandic has lost these accents. According to this scenario, the lack of deaccentuation may be seen as a reminiscent of an historically present behaviour of a pitch accent language. We contrast this claim with the other historical possibility: i. e. the idea that Old Norse did not have any tonal accents, and languages like Swedish and Norwegian developed tonal accents, but Icelandic did not. Alternatively, as for (ii), we will discuss the possibility of viewing the post-focal prosodic events as a consequence of eurythmic organisation, in which post-focal accentuation is independent of information structural, and word-level prosodic constraints, and that post-focal accentuation is present solely as a result of a strong phonological requirement to maintain rhythmic grouping, even when the items are given.

Selected References: Dehé, N. (2009), An intonational grammar for Icelandic. *Nordic Journal of Linguistics* 32: 5-34. ■ Nolan, F. & H. Jónsdóttir. (2001), Accentuation patterns in Icelandic. In W. A. van Dommelen & T. Fretheim (eds), *Nordic Prosody: 8th Conference, Trondheim 2000* Frankfurt, Berlin: Peter Lang, 187-198. ■ Ott, D. & V. Struckmeier. (2016), Deletion in clausal ellipsis: Remnants in the middle field, *Proceedings of the 39th Annual Penn Linguistics Conference* 22: 225-234. ■ Schwarzschild, R. (1999), GIVENness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics* 7: 141-177. ■ Tancredi, C. (1992), Deletion, deaccenting and presupposition. PhD thesis, MIT. ■ Wood, J., M. Barros & E. F. Sigurðsson. (2020), Case mismatching in Icelandic clausal ellipsis. *Journal of Linguistics* 56: 399-439.

L1-Acquisition of Deontic and Epistemic Meanings of Czech *muset* [must]

Edita Schejbalová, Charles University, schejbae@ff.cuni.cz
Radek Šimík, Charles University, radek.simik@ff.cuni.cz

Background The interpretation and acquisition of modal verbs in Czech, and indeed most Slavic languages, remain relatively underexplored compared to their counterparts in English (Cournane & PérezLeroux 2020, Cournane 2021, Cournane & Veselinović 2022). Our study investigates how Czech-speaking children acquire the modal verb *muset* [must], which as flavor variable modal conveys deontic (obligation) and epistemic (certainty) interpretations. We explore if Czech children undergo developmental phases like those in English and Bosnian/Croatian/Serbian (BCS), such as a deontic preference around age 3 (in English), followed by a commitment to epistemic meanings around age 5, which is believed to be a result of a type of pragmatic inference: *Obligation implies that the desired activity is highly likely to occur* (Cournane & Veselinović 2022). We also examine the influence of morphosyntactic factors on adults' epistemic interpretations of *muset*: the tense of *muset* and the aspect of its infinitive complement. The main **research questions** are: What morphosyntactic cues support the epistemic interpretation of the verb *muset* in Czech? Are Czech children sensitive to these cues? Do Czech children go through phases of deontic preference and epistemic commitment? The **hypotheses** that we operate with are: for adults **1)** Past tense on *muset* enhances epistemic interpretation. **2)** Imperfective infinitive enhances epistemic interpretation; for children: **1)** At age 3, a preference for deontic interpretations, **2)** At age 5, overcommitment to epistemic interpretations.

Method We employed the picture preference task (Cournane & PérezLeroux 2020) to assess how participants interpret sentences using accompanying images, which represent deontic or epistemic meanings of modal verb sentences. In the experiment, we manipulated the TENSE of *muset* (PRES/PAST) and ASPECT of its infinitival complement (PVF / IPVF) and tested the impact of these variables on picture selection (deontic/epistemic). TENSE was manipulated within items and ASPECT between items (8 IPVF, 8 PFV). The task of the participants was to select one of 2 pictures (a proxy for epistemic vs. deontic reading) for 16 sentences (plus 8 fillers) in a randomized order. We tested 43 monolingual Czech children in two age groups (3-4-year-olds and 5-year-olds) and 26 monolingual Czech adults. Interviews were conducted with participants during the experiment to gather insights into their choices.

Materials Below is an example of two of the 16 items, representing all the 4 conditions tested.

Present *muset* + perfective / imperfective complement → 3.PRES + PFV / 3. PRES + IPVF

Zajíc musí rychle utéct.	*Micka musí chytat myš.*	
Rabbit must.3SG.PRES fast escape.PFV.INF.	Micka must.3SG.PRES. catch.IPFV.INF mouse.	
'Rabbit must escape quickly.'	'Kitty must be catching a mouse.'	

Past *muset* + perfective / imperfective complement → 3.PAST + PFV / 3. PAST + IPVF

Zajíc musel rychle utéct.	*Micka musela chytat myš.*	
Rabbit must.3SG.PAST fast escape.PFV.INF.	Micka must.3SG.PAST catch.IPVF.INF mouse.	
'Rabbit must have escaped quickly.'	'Kitty must have been catching a mouse.

Results We employed a series of generalized linear mixed models to assess the impact of manipulated variables, including age and tense. Model (i) demonstrated a significant main effect of TENSE ($z = 4.825$, $p < .001$) and an interaction between TENSE and GROUP (adult x child) ($z = 3.278$, $p = .001$), revealing that children aged 3 to 5 had not fully acquired the sensitivity to grammatical tense for modality, unlike adults. Model (ii) showed a main effect of AGE ($z = -3.663$, $p < .001$) and TENSE ($z = 2.249$, $p = .025$), with 5-year-old children favoring epistemic interpretations over 3&4-year-olds. Model (iii), comparing adults to 3&4-year-olds, revealed an interaction between TENSE and AGE ($z = -3.301$, $p = .001$), suggesting that adults, but not children are sensitive to the tense cue. Model (iv) further supported the influence of TENSE ($z = 5.053$, $p < .001$) and its interaction with AGE ($z = -2.356$, $p = .018$). We found no significant effect of ASPECT in the adult model ($z = 0.773$, $p = .44$) or an interaction between TENSE and ASPECT ($z = -0.458$, $p = .65$).
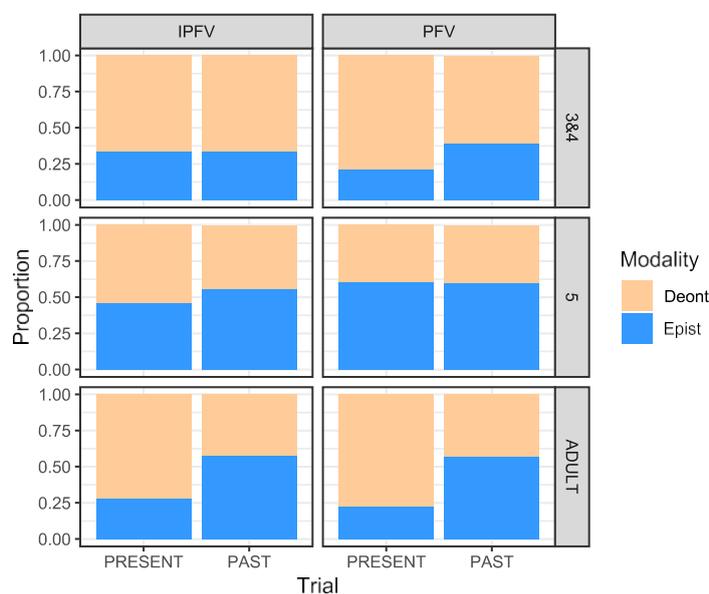
Figure 1. Proportion of deontic and epistemic choices across conditions for each age group.

Discussion The results confirm the past tense hypothesis among adults, making epistemic readings more likely in the past tense. The aspect hypothesis for adults has not panned out. Children exhibit a deontic preference at the age of 3/4, transitioning to overselecting epistemic interpretations by age 5. The qualitative findings suggest that the initial epistemic gap followed by the epistemic overcommitment may be a crosslinguistic phenomenon, with the latter stemming from a pragmatic modal inference. Overall, the study reveals a developmental pattern in which children shift from a deontic preference to overselecting epistemic interpretations and aligns with patterns observed in English and BCS children, suggesting potential crosslinguistic universality in modal acquisition.

References Cournane, A. and Veselinović, D. (2022) 'If they must, they will': Children overcommit to likeliness inferences from deontic modals'. Glossa, 7(1). | Cournane, A. and Pérez-Leroux, A. (2020) 'Leaving obligations behind: Epistemic incrementation in preschool English'. Language Learning and Development, 16(3), pp. 1–22. | Cournane, A. (2021) 'Revisiting the Epistemic Gap: It's not the thought that counts'. Language Acquisition, 28(3), pp. 215–240

Ja or Jaaaaa? The Influence of Iconically Lengthened Response Particles on the Scalar Interpretation of Adjectives

Lennart Fritzsche, Goethe University Frankfurt

Summary. The traditional view that language is arbitrary (Hockett, 1960) has become increasingly challenged recently (e.g., Blasi et al., 2016): Iconic mappings between form and meaning are found throughout language, as for example in iconic modulations of length (Fuchs et al., 2019). In German, it is possible to modulate the length of response particles, particularly in written language, in responses to polar questions containing a gradable adjective. This work presents results from a rating study showing that iconic lengthening of response particles in German directly influences the scalar interpretation of adjectives.

Background. Gradable adjectives such as *long*, *expensive* or *pretty* map their arguments onto abstract representations of measurement, or *degrees*, which are defined as points or intervals ordered along *dimensions* such as length, cost or prettiness. A set of ordered degrees constitutes a *scale* (Kennedy and McNally, 2005). For example, the adjective *long* is a function that maps individuals with (temporal) length to degrees of duration. A speaker uttering *The talk was long* conveys that the length of the talk exceeds a certain standard of length for talks. The same speaker could stress the meaning of the adjective by using a degree modifier, for example *very* as in *The talk was very long*, thereby raising the degree to which the adjective holds of its argument (Kennedy and McNally, 2005). Another way of strengthening the meaning of the adjective would be to lengthen the adjective, as in (1).

(1) *I am normally rather patient. But if the talk is loooong, I'll leave before the end.*
  $\nRightarrow$ if the talk is long, the speaker will leave before the end
  $\Rightarrow$ if the talk is very long, the speaker will leave before the end. (Guerrini, 2020)

This example illustrates that language can make use of iconic means. While traditional views assume that the relation between form and meaning is exclusively arbitrary (Hockett, 1960), this notion has become increasingly challenged recently (see e.g., Blasi et al., 2016). *Iconicity*, i.e., the resemblance between linguistic form and meaning (Dingemanse et al., 2015; cf. (1) where the speaker modulates the form of *long* to modulate its meaning), has been argued to be recognized as a universal feature of natural languages (Perniss et al., 2010). Particularly scalar dimensions such as duration or size are domains of iconic manipulation—which is why the domain of gradable adjectives is of prime interest in research on iconicity. Fuchs et al. (2019) conducted a corpus study on an English social media corpus, investigating iconic lengthening in adjectives in written language. Lengthening was operationalized by means of letter replications, analogous to (1). With a focus on antonyms describing differences in a certain dimension (e.g., spatial: *short-long*), they find that larger dimensions are more likely to be mapped via lengthening, i.e., the occurrence of, for example, *loooong* is more frequent than its small dimension counterpart *shoooort*. To my knowledge, there is no previous work on iconic modulations of *response particles*, such as German *Ja* ('Yes'). In German, instances of particle lengthening are found particularly in online (written) communication. Particle lengthening can be employed in responses to polar questions containing a gradable adjective, as in *Is Maria pretty?*. Here, speakers or texters might respond with a simple *Ja* ('Yes') or choose to lengthen the particle as in *Jaaa*. There is another variant of the positive response particle in German, *Joa* (roughly 'Yup') that expresses mild or uncertain agreement. There is a lack of empirical work on whether these instances of particle lengthening are indeed iconic. It is hypothesized in this work that particle lengthening is employed to convey iconic meaning that influences the scalar interpretation of an adjective. Lengthening of *Ja* is assumed to map the argument of an adjective higher on the scale, while lengthening of *Joa* should result in an inverse effect.

Experimental study. An experimental rating study was conducted. Participants were presented short chat conversations between two interlocutors A and B. In these, A would always ask a polar question containing a gradable adjective denoting properties that apply to persons, cf. (2).

(2) A: *Darf ich dich was fragen?* ('Can I ask you something?')
  B: *Klar, schieß los.* ('Of course, shoot.')
  A: *Findest du Maria schön?* ('Do you find Maria pretty?')

B always answered with a response particle (*Ja* or *Joa*) that was manipulated corresponding to a 3x2

design with the factors LENGTH (short vs. medium vs. long) and VOWEL (/a/ vs. /oa/), cf. (3). The length manipulation was achieved through letter replication: response particles in the short conditions had one vowel, medium conditions had three vowels and long conditions five (/oa/ was treated as one vowel in the short condition, for the medium and long conditions only the /o/ was replicated). The response particle was always follwed with a counter question (*Und du?* ('How about you?')) to increase the naturalness of the chat conversations.

(3) B1: *Ja - und du?*   B3: *Jaaaaa - und du?*   B5: *Joooa - und du?*
     B2: *Jaaa - und du?*   B4: *Joa - und du?*   B6: *Joooooa - und du?*

The task was to rate the extent to which B believes that the property of the adjective applies to the subject of the polar question (e.g., *Wie schön denkt B ist Maria?* ('How beautiful does B think Maria is?')) by means of a slider ranging from 1 (*Überhaupt nicht* ('Not at all')) to 101 (*Wahnsinnig* ('Insanely')). An interaction of the two factors was predicted since the effect of *vowel* on the rating was assumed to vary with different levels of *length*. For the *Ja* conditions B1, B2, and B3 (cf. (3)), higher ratings were predicted with increasing particle length, while for the *Joa* conditions B4, B5 and B6 it should be inverse, and ratings should be lower the longer the particle. Additionally, it was predicted that overall, the property should be rated higher for all *Ja* conditions as opposed to all *Joa* conditions. 24 items were distributed across six lists according to a Latin square design and interspersed by 24 filler items. 78 native speakers of German participated in the study. The results show that ratings for the *Ja* conditions were higher with increasing particle length (*Ja*: mean = 77.45, SD = 13.70; *Jaaa*: mean = 83.88, SD = 14.96; *Jaaaaa*: mean = 88.54, SD = 15.44). Ratings were lower with increasing particle length in the *Joa* conditions (*Joa*: mean = 49.93, SD = 16.79; *Joooa*: mean = 46.46, SD = 18.67; *Joooooa*: mean = 43.27, SD = 22.10). An ordinal mixed-effects model was fitted onto the data, specifying vowel and length:vowel as fixed and participants and items as random effects. It yielded significant interactions, cf. Table 1.

Table 1: Ordinal mixed-effects model with vowel and length:vowel as fixed effects, and participants and items as random effects.

	Estimate	Std. Error	z value	Pr(>|z|)
Vowel	-4.2218	0.1242	-33.993	<2e-16 ***
/a/:length - long vs. medium	0.8697	0.1456	5.972	2.35e-09 ***
/oa/:length - long vs. medium	-0.3012	0.1359	-2.217	0.0266 *
/a/:length - medium vs. short	0.9108	0.1395	6.529	6.60e-11 ***
/oa/:length - medium vs. short	-0.3282	0.1317	-2.492	0.0127 *

Discussion. The results indicate that iconic lengthening of response particles has a direct influence on the scalar meaning of adjectives. Strengthened meaning can be expressed iconically via the amount of letters used in response particles. Interlocutors can lengthen the *Ja* to the extent that they think an individual is placed on a scale formed by an adjective, with longer length iconically expressing higher positioning on the scale. Inversely, the *Joa* can be lengthened to express that the interlocutor believes the individual to be placed on lower positions on the constituted scale.

References

Blasi, Wichmann, Hammarström, Stadler, & Christiansen. (2016). Sound–meaning association biases evidenced across thousands of languages. *PNAS*, *113*(39).

Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *TGS*, *19*(10).

Fuchs, Savin, Solt, Ebert, & Krifka. (2019). Antonym adjective pairs and prosodic iconicity: Evidence from letter replications in an english blogger corpus. *LV*, *5*(1).

Guerrini. (2020). Vowel quality and iconic lengthening. *Proceedings of SuB*, *24*(1).

Hockett. (1960). The Origin of Speech. *SA*, *203*(3).

Kennedy & McNally. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*.

Perniss, Thompson, & Vigliocco. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *FP*, *1*.

(Un)marked indefinites in Russian and Bulgarian:
An experimental investigation

Luca Molinari (University of Warsaw / Ca' Foscari University of Venice) and Daria Seres (University of Graz)

1. Introduction & RQs | The distribution of bare vs non-bare NPs and the underlying (in)definite nature of NPs (connected with the (non-)obligatoriness of (in)definiteness marking) in languages with and without articles is subject to continuous debate (e.g., Chierchia, 1998; Dayal, 2004; Šimík & Demian, 2020; Seres & Borik, 2021, i.a.). We contribute to this ongoing discussion by bringing new data collected through an experimental investigation using a gap-filling task for Russian (articleless) and Bulgarian (definite article only). The aim of our research is twofold: (i) verify the extent to which (in)definite marking is optional/available to Russian and Bulgarian speakers; (ii) test the reliability of the gap-filling task applied to the issue under investigation (to the best of our knowledge, such task has never been used in related issues). Our RQs are as follows: What is the choice of the NP form (bare vs. overtly marked as (in)definite) for a discourse-new referent in subject position in Russian and Bulgarian? What are the factors that affect this choice (e.g., type of the sentence (existential vs non-existential), linear word order (preverbal vs postverbal subject), modification by a relative clause)? Is there optionality (in Sorace's (2000) terms) between bare vs overtly marked NPs? Can data collected via a gap-filling task be reliable?

2. Methods | The experiment was administered online by means of the LimeSurvey software and consisted of 24 experimental items (4 x 6 item types) + 12 filler items with only gap-filling tasks. Anonymous Russian (n=85) and Bulgarian (n=81) L1 speakers were recruited on social networking sites. They were instructed to fill in the gap before the subject NP with any suitable expression only if necessary (otherwise they should leave the gap blank). The NPs were discourse-new, not easily context-inferable, and non-topical (which was supposed to trigger an indefinite reading). All experimental items had intransitive and imperfective verbs, producing the following types of sentences: (i) existential, (ii) thetic with postverbal subject, (iii) thetic with preverbal subject. The subject in each condition was either unmodified or modified by a relative clause. This resulted in 6 types of items, as illustrated (only for Russian for the reasons of space) below: (1) Existential_MOD; (2) Existential; (3) Postverbal_MOD; (4) Postverbal; (5) Preverbal_MOD; (6) Preverbal.

(1) Žila-byla v sele ___ *staruxa*, kotoraja sobirala koren'ja i list'ja. Nikto ne ponimal, dlja čego oni byli nužny.
'In the village, there lived an old lady, who gathered roots and leaves. Nobody understood why they were needed.'
(2) Davnym-davno v zamke žila ___ *koroleva*. Ona byla nevežliva so vsemi.
'A long time ago there lived ___ queen in a castle. She was impolite to all.'
(3) Tvoi druz'ja črezvyčajno volnovalis'. Počemu? -Na koncerte vystupal ___*pevec*, kotoryj nedavno stal izvestnym.
'You friends were extremely excited. Why? - ___ singer who recently became famous performed in the concert.'
(4) Tvoja koška volnovalas'. Počemu? - Na ulice lajala ___ *sobaka*.
'Your cat was worried. Why? - In the street, ___ dog was barking.'
(5) Pered cerkov'ju sobralas' kuča ljudej. Čto tam bylo? -___*svjaščennik*, kotoryj prijexal iz Vatikana, propovedoval.
'A crowd gathered in front of the church. What was happening there? - ___ priest who came from the Vatican was preaching.'
(6) Vse v ofise smotreli v okno. Počemu? -___ *mužik* ležal posredi dorogi.
'Everybody in the office was looking outside of the window. Why? - ___ man was lying in the middle of the road.'
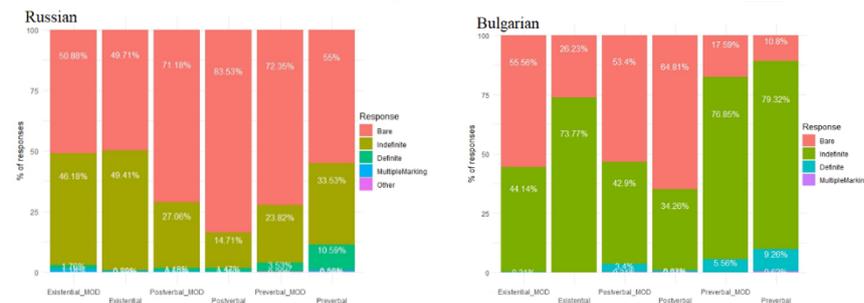
3. Results | The results are summarized in the plots which represent the percentage of different markings for each item type. As expected, Russian displays an overwhelming preference for bare NPs in all the contexts (oscillating from 49.71% to 83.53%), with a tendency to optionally mark the NP overtly in existential sentences

(with almost 50% of indefinite marking). Fisher's Exact Test revealed that the difference in the overt marking between Preverbal and Preverbal_MOD is significant (*p*<.001) as well as the difference between Postverbal and Postverbal_MOD (*p*=.001). However, the difference Postverbal_MOD vs Preverbal_MOD is not significant (*p*=.081), while Postverbal vs Preverbal is (*p*<.001). Bulgarian privileges instead overt marking in many contexts, although bare NPs prevail in Existential_MOD and in postverbal position. Significant differences are found between Existential vs Existential_MOD (*p*<.001), Postverbal vs Postverbal_MOD (*p*<.001), as well as for Preverbal vs Preverbal_MOD (*p*=.011). The differences between preverbal and postverbal position are also significant (with modification: *p*<.001; without modification: *p*<.001). In both languages, the difference in marking between all sentences with preverbal vs postverbal subject is significant (*p*<.001).

4. Discussion | The experimental results have shown that Russian L1 speakers prefer bare NPs in any context, while Bulgarian L1 participants predominantly use them only for postverbal subjects of thetic sentences and for modified subjects of existential sentences. In other cases, the overtly marked indefinite is used in Bulgarian. In Russian, the type of sentence (and the discourse prominence of the referent) seems to play a role in the use of overtly marked indefinites: they are more likely to appear in existential sentences and in sentences with preverbal subjects. In Bulgarian, NPs in preverbal subject position are most likely to get overt indefinite marking, followed by existential sentences, where modification increases the likelihood of bare NPs. Moreover, Russian demonstrates a higher degree of optionality of indefiniteness marking, which is instead strongly preferred in Bulgarian (with limited optionality). This gives empirical support for the hypothesis proposed in the literature (Šimík & Demian, 2020; Seres & Borik, 2021, i.a.) that bare NPs are inherently indefinite, or at least that they are compatible with the indefiniteness reading or underspecified for (in)definiteness (Chierchia, 1998; Geist, 2010, i.a.), however, the definite interpretation of bare NPs in Russian is never blocked. The relatively low use of overt indefinites also indicates that indefiniteness marking cannot be obligatory in Russian (*contra* Dayal, 2004). This finding is in line with the results of the recent experimental study by Seres et al. (2023) for bare vs. non-bare NPs in bridging contexts. At the same time, we also have evidence that overt indefinite marking is available (although to a different extent) in all contexts. This indicates a limited degree of optionality, hinting at the fact that bare NPs may be headed by a null indefinite.

As Bulgarian has a definite article, the definite interpretation of a bare NP is blocked. There is instead a competition between bare and overtly marked indefinites in the indefiniteness domain only. As pointed out in Gorishneva (2016), a bare NP denotes a non-specific/non-identifiable referent, while a 'one'-marked indefinite denotes a specific one. The preverbal position may favor a higher degree of discourse salience and, thus, potential identifiability. In Russian, the use of 'one'-marked indefinites was rather high in existential sentences (corresponding to almost the totality of "indefinite" answers), which can be explained by the identifiability condition (cf. Ionin, 2013). 'One' as a specificity marker (Pereltsvaig, 2007; Geist, 2008; Borik, 2016) can only appear on NPs in contexts where the referent can be identified by the speaker, i.e., the referent is *potentially* epistemically specific (in terms of von Heusinger, 2002; Ihsane, 2021).

5. Conclusions & further research | This study has shown that a gap-filling task is a powerful tool to investigate the distribution of (non-)bare NPs in languages belonging to the same group (Slavic) but being typologically different (no article vs definite article only) as the results are in line with the expectations (e.g., the overwhelming use of bare NPs and use of 'one' in existential contexts in Russian, and the use of overtly marked indefinites in Bulgarian). The reliability of the data lies in the fact that the "bare" option was presented as a default, i.e., overt marking was deemed necessary by the subjects that filled out the blank.

Future research will apply this method to other Slavic languages (starting with Ukrainian and Polish). Furthermore, an experimental study which measures the acceptability of bare vs. non-bare NPs is planned.

References (selected) | **Chierchia, G. (1998).** Reference to Kinds across Language. *Natural Language Semantics* 6, 339–405. ● **Dayal, V. (2004)**. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, *27*(4), 393–450. ● **Geist, L. (2010)**. Bare Singular NPs in Argument Positions. *International Review of Pragmatics* 2, 191–227. ● **Seres, D. & Borik, O. (2021)**. Definiteness in the absence of uniqueness, in Advances in Formal Slavic Linguistics 2018, pp. 339–363. **Šimík, R., & Demian, C. (2020)**. Definiteness, Uniqueness, and Maximality in Languages With and Without Articles. *Journal of Semantics*, *37*(3), 311–366. ● **Sorace, A. (2000)**. Syntactic optionality in non-native grammars. *Second Language Research*, *16*(2), 93–102.

Experimenting with principle C in German ATB movement
Timea Szarvas, University of Potsdam

Claim: We report on a co-reference judgment study of German ATB constructions. Our preliminary results (i) show that German ATB movement reconstructs asymmetrically based on principle C violations only in the initial gap, (ii) supporting a derivation where material is extracted from the initial conjunct and elided in the non-initial one, obviating the violation via vehicle change. The relatively low magnitude of the effect prompts a discussion about suitable experimental designs to be tested in follow-up studies.

Background: Approaches to ATB movement predict (a-)symmetries in syntactic, morphological and semantic phenomena between the conjuncts based on how the filler is related to each gap. Parallel extraction approaches (Williams 1978; Hein & Murphy 2020) and multidominance (Citko 2005) predict symmetrical reconstruction, while the predictions of asymmetrical approaches depend on the operation applied in the conjunct without extraction. The evidence is mostly

	Initial	Non-initial
Multidom; parallel extr.	✓	✓
Sideward mvmt; non-initial extr.	✗	✓
Initial extr. + ellipsis	✓	✗

Table 1: Reconstruction predictions.

based on individual judgments. There exists experimental work on the repair of case mismatches by syncretism for Polish (Rothert 2022) and German (though against it, by Hartmann et al. 2016). In English, variable binding, idioms and strong crossover are reported to reconstruct symmetrically, but principle A and C as well as weak crossover induce violations only in the initial gap (based on individual judgments, Citko 2005; Salzmann 2012). Focusing on principle C, extracted objects with an R-expression should evoke disjoint reference with the pronominal subject c-commanding their base position:

(1) a. She$_{*i/j}$ read the story about Mary$_i$.
     b. [Which story about Mary$_i$] did she$_{*i/j}$ read ___$_i$?

Principle C is argued to be a soft constraint without a pronounced argument-adjunct asymmetry in German (Salzmann et al. 2022).[1] Further, only the base position of the R-expression (or pronoun) seems to matter, i.e. if the configuration arises in the surface or an intermediate position, it does not lead to a principle C violation (Nissenbaum 2000; Sportiche 2017). This allows us to identify where the filler originates, and it is precisely where predictions about ATB constructions diverge, with accounts positing movement from all gaps (Ross 1967; Williams 1978; Citko 2005; Hein & Murphy 2020), either gap (Munn 1992; An 2007; Ha 2008; Zhang 2010; Salzmann 2012; Larson 2013), or from one through the other (Nunes 2001). Rather than proposing yet another theory, we thus argue that the evidence needs to be tested experimentally to evaluate its robustness, simultaneously exploring the possibility that there is cross-linguistic variability.

Study: 300 participants were recruited for an online co-reference judgment study. They were shown 12 target items and 36 distractors. In the target items, a pronoun preceded either the initial or non-initial gap and matched the ϕ-features of the R-expression in the extracted *wh*-phrase. The dependency was embedded into a matrix clause with a referent also matching the pronoun. Successful reconstruction should reverse c-command relations only in conditions where the extracted element is an object:

(2) Ich habe Marie gefragt, *welche Geschichte über **Laura**…*
     a. ___ **sie** entzückt und ___ Michael überrascht hat. SUBJECT, INITIAL
         her delighted and   Michael surprised has
     b. ___ Michael überrascht und ___ **sie** entzückt hat. SUBJECT, NON-INITIAL
         Michael surprised   and   her delighted has.
     c. **sie** ___ gehört und Michael ___ weitererzählt hat. OBJECT, INITIAL
         she   heard and Michael   passed.on   has

---

[1] C-command does not guarantee co-reference (Büring 2005; Cunnings et al. 2014; Kush et al. 2015) and binding principles can be circumvented if non-syntactic requirements are met (Cunnings et al. 2015), emphasizing the need for experimentation.

d.   Michael ___ weitererzählt und **sie** ___ gehört hat. *OBJECT, NON-INITIAL*
       Michael   passed.on   and she   heard has
       'I asked Marie which story about Laura delighted her and surprised Michael.'
       'I asked Marie which story about Laura surprised Michael and delighted her.'
       'I asked Marie which story about Laura she heard and Michael passed on.'
       'I asked Marie which story about Laura Michael passed on and she heard.'

All items were preceded by a context sentence introducing all three referents to avoid making either more prominent than the others. In each trial, participants had to answer two binary choice questions: whether the pronoun could refer to the matrix and the embedded referent, respectively. The matrix referent was included to avoid a bias towards the embedded referent due to the lack of an alternative (Gordon & Hendrick 1998; cf. Adger et al. 2017). A strong bias in favour of the matrix referent was expected in a forced choice task between the matrix and embedded referent (Bruening & Al Khalaf 2019; vs. Stockwell et al. 2021, 2022). Non-syntactic factors largely influence co-reference, including information structure (Cowles et al. 2007; Kaiser 2011), subjecthood (Kaiser 2011) and mention-first (Järvikivi et al. 2005). The matrix referent was thus not a subject nor a topic. Reconstruction in object conditions should yield a principle C violation, inducing disjoint reference with the pronoun preceding the respective gap.

Results and discussion: Modeling will be carried out in the Bayesian framework using a categorical model from the brms package (Bürkner 2021). The raw data indicate reconstruction to the initial gap based on the proportion to which co-reference between the pronoun preceding the initial gap was allowed in object vs. subject conditions. This contrast is absent in the non-initial gap. Approaches positing extraction from the initial gap predict this pattern, though only if the material in the non-initial conjunct is subject to ellipsis and thus allows for vehicle change (Salzmann 2012; Fiengo & May 1994). Though the same reconstruction pattern is predicted if there is an empty operator (Munn 1992) or no syntactic material present at all (Larson 2013), differences between parasitic gaps and ATB movement make the former approach unlikely (Salzmann 2012), while various case and agreement phenomena sensitive to the non-initial gap are problematic for the latter. PF-deletion is likewise out due to not allowing for vehicle change (An 2007). Nevertheless, the contrast between subject and object conditions for the initial gap is numerically too small (0.10) for a purely syntactic explanation to be viable. Further, as for the non-initial gap, syntactic factors cannot explain that participants seemed to allow co-reference with the embedded referent at chance level in both the subject and object condition. If this was a matter of (the lack of) reconstruction, we would expect higher levels of co-reference. Assuming that the experimental design was too complex, we need to (a) eliminate confounding factors that reduce co-reference with the subject in the initial gap condition and (b) examine why contrasts vanish in the non-initial gap entirely, yielding chance level performance.

Coreference with embedded referent

Outlook: We will pilot two designs, suspecting interference from the matrix referent to be the cause of (a) and partially also of (b). In one study, we will omit the question about the matrix referent. In the other, we will also leave out the embedding containing it, providing a neutral context, and offering a binary choice between the embedded referent or 'someone else' (cf. Stockwell et al. 2021, 2022). Addressing the effect of distance potentially causing (b), we will carry out the same study on parasitic gaps where linear and structural distance from the filler can be distinguished by manipulating extraposition. To avoid lexical effects, we will contrast patterns like 'X delighted Y' with 'Y found X delightful', X being the filler with the R-expression, Y the pronoun, instead of using distinct verbs (cf. Salzmann et al. 2022). Crucially, due to the impact of non-syntactic factors, an adequate level of (experimental) control is necessary to make principle C reconstruction the reliable and powerful diagnostic of movement that it can be.

Selected references: Adger, D. et al. 2017. Is there Condition c reconstruction? * An, D. 2007. Asymmetric T-to-C movement in ATB constructions. * Bruening, B. & E. Al Khalaf. 2019. No argument-adjunct asymmetry in reconstruction for Binding Principle C. * Gordon, P. C. & R. Hendrick. 1998. The representation and processing of coreference in discourse. * Hartmann, J. et al. 2016. On the limits of non-parallelism in ATB movement. * Larson,
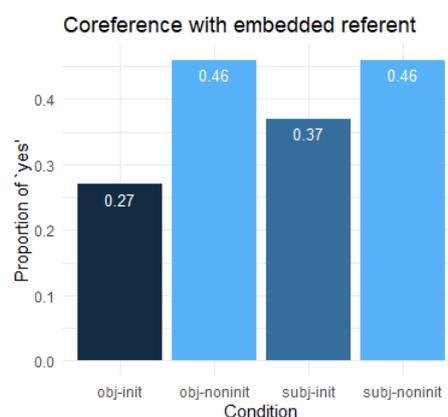
B. 2013. The syntax of non-syntactic dependencies * Munn, A. 1992. A null operator analysis of ATB gaps. * Nissenbaum, J. 2000. Investigations of covert phrase movement. * Nunes, J. 2001. Sideward movement. * Rothert, J. 2022. An experimental investigation of syncretism and proximity effects in Polish ATB topicalization and RNR. * Salzmann, M. 2012. A derivational ellipsis approach to ATB-movement. * Salzmann, M. et al. 2022. Condition C in German A'-movement. * Stockwell, R. et al. 2021. There is reconstruction for Condition C in English questions. * Stockwell, R.et al. 2022. Experimental evidence for the Condition C argument-adjunct asymmetry in English questions. * te Velde, J. R. 2005. Deriving coordinate symmetries.

Acquisition of auxiliary selection in French and Italian

Balthazar Lauzon, Raffaella Folli, Juliana Gerard, Christina Sevdali

In a variety of languages, two auxiliaries, equivalent to *be* and *have*, are used in perfect tenses. The choice of either auxiliary is termed *auxiliary selection*. In French and Italian, *have* is used with all transitive verbs, while *be* is used with reflexives (1a) and a subset of intransitives: unergative verbs, typically denoting activities are used with *have* (1b), while unaccusative verbs typically denoting a change in location or state take *be* (1c). Italian has a much larger number of *be*-selecting unaccusatives than French.

(1) a. Jean **a** nagé dans la piscine.   b. Le train **est** arrive   c. La branche **s'est** cassée
     J. has swum in the pool.   The train is arrived   The branch is broken
     'J. swam in the pool.'   'The train has arrived'   'The branch broke'

Prior research in French has argued that children are initially accurate in their use of *be* in reflexive constructions [1], while fundamental differences have also been observed between intransitives and reflexive verbs [2], including a difference between 1st and 3rd person forms with reflexives.

The present study aims to better understand the acquisition of auxiliary selection in French and Italian while accounting for the role of language input. We analysed auxiliary use in 18 French-speaking children (1;0-4;0) and corresponding parental speech in the Leveillé, Champaud, Lyon, Paris, York, and Geneva corpora [3–8] from CHILDES [9]. We also analysed auxiliary use for 9 Italian-speaking children from the Antelmi, Calambrone and Tonelli corpora [10–12]. We counted unambiguous auxiliary-past participle combinations (1a-c). French verbs were categorized as categorical unaccusatives (*aller* 'go'), variable unaccusatives (*tomber* 'fall'), reflexives (*se tromper* 'make a mistake') and unergative verbs (*jouer* 'play'). Italian verbs were categorized as unaccusative (all *be* intransitives), reflexive or unergative.

Our results first confirm that French-speaking children make errors with a significant percentage of *be*-selecting verbs (reflexive and unaccusative) but do not overuse *be* with *have*-selecting intransitives (unergatives) (Table 1). The rate of non-adult-like auxiliaries is highest in reflexive constructions (20%), and lowest with categorical unaccusatives (7%), with an intermediate rate for variable unaccusatives (13%). In Italian, children are generally very accurate with both *be* and *have*-selecting verbs (Table 2). Secondly, we analysed parental speech to determine whether the observed errors are explained by input variability. This is important since in at least some varieties of spoken French, *have* is used with verbs that require *be* in standard French [13]. The results show that parents use *be* with all unaccusative and reflexive verbs (Table 3), which means that the observed errors in child speech (Table 1) are innovations. The lack of variability may be due to the linguistic varieties represented in the corpora or could be explained by the high socio-economic status or level of education of these families. Finally, we take a closer look at the distribution of child non-adult-like productions in French (Table 4). With reflexives, errors are very frequent with 1-SG auxiliaries (70% of present 1-SG verbs) but are not observed in other persons; with unaccusatives, errors are frequent with 1-SG (41% of 1-SG verbs) and much lower with 3-SG and 3-PL (5 and 17%). The higher rates with 1-SG may be due to another source of error: since 1-SG *have* and 3-SG *be* are homophonous [e] in spoken French, 1-SG errors may represent 3-SG *be* auxiliaries extended to 1-SG verbs, rather than true *have/be* selection errors. We also consider the significance of these results in a diachronic context where acquisition is a driver of language change, since French (but not Italian) may be in a process of replacement of *be* with *have* [14]. Our results suggest that there may additionally be a pressure towards levelling of the *be* paradigm in French through a process of analogy.

[1] Snyder W, Hyams N and Crisma P 1995 Romance auxiliary selection with reflexive clitics: Evidence for early knowledge of unaccusativity *Proceedings of the 26th annual child language research forum* pp 127–36
[2] Boyce V, Aravind A and Hackl M 2017 Lexical and Syntactic Effects on Auxiliary Selection: Evidence from Child French *Proceedings of the 41st annual Boston University Conference on Language Development*
[3] Leveillé M and Suppes P 2004 CHILDES French Leveillé Corpus
[4] Champaud C 2004 CHILDES French Champaud Corpus
[5] Demuth K and Jisa H 2004 PhonBank French Lyon Corpus

[6] Morgenstern A and Parisse C 2009 PhonBank French Paris Corpus

[7] Rasetti L, Hamann C, Frauenfelder U, Rizzi L and Zesiger P 2004 CHILDES French Geneva Corpus

[8] De Cat, C. & B. Plunkett 2002. CHILDES French York Corpus

[9] MacWhinney B 2000 *The CHILDES project: The database* vol 2 (Psychology Press)

[10] Antelmi D 2004 CHILDES Italian Antelmi Corpus

[11] Cipriani P and Cappelli G 2004 CHILDES Italian Calambrone Corpus

[12] Tonelli L 2004 CHILDES Italian Tonelli Corpus

[13] Sankoff G 2019 Language change across the lifespan: Three trajectory types *Language* **95** 197–229

[14] Vincent N. 1982, The development of the auxiliaries habere and esse in Romance.

Table 1. Child auxiliary use in French intransitive and reflexive verbs. AL: adult-like, NAL: non-adult-like

child	categorical unaccusative		variable unaccusative		reflexive		unergative	
	AL	NAL	AL	NAL	AL	NAL	AL	NAL
Adrien	5	0	16	1	4	0	28	0
Anae	31	5	24	2	26	0	41	1
Anais	7	0	38	5	2	0	64	0
Anne	13	0	48	2	1	2	47	0
Antoine	15	0	18	7	9	5	48	0
G-Marie	4	0	2	3	1	0	9	0
Gregoire	14	0	48	5	8	7	16	0
Julie	5	0	1	1	3	0	7	0
L-Marie	22	4	21	1	8	2	46	0
Lea	72	0	13	0	32	10	90	1
Leonard	15	1	10	1	4	1	17	0
Madeleine	34	5	36	6	16	8	69	0
Marilyn	1	0	18	0	3	1	9	0
Max	12	0	7	2	6	4	14	0
Nathan	4	1	1	0	0	0	7	0
Philippe	80	5	59	9	42	3	63	0
Theophile	13	3	11	9	11	7	46	0
Theotime	19	3	55	8	25	0	34	0
Total	366	27	426	62	201	50	655	2
Percent	93	7	87	13	80	20	100	0

Table 2. Child auxiliary use in Italian intransitive and reflexive verbs. AL: adult-like, NAL: non-adult-like.

child	unaccusative		reflexive		unergative	
	AL	NAL	AL	NAL	AL	NAL
Cam	5	0	2	0	14	0
Diana	26	3	14	1	20	1
Elisa	5	0	5	0	20	0
Guglielmo	22	0	13	0	11	0
Marco	31	0	14	0	15	1
Martina	1	0	3	0	4	0
Raffaello	18	1	2	0	3	0
Rosa	6	0	2	0	5	0
Viola	2	0	0	0	7	0
Total	116	4	55	1	99	2
Percent	97	3	98	2	98	2

Table 3. Auxiliary use in French unaccusative and reflexive verbs in parental speech.

child	categorical unaccusative		variable unaccusative		reflexive	
	be	have	be	have	be	have
Adrien	85	0	30	0	49	0
Anae	104	0	56	0	66	0
Anais	81	0	82	0	82	0
Anne	23	0	8	0	16	0
Antoine	44	0	29	1	58	0
G-Marie	13	0	0	0	3	0
Gregoire	27	0	21	0	22	0
Julie	35	0	30	0	12	0
L-Marie	91	0	76	0	86	0
Lea	17	0	2	0	12	0
Leonard	45	0	16	0	45	0
Madeleine	77	0	26	0	44	0
Marilyn	20	0	20	0	23	0
Max	19	0	4	0	5	0
Nathan	64	0	50	0	65	0
Philippe	163	0	48	0	91	0
Theophile	101	0	33	2	97	0
Theotime	124	0	65	0	88	0
Total	1,133	0	596	3	864	0

Table 4. Distribution of child present tense auxiliaries with be-selecting verbs by person in French (non-adult-like productions shaded).

Verb type	person and auxiliary	counts	% per person	% per verb type
unaccusative	1sg be	58	59%	7%
	1sg have	41	41%	5%
	2sg be	11	100%	1%
	2sg have	0	0%	0%
	3sg be	633	95%	75%
	3sg have	34	5%	4%
	3pl be	55	83%	7%
	3pl have	11	17%	1%
reflexive	1sg be	21	30%	9%
	1sg have	49	70%	20%
	2sg be	14	100%	6%
	2sg have	0	0%	0%
	3sg be	159	100%	65%
	3sg have	0	0%	0%
	3pl be	2	100%	1%
	3pl have	0	0%	0%

Production and comprehension of case marking after local two-case prepositions in German-speaking preschoolers

Tanja Diederich and Flavia Adani – Free University of Berlin

The adult-like use of case marking on definite articles poses challenges to children learning German as their first or second language (e.g., Meisel, 1986; Ulrich et al., 2016), particularly after the so-called local two-case prepositions (PP). Contrary to local one-case PP, which assign a single case (either accusative or dative) to the following noun phrase, two-case PP allow the use of both, which leads to different sentence meanings (see Fig. 1). These PP can be embedded in a canonical or non-canonical word order in German sentences (see Fig. 1). There are inconsistencies in previous research in regard to a) which case is acquired first: accusative (Turgay, 2011) or dative (e.g., Parodi, 1990) or if both cases are mastered simultaneously (e.g., Meisel, 1986), b) if case marking after one- and two-case PP emerges at the same time (Turgay, 2011) or if it emerges first after one-case PP and later after two-case PP (Baten & Willems, 2012), c) whether children first develop the comprehension of a grammatical phenomenon or if production precedes comprehension with more consensus on comprehension preceding production (e.g., Lovell & Dixon, 1967). Sentences which do not follow the typical German word order are generally more difficult for children than canonical ones (Schipke et al., 2012). Given this fragmented picture, we designed a sentence scene verification task (comprehension) and a sentence completion task (production) to test children's ability to understand and mark case after two-case PP in German.

We manipulated CASE (accusative vs. dative), PP (one- vs. two-case preposition), WORD ORDER (**s**ubject-**v**erb-**o**bject (SVO) vs. **o**bject-**v**erb-**s**ubject (OVS)) and MODALITY (production vs. comprehension). Eighty-three German-speaking preschoolers (M_{age}=5;7, range=4;7-6;5) watched short video clips while interacting with a puppet. In the comprehension task, the child had to tell the puppet if the auditorily presented sentence matches the video. To test the child's ability to produce case marking, it had to complete the sentence that the puppet had started (e.g., The chick walks into…). We assume that children will be more accurate in accusative case, after one-case PP, in SVO-sentences, and in comprehension.

All responses (coded as "1" when adult-like and "0" otherwise) entered a GLMM (see Tab. 1 for model specifications). We found main effects of CASE (accusative > dative, whereby ">" means more accurate, $p<0.0001$), PP (one > two, $p<0.0001$) and MODALITY (comprehension > production, $p<0.05$), but not of WORD ORDER. Moreover, a WORD ORDER by CASE interaction depicts that children are more accurate in accusative case in SVO-sentences ($p<0.01$, see Fig. 2).

Acquiring the German case system is a long-lasting process, which seems to proceed stepwise and is not completed by school entry, especially dative marking. Our results are consistent with claims that children have difficulties in marking case at least up to the age of 5;7 and that two-case PP are more difficult for children than one-case PP. These results support the prediction of comprehension preceding production.

Fixed factors	Model specification
case	
type of preposition	accuracy~case*pp*word_order*modality+(1+case|subject_id)+(1+case|item)
word order	
modality	

Tab. 1. Model used for statistical analysis.

Das Küken	läuft	in	die	Box. (SVO)	Das Küken	läuft	in	der	Box. (SVO)
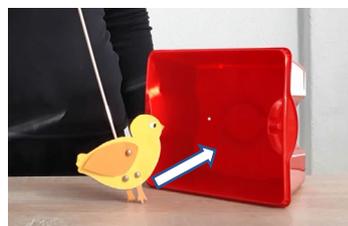
Das Küken läuft in die   Box. (SVO)   Das Küken läuft in der   Box. (SVO)
The chick walks into the-ACC box.   The chick walks inside the-DAT box.

In die   Box läuft das Küken. (OVS)   In der   Box läuft das Küken. (OVS)
Into the-ACC box walks the chick.   Inside the-DAT box walks the chick.
'The chick walks into the box.'   'The chick walks inside the box.'

Fig. 1. Example of a visual stimulus and target sentences used in the experiment.
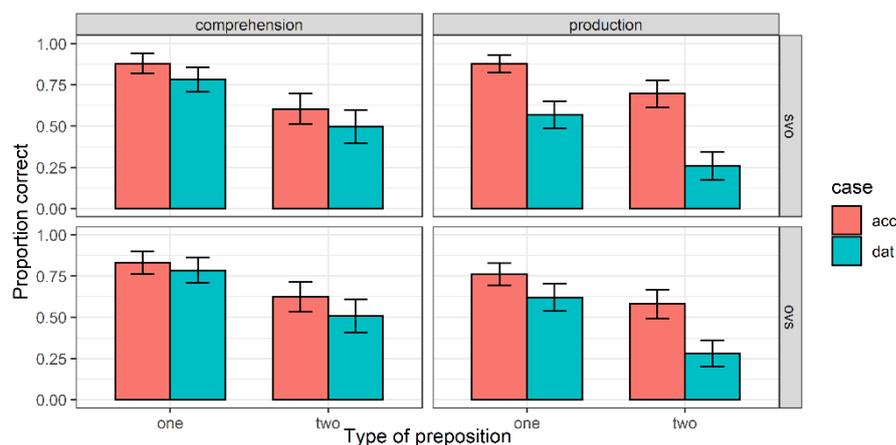
Fig. 2. Mean accuracy on understanding and marking accusative and dative case after local one-case and two-case prepositions in SVO- and OVS-sentences.

acc=accusative   dat=dative   one=one-case preposition
two=two-case preposition SVO=subject-verb-object word order   OVS=object-verb-subject word order

References

Baten, K. & Willems, K. (2012). Kasuserwerb in der Präpositionalphrase vom Standpunkt der Verarbeitbarkeitstheorie. *Deutsche Sprache*, *40(3)*, 221–239.

Lovell, K. & Dixon, E. M. (1967). The growth of the control of grammar in imitation, comprehension, and production. *Journal of Child Psychology and Psychiatry, 8(1)*, 31– 39.

Meisel, J. (1986). Word order and case marking in early child language. Evidence from simultaneous acquisition of two first languages: French and German. *Linguistics, 24*, 123- 183.

Parodi, T. (1990). The acquisition of word order regularities and case morphology. In J. Meisel (ed.), *Two first languages: early grammatical Development in Bilingual Children* (S. 157-190). Dordrecht: Foris.

Schipke, C. S., Knoll, L. J., Friederici, A. D. & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: neural correlates of their behavioral performance. *Developmental Science, 15(6)*, 762–774.

Turgay, K. (2011). Der Zweitspracherwerb des deutschen Kasus in der Präpositionalphrase. G*ermanistische Linguistik, 39(1)*, 24–54.

Ulrich, T., Penke, M., Berg, M., Lüdtke, U. & Motsch, H.-J. (2016). Der Dativerwerb – Forschungsergebnisse und ihre therapeutischen Konsequenzen. *Logos, 24(3)*, 176- 190.

Agreement attraction effects depend on the goal of processing
Anna Laurinavichyute, Himanshu Yadav, Titus von der Malsburg, Shravan Vasishth

When people read ungrammatical sentences like (1a) and (1b), their reading times at the verb are consistently faster in (1a) compared to (1b):

> Example 1: (a) *The key to the cabinets are rusty.
> (b) *The key to the cabinet are rusty.

This speedup is referred to as agreement attraction and attributed to the presence of the plural attractor, *cabinets,* that matches the number of the verb [1, 2]. However, the opposite effect is observed in acceptability judgment times [3] and in the decision tasks where participants have to select a correct verb form [4, 5, 6]: The judgment/decision times are slower for the number mismatch condition (1a) compared to (1b). One potential explanation for the opposite effects is that the act of judging sentence acceptability requires additional processes on top of those involved in reading, and those additional processes lead to the reversal of the effect. Another potential explanation that we pursue here is that the observed effect depends not on the particular action but rather on the participants' processing goal. The key role of goal for the processing of agreement attraction errors has been first suggested by [6] and further corroborated by [7], who showed that for grammatical sentences, a slowdown in number mismatch condition is present when participants expect to judge sentence acceptability and disappears when participants expect to answer comprehension questions. For the ungrammatical sentences tested here, we predict a similar, judgment-typical, slowdown in number mismatch condition, (1a) vs. (1b), when participants expect to judge sentence acceptability. When participants expect to answer comprehension questions, we predict the well-attested attraction effect, i.e., a speedup in number mismatch condition (1a) vs. (1b).

   Methods. Two web-based experiments were conducted using the same set of stimuli: 16 experimental items in two conditions (attractor matching/mismatching the subject in number), such as Ex. (2). Experiments differed only in the training procedure: in Experiment 1, participants saw three simple training sentences (two well-formed and one with an apparent agreement error), and were asked to rate their acceptability on a 1-7 Likert scale. In Experiment 2, participants saw three complex training sentences (one with a nested object relative clause and two with subject relative clauses), and were asked to answer comprehension questions by choosing one of five response options. After the training sentences, each participant saw one experimental item (experimental items were the same across experiments), and was asked to rate it in Experiment 1 (N = 3,856) or to answer a comprehension question in Experiment 2 (N= 3,761).

   Results. On the word following the critical verb, a typical attraction effect, i.e., a speedup in number mismatch condition, was observed in both experiments (see Fig. 1; Ratings: -36 ms with 95% credible interval (CrI) of [-72, 0.8] ms; Questions: -86 ms with 95%-CrI [-127, -45] ms). There was an interaction between the attraction effect and task: the attraction effect was greater when questions were expected. On the second word following the verb, there again was an interaction between the attraction effect and task (see Fig. 1): in Experiment 1 with judgments, in the number mismatch condition, there was a rating-typical slowdown of 20 ms (CrI [1, 40] ms); in Experiment 2 with questions, in the number mismatch condition, there was a reading-typical speedup of -20 ms (CrI [-40, 1] ms).

   Discussion. The present data demonstrates the effects of goal on sentence processing: A well-attested attraction effect in ungrammatical sentences was smaller in the verb + 1 and reversed in the verb + 2 region when participants expected to rate sentence acceptability. The reversed attraction effect, a slowdown in the number-mismatching condition, is typical for the

reaction times when rating attraction configurations. Thus, the rating-typical slowdown can manifest already in reading times, as long as the goal to rate the sentence is known in advance.

Example 2:   (a) The candidate that the lobbyist openly *support ...
               (b) The candidates that the lobbyist openly *support …
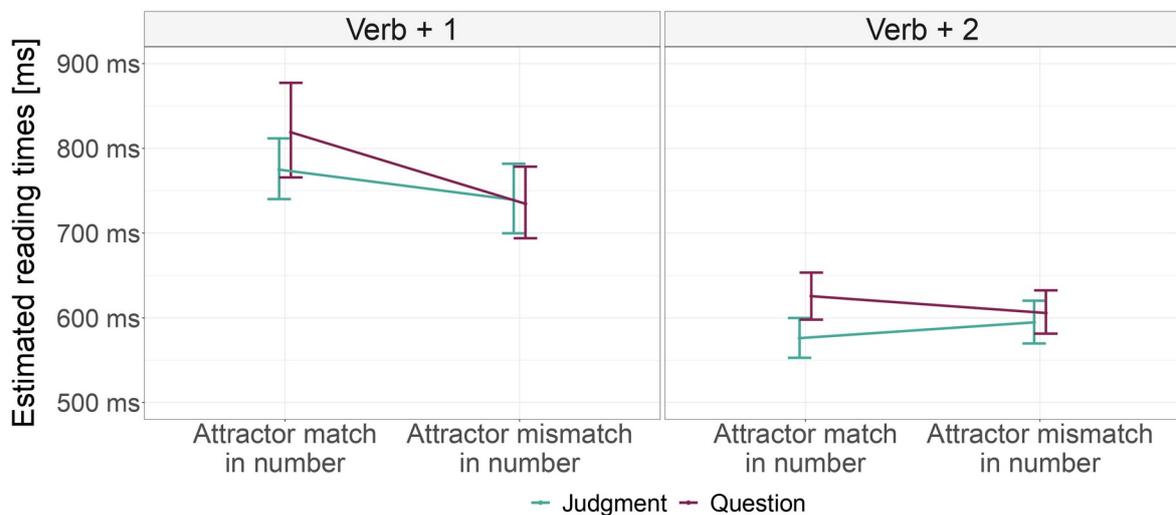
Figure 1. Estimates of reading times across conditions and sentence regions.

References
[1] Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. Journal of Memory and Language, 82, 133-149.
[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] Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension?. Journal of Memory and Language, 112, 104087.
[4] Schlueter, Z., Parker, D., & Lau, E. (2019). Error-driven retrieval in agreement attraction rarely leads to misinterpretation. Frontiers in psychology, 10, 1002.
[5] Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. Journal of memory and language, 60(2), 308-327.
[6] Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. Frontiers in psychology, 349.
[7] Laurinavichyute, A., & von der Malsburg, T. (2023, March 20). Agreement attraction in grammatical sentences and the role of the task. Preprint: https://doi.org/10.31234/osf.io/n75vc.

'You don't hang a Frida Kahlo next to a Jackson Pollock.' The effect of referential features and gender congruency on the comprehension of unfamiliar artist-for-work metonymies in German

Franziska Kretzschmar, Sandra Hansen, Anna Volodina & Christian Lang (Leibniz Institute for the German Language)

Previous research has shown that context – either statistical patterns created through conventionalization or immediate linguistic context – modulates the ease of comprehending systematic metonymies (e.g. Weiland-Breckle & Schumacher 2017, Piñango et al. 2017, Lowder & Gordon 2013). For example, novel producer-for-product metonymies are harder to process than familiar ones without supportive context (e.g. *read Needham* vs. *read Dickens*; Frisson & Pickering 2007), because neither conventionalized knowledge about the proper name nor immediate linguistic context helps in sense selection. This suggests that familiarity is a precondition for successful meaning transfer, because it makes available the feature(s) necessary for metonymic sense selection. Here, we tested a hitherto understudied cultural-historical aspect of familiarity and its effect on grammatical gender marking of artist-for-work metonymy in German: whether humans are sensitive to the sex/gender distributions in relation to the painter in artist-for-work metonymy of the type *ein echter Rembrandt* ('a real Rembrandt'). In the domain of (European) painting, there is a clear dominance of male painters, which predicts that familiarity with the metonymic NP 'a real X' should be biased toward the more common sex – and, hence, masculine gender marking – via conventionalization.

Linguistic accounts assume that the artist's sex is typically preserved in the grammatical gender of the metonymic expression, which – intriguingly – is often illustrated by male referents (cf. Nübling et al. 2015), so that this particular referential feature may be regarded as relevant for metonymic sense selection. This predicts that for the less common sex (here: female painters) there may be more variability in gender marking on the metonymic expression or language users may resort to alternative strategies such as assuming an elided NP head (e.g. *eine echte Kahlo* [-*Zeichnung*], 'a real [drawing of] Kahlo'; cf. Fahlbusch & Nübling 2016).

Because there is little empirical data on the impact of sex/gender distributions on artist-for-work metonymy, we conducted a corpus study and two acceptability studies on German, focusing on whether or not the grammatical gender marking of the metonymic NP conforms with the artist's sex. In the corpus study, we searched for two male painters (Rembrandt van Rijn, Wassily Kandinsky) and six female painters (e.g. Frida Kahlo, Kätze Kollwitz) – with an extended set of women because of a very low hit rate for this group. For the male painters, we found 861 hits (i.e. metonymic expressions of the type 'a (real) X'), 439 of which were unambiguously marked for masculine gender and 1 showed feminine marking. For female painters, there were 9 hits, with 7 cases unambiguously marked for feminine gender and two ambiguous cases. These results support that the painter's sex is preserved in the grammatical gender of the metonymic expression and that the construction as a whole is biased toward the culturally and historically dominant male sex.

Next, we ran two implicit acceptability experiments online to examine how participants comprehended the metonymic expression, when sex and grammatical gender features either matched or mismatched. The experiments were identical regarding stimuli and design except for the sex of the artist. Exp. 1 tested the acceptability of the metonymy with a fictitious female artist, while Exp. 2 tested the metonymy with a fictitious male artist. We constructed a short text that was designed as a letter about the coverage of a national German newspaper. In the letter, the painter was introduced as a famous artist, and two sentences separated the contextual introduction from the target sentence that contained the metonymic expression with grammatical gender marking in three conditions (Example 1c). Gender marking was either congruent with the artist's sex (e.g. female sex – feminine gender or male sex – masculine gender) or incongruent in two different variants: the grammatical gender was incongruent to the sex of the previously introduced artist (e.g. female sex – masculine gender or male sex – feminine gender) or the grammatical gender of the target expression was neuter to allow for readings of an elided NP head with neuter gender (e.g. *das Gemälde*neut *von X*, 'the painting of X').

Example 1: referent introduction and target sentence with three gender conditions (only one was displayed to a participant)
a. Exp. 1: female artist introduced in the linguistic context: *Irina Maximowa*
b. Exp. 2: male artist introduced in the linguistic context: *Ilias Maximowa*

c. Target sentence (Exp. 1 & 2) : *Sie haben* [*eine echte | einen echten | ein echtes*] *Maximowa gestohlen.*

They have [a fem | masc | neut real fem | masc | neut] Maximowa stolen. 'They stole a real Maximowa.'

In addition, we varied (and pre-tested) the proper name: It could be considered female (*Maximowa*), male (*Christophson*), or neutral (*Gutleb*). Participants' task was to read the text for comprehension and to correct errors. The stimulus text contained task-relevant grammatical and orthographic errors, none of which occurred in the critical context and target sentences.

Data from 265 participants were analyzed in Experiment 1. Data collection for Experiment 2 is ongoing, here we present descriptive analyses from 92 participants. We statistically analyzed the ratio of corrections with respect to grammatical gender marking in the target sentence, i.e. whether participants changed the original gender marking of the metonymic expression and if so what form they chose. Descriptive analyses show that in Experiment 1 (female painter), for both feminine and masculine gender marking, the majority of cases of the original forms remained unchanged (unchanged/changed: 81%/19% and 86%/14% respectively; see Figure 1A). Neuter gender marking was changed in 55% of all cases (vs. 45% unchanged). In addition to changing neuter gender marking to other forms, participants often used cognate paraphrases (e.g. 'a masterpiece') without the proper name. A multinomial logistic regression with the predictor variables original gender marking and sex association of the proper name supported the descriptive analysis and also revealed that names that were not associated with the female sex (*Christophson*, *Gutleb*) increased the probability of masculine gender marking after correction. Experiment 2 (male painter), by contrast, showed a different pattern in that original feminine gender marking was changed more often and showed a trend toward masculine gender marking, while masculine gender marking was not changed at all during error correction (Figure 1B). The neuter condition remained unchanged to a larger extent, but when it was changed, there was a clear trend towards changes to masculine gender.

Figure 1. A: Gender-marking corrections in Experiment 1. B: Gender-marking corrections in Experiment 2.
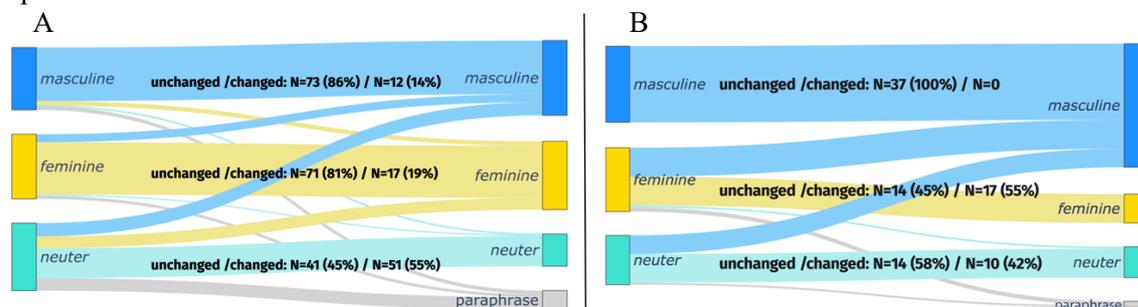
Note. Left side of each Sankey plot: Proportions of original gender marking. Right sight of each Sankey plot: Proportions of gender marking after correction

In sum, our results support the notion of a moderate male bias for the artist-for-work metonymy 'a real X', in line with the cultural-historical dominance of men in the domain of painting. The metonymy is used more frequently with male artists, as revealed by our corpus study. The bias is also reflected in Experiment 2 focusing on male artists where gender marking was changed more often when it was not masculine. However, the findings of Experiment 1 suggest that the male bias does not influence responses in an implicit acceptability task when it involves a female artist (a member of the less common sex associated with the metonymy), i.e. cultural-historical dominance does not seem to lead to a general bias toward the male sex. Otherwise, we should have found a higher probability of changes to masculine gender marking in Exp. 1. Intriguingly, we also did not find evidence for a shift toward feminine gender marking, which would have been expected if the sex of the referent introduced in the immediate linguistic context biases gender marking. This weakens the assumption that the sex feature is preserved during meaning transfer, at least in highly infrequent metonymic expressions. Finally, the proportions of changes for the neuter gender condition in both experiments suggest that participants may adopt an interpretation involving an elided NP head with neuter gender, but this seems to occur less often than associating the metonymic name to the other genders.

Prosodic Prominence and Negation

Frank Kügler & Markus Bader (Goethe University Frankfurt)

According to a widespread assumption, negation particles usually carry prosodic prominence, and in particular empirical evidence from English read speech suggests that the majority of negations show higher F0, meaning higher prominence (e.g. O'Shaughnessy & Allen 1983). It is assumed that negative information is accompanied by higher prominence since negation is regarded as semantically being focal information. Evidence for this assumption is however mixed: Yaeger-Dror (1997) and Kaufmann (2002) reported on English corpus data of spontaneous conversation that the majority of negated elements did not receive higher prominence. In general, according to the Handbook article of Prieto & Espinal (2020), the role that intonation plays with respect to negation is rather unexplored. It remains yet to be empirically shown if and how negation systematically is accompanied by prominence in plain sentential negation, and which conditions render a negative element less prominent. Therefore, we conducted a production study to investigate the variation of prominence on negation in German. Participants read out aloud sentences as in (2)-(4) after having heard contexts as in (1).

1. **Context**
    Marie war heimlich im Zimmer ihrer Schwester und hat dort eine gelbe Birne entdeckt. (_{doch} Marie wollte eigentlich fasten, aber sie hat es sich anders überlegt.) Was macht Marie wohl als nächstes?
    'Mary was secretly in her sister's room and discovered a yellow pear there. (Mary actually wanted to fast, but she changed her mind.) What will Mary do next?'

2. **SO target sentence with late negation**
    Marie wird die gelbe Birne ihrer Schwester **(Ø/nicht/doch)** essen.
    Mary will the yellow pear her sister Ø/not/after all eat
    'Mary will eat her sister's yellow pear (Ø/not/after all).'

3. **SO target sentence with early negation**
    Marie wird **(Ø/nicht/doch)** die gelbe Birne ihrer Schwester essen.
    Mary will Ø/not/after all the yellow pear her sister eat
    'Mary will eat her sister's yellow pear (Ø/not/after all).'

4. **OS target sentence**
    Die gelbe Birne ihrer Schwester wird Marie **(Ø/nicht/doch)** essen.
    the yellow pear her sister will Mary Ø/not/after all eat
    'Her sister's yellow pear, Mary will eat (Ø/not/after all).'

The context introduces three referents that are mentioned again in the target sentence, thus rendering them as given. The target sentence, which answers the question posed in the context, varies in two dimensions. First, target sentences were negated, simple affirmative or affirmative with the particle *doch* 'after all'. This adverbial occurs at the same sentential position as the negation *nicht* 'not' but without negating the sentence. Sentences with *doch* are included alongside sentences without *nicht* (affirmative) and sentences with *nicht* (negating) as a control for intonational effects related to material between the object and clause-final verb. In order to satisfy the special contextual requirements of the adverbial *doch*, the context included the additional sentence in parentheses for sentences with *doch* whereas the contexts for simple affirmative and negative sentences did not. The second dimension in which target sentences vary is the order between subject and object (subject-before-object/SO versus object-before-subject/OS) and for sentences with SO order, the order between negation/*doch* and the object (early negation/*doch* versus late negation/*doch*). Because all three referents mentioned in the target sentence are already introduced in the context, the new information in the answer is either the verb alone (affirmative sentences) or the verb together with the negation (negated sentences). Since new information in front of the verb attracts prominence in German we predict the highest prominence on the negative particle.

Seven participants heard twelve distinct contexts as in (1) and then read out aloud the target sentence (either (2), (3), or (4)) to the question posed in the context (7 speakers x 12 items x 3 conditions x 3 word orders = 756 sentences). All target sentences were digitally recorded in a sound-proof booth with a condenser microphone and acoustically analyzed using the Praat software (Boersma & Weenink,

2023). After word segmentation, a customized Praat script ran an f0 analysis collecting f0 values in Hertz at five equidistant points per word. f0 means were calculated aggregating over speakers and items per condition and word order. In addition, prominence rating according to DIMA (Kügler et al. 2022) were performed on the negative particle on a three-level prominence scale with normal prominence as level 2, reduced prominence as level 1, and emphatic prominence as level 3.

The results show that a late negation in a SVO or OVS sentence does not exhibit any particular prosodic pattern compared to the affirmative or the affirmative with the particle *doch* (see Fig's 1 & 2). Before the negation, the f0-contours are almost identical across conditions, and the negation as well as the adverbial *doch* in (2) are consistently realized with a falling pitch accent. A clear difference arises in early negation sentences (Fig 3). The baseline (affirmative) shows regular accentuation while both the negation and the affirmative with the particle *doch* show an effect of "deaccentuation" after the particle *nicht/doch*. The prominence ratings reveal that *doch* carries regular prominence while the prominence on *nicht* is almost evenly distributed on two prominence levels, regular and weak prominence (Fig 4).

These results point to two facts. First, the negative particle as well as the adverbial *doch* regularly receive prominence (pitch accent). Second, the prosodic pattern before an upcoming negation does not seem to differ from its positive counterpart. However, the strength of prominence on *nicht* varies. This indicates a certain amount of optionality in the accentuation of negation, which may contribute to the seemingly contradictory results in the prior literature.
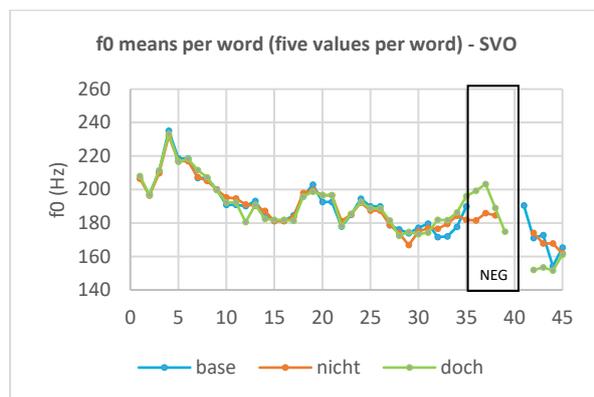
Fig 1. Time-normalized f0, SVO late negation
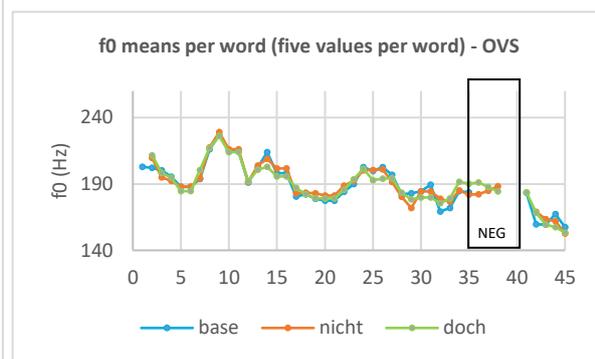
Fig 2. Time-normalized f0, OVS late negation
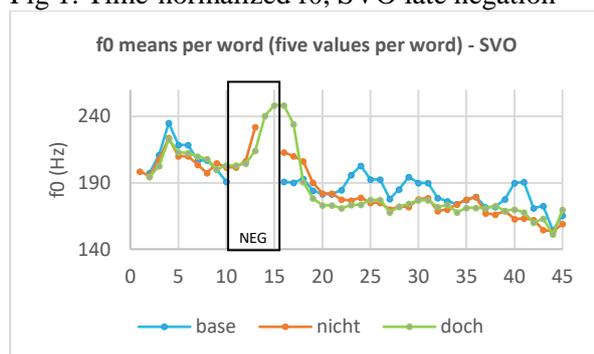
Fig 3. Time-normalized f0, SVO early negation
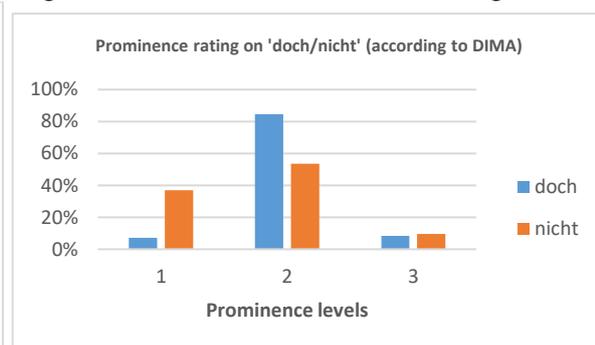
Fig 3. Prominence ratings for particles *nicht/doch*

References

Boersma, Paul & David Weenink. 2023. Praat: doing phonetics by computer [Computer software]. http://www.praat.org/

Kügler, Frank, Baumann, Stefan & Röhr, Christine T. (2022) Deutsche Intonation, Modellierung und Annotation (DIMA) – Richtlinien zur prosodischen Annotation des Deutschen. In Schwarze, Cordula & Grawunder, Sven (eds.) Transkription und Annotation gesprochener Sprache und multimodaler Interaktion. Tübingen: Narr, 23-54.

Prieto, Pilar & Maria Teresa Espinal. 2020. Negation, Prosody, and Gesture. In Viviane Déprez & Maria T. Espinal i Farré (eds), The Oxford handbook of negation. Oxford: OUP. 676–693.

O'Shaughnessy, Douglas & Jonathan Allen. 1983. Linguistic modality effects on fundamental frequency in speech. *The Journal of the Acoustical Society of America* 74(4):1155–1171.

Yaeger-Dror, Malcah L. 1997. Contraction of negatives as evidence of variance in register-specific interactive rules. *Language Variation and Change* 9:1–37.

Is there a conjunctive default in the interpretation of disjunction? A nonce word approach

Adina Camelia Bleotu (University of Bucharest), Andreea Nicolae (ZAS Berlin), Gabriela Bîlbîie (University of Bucharest), Mara Panaitescu (University of Bucharest), Anton Benz (ZAS Berlin), Lyn Tieu (University of Toronto)

Contact: adina.bleotu@lls.unibuc.ro

The current paper addresses the question of whether there is a conjunctive default in the interpretation of disjunction by probing into Romanian adults' understanding of nonce functional words. Previewing the results, we find that, when exposed for the first time to sequences of words containing nonce connectives such as *A mo B* or *mo A mo B*, potentially corresponding to '(both) A and B'/ '(either) A or B'/ 'A not B/neither A nor B', participants tend to associate them with a conjunctive interpretation rather than a disjunctive or negative one. Our findings support the idea that a possible reason for why children interpret disjunction as conjunction is the existence of a conjunctive default in interpreting operators linking A and B. Our findings also raise deeper questions about why speakers default to one interpretation over another, what the set of logical primitives are, and the possible role of frequency in assigning interpretations.

Background on the interpretation of disjunction Adults have been found to interpret simplex disjunction ('The hen pushed the train or the boat'*) inclusively* (*The hen pushed one, possibly both*) or *exclusively* (*The hen pushed either one but not both*), while they tend to associate complex disjunction (e.g., *either…or*) only with exclusive interpretations (Spector 2014, Nicolae et al. 2023). In contrast, children, have been found to treat simplex and complex disjunctions alike, providing *inclusive, conjunctive* (*The hen pushed both*) or *exclusive* interpretations: French and Japanese children were shown to be *inclusive* or *conjunctive* (Tieu et al. 2017), while German children were shown to be *inclusive* or *exclusive* (Sauerland & Yatsushiro 2018).

Background on the acquisition of disjunction in Romanian Recently, this line of investigation has been extended to Romanian, which makes use of multiple commonly used disjunctions: the complex disjunction (i) *sau…sau* which is built off the simplex *sau*, and (ii) *fie…fie*, which lacks a simplex disjunctive counterpart. Romanian also employs two distinct prosodic patterns for *sau*: (iii) a neutral prosody with no prosodic boundary after the first disjunct, and (iv) a marked prosody, where both disjuncts are stressed (as in complex disjunctions). In two studies by Bleotu et al. (2023a, b), building on the design in Tieu et al. (2017), Romanian 5-year-olds were shown to be inclusive with all *sau*-based disjunctions, but conjunctive and inclusive with *fie…fie.*

The source of conjunctive interpretations in child language While children's inclusivity is typically explained as a logical interpretation of disjunction, the conjunctive interpretation of disjunction has been a matter of debate. Skordos et al. (2020) and Huang & Crain (2020) have argued that the conjunctive behavior is an experimental artifact, which arises when the background only contains the disjuncts mentioned in disjunctive utterances but no other objects. In this context, disjunction would not be informative, which is why children default to conjunction. However, Bleotu et al. (2023b) have shown that conjunctive behavior persists even when the background contains additional objects, thus casting doubt on this explanation. An alternative explanation is that children's conjunctive interpretation of disjunction is a genuine semantic-pragmatic interpretation, which may originate either as a default (Roeper 2011), as an implicature (Singh et al. 2016), or as an additional meaning of disjunction alongside inclusive disjunction (Sauerland & Yatsushiro 2018).

We focus on the conjunctive default hypothesis, probing into whether, when participants are exposed to an operator unknown to them, they default to conjunction. We employ a nonce paradigm, following a long tradition of linguistic research (Brown 1957, Gleitman 1990, Gillette et al. 1999, a.o.)

Nonce words paradigms Nonce words have been employed in linguistics from as early as the 1950s in order to probe into children's ability to get to the meaning of words by means of syntactic cues, also known as *syntactic bootstrapping* (Gleitman 1990). In 1957, Robert Brown showed experimentally that preschool-aged children could use their knowledge of different parts of speech to distinguish the meaning of nonsense words in English (*Do you see any/ a **sib**?, What is **sibbing**?*). Jean Berko Gleason's Wug Test (1958) used

nonce words to explore children's acquisition of plural morphology (*one wug-two wugs*), possessives (*wug's, wugs'*) and verbal morphology (*He zibs*). Interesting experimental work further ensued (Naigles, 1990; Soja, 1992; Höhle et al., 2004; Cristophe et al., 2008; Syrrett et. al., 2010; Yuan & Fisher, 2012; Jin & Fisher, 2014; Cao & Lewis, 2021; Huang et al., 2021; a.o.). In the current experimental landscape, many studies employ nonce words to investigate people's biases in interpretation. Some interesting paradigms that have emerged are the *Human Simulation Paradigm* (HSP; Gillette et al. 1999), testing whether adults can infer meaning from context, and *Artificial Language Learning Paradigms* (Culbertson & Schuler 2019, Maldonado & Culbertson 2021 a, b), testing whether adults and children can learn artificial words and what their biases are when doing so. These paradigms have recently been employed to probe into logical words such as modals (see Dieueleveut et al. 2022) and negation (Maldonado & Culbertson 2021b).

In our investigation, we look at what kinds of meanings adults ascribe to a nonce word linking A and B, using the materials from Tieu et al. (2017). If adults default to conjunction, we reason that, assuming that universal interpretive biases persist at different stages of developments, when a disjunctive word is infrequent or yet unknown by children, children might also default to a conjunctive interpretation.

Experiments on *mo* and *mo…mo…*.: We tested 21 adult native speakers of Romanian on their interpretation of the nonce expressions *mo* and *mo…mo*. The same participants took part in the *Mo* experiment first and the *Mo…mo…* experiment second. Following Tieu et al. (2017), we used a modified TVJT presented in Prediction rather than Description Mode (Singh et al. 2016) in order to license *ignorance inferences*, which often characterize disjunctive statements. Participants were introduced to a puppet, Bibi, who would make guesses about what would happen in various situations. They were told that Bibi would sometimes make use of an unknown word, and they had to decide what it meant for Bibi. Importantly, they were also told that the unknown word does not refer to something that one can point to. Bibi would be familiarized with a situation involving an animal and two objects (see Fig. 1a). Bibi would make a guess about what would happen (*The hen pushed the train **mo** the boat/ The hen pushed **mo** the train **mo** the boat*). Participants then saw the outcome (Fig. 1b) and had to say whether Bibi had guessed well. At the end of each experiment, participants were asked what they thought *mo/ mo…mo…* meant. Each participant saw a total of 15 sentences: 2 practice trials and 13 experimental items (8 targets, 2 controls, 3 fillers). Target test sentences (*The hen pushed the train mo the boat/ The hen pushed mo the train mo the boat*) were presented in 1-disjunct-true (1DT) contexts (x4 items) where only one disjunct was true (*The hen pushed only the train*), and 2-disjunct-true (2DT) contexts (x4 items) where both disjuncts were true (*The hen pushed both objects)*. In control items neither disjunct was true.

a. Start   b. Outcome
Fig.1 Example of an experimental item in the 2DT

Results 1 participant was removed from the analysis given inaccuracy in the fillers. Adult participants were overwhelmingly conjunctive in their interpretation of utterances containing *mo* and *mo…mo* (i.e., they accepted the target sentences in 2DT but rejected them in 1DT). In the *Mo* Experiment, 13 of 20 participants were conjunctive, while two thought it meant 'A not B', and five oscillated between a conjunctive and a negative interpretation. In the *Mo….mo…* experiment, 16 of 20 participants were conjunctive, while two thought it meant 'neither A nor B', and two oscillated between a conjunctive ('and A and B') and a negative interpretation of *mo* ('neither A nor B').

Discussion Our results suggest that, when adult participants are exposed to nonce words connecting A and B, their default interpretation seems to be conjunctive. Even more strikingly, they seem to default to conjunction even within an experiment where Bibi does not always make correct guesses. These findings can be interpreted in multiple ways. Under *a frequency approach*, it could be argued that adults simply associate the unknown connectors with the interpretation corresponding to the most frequent logical operator linking two elements, namely, conjunction (see Jasbi et al. 2018, 2022 for a discussion of corpus evidence that conjunction is more frequent than disjunction). Under *a logical universal primitives approach*, it could be argued that conjunction is more basic than disjunction, since disjunctive interpretations can be reduced to the conjunction of two modalized elements (Zimmerman 2000): *possible A* and *possible B*. Conjunction would also have the advantage of conceptual simplicity: (A and B) is simpler

than (*possible A* and *possible B*). It is difficult to distinguish between these two approaches, given that frequency may also be a consequence of this bias. In the context of the discussion on children's interpretation of disjunction, our findings suggest that a conjunctive default could be a possible source for children's interpretation of *fie…fie* as conjunctive, if *fie…fie* is perceived as infrequent/unknown. This is partly supported by the lower frequency of *fie…fie* compared to *sau…sau…* from adult corpora (see Bleotu et al. 2023a), but future studies should also examine corpus data from child language.

It's not just all in the head: Towards a processing model of German adjective-noun-noun constructions and the bracketing paradox

Anna Prysłopska and Titus von der Malsburg, Institute of Linguistics, University of Stuttgart

German nominal compounds modified by an adjective typically have an canonical reading (1a) in which the adjectives modifies the second noun of the compound. However, in some of these constructions, the adjective can equally plausibly or even preferentially attach to the first noun (1b). The second reading is referred to as a *bracketing paradox* (Bergmann 1980; Winkler 2015). These constructions have been hypothesizes to have the same syntactic bracketing but different semantic bracketing.

From a grammatical standpoint, the adjective should apply to the second noun or to the compound as a whole (2a), but crucially not to the first noun (Duden Online 2023; Bergmann 1980). How, then, are bracketing paradoxes licensed, whether odd (2b) or unremarkable (2c)? The context, world knowledge, and pragmatic factors are potential contributors to interpretation preferences, along with the semantic compatibility between the adjective and individual nouns, and morphosyntactic agreement. Language economy and how lexicalized the compound is could play a role as well (Bär 2007; Spalding et al. 2010; Schlücker 2014; Maienborn 2020). Despite a large body of theoretical literature, the empirical study of bracketing paradoxes has been largely neglected (Dima et al. 2017). Noun-noun compounds have garnered more attention, but the relationship between the individual nouns and the adjectival modifier would benefit from more inquiry.

Methods: In a questionnaire study, we evaluated adjective and nominal compound phrases, as in (3), compiled based on the theoretical literature and newspaper articles. 204 items in 3 conditions were divided into lists using a Latin square design. 36 participants recruited on Prolific were randomly assigned to one of three lists. They assigned 1–5 values to the items on the dimensions of naturalness, comprehensibility, and stylistic form. The study aimed to reveal how the compatibility between an adjective and the individual nouns affect the adjective-nominal-compound construction as a whole. Thus, we take first steps towards an processing model of bracketing paradoxes.

(1) a. [Deutsche [Sprachwissenschaft]] (*German* language.*science*) canonical reading
    b. [[Deutsche Sprach]wissenschaft] (*German language*.science) bracketing paradox reading

(2) a. Verrückter Chemieprofessor (Crazy chemistry.professor) ✖ Chemie ✔ Professor
    b. ?Verstöckiger Hausbesitzer (Four.story house.owner) ✔ Haus ✖ Besitzer
    c. Schwere Unwetterwarnung (Severe weather.warning) ✔ Unwetter ✖ Warnung

(3) a. Psychologische Beratungsstelle (Psychological counseling.center) AN_1N_2
    b. Psychologische Beratung (Psychological counseling) AN_1
    c. Psychologische Stelle (Psychological center) AN_2

Results: The ratings across scales were highly correlated (lowest $r = 0.95$, $p < 0.001$). Therefore, we used the mean of the three ratings, which we also scaled to the interval $[0, 1]$ for convenience. All but three items received good ratings for either AN_1 or AN_2 or for both AN_1 and AN_2 (Fig. 1a). This follows from our attempt to exclude AN_1N_2 constructions where the adjective was a poor match for both of the nouns as these are unlikely to be used. As a consequence, AN_1 and AN_2 ratings were negatively correlated ($r = -0.5$). A Bayesian Beta regression was fit that modeled the averaged and scaled ratings of the AN_1N_2 constructions as a function of the corresponding AN_1 and AN_2 ratings along with their interaction. Predictors' effects with 95% credible intervals are shown in Table 1 and Figure 1. As expected, high AN_2

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	-4.02	0.76	-5.62	-2.60	1.00	3357	3907
AN_1	3.32	0.89	1.67	5.17	1.00	3383	3757
AN_2	6.34	0.89	4.69	8.19	1.00	3291	3916
$AN_1:AN_2$	-4.01	1.07	-6.20	-1.99	1.00	3378	3748

Table 1: Summary of the model effects.

(a) Relationship between AN_2 and AN_1.

(b) Relationship between AN_1N_2 and AN_2.

(c) Relationship between AN_1N_2 and AN_1. Lines correspond to AN_2 value groups.
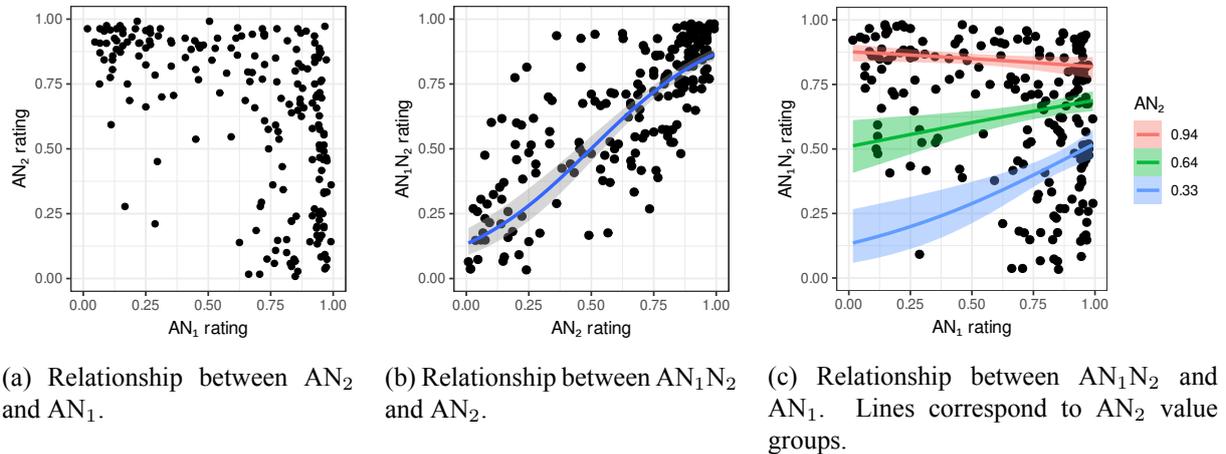
Figure 1: Overview of ratings between conditions in (3) scaled from 1–5 to 0–1 values.

ratings were predictive of high AN_1N_2 ratings ($\beta = 6.3$, Fig. 1b). However, AN_1 ratings, too, had a positive, albeit smaller effect on AN_1N_2 ratings ($\beta = 3.3$). Crucially, there was an interaction of the AN_1 and AN_2 ratings ($\beta = -4$, Fig. 1c): When AN_2 ratings were low, AN_1 ratings had a substantial positive effect. When AN_2 ratings were high, higher AN_1 ratings slightly *reduced* the predicted AN_1N_2 ratings.

Conclusions: Contrary to grammatical and strictly compositional constraints on their relationship, the first noun plays an important role in the overall acceptability of a nominal compound modified by an adjective. This is in spite of the second noun's dominance over the adjective and compound. This result aligns with the role of semantic and pragmatic factors on such constructions, which can favor an otherwise grammatically unavailable attachment site. When both nouns are good matches for the adjective, acceptability is slightly reduced suggesting a perceived conflict between possible attachment sites. Thus, even though both nouns have a positive influence on the compound's acceptability, their effects are not strictly additive. In the absence of a suitable head noun candidate, the first noun becomes an attractive modification target for the adjective. More broadly, this work suggests that the processing of adjective noun-noun constructions is more complex than previously thought. In follow-up work, we investigate the role of morphosyntactic information (gender, number), animacy, lexicalization, and discourse information, which may influence the acceptability and interpretation of these constructions.

Bär, Jochen A (2007). "Kürze als grammatisches Problem: determinative Verschränkungen." In: *Sprachliche Kürze. Konzeptuelle, strukturelle und pragmatische Aspekte*. Berlin & New York: de Gruyter, pp. 310–338.

Bergmann, Rolf (1980). "Verregnete Feriengefahr und Deutsche Sprachwissenschaft. Zum Verhältnis von Substantivkompositum und Adjektivattribut." In: *Sprachwissenschaft* 5.3, pp. 234–265.

Dima, Corina et al. (2017). "A Corpus-Based Model of Semantic Plausibility for German Bracketing Paradoxes." In: *CDH*, pp. 64–70.

Duden Online (2023). *Attribute vor zusammengesetzten Substantiven*. URL: `https://www.duden.de/sprachwissen/sprachratgeber/Attribute-vor-zusammengesetzten-Substantiven` (visited on 10/30/2023).

Maienborn, Claudia (2020). "Wider die Klammerparadoxie: Kombinatorische Illusionen beim Adjektivbezug auf NN-Komposita." In: *Zeitschrift für Sprachwissenschaft* 39.2, pp. 149–200.

Schlücker, Barbara (2014). *Grammatik im Lexikon: Adjektiv-Nomen-Verbindungen im Deutschen und Niederländischen*. Berlin, München, Boston: De Gruyter.

Spalding, Thomas L. et al. (2010). "Relation-based interpretation of noun-noun phrases: A new theoretical approach." In: *New Impulses in Word-Formation*. Ed. by Susan Olsen, pp. 283–315.

Winkler, Julia (2015). "Kleine Geschichte der 'schiefen Attribute'." In: *ZAS Papers in Linguistics* 58, pp. 124–139.

Experiments on anaphora resolution of generic masculine nouns in German

Philipp Rauth, Robin Lemke, Lisa Schäfer (Saarland University)

We present two experiments on the perception of generic masculine forms in German. They address potential shortcomings of previous research and suggest that the alleged male bias of these generics can to some extent be traced back to methodical choices and item design.

Motivation. In several languages with overt morphological gender marking, so-called generic masculine forms are used to refer to potentially gender-mixed groups of people. For instance, in German, a masculine plural DP like *die Lehrer* 'the teachers' can denote a group consisting only of male teachers or a gender-mixed group, whereas the feminine form *die Lehrerinnen* can only denote a homogeneously female group. The use of such morphological masculine forms to denote mixed groups has been criticized from a feminist perspective because it leads to women being cognitively under-represented (e.g., Pusch 1984). It is an empirical question to what extent interlocutors actually interpret masculine forms as referring more to men despite their presumably generic use. So far, several experiments suggest that although these form are often interpreted generically, they are significantly more likely to be understood as referring to an all-male than to a gender-mixed group. For instance, Gygax et al. (2008) conducted a binary sentence continuation plausibility judgment task and found that for German sentence pairs such as (1), participants were significantly more likely to judge the continuation with the anaphoric male DP *Männer* 'men' as plausible in the context with the antecedent *Sozialarbeiter* 'social workers' than the continuation with *Frauen* 'women', and were also faster in doing so. Similar results were obtained in further studies using variations of the design (e.g., Garnham et al. 2012, Körner et al. 2022).

(1) Die **Sozialarbeiter** liefen durch den Bahnhof. Wegen der schönen Wetterprognose
     the social.workers walked through the station because.of the nice weather.forecast
     trugen **mehrere der** (**Frauen**|**Männer**) keine Jacke
     wore several of.the women men no coat
     'The social workers were walking through the station. Since sunny weather was forecast several of the (women|men) weren't wearing a coat.'

These results are convincing, but the partitive *der Frauen* might be ambiguous: It could be interpreted as being coreferent with the antecedent *die Sozialarbeiter* (coreferent interpretation), but in this case the feminine *Sozialarbeiterinnen* would be pragmatically preferred to denote an all-female group. It could also be interpreted as constituting a subgroup of a larger gender-mixed group described by the antecedent *die Sozialarbeiter* (subset interpretation). Only for the subset interpretation, the lower plausibility ratings for the continuation with *der Frauen*, which have been found in previous research, can be interpreted as informative with respect to a possible gender bias of generic masculines and not as a consequence of a pragmatically inappropriate antecedent. We therefore conducted two experiments in order to disambiguate the interpretation of the anaphoric DP. In experiment 1, we presented the stimuli not in written form as the previous studies did but auditorily to control the intended interpretation of the anaphoric DP by means of prosody. In experiment 2, we used written stimuli but ensured the intended interpretation syntactically by using the unambiguous partitive *die Frauen unter ihnen* 'the women among them'.

Exp. 1 – Prosodic disambiguation. (preregistered). We used the established binary plausibility judgment task and the stimuli of Körner et al. (2022), which we presented auditorily in one of two prosodic conditions respectively: an *unmarked* condition (strongest pitch accent on the last content word, (2a)) and a *contrastive* condition (hat contour marking a contrastive topic, see Büring (1997), (2b)).

(2) a. Wegen der schönen Wetterprognose trugen mehrere der Frauen keine JAcke.
     b. Wegen der schönen Wetterprognose trugen mehrere der /FRAUen keine JA\cke.

We expect that the unmarked prosody favors the coreferent interpretation because *Frauen* is marked as given by deaccentuation and, therefore, interpreted as identical to the group denoted by *Sozialarbeiter*. Since in this case it would be pragmatically more appropriate to use the distinct feminine form *Sozialarbeiterinnen*, the generic masculine form *Sozialarbeiter* may reduce the plausibility of the continuation.

Conversely, we expect that contrastive prosody increases the plausibility of the *Frauen* continuation because it evokes a subset of male individuals contrasting with the subset *die /FRAUen*, both contained within the gender-mixed group of *Sozialarbeiter*. If participants in the study by Gygax et al. (2008) accommodated the unmarked prosody (Féry 2006), this could have biased them towards a coreferent interpretation more often and therefore to reject the female continuation for pragmatic reasons. We expect that controlling for prosody reduces the male bias in the contrastive condition, which would be supported by an interaction between PROSODY and GENDER of the anaphoric DP. A main effect of GENDER would evidence a general male bias independent of prosody.

Fig. 1 Mean ratings of exp. 2.

109 native German speakers heard a total of 72 audio files (36 critical items, 36 fillers) and judged whether the second sentence was a sensible continuation of the first one (binary task). We report analyses of 69 subjects who passed the threshold of rejecting more than half of 13 attention checks and are currently replacing those who did not. The data (see Fig. 1) were analyzed with mixed effects logistic regressions (Bates et al. 2015) in R (R Core Team 2023). We find a marginal PROSODY:GENDER interaction ($\chi^2 = 3.2, p = 0.07$) indicating that contrastive prosody increases the acceptability of a feminine continuation.

Fig. 2 Mean ratings of exp. 2.

We also find a highly significant GENDER main effect, which shows that subjects overall reject the feminine continuation more often ($\chi^2 = 126.05, p < 0.001$). This tentatively supports our hypothesis that subjects accommodate neutral a prosodic contour when interpreting the anaphor in previous studies but it also confirms a male bias in the interpretation of masculine plurals in German.

Exp. 2. – Syntactic disambiguation. Experiment 2 disambiguated the anaphoric DP syntactically by using the unambiguous partitive *unter ihnen* 'among them' (3b). In a written acceptability rating study, 47 native German speakers judged 16 critical items such as (3) and 72 fillers on a 7-point Likert scale (7 = totally natural). We analyzed the data with CLMMs (Christensen 2022) in R and found a marginally significant main effect of GENDER. There was a tendency for sentences with 'men' being prefered over 'women' ($\chi^2 = 3.65$, p < 0.057) (see Fig. 2).

(3) a. 'The **customers** were waiting at the checkout.'
     b. 'Because of the long queue, **the (women|men) among them** were very annoyed.'

The effect is consistent with those found by Gygax et al. (2008) and subsequent studies but much weaker. This tentatively suggests that the bias toward a non-mixed interpretation of the generic masculine forms can be reduced by ruling out the possibility of a coreferent interpretation.

Discussion. In sum, our studies suggest that the male bias of generic masculine plural forms in German repeatedly found in previous studies is partially explained by properties of the experimental design, in particular the ambiguity of the anaphor between a coreferent and a subset interpretation. When this ambiguity is ruled out by prosodic or syntactic means, the male bias is reduced in both of our experiments. However, we still find evidence for a male bias in both experiments (marginal in exp. 2), which suggests that the bias is not only caused by confounds in the design.

Selected references Büring, D. (1997). *The meaning of topic and focus.* Routledge. ● Féry, C. (2006). Laute und leise Prosodie. In H. Blühdorn et al. (Eds.), *Text—Verstehen* (pp. 164–186). De Gruyter. ● Garnham, A. et al. (2012). Gender representation in different languages and grammatical marking on pronouns: When beauticians, musicians, and mechanics remain men. *Discourse Processes*, 49(6), 481–500. ● Gygax, P. et al. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, 23(3), 464–485. ● Körner, A. et al. (2022). Gender representations elicited by the gender star form. *Journal of Language and Social Psychology*, 41(5), 553–571. ● Pusch, L. (1984). *Das Deutsche als Männersprache.* Suhrkamp.
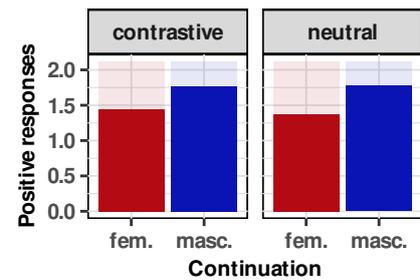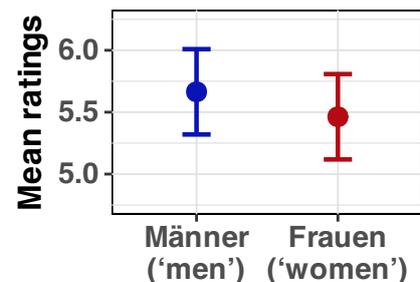
The psycholinguistics of historical language change – Structural priming and the dative alternation in Middle English

Gunnar Jacob and Mirja Maier
University of Mannheim

It has been suggested that structural priming, i.e. a speaker's tendency to re-use syntactic structures encountered or produced shortly before, may constitute a key psycholinguistic mechanism behind grammatical language change (e.g. Kootstra & Muysken, 2019). Specifically, new grammatical structures may consolidate themselves and spread further within a language through structural priming. However, experimental studies on this issue are faced with a principled methodological problem: For well-documented cases of historical language change, it is obviously impossible to conduct experimental psycholinguistic research, while for phenomena which are currently in the middle of changing, it is not yet clear how this process will develop in the future.

The present study attempts to avoid this problem by investigating structural priming effects in historical corpus data. We investigate the possible role of structural self-priming in a well-documented instance of historical language change, i.e. the emergence of prepositional-object (PO) ditransitives in Middle English (e.g. McFadden, 2002; Trips & Stein, 2019). While only sentences with the double-object (DO) structure, such as (1), existed in Old English, sentences with the alternative prepositional-object (PO) structure, such as (2), started to emerge in Middle English from around 1250 AD, and consistently increased in frequency between 1250-1350 AD.

We relied on Gries' (2005) approach for investigating structural (self-) priming in contemporary corpora to "simulate" a structural priming experiment with corpus data from the M2 (1250-1350 AD) and M3 (1350-1420 AD) sections of the Penn-Helsinki Parsed Corpus of Middle-English (PPCME2). A corpus search for ditransitive sentences with a PO or DO structure yielded a total of 2927 ditransitive prime sentences. These were categorized as either 'PO', 'DO', or 'PO-similar'. The 'PO-similar' category consisted of sentences such as (3), which superficially resembled a PO in purely structural terms, but did not constitute an alternative for a DO. For each prime sentence, we determined the grammatical structure of the next ditransitive sentence following the prime within the same historical text (i.e. the equivalent of a target sentence in an experimental study).

Figure 1 shows the proportions of PO, DO, and PO-similar target sentences, separately for each of the three prime types. The results revealed a self-priming effect, with a significantly higher proportion of PO targets following PO primes than following DO primes (and vice versa). This suggests that authors were indeed primed by the structure of the preceding prime sentence while choosing a structure for the target sentence. Interestingly, the results showed an increased proportion of PO targets not only after PO primes, but also after PO-similar primes. This indicates that the PO structure can also be primed by structures which only superficially resemble a PO structure. Thus, it is possible that such structures may have contributed to the consolidation of the PO structure in the English language, through structural priming.

Priming effects were strongest when prime and target were separated by only a small number of intervening non-ditransitive sentences, and gradually decreased with more sentences between prime and target. This result is consistent with the findings from experimental studies on structural priming, in which the magnitude of the priming effect was also affected by the number of intervening sentences (e.g. Branigan, Pickering & Cleland, 1999).

With respect to the role of structural priming in language change, our results are consistent with the claim that the new PO structure consolidated itself within the English language through structural self-priming in medieval authors. The results also suggest that it is possible to detect traces of the psycholinguistic mechanisms driving language change in historical corpora. Practical problems with investigating structural priming in historical corpus data, such as the repetition of structures for stylistic reasons, are discussed.

(1) (…) I wyll +geue you tresur wythout nombyr (CMMIRK,113.3103)

(*I will give you treasure without number*)

(2) (…) and gave the scabbard Excaliber to her lover (CMMALORY, 655.4476)

(*and gave the scabbard Excalibur to her lover*)

(3) (…) he +trewe doun +te emperour to +te er+te (CMPOLYCH, VI,435.3188)

(*he threw down the emperor to the earth*)

Figure 1. Proportions of 'PO', 'DO', and 'PO-similar' target sentences by prime type.
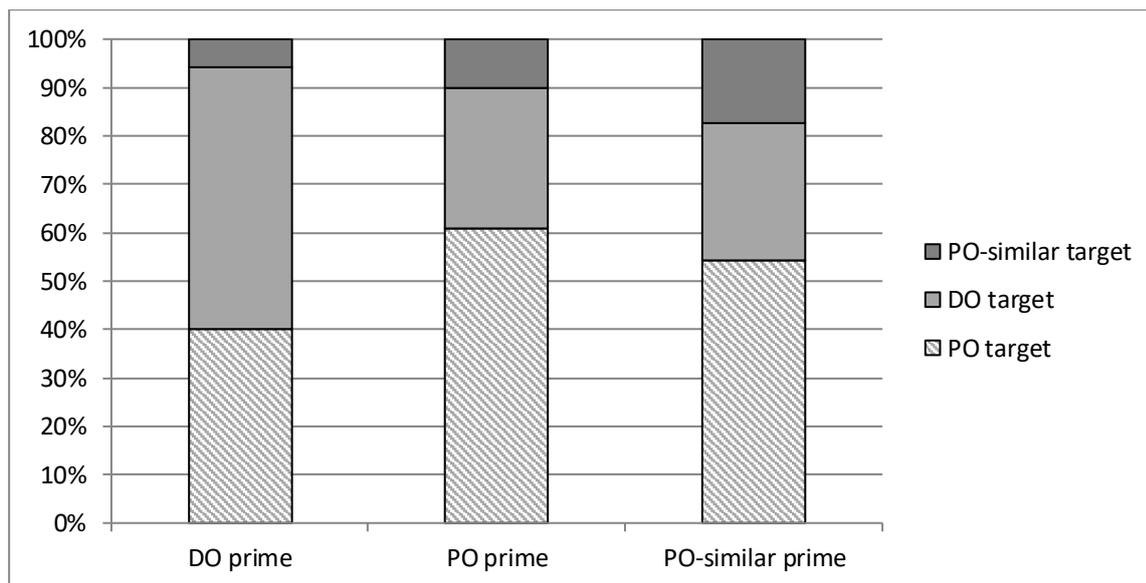
References

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, *6*, 635-640.

Gries, S. Th. (2005). Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research* , *34* (4), 365–399.

Kootstra, G. J. & Muysken, P. (2019). Structural Priming, Levels of Awareness, and Agency in Contact-Induced Language Change. *Languages*; 4(3). 65.

McFadden, T. (2002). The rise of the to-dative in Middle English. In D. W. Lightfoot (Ed.), *Syntactic effects of morphological change* (pp. 107–123). Oxford University Press.

Trips, C., & Stein, A. (2019). Contact-Induced Changes in the Argument Structure of Middle English Verbs on the Model of Old French. *Journal of Language Contact*, *12* (1), 232–267.