

Linguistic Evidence 2022

6-8 October

Paris, France

Organizers (in alphabetical order):

Anne Abeillé
Lisa Brunetti
Barbara Hemforth
Philip Miller
Gabriel Thiberge
Elodie Winckel

Thanks to all our reviewers:

Evangelia Adamou, Markus Bader, Gabriela Bilbiie, Charles Clifton, Emmanuel Chemla, Saveria Colonna, Israel de la Fuente, Vera Demberg, Emilie Destruel, Pegah Faghiri, Claudia Felser, Lyn Frazier, Michael Franke, Lyn Frazier, Alan Garnham, Edward Gibson, Martine Grice, Jana Haeussler, Barbara Hoehle, Gerhard Jaeger, Johan Rooryck, Elsi Kaiser, Jong-Bok Kim, Andreas Konietzko, Andy Luecking, Johanna Nyckel, Oliver Bott, Dario Paape, Céline Pozniak, Janina Rado, Ingo Reich, Leah Roberts, Sophie Repp, Rui Chaves, Christoph Scheepers, Florian Schwarz, Torgrim Solstad, Katharina Spalek, Shari Speer, Britta Stolterfoht, Patrick Sturt, Shravan Vasishth, Titus von der Malsburg, Albert Wall, Isabell Wartenburger, Thomas Weskott, Susanne Winkler

Thursday, October 6, 2022

TIME	EVENT
08:30 - 08:50	Registration + coffee - Registration + coffee
08:50 - 09:00	Welcome - Welcome
09:00 - 09:40	Predictability effects on the acceptability of antecedent-target mismatches under verb phrase ellipsis (Amphi Buffon) - Tyll Robin Lemke, Lisa Schäfer, Heiner Drenhaus, Ingo Reich
09:40 - 10:20	Preposition omission in French sluicing: An empirical approach (Amphi Buffon) - Amal Hassen, Anne Abeillé
10:20 - 10:40	Coffee break (Amphi Buffon)
10:40 - 11:20	How specific are linguistic structures? Mathematical priming on relative clause attachment in French (Amphi Buffon) - Céline Pozniak, Mireille Copin, Giuseppina Turco, Barbara Hemforth
11:20 - 12:00	Priming Prosodic Boundaries Across Constructions and Languages (Amphi Buffon) - Dorotea Bevivino, Giuseppina Turco
12:00 - 12:40	Investigating lexical bias in Persian light verb constructions: What can we learn from priming experiments? (Amphi Buffon) - Pegah Faghiri, Monique Flecken and Eva van Lier
12:40 - 13:40	Lunch (Amphi Buffon)
13:40 - 13:55	No stone was left unturned: the passivizability of Dutch idioms (Amphi Buffon) - Michelle Suijkerbuijk & Ferdie Hubers
13:55 - 14:10	"Spinnt sie" or "Spinnt die"? Empirical studies on d- and p-pronouns in German. (Amphi Buffon) - Luise Ehrmantraut
14:10 - 14:25	The interplay between quotation and referentiality: An empirical investigation into name-mentioning constructions (Amphi Buffon) - Natascha Raue & Álvaro Cortés Rodríguez
14:25 - 14:40	Internet demonstrations: Using quotation and demonstration in written language (Amphi Buffon) - Kathryn Barnes & Cornelia Ebert
14:40 - 14:50	Break (Amphi Buffon)
14:50 - 15:05	Iconic performances: The information structure and semantic contributions of ideophones and gestures (Amphi Buffon) - Kathryn Barnes
15:05 - 15:20	Categorical speaker-memory in native and non-native listeners (Amphi Buffon) - Sara Beck & Andrea Weber
15:20 - 15:35	Dimensions of judgment in stigmatized and non-stigmatized variation (Amphi Buffon) - Gert-Jan Schoenmakers
15:35 - 15:50	Embedded questions: Evidence in a decision-theoretic paradigm for 'surprise' & 'agree' (Amphi Buffon) - Emilie Destruel, Lea Fricke, Edgar Onea, Malte Zimmermann
15:50 - 16:05	Processing past time reference, but which one? An ERP study on the Mandarin Chinese aspect morphemes guo1 and guo2 with definite and indefinite time adverbs (Amphi Buffon) - Aymeric Collart
16:05 - 16:25	Coffee break (Amphi Buffon)
16:25 - 17:05	Je peux ou je dois ? Faudrait savoir ! - Acquiring modals' force: evidence from French (Amphi Buffon) - Anouk Dieuleveut
17:05 - 17:45	Interdiction as a primary reading of negated deontic necessity in child Romanian (Amphi Buffon) - Adina Camelia Bleotu, Anton Benz & Gabriela Brozba

17:45 - 18:45 Keynote: The language system in the human brain (Amphi Buffon) - Evelina Fedorenko

Friday, October 7, 2022

TIME	EVENT
08:30 - 09:00	Coffee (Amphi Buffon)
09:00 - 10:00	Keynote: Frequency and acceptability as evidence for the grammar (Amphi Buffon) - Sam Featherston
10:00 - 10:40	Dutch long passive was tried to reject (Amphi Buffon) - Iva Kovač, Gert-Jan Schoenmakers
10:40 - 11:00	Coffee break (Amphi Buffon)
11:00 - 11:40	Contrastive topic marking with German dagegen ('in contrast') and wiederum ('in turn') (Amphi Buffon) - Regina Zielege
11:40 - 12:20	Topic position or prefield? – Disentangling the positional restriction of topic drop in German based on acceptability rating data (Amphi Buffon) - Lisa Schäfer
12:20 - 13:00	The effect of scope of negation on the choice of word order during sentence production (Amphi Buffon) - Markus Bader & Yvonne Portele
13:00 - 14:00	Lunch
14:00 - 14:15	Incremental negation processing with positive questions under discussion (Amphi Buffon) - Oksana Tsaregorodtseva, Elena Albu, Barbara Kaup
14:15 - 14:30	French evaluative adjectives: position and interpretation (Amphi Buffon) - Lisa Brunetti, Gabriel Thiberge
14:30 - 14:45	The (not-)at-issue status of character viewpoint gestures (Amphi Buffon) - Sebastian Walter
14:45 - 15:00	Gapping in Romanian comparatives: an experimental perspective on case marking and animacy (Amphi Buffon) - Gabriela Bilbiie
15:00 - 15:10	Break (Amphi Buffon)
15:10 - 15:25	What the use of the German focus particle 'auch' can tell us about the influence of structural properties of the context (Amphi Buffon) - Laura Reimer, Bettina Braun, Christine Dimroth
15:25 - 15:40	Comprehending non-canonical and indirect speech acts in German (Amphi Buffon) - Andreas Trotzke & Laura Reimer
15:40 - 15:55	Reading non-canonical sentences in context: Identity vs. Poset (Amphi Buffon) - Yvonne Portele, Markus Bader
15:55 - 16:10	The effects of information structure and sentence structure on sentence processing (Amphi Buffon) - Hans Wilke, Jet Hoek, Hannah Rohde
16:10 - 16:40	Coffee break (Amphi Buffon)
16:40 - 17:20	Anaphoric Potential of Weak Definites in Contrast with Implicit Entities and Indefinites in German (Amphi Buffon) - Fereshteh Modarresi & Manfred Krifka
17:20 - 18:00	Referring to someone using only their last name: Insights from gender-marked pronouns (Amphi Buffon) - Elsi Kaiser, Deborah Ho, Haley Hsu, Claire Post, Madeline Rouse
18:00 - 18:40	The Position of Antecedent Hypothesis in Romanian subject alternation (Amphi Buffon) - Fabian Istrate, Anne Abeillé & Barbara Hemforth

19:00 - 22:00 Conference dinner - Conference dinner

Saturday, October 8, 2022

TIME	EVENT
08:30 - 09:00	Coffee (Amphi Buffon)
09:00 - 10:00	Keynote: Discourse and processing approaches to syntactic "island" effects (Amphi Buffon) - Ted Gibson
10:00 - 10:40	Island bridges? Extraction from adjunct clauses in Danish (Amphi Buffon) - Anne Mette Nyvad, Ken Ramshøj Christensen, Christiane Müller
10:40 - 11:00	Coffee break (Amphi Buffon)
11:00 - 12:00	Invited talk: The Radical Unacceptability Hypothesis (Amphi Buffon) - Susanne Winkler and Giuseppe Varaschin
12:00 - 12:40	The Discourse Function of Constructions Predicts Island Status (Amphi Buffon) - Nicole Cuneo, Adele E Goldberg
12:40 - 13:40	Lunch
13:40 - 14:00	Dependency formation during real-time processing: Evidence from webcam-based eye-tracking with subjective and objective adjectives (Amphi Buffon) - Elsi Kaiser & Jesse Storbeck
14:00 - 14:20	Extraction asymmetries in complex participle adjuncts (Amphi Buffon) - Andreas Kehl
14:20 - 14:40	Testing extraction (out) of subjects and objects in Mandarin Relative Clauses (Amphi Buffon) - Ruihua Mao
14:40 - 15:00	The sensitivity of native Dutch speakers to wh- and coordination island constraints (Amphi Buffon) - Michelle Suijkerbuijk, Frank Stefan, Peter De Swart
15:00 - 15:20	Coffee break (Amphi Buffon)
15:20 - 16:20	Invited talk: Resumptive pronouns and processing difficulties (Amphi Buffon) - Mayaan Keshev
16:20 - 17:20	Invited talk: Satiation and Syntactic Adaptation (Amphi Buffon) - Rui Chaves
17:20 - 17:30	Good byes - Good byes



CCSD

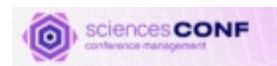


Table of contents

The effect of scope of negation on the choice of word order during sentence production, Bader Markus [et al.]	1
Iconic performances: The information structure and semantic contributions of ideophones and gestures, Barnes Kathryn	6
Internet demonstrations: Using quotation and demonstration in written language, Barnes Kathryn [et al.]	10
Categorical speaker-memory in native and non-native listeners, Beck Sara [et al.]	14
Priming Prosodic Boundaries Across Constructions and Languages, Bevivino Dorotea [et al.]	18
Gapping in Romanian comparatives: an experimental perspective on case marking and animacy, Bilbiie Gabriela	22
Interdiction as a primary reading of negated deontic necessity in child Romanian., Bleotu Adina Camelia [et al.]	25
French evaluative adjectives: position and interpretation, Brunetti Lisa [et al.]	29
Processing past time reference, but which one? An ERP study on the Mandarin Chinese aspect morphemes <i>guo1</i> and <i>guo2</i> with definite and indefinite time adverbs, Collart Aymeric	33

The Discourse Function of Constructions Predicts Island Status, Cuneo Nicole [et al.]	37
Embedded questions: Evidence in a decision-theoretic paradigm for ‘surprise’ & ‘agree’, Destruel Emilie [et al.]	41
Je peux ou je dois ? Faudrait savoir ! - Acquiring modals’ force: evidence from French, Dieuleveut Anouk	45
”Spinnt sie” or ”Spinnt die”? Empirical studies on d- and p-pronouns in German, Ehrmantraut Luise	49
Investigating lexical bias in Persian light verb constructions: What can we learn from priming experiments?, Faghiri Pegah [et al.]	53
Preposition omission in French sluicing: An empirical approach, Hassen Amal [et al.]	57
The Position of Antecedent Hypothesis in Romanian subject alternation, Istrate Marius Fabian [et al.]	61
Dependency formation during real-time processing: Evidence from webcam-based eye-tracking with subjective and objective adjectives, Kaiser Elsi [et al.]	65
Referring to someone using only their last name: Insights from gender-marked pronouns, Kaiser Elsi [et al.]	69
Extraction asymmetries in complex participle adjuncts, Kehl Andreas	73
Dutch long passive was tried to reject, Kovač Iva [et al.]	77
Predictability effects on the acceptability of antecedent-target mismatches under verb phrase ellipsis, Lemke Tyll Robin [et al.]	81
Testing subject preference and subject island in Mandarin Chinese relative clauses, Mao Ruihua	85

Anaphoric Potential of Weak Definites in Contrast with Implicit Entities and Indefinites in German, Modarresi Fereshteh [et al.]	89
Islands with bridges? Extraction from adjunct clauses in Danish, Nyvad Anne Mette [et al.]	93
How specific are linguistic structures? Mathematical priming on relative clause attachment in French, Pozniak Céline [et al.]	97
Reading non-canonical sentences in context: Identity vs. Poset, Portele Yvonne [et al.]	101
The interplay between quotation and referentiality: An empirical investigation into name-mentioning constructions, Raue Natascha [et al.]	105
What the use of the German focus particle 'auch' can tell us about the influence of structural properties of the context, Reimer Laura [et al.]	109
Dimensions of judgment in stigmatized and non-stigmatized variation, Schoenmakers Gert-Jan	113
Topic position or prefield? – Disentangling the positional restriction of topic drop in German based on acceptability rating data, Schäfer Lisa	117
The 'human-like' learnability of the wh- and coordination island constraints by artificial neural networks in Dutch, Suijkerbuijk Michelle [et al.]	121
No stone was left unturned: the passivizability of Dutch idioms., Suijkerbuijk Michelle [et al.]	125
Comprehending non-canonical and indirect speech acts in German, Trotzke Andreas [et al.]	129
Incremental negation processing with positive questions under discussion, Tsaregorodtseva Oksana [et al.]	133
The (not-)at-issue status of character viewpoint gestures, Walter Sebastian	136

The effects of information structure and sentence structure on sentence processing, Wilke Hans [et al.]	140
Contrastive topic marking with German dagegen ('in contrast') and wiederum ('in turn'), Zieleke Regina	144
Author Index	148

The effect of scope of negation on the choice of word order during sentence production

Markus Bader & Yvonne Portele (Goethe University Frankfurt)

In two experiments using the production-from-memory paradigm, we investigated how German sentences with an indefinite object in the scope of negation are linearized. Memorized sentences had either canonical order (negation and indefinite determiner contracted to “kein/no”) or non-canonical order, either a clause-medial object before the negation (S-O-neg) or a clause-initial object before the negation (O-S-neg). Participants recalled non-canonical sentences often with canonical order, but S-O-neg sentences were also recalled as O-S-neg, especially when the object stood in a poset relation to the preceding context.

1 Introduction

A large body of language production research has been concerned with the choice between canonical and non-canonical word orders during syntactic encoding. So far, this research has focused on how conceptual accessibility (e.g., animacy, givenness) affects the choice of word orders, a recurrent finding being that more accessible referents tend to be produced before less accessible referents (e.g., McDonald, Bock, & Kelly, 1993). In this study, we go beyond conceptual accessibility by investigating the effect of scope relations on the choice between canonical subject-object (SO) and non-canonical object-subject (OS) order in German. The scope relation under investigation is the relation between an existential quantifier in the scope of negation ($\neg\exists$). The most common way to express this relation is by contracting *nicht* ‘not’ and *ein* ‘a’ to *kein* ‘no’, as illustrated in (1a).¹ However, corpus data show that $\neg\exists$ is sometimes also produced with inverse order of negation and quantifier, as in (1b), although this is clearly less common than the variant with *kein*. Finally, the object in the scope of negation may also be fronted to the sentence initial position, as in (1c), resulting in a sentence with OS order. In this case, the negation must stay in situ because fronting “*kein* N” is not licit here.

- (1) Der Dozent ist mit der Seminarvorbereitung schon weit fortgeschritten.
the lecturer is with the seminar-preparation already far proceeded
‘The lecturer is already done with the preparation of the seminar.’
- a. Vermutlich wird er kein Skript bereitstellen.
presumably will he no script provide
‘Presumably, he will provide no script.’ SO-*kein*
- b. Vermutlich wird er ein Skript nicht bereitstellen.
presumably will he a script not provide
‘Presumably, he will not provide a script.’ SO-*ein*
- c. Ein Skript wird er vermutlich nicht bereitstellen.
a script will he presumably not provide
‘A script, he will presumably not provide.’ OS-*ein*

Work on quantifier scope ambiguities in general (Kiss & Pafel, 2017) and on scope in German in particular (e.g., Pafel, 2006; Wurmbrand, 2006) has revealed a preference for surface scope, as captured in the Scope Transparency Principle of Bobaljik and Wurmbrand (2012, 3):

- (2) Scope Transparency (ScoT)
If the order of two elements at LF is A»B, the order at PF is A»B.

¹An indefinite NP following the negation without contraction results in constituent instead of sentence negation.

As a violable constraint, Scope Transparency can be offset by other constraints favoring a surface order with inverted scope. In (1b) and (1c), realizing the indefinite object in front of the negation is weakly motivated insofar as the object's referent is situationally given by the prior context (e.g., preparing for a class often involves distributing a script). Furthermore, clause-medial fronting of an indefinite object (within the so-called middlefield) is more restricted than fronting to the clause-initial position. Thus, although sentence (1c) involves a non-canonical OS structure, it should be preferred to sentence (1b) with its canonical SO structure. In sum, we hypothesize the preference ranking in (3) for the three structures in (1):

- (3) Ranking of $\neg\exists$ structures:
 SO-*kein* \gg OS-*ein* \gg SO-*ein*

Experiment 1 investigated this hypothesis using the production-from-memory (PfM) paradigm, which has a long tradition in language production research. Beginning with Bock and Warren (1985), PfM has been a major tool in establishing the central role of conceptual accessibility for fixing the order of arguments and/or grammatical function assignment. In PfM experiments, sentences with non-canonical word/argument order are often recalled with canonical word order, whereas the reverse pattern occurs only rarely when at all. Experiment 2 is like Experiment 1, but strengthens the discourse relation between the object and the preceding discourse by means of a *partially ordered set relation* (poset; e.g., Ward & Prince, 1991), indicated by the adjective *weiteres* 'further' in (4). For German, corpus research has shown that poset elements occur preferentially in clause-initial position (e.g., Speyer, 2010).

- (4) Der Dozent hat schon mehrere Skripte im Seminar verteilt.
 the lecturer has already several scripts in-the seminar distributed
 'During the seminar, the lecturer already distributed several scripts.'
- a. Vermutlich wird er kein weiteres Skript bereitstellen.
 presumably will he no further script provide
 'Presumably, he will provide no further script.' SO-*kein*
- b. Vermutlich wird er ein weiteres Skript nicht bereitstellen. SO-*ein*
- c. Ein weiteres Skript wird er vermutlich nicht bereitstellen. OS-*ein*

Like Experiment 1, Experiment 2 should find evidence for the ranking in (3). In addition, due to the use of a poset relation, the number of sentences with non-canonical structure should increase.

2 Experiments

2.1 Method

2.1.1 Participants

24 native speakers of German were recruited via Prolific for each experiment.

2.1.2 Materials

We created 18 experimental item sets consisting of a context sentence, which was identical across all conditions and served as later recall prompt, and a target sentence, which had to be recalled and varied depending on the factor *Structure* (see (1) for Experiment 1 and (4) for Experiment 2). In Experiment 1, the context sentence set up a scene including an animate referent. This referent was always taken up in the second sentence with the subject pronoun *er* 'he' that occurred immediately after the finite verb in verb-second position. The object of the second sentence was an inanimate NP of varying form. In condition "SO-*kein*", the object was of the form *kein N* 'no N' and followed the subject. The sentence-initial position was filled by an adverbial. In condition "SO-*ein*", the object had the form *ein N* 'a N' and again followed

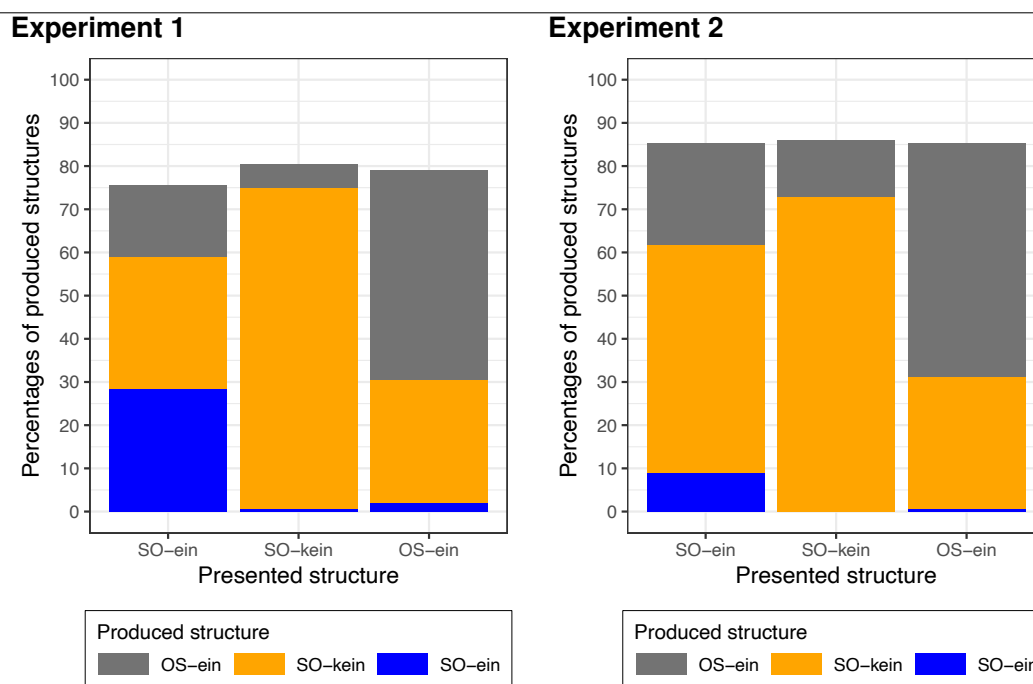


Figure 1: Percentages of produced structures depending on presented structures

the subject. This condition also contained the negation word *nicht* ‘not’ following the object. In condition “OS-*ein*”, an object of the form *ein N* occurred clause-initially and the negation *nicht* occurred at the same position as in condition “SO-*ein*”. The clause-initial adverbial of the other two conditions now occurred after the subject. In Experiment 2, the first sentence also set up a scene, but now contained an inanimate object in addition to the animate subject. The inanimate noun was repeated across the first and the second sentence. In each condition, the second sentence was identical to the one of Experiment 1 with the exception that the object now also contained either the adjective *anderes* ‘other’ or *weiteres* ‘further’. The 18 experimental sentences were combined with 54 fillers for a total of 72 sentences.

2.1.3 Procedure

The 72 sentences were randomly distributed onto 12 blocks of 6 sentences, each including 1 or 2 experimental sentences. Sentences were presented according to a Latin Square design. Each block consisted of a learning phase followed by a recall phase. In the learning phase, the six sentence pairs of a block were presented one-by-one on a single screen for reading; reading was self-paced with a 10 sec deadline. In the recall phase, the first sentence of each pair was presented and participants had to write down the second sentence from memory. Items were presented in different orders in the learning and the recall phase. The study was conducted online via *PClbexFarm* (Zehr & Schwarz, 2018). All sentences were scored according to the following categories (i) SO-*kein* (ii) SO-*ein* (iii) OS-*ein* (iv) other, including omissions and non-target like syntactic structures.

2.2 Results

Figure 1 shows how often each structure was produced in Experiment 1 and Experiment 2 depending on the structure of the memorized sentence. The bars do not sum to a 100% because of a certain amount of missing responses and responses that did not preserve the meaning of the memorized sentence (e.g., omission of the object or the negation). The results were analysed using generalized linear mixed-effects models (not shown for reasons of space). SO-*kein* sentences were most of the time recalled as SO-*kein* sentences, as expected given that this structure can be considered as canonical structure. OS-*ein* sentences were recalled

with OS-*ein* in the majority of cases, but were also restored to SO-*kein* in about a third of all cases. Most deviations from the memorized structure were found for SO-*ein* sentences, which were most of the time recalled as canonical SO-*kein* sentences. However, they were also produced with OS-*ein* structure in a number of cases, thus switching from canonical SO to non-canonical OS order. For all three structures, OS-*ein* sentences, that is, sentences with non-canonical OS word order, were produced more often in Experiment 2 where the object was related to the preceding context sentence by a poset relation. In a mixed-effects model fitted to both experiments with experiment as fixed effect, this effect was marginally significant with $p = .051$. For SO-*ein* sentences, the rate of SO-*kein* was even higher than the rate of SO-*ein* in Experiment 2.

3 Discussion

Two production-from-memory (PfM) experiments investigated syntactic encoding of sentences with an indefinite NP in the scope of negation. The results provide evidence for the structure ranking in (3). Surface scope (neg before indefinite, realized as *kein*) is the canonical structure and was produced most often. Sentences with the indefinite object in sentence-initial position were also produced with some regularity, even when not given in the input, showing that non-canonical orders can be triggered for scope reasons. SO-*ein* were only produced when this was the input structure. An object in a poset relation to the context increased the number of OS sentences, confirming that poset-given objects favor the production of non-canonical structures. In sum, our experiments show that PfM can be used to investigate syntactic encoding beyond conceptual accessibility.

Because our experiments involved written sentence presentation and production, prosody could not be taken into account. According to our intuition, neither of the sentences in (1) requires any kind of special intonation (e.g., rise-fall contour, emphatic stress on the negation). We assume that participants assigned a default prosody during reading and therefore computed the readings with the indefinite in the scope of negation. Experiments in the spoken modality are currently in preparation to test these assumptions.

References

- Bobaljik, J. D., & Wurmbrand, S. (2012). Word order and scope: Transparent interfaces and the 3/4 signature. *Linguistic inquiry*, 43(3), 371–421.
- Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47–67. doi: 10.1016/0010-0277(85)90023-X
- Kiss, K. E., & Pafel, J. (2017). Quantifier scope ambiguities. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (2nd ed., Vol. 5, pp. 1–36). Oxford: Blackwell. doi: <https://doi.org/10.1002/9781118358733.wbsyncom005>
- McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25, 188–230.
- Pafel, J. (2006). *Quantifier scope in German*. Amsterdam: John Benjamins.
- Speyer, A. (2010). Filling the German vorfeld in written and spoken discourse. In S.-K. Taniskanen, M.-L. Helasvuo, & M. Johansson (Eds.), *Discourses in Interaction* (Vol. 203, pp. 263–290). John Benjamins Publishing.
- Ward, G. L., & Prince, E. F. (1991). On the topicalization of indefinite NPs. *Journal of Pragmatics*, 16(2), 167–177.
- Wurmbrand, S. (2006). Verb clusters, verb raising, and restructuring. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (1st ed., Vol. 5, p. 229–343). Oxford: Blackwell.
- Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. doi: 10.17605/OSF.IO/MD832

Iconic performances: The information structure and semantic contributions of ideophones and gestures

Kathryn Barnes

Goethe Universität, Frankfurt am Main

Kita (1993) found that up to 94% of ideophones in a Japanese corpus were accompanied by a gesture, with the majority being iconic gestures. Researchers now generally agree that gestures contribute information in addition to the accompanying speech and this also appears to be the case for ideophones. Dingemanse (2012) has argued that ideophones and iconic gestures form two parts of one multimodal iconic performance, with the speaker using both modalities in order to maximise the iconic potential of the performance. While in recent years researchers such as Ebert, Ebert & Hörnig (2020), Esipova (2019) and Schlenker (2018) have proposed semantic analyses of the meaning contributions of iconic gestures and Ebert & Barnes (in draft), Henderson (2016) and Kawahara (2020) have provided formal semantic analyses of ideophones in German, Tselal and Japanese; there has to date been no research into the combined meaning contributions of ideophones and gestures.

The key research questions for ideophone and gesture combinations concern what information the gestures adds to the ideophone expression and what the semantic contribution of these iconic performances is. The goal of this research is to provide an initial semantic analysis for ideophones and gestures occurring together, while also proposing future research to investigate the exact meaning contributions of gestures alongside ideophones. Both Dingemanse (2015) and Nuckolls (2019) have shown evidence that gestures appear to contribute additional information on top of the ideophone. This can be seen in (1), elicited by Dingemanse (2015, p.219) during fieldwork on Siwu. In this example, the participant performed a gesture described as “right hand flat, moves from upper right down to alongside body depicting flow of water” (p.219).

(1) The water goes [yááá]-GUSHING

The gesture contributes additional information about the path of the water and could potentially include further information such as manner or speed. As such then, iconic performances with ideophones and gestures seem to allow speakers to not only give multisensory information about the described events through the spoken modality, but to enhance this depiction by using the visual medium to give additional information about said event which is not already encoded in the ideophone. As Dingemanse (2013) notes, this makes the most of the ideophone’s iconic potential and allows speakers to give a more embellished performance of the event they are discussing.

Nevertheless, the combination of the two iconic components does not seem to impact the at-issue status of the co-speech gesture, which remains not-at-issue (cf. Ebert, Ebert & Hörnig 2020; Schlenker 2018). This can also be seen in attempting to deny the gesture contribution in (2), taken from Dingemanse (2015), where the gesture is described as “both hands flat, palm down, moving and meandering horizontally while body is turning” (p. 219):

(2) A: The water just goes [yááá]-GUSHING
B: No, that’s not true, the water was moving quite slowly.
B: # No, that’s not true, the water went straight past.

It seems that the gesture behaves as a normal co-speech gesture; it is not at-issue and cannot be

targeted by a direct denial. In this case then, the combination of gesture and ideophone does not appear to be able to shift the gesture towards at-issue status.

Based on these observations, an initial semantic account of ideophone and gestures can be provided. I follow Ebert, Ebert & Hörnig (2020) and assume that the lexical meaning of a gesture is reference to an individual, the gesture referent. Depending on the temporal alignment of the gesture and speech, this gesture referent stands in varying relations to the verbal referent and this meaning contribution is default not at-issue. An ideophone, on the other hand has two meaning contributions; the first being its conventionalised meaning which behaves in the same manner as other arbitrary items from the same syntactic category. For example the conventionalised meaning of an ideophone used predicatively will provide an event argument, whereas an ideophone used adverbially will function as an event modifier. The second meaning component of an ideophone is its iconic meaning, which comes about due to varying aspects of the ideophone utterance itself, which is formalised using demonstrations, d (cf. Davidson 2015). The ideophone utterance as a demonstration then stands in a *SIM* relation to the main event. The iconic meaning component of an ideophone is default not at-issue, whereas the conventionalised meaning component can be shifted towards at-issue status given the right conditions.

Assuming demonstrations to be a proper subset of events, then we can also assume that demonstrations are connected to the domain of individuals via theta roles (cf. Henderson 2016). As such, when an iconic gesture co-occurs with an ideophone, the gesture referent fills a theta role in the demonstration argument structure and in doing so generates an implicature about properties or actions of a salient individual in the main event.

This approach can be applied to the gesture ideophone combination in (1). Due it being used predicatively, the conventionalised meaning component of the ideophone is necessarily at-issue and contributes that the reported event is some sort of whooshing event. However, the iconic part of the predicative ideophone remains not at-issue (cf. Barnes et al. 2022). This iconic part contributes as its meaning a demonstration, namely the utterance of $\gamma\acute{a}\acute{a}\acute{a}$ event, which is *SIM* in the relevant dimensions to the event of the water gushing. The gesture on the other hand refers to the agent of the demonstration and iconically depicts its path. The combination of the ideophone and gesture therefore makes the not-at-issue contribution that the gushing event is similar in the relevant dimensions to the demonstration and as such that the movements of the agent in the demonstration is similar to the movements of the water in the reported event.

It is then possible to provide the analysis in (3-b) for the adapted version of (1) in (3-a). (Rough gloss given in (3-c)).

- (3)
- a. The water goes [whoosh]- GUSHING.
 - b. $[e] \wedge \text{agent}(e, x) \text{water}(x) \wedge \text{goes whoosh}_p(e) \wedge [d] d = d_{\text{whoosh}} \wedge [z] \wedge z = g \wedge \text{agent}(d, z) \wedge \text{SIM}_p^*(d, e)$
 - c. There is an event, e and the agent of e is the water and e is an event of whooshing. There is a demonstration, d , namely the utterance of *whoosh* or d_{whoosh} , and there is a gesture, z , whose referent is g and z is the agent of d_{whoosh} and d_{whoosh} is similar in the relevant dimensions to e .

This preliminary analysis provides a basis from which we can begin to formalise the meaning contributions of iconic performances containing multiple iconic enrichments.

There are however many remaining questions, particularly concerning what kind of information

a gesture can contribute when it co-occurs with a gesture. Kita & Özyürek (2003) showed that gestures tend to be shaped by how the information is structured in the verbal domain, but may add information to this linguistic expression. For example, when recounting the plot of cartoons, English speakers often accompanied the verb *swing* with gestures showing an agent's trajectory while swinging or the manner in which they swung. In German, ideophones predominantly encode sound and movement (cf. Ćwiek submitted), which, as it is difficult to imagine how a gesture could encode sound, suggests that the majority of gestures will be restricted to depicting aspects of movement when co-occurring with such ideophones. As such, the hypothesis in this study is that iconic gestures accompanying ideophones in German will encode information around trajectory or manner in a similar manner to gestures which accompany English verbs such as *swing*. A further prediction is that these gestures will have the same at-issue status as other co-speech iconic gestures and also contribute not-at-issue information.

In order to test the first of these predictions, I am in the process of conducting a production study on ideophones and gestures in German. The study will be conducted with native speakers of German, primarily primary school teachers or others who regularly work with children, as such participants will be more likely to produce ideophone and gesture combinations. Its aim will be to provide a systematic review of the combined meaning contributions of ideophone and iconic gestures, allowing for a better understanding of the kinds of information contributed by both ideophones and gestures. The methodology will partly adapt the fieldwork conducted by Dingemanse (2015) by asking participants to give definitions for ideophones and to provide example sentences with each ideophone. The ideophones to be used will be predominantly those which have been shown to encode both sound and movement in German (cf. Ćwiek submitted), for example *holterdiepolter* 'helter-skelter', *plums* 'thud', *schnippschnapp* 'snip snap', *wusch* 'woosh'. Participants will be asked to imagine they are explaining the ideophone to a 6 year old child, which it is hoped, will result in a greater amount of gestures. I would hope to have initial data from this study to present alongside the theoretical analysis at Linguistic Evidence.

References

- 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). DOI: 10.16995/glossa.5827.
- Ćwiek, Aleksandra (submitted). 'Iconicity in Language and Speech'. PhD thesis. Berlin: Humboldt-Universität zu Berlin.
- Davidson, Kathryn (2015). 'Quotation, Demonstration, and Iconicity'. *Linguistics and Philosophy* 38: 477–520. DOI: 10.1007/s10988-015-9180-1.
- Dingemanse, Mark (2012). 'Advances in the Cross-Linguistic Study of Ideophones'. *Language and Linguistics Compass* 10 (6): 654–672. DOI: 10.1002/ln3.361.
- (2013). 'Ideophones and gesture in everyday speech'. *Gesture* 13 (2): 143–165. DOI: 10.1075/gest.13.2.02din.
- (Oct. 2015). 'Folk definitions in linguistic fieldwork'. In: *Language Documentation and Endangerment in Africa*. Ed. by James Essegbey, Brent Henderson & Fiona McLaughlin. Cul-

-
- ture and Language Use. Amsterdam: John Benjamins Publishing Company, 215–238. DOI: 10.1075/clu.17.09din.
- Ebert, Christian, Cornelia Ebert & Robin Hörnig (2020). ‘Demonstratives as dimension shifters’. *Proceedings of Sinn und Bedeutung*. Ed. by Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab. Vol. 24. 1, 161–178. DOI: 10.18148/sub/2020.v24i1.859.
- Ebert, Cornelia & Kathryn Barnes (in draft). ‘The information status of iconic enrichments: Evidence for gradient at-issueness. (Invited contribution to Theoretical Linguistics)’. *Theoretical Linguistics*.
- Esipova, Maria (2019). ‘Composition and projection in speech and gesture’. PhD thesis. New York, NY: New York University.
- Henderson, Robert (Oct. 2016). ‘A demonstration-based account of (pluractional) ideophones’. *Proceedings of Semantics and Linguistic Theory*. Ed. by Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf. Vol. 26, 664–683. DOI: 10.3765/salt.v26i0.3786.
- Kawahara, Koji (2020). ‘Subjective Ideophones and Their Core Meanings’. *Japanese/Korean Linguistics*. Ed. by Shoichi Iwasaki, Susan Strauss, Shin Fukuda, Sun-Ah Jun, Sung-Ock Sohn & Kie Zuraw. Vol. 26, 1–10.
- Kita, Sotaro (1993). ‘Language and thought interface: A study of spontaneous gestures and Japanese mimetics’. PhD thesis. Chicago, IL: University of Chicago.
- Kita, Sotaro & Aslı Özyürek (2003). ‘What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking’. *Journal of Memory and Language* 48: 16–32.
- Nuckolls, Janis B. (May 2019). ‘The sensori-semantic clustering of ideophonic meaning in Pastaza Quichua’. In: *Ideophones, Mimetics and Expressives*. Ed. by Kimi Akita & Prashant Pardeshi. Amsterdam: John Benjamins Publishing Company, 167–198. DOI: 10.1075/i11.16.08nuc.
- Schlenker, Philippe (2018). ‘Gesture Projection and Cosuppositions’. *Linguistics and Philosophy* 41 (3): 295–365. DOI: 10.1007/s10988-017-9225-8.

Internet demonstrations: Using quotation and demonstration in written language

Kathryn Barnes and Cornelia Ebert

Goethe Universität, Frankfurt am Main

Davidson (2015) first noted that analyses of quotation have neglected to account for how quotation actually occurs in spoken language. She provides a demonstration based analysis of quotation in spoken language, which accounts for iconic enrichments, such as gestures and prosodic modulations, that often accompany spoken quotations. The goal of this research is to extend this analysis to examples of reported speech on the internet, which also frequently feature reports of action alongside the reported speech. Typical examples can be seen in Figure 1. Although these kinds of speech and action reports are very common on social media platforms like Twitter, Instagram and Tumblr, as far as we are aware, there have been no previous attempts to provide a semantic analysis of such instances of reported speech. In this analysis, we propose to follow Davidson (2015) and analyse the action reports as a form of internet demonstration, somewhat comparable to gestures and prosodic modulation that may accompany reported speech in spoken language.

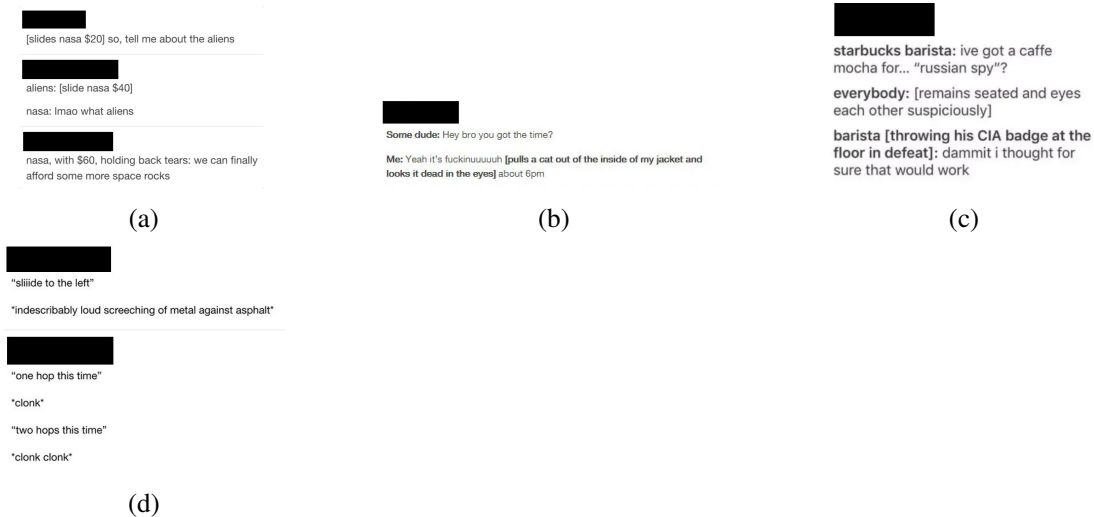


Figure 1: Examples of internet demonstrations from Golder (2019)

The general form of social media speech reports is the speaker's name followed by a colon and then the reported speech and/or action. The action reports may be interspersed throughout the quoted speech, as in (1a) and (1b) to indicate actions or sounds occurring at different points during the reported event, or alongside the speaker's name before the colon, as in (1a) and (1c), and sometimes they occur after the colon, but without any accompanying quotation, also as in (1a). The action reports are generally marked with asterisks or square brackets and use distinct morphology, in contrast to the reported speech which is direct quotation. This morphology consists of the third person inflection or the gerund in English, as in (1a),(1b) and (1c), for example, or reduced conjugation in German. There are also cases where the action report occurs without any quotation present, as in (1d).

Davidson (2015) analyses quotation in spoken language as a demonstration, *d*, which allows for the iconic enrichments such as gestures or prosodic modulation, which frequently occur alongside

these quotations, as can be seen in (1).

- (1) a. And Bill was like [that’s not fair]_WHINY VOICE.
b. And Bill was like + GOBBLING GESTURE

Davidson (2015) argues that the demonstrative *like* provides the demonstration argument in quotations and that the demonstration stands in a *demonstration-of* relation to the reported speech event, where *demonstration(d,e)* applies just in case *d* reproduces relevant aspects of *e*.

The examples of iconic enrichments alongside reported speech in (1) clearly resemble the internet demonstrations observed alongside reported speech on social media. They seem to behave in a similar manner and to add similar information to their accompanying reports of speech. However, there is a crucial difference between internet demonstrations and iconic enrichments in spoken quotation; internet demonstrations are not iconic, as they are written and therefore employ a conventionalised system of arbitrary signs. As such, they are not directly gestural in the same manner that prosodic modulation or manual gestures would be in spoken language and cannot be modelled via a *demonstration-of* relation as proposed by Davidson (2015). As such, the demonstration argument is not necessary for this analysis, as there is no need to model an iconic relation.

Instead, I propose an analysis of reported speech and action on social media using events and subevents. The conventionalised form of introducing the speech and/or action report, namely a speaker or agent’s name followed by a colon, introduces an main event argument. Quoted speech, which is generally direct speech and therefore marked by first person inflection, introduces a subevent of the main event. This is then an event of speaking, which will also result in the introduction of the covert predicate *speak*, which takes the reported speech as its complement. The conventionalised markers of internet demonstrations, such as an asterisk or square brackets or third person inflection, then introduce a further subevent of the main event, which is the reported action event. These subevents can stand in differing relations to each other within the main event.

Based on this, we can give the following analysis of the reported speech and demonstration in (2-a) (taken from 1c) as in (2-b), which is roughly glossed as in ??.

- (2) a. Barista: [throwing his CIA badge on the floor in defeat] Dammit, I thought for sure it would work.
b. $[e] \wedge \text{agent}(e,x) \wedge \text{barista}(x) \wedge [e'] \wedge e' \subseteq e \wedge \text{throws-CIA-badge-on-the-floor-in-defeat}(e') \wedge [e''] \wedge e'' \subseteq \text{espeak}(e'') \wedge \text{form}(e'') = \text{“Dammit, I thought for sure it would work”} \wedge$

Here the construction *Barista:* introduces the main event *e* and posits the barista as the agent of this event and therefore all subevents. The square brackets then introduce the subevent *e'*, which contains the action report that the barista throws his CIA badge on the floor in defeat. Finally, the speech report beginning with *Dammit, I thought* introduces a second subevent, an event of speaking and therefore also introduces the covert predicate *speak*, whose complement is the speech report of what the barista said.

The alignment of the quoted speech and the demonstration also seems to play a role in the interpretation of the demonstration. This can be seen in Figure (1b), where the event of pulling the cat out of the jacket is clearly supposed to be interpreted as occurring part way through the speech event. The analysis proposed here then allows for the main event to be split into several subevents, which can then stand in differing temporal relations to each other. For example, (2-a), the subevent

of the barista speaking and the subevent of throwing their badge on the floor occur simultaneously, whereas in Figure (1b), the action event of removing the cat from the jacket occurs partway through the speech event. Ironing out the exact details of this temporal alignment is left to future research.

This analysis can furthermore account for cases where the agent’s name followed by a colon is used in combination with an image or gif, in place of written language, in order to indicate the actions of the agent. In this case, the action report clearly relies upon an iconic relation between the image and the reported agent’s actions. While the *agent:* still introduces a main event, the use of an image generates two implicatures; firstly that there is a subevent of e , e' , which is the event of the agent’s actions and secondly that the gif is a demonstration of this subevent e' . This is best illustrated with the example in 2, where the gif is a short clip of man jumping off a balcony before landing on a stage transformed as a popstar singing. The Bonnie Tyler song “Holding out for a Hero” plays in the background.

Me:



Figure 2: Source: @itslitgayshit Instagram Reel

This can be analysed as in (3-b).

- (3) a. *me:* [Gif of man jumping over balcony as “Hero” plays in the background]
 b. $[e] \wedge \text{agent}(e, \text{SPEAKER}) \wedge [e'] \wedge e' \subseteq e \wedge [d] \wedge d = d_{\text{gif}} \wedge \text{demonstration}(d, e)$

Here the construction *me:* again introduces a main event. The use of a gif within this construction then implies that there is a subevent of e , e' , which is the event of the speaker’s actions and that the gif is a demonstration of this event and therefore these actions.

An additional question regarding these internet demonstrations concerns their at-issue status. Ebert & Hinterwimmer (in draft) argue that iconic enrichments, such as gestures and prosodic modulations, which accompany reported speech are by default not at-issue. Whether this is also the case for internet demonstrations also requires further empirical investigation. When they occur alongside reported speech, the action reports do appear to be not at-issue. Take the final part of Figure 1c: while it seems acceptable to directly target the reported speech of the barista with a denial, it seems somewhat odd to target the information given in the demonstration, as can be seen

in (4).

- (4) a. A: Barista: [throwing his CIA badge on the floor in defeat] Dammit, I thought for sure it would work.
b. B: That's not true, the barista didn't say that!
c. B': # That's not true, the barista didn't throw his badge on the floor!

In cases where the demonstrations occur within such quotation structures, but without any accompanying reported speech as in (1a), we argue that they resemble *pro-speech gestures* (cf. Schlenker 2018) and are necessarily at-issue due to the lack of any other at-issue content in the utterance and that this would also be the case for examples such as 2. There are also likely to be cases where the demonstration does occur with reported speech and is still at-issue, though when and how this could be the case remains to be determined.

These social media demonstrations are very similar to cases of spoken quotation with *be like*, which not only indicates the resemblance of internet language to spoken language, but also highlights how speakers compensate for the one dimensional nature of internet communication. Faced with a lack of multimodality and access to iconicity, which they would normally exploit when reporting speech in spoken language, social media users have developed new conventions for introducing descriptions of non-verbal actions alongside speech reports, allowing them to enhance their narratives and to include these aspects, even when restricted to the written medium. This in turn provides evidence for the argument initially made by Davidson (2015) that demonstrations play a key role in quotation; even in the written modality, speakers will seek to incorporate demonstrations in their utterances.

References

- Davidson, Kathryn (2015). 'Quotation, Demonstration, and Iconicity'. *Linguistics and Philosophy* 38: 477–520. DOI: 10.1007/s10988-015-9180-1.
- Ebert, Cornelia & Stefan Hinterwimmer (in draft). 'Free Indirect Discourse Meets Character Viewpoint Gestures'. *Proceedings of Linguistic Evidence 10*.
- Golder, Andy (Oct. 2019). '100 Tumblr Posts From The 2010s That Are All Funny, No Filler, Total Classics.' *Buzzfeed*. URL: <https://www.buzzfeed.com/andyneuschwander/100-of-the-%20best-funniest-tumblr-posts-of-the-2010>.
- Schlenker, Philippe (2018). 'Iconic Pragmatics'. *Natural Language & Linguistic Theory* 36: 877–936. DOI: 10.1007/s11049-017-9392-x.

Categorical speaker-memory in native and non-native listeners

Sara D. Beck & Andrea Weber, University of Tübingen

While listeners associate speech pattern signals with specific speakers to remember what was said, less is known about memory for speakers and how listener proficiency affects such memory. In a recognition memory task, we tested non-native and native participants' ability to associate objects of prototypical colors (e.g., red lobster, green frog) with speakers. One of two speakers referred to a subset of 30 items (instructions: "Click on the ..."), presented in either a categorical (one color per speaker) or random (mixed speaker-color) condition. While native listeners showed significant improvement in speaker recognition in the categorical condition, non-native listeners did not.

1 Introduction

Listeners can learn to associate patterns in the speech signal with specific speakers (e.g., higher-pitched voices for female speakers) and use this information for understanding spoken input (e.g., Lattner & Friederici, 2003). These learned associations can furthermore help to memorize what has been said (e.g., Clopper, Tamati & Pierrehumbert, 2016), also in cases when associations are learned between objects and speakers who refer to them rather than between the speech signal and speakers (e.g., McKinley, Brown-Schmidt, & Benjamin, 2017). In addition to information about what has been said, information about who has said it is also available for storage, but less is known about whether learned associations also help to memorize who the speaker was. Remembering both what has been said as well as who said it can help in the long run to build general knowledge about stereotypical correspondences between objects and speakers (e.g., small children and toys). Such information may be particularly critical for listeners with less social experience in a particular speech environment, such as non-native speakers, as making and remembering these connections may help to bridge social or cultural gaps. However, less proficient listeners may be at a disadvantage as they need to deal with the additional cognitive demands of L2 processing (e.g., Morishima, 2013) and may have more difficulty memorizing categorically associated information (e.g., Waring, 1997).

The current study investigates how this process of associating speakers with object categories impacts memory for speakers and how this relationship varies as a function of participants' language proficiency. To our knowledge, there are no current studies that have examined this type of speaker-item-based relationship for non-native listeners; though, there is some research for native listeners. Horton and Slaten (2012), for instance, investigated how newly learned speaker-item associations are used online to predict linguistic behavior in eye-tracking. Native participants first listened to speakers referring to various tangram objects, and subsequently a speaker's voice predicted participants' looks to objects that had previously been referred to by the speaker. McKinley and colleagues (2017) used a referential communication task in which participants first had to label pictures of objects before they were asked *inter alia* who had used the label. Note that neither in Horton and Slaten (2012) nor in McKinley et al. (2017) was there an ascertainable pattern between speakers and objects. Considering talker-recognition studies, however, there is evidence that listeners are less able to identify talkers in their L2 compared to their L1 (e.g., Bregman & Creel, 2014), even when memory for the speaker is in focus, as is not the case in the current study.

The current study incorporates the aspect of a categorical pattern, particularly the prototypical color associated with an item (e.g., frogs are prototypically green), such that speakers either showed a pattern of referring only to objects of one color or referred to objects of several colors (no pattern). While there is little research following this schema,

research on categorical memory suggests that semantic patterns, for example, can aid memory for words in lists (e.g., Poirier & Saint-Aubin, 1995), and even that patterns need not be part of a conscious memory strategy to contribute to processing (see e.g., Schacter et al., 2004). For non-native participants, the influence of such categorical information is less clear, but such categories may even hinder memory performance (e.g., Waring, 1997, but see Hoshino, 2010).

2 Experiment

Using a recognition memory task, the ability of non-native and native participants to associate objects that have prototypical colors with speakers was tested. Based on the processing research discussed above, we expected that native participants will show an increase in memory performance when color patterns are associated categorically with specific speakers in comparison to a random association. While predictions for non-native participants were less clear, we expected that this increase may be limited if present at all.

2.1 Methods

2.1.1 Participants

Sixty native speakers of American English (18-35, mean: 26.97, 32 male, 27 female) and 62 highly proficient non-native speakers (German L1, 18-49, mean:25.26, 43 female, 18 male) participated in the experiment online via Gorilla Experiment Builder (Anwyl-Irvine et al., 2020).

2.1.2 Materials

Experimental items were 30 pictures of objects that prototypically belong to the perceptual categories of either red or green objects (e.g., red lobster, green frog). Strong associations between the objects and their prototypical color were confirmed via association strength norms from the Small World of Words database (Deyne et al., 2019). All pictures were shown in black and white during the experiment. Two native speakers of American English (male, 31 and female, 33) were recorded referring to all objects with the carrier phrase “Click on the ...”.

2.1.3 Procedure

In an association phase, two objects were shown on the screen in a trial, and a recording from either the male or female speaker instructed participants to click on one of the objects. In a categorical condition, all objects referred to by one speaker had the same prototypical color (i.e., the female speaker only referred to green objects and the male speaker only to red objects, or vice versa), and in a random condition, both speakers referred to objects of both colors arbitrarily. The 30 objects were referred to twice by the same speaker across two blocks.

In the testing phase, pictures of the objects were shown individually, and participants had to indicate whether the male or female speaker had referred to the object previously. All participants ended the experiment with a short questionnaire, including some language background information and a question about whether they had noticed a pattern in the speakers’ referential expressions; non-native participants additionally completed the LexTale task (Lemhöfer, 2012) to assess their proficiency in English.

2.2 Analysis and Results

Linear mixed-effects regression models were performed with *correctness* (1 = correct, 0 = incorrect) as the dependent variable and *category* (categorical and random, coded as 0.5 and -0.5, respectively) and *language* (native, non-native, coded as 0.5, and -0.5, respectively) as fixed effects. *Subjects* and *Items* were also included as random factors with random slopes, where justified. Other effects considered in model-building were *trial order*, *speaker gender*, *item color*, *participant gender*, *participant age*, and indicated *use of headphones*. All factors were numerically centered around zero, and binary factors sum-

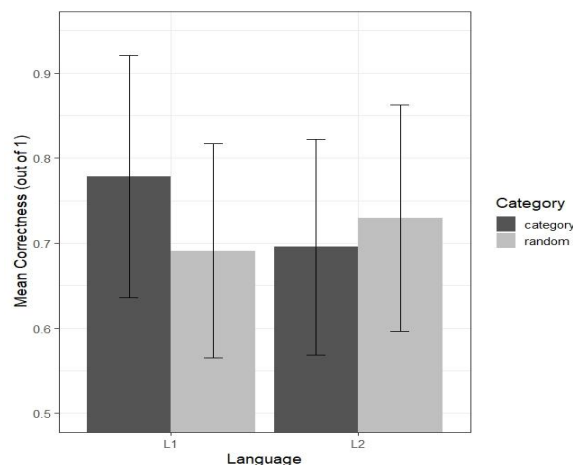
coded as above. One participant did not follow instructions and four participants performed very poorly on the task and were removed from the data analysis, leaving 117 participants in the final analysis (58 L1, 59 L2).

Table 1

Fixed Effects	β	SE	t	Pr(> t)
(Intercept)	0.724	0.018	41.017	<2e-16 ***
Category	0.027	0.030	0.904	0.3684
Language	0.022	0.028	0.801	0.4248
Category x Language	0.122	0.055	2.201	0.0298 *
Random Effects	Variance	SD	Correlation	
Subject	0.016264	0.12753		
Item	0.003626	0.06022		
Category	0.003296	0.05741	-0.11	

The results are summarized in Table 1, including only the factors that improved model fit. A significant interaction between *language* and *category* suggests that recognition memory for speaker-item associations was impacted differently by perceptual color patterns for native and non-native participants. The mean correctness by *language* and *category* is displayed in Figure 1 (whiskers display standard error of the mean).

Figure 1



Further analyses of each language group individually confirm that native participants show improved retrieval of speaker-item associations in the categorical condition significantly ($\beta = .087$, $t = 2.167$, $p < .05$) whereas non-native participants did not show an effect ($\beta = -.040$, $t = -1.051$, $p = .299$).

3 Conclusion

This study found differences in retrieval between native and non-native participants for speaker-item associations. Specifically, category-based speaker preferences influenced native participants' memory, while memory of highly proficient non-native participants was not influenced by that information. The questionnaire also suggested that this memory advantage was not dependent on consciously noticing speaker preferences, as participants had not reliably identified the perceptual pattern in the categorical condition.

The results were in line with expectations that the processing abilities of native speakers to use referential information in online processing (e.g., Horton & Slaten, 2012; MicKinley et al., 2017) translates to benefits for memory. The lack of memory benefit in non-native listeners

is in line with both decreased abilities in talker recognition (e.g., Bregman & Creel, 2014) as well as the possibility that fewer free processing resources may prevent L2 listeners from taking advantage of the same information (e.g., Morishima, 2013; Sorace & Filiaci, 2006). Thus, while a steady category preference can improve native participants' speaker memory, the same is not true for non-native participants.

References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Bregman M.R., Creel, S.C. (2014). Gradient language dominance affects talker learning. *Cognition*, 130(1), 85-95. <https://doi.org/10.1016/j.cognition.2013.09.010>
- Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of Phonetics*, 58, 87-103. <https://doi.org/10.1016/j.wocn.2016.06.002>
- Deyne, S. de, Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods*, 51, 987–1006. <https://doi.org/10.3758/s13428-018-1115-7>
- Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition*, 40, 113–126. <https://doi.org/10.3758/s13421-011-0135-7>
- Hoshino, Y. (2010). The Categorical Facilitation Effects on L2 Vocabulary Learning in a Classroom Setting. *RELC Journal*, 41(3), 301–312. <https://doi.org/10.1177/0033688210380558>
- Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—evidence from event-related brain potentials. *Neuroscience Letters*, 339(3), 191-194. [https://doi.org/10.1016/S0304-3940\(03\)00027-2](https://doi.org/10.1016/S0304-3940(03)00027-2)
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325-343. <https://doi.org/10.3758/s13428-011-0146-0>
- McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, 45(8), 1281-1294. <https://doi.org/10.3758/s13421-017-0730-3>
- Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, 50(8), 577-597. <https://doi.org/10.1080/0163853X.2013.846964>
- Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *QJEP*, 48(2), 384–404. <https://doi.org/10.1080/14640749508401396>
- Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, 5(11), 853–862. <https://doi.org/10.1038/nrn1534>
- Sorace A, Filiaci F. Anaphora resolution in near-native speakers of Italian. *Second Language Research*. 2006;22(3):339-368. <https://doi.org/10.1191/0267658306sr2710a>
- Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, 25, 261-274. [https://doi.org/10.1016/S0346-251X\(97\)00013-4](https://doi.org/10.1016/S0346-251X(97)00013-4)

Priming Prosodic Boundaries Across Constructions

Dorotea Bevivino & Giuseppina Turco – Université Paris Cité, LLF, CNRS
dorotea.bevivino@u-paris.fr

Unlike perception studies, production studies have so far failed to prime intonational phrase structure. One reason for this might be syntactically-biased materials. This study aims to assess whether prosodic boundaries can be primed in production when their saliency is not reduced by a strong syntactic bias. We replicated the prosodic priming paradigm from Tooley et al. (2014, 2018) in an online production study testing two constructions: the original PP-attachment (transfer of location) sentences used by Tooley et al. and newly constructed relative clause attachment constructions. A norming study confirmed a strong syntactic bias in the original materials while showing more balanced preferences for RC attachment. Prosodic analyses revealed a priming effect for RC attachment constructions only, suggesting that intonational phrase structure can be primed in production in proper conditions

1 Introduction

In the last decades, research in online sentence comprehension has pointed to the key role played by prosody in language processing (for a review, Pratt, 2017). In a series of perception studies, Jun and Bishop (2015) (see also Mills, 2020) have shown that by priming the intonational phrase structure (i.e., prosodic boundaries), it is possible to promote one interpretation of syntactically ambiguous sentences over the other. Surprisingly, the quite robust prosodic priming effect found in perception has not been replicated in production experiments, where intonational phrase boundaries were not primed (Tooley et al., 2014, 2018). One possible explanation to this asymmetry is that the production studies conducted so far tested the effect of prosodic boundaries under the hardest possible conditions – that is, when the intonational phrase structure is redundant to the internal structure of the sentence and/or when there is a strong syntactic preference boosting one interpretation of an ambiguous sentence over the other.

In light of these findings and their potential limitations, we replicated the prosodic priming paradigm from Tooley et al. in online studies testing two different structures which we normed for their underlying interpretational bias. The goal of the experiments was to assess whether prosodic boundaries can be primed in production, when the saliency of prosodic cues is not heavily reduced by strong syntactic biases. Our overall hypothesis was that, when addressing some methodological limitations in the design, it would be possible to see a priming effect of prosodic boundaries in production as well.

2 Experiment 1: Norming study

Experiment 1 was a norming study with the double purpose of (i) testing our hypothesis of a bias in the materials originally tested, and (ii) validating the ambiguity of newly-created materials. If our hypothesis is correct, we can expect a difference in the ratings for the two interpretations of the original sentences, but not for the new set of sentences.

2.1 Methods

Participants: Twenty-eight young adult (18-35 years old, M age = 26.86) English speakers from the same UK area were recruited online via Prolific. All participants had acquired the target language as (one of) their first language(s), had been exposed to it for the most part of their lives while growing up, and still has it as their strongest and dominant language (Cheng et al., 2021). Participants were controlled for gender, age, and educational level. All participants had normal or corrected-to-normal vision and no hearing impairments. No participants presented any known neurological, speech, or communication disorders at the time of testing. All procedures were performed in accordance with standard ethical guidelines and protocols, as approved by the Ethics Committee at the Université Paris Cité and the University of York. All participants provided informed consent prior to testing.

Materials: The experimental stimuli consisted of 80 ambiguous sentences: 40 sentences describing transfer-of-location events and 40 relative clause attachment sentences. For the transfer-of-location sentences, we used the same PP-attachment sentences from Tooley et al.'s (2018) second

experiment. An example of an ambiguous transfer sentence with its two alternative readings can be seen in (1a-b). For the relative clause attachment ambiguities (2a-b), we created globally-ambiguous sentences, controlled for length, semantic plausibility, and all adhering to a constant structure to avoid any preferences for low or high attachment (Grillo et al., 2015; Hemforth et al., 2013, 2015).

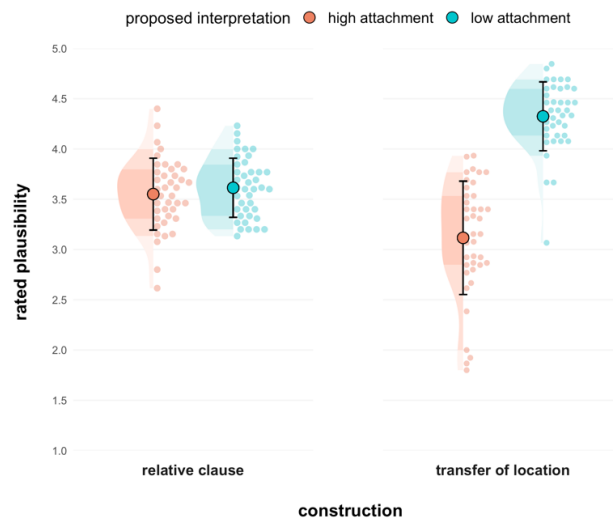
- (1) She put the money in the basket on the table.
 - a. *The money was in the basket and she put it on a table* (high attachment)
 - b. *She put money in the basket that is on the table* (low attachment)
- (2) She stayed with the patient of the doctor who waits for the blood results.
 - a. *The patient of the doctor waits for the blood results* (high attachment)
 - b. *The doctor waits for the blood results* (low attachment)

Procedure: The experiment consisted in an online acceptability judgement task created in PCibex (Zehr & Schwarz, 2018) and run on the university-hosted Ibex farm server. Participants were shown an ambiguous sentence and asked to rate on a 1-5 Likert scale the plausibility of one of its possible interpretations. The proposed interpretations (high vs. low attachment) were distributed across lists. A simple Y/N comprehension question followed roughly 25% of the experimental sentences. Filler sentences with either only a plausible reading or a totally implausible reading were added at the end of the task as a baseline for judgements.

2.2 Results

The plausibility ratings for the experimental materials were analyzed fitting a cumulative link mixed model, including main effects of construction and proposed attachment (both sum-coded), as well as the two-way interactions between them. Participant and item were entered as random effects. All analyses were performed in the R environment (R Core Team, 2021.09) using the ordinal package (Christensen, 2019). The interaction effect is plotted in Figure 1. In line with our initial prediction, the results of the model revealed a significant interaction between construction and proposed attachment ($\beta = 0.49373$, $z = 0.04046$, $SE = 12.202$, $p < 0.001$), with low-attachment interpretations of transfer-of-location sentences being preferred on average of 1.97 rating points to the alternative high-attachment interpretations. The results showed no effect of attachment on the plausibility of the relative clause sentences.

Figure 1. *Plausibility ratings on the interpretations of ambiguous sentences*



2.3 Discussion

The results of the norming study on the ambiguity of the materials showed a strong preference for one interpretation over the other in the transfer-of-location sentences but not in the relative clause attachment constructions. These findings substantiate the claim of a strong syntactic bias in the material originally tested (Tooley et al., 2014, 2018); whereas they confirm the more balanced ambiguity of the newly-created materials.

3 Experiment 2: Prosodic Priming

Experiment 2 replicated Tooley et al.'s prosodic priming paradigm with the two constructions to test whether there is a prosodic priming effect in production, when the saliency of prosodic cues is not heavily reduced by syntactic attachment preferences. If intonational phrase boundaries can be primed, we expect to see an effect of boundary location in the less biased materials – i.e., the relative clause attachments. Specifically, we expect sentences primed for an early boundary to lead speakers to produce longer NP1s than NP2s (presenting an early boundary as well) both in the repetition and in the production of a new relative clause; whereas we expected sentences primed in the late boundary to lead speakers to produce longer NP2s than NP1s (presenting a late boundary as well) both in repetition and in production.

3.1 Methods

Participants: Forty participants with the same inclusion/exclusion criteria as the norming study were recruited online via Prolific. Seventeen participants were subsequently excluded from the final data analyses due to bad audio quality and/or production accuracy, resulting in a final sample of 23 participants and 1608 recorded sentences.

Materials: The experimental sentences were the same as tested in Experiment 1. Each item was manipulated to be either ambiguous or not, and to either include a prosodic boundary or not. Half of the sentences with boundaries presented an early boundary (after NP1); half a late boundary (after NP2). Following Tooley et al.'s (2014, 2018) procedure, the sentences in each of the sets (transfer-of-location and relative clauses) were paired two by two, to form prime-target pairs. The prime sentences were assigned by Latin square to one of the four experimental conditions (Y/N ambiguous by Y/N boundary); whereas the target sentences were always in the ambiguous condition, and always without boundary, being visual stimuli (see below). Ambiguity (present or absent), boundary (present or absent), and sentence position (prime or target) were counterbalanced across items and across participants. The two sets were then intertwined to alternate the two constructions. As in the original experiment, each participant was presented with 20 sentences of each set (five sentences per prime-condition), with no more than two sentences in the same condition in a row.

Procedure: Tooley et al.'s (2014, 2018) prosodic priming paradigm, the task consisted in repeating back out loud an auditorily or visually presented sentence. For each trial, participants listened to (and repeated) a prime sentence, and then silently read and then produced a novel target sentence. One or two filler sentences presenting various syntactic structures and intonational boundaries at various locations were added as audio or visual stimulus between each prime-target pair. The task was created in PCLbex (Zehr & Schwarz, 2018) and run on the university-hosted Ibex farm server.

Data Processing and Analyses: The audio-recorded data were transcribed, force-aligned using the Montreal Forced Aligner (McAuliffe et al., 2017), manually checked and analyzed offline using the Praat Software (Boersma & Weenink, 2021). For each repeated sentence, the 'word-and-pause' duration at the two critical regions (NP1 and NP2) was measured to assess whether or not a boundary was produced. Each absolute duration was hence normalized as a function of the noun length and the total duration of the sentence.

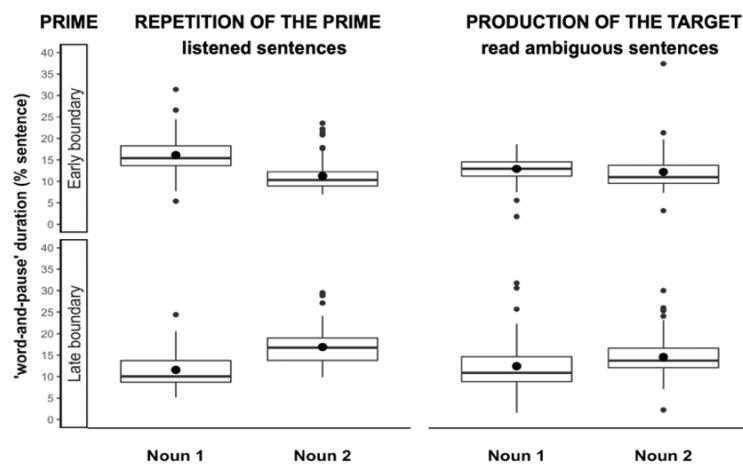
3.2 Results

We used a Bayesian statistical approach to test our prediction of a prosodic priming effect depending on the boundary location. For the purpose of the current analysis, we only included items primed with a boundary (early vs. late), both in the ambiguous and unambiguous prime conditions. Since we expected a between-construction difference (relative clauses > transfer-of-location), supported by the findings of Experiment 1, we fitted separate models for each of the two constructions. The models were constructed and performed in the R environment (R Core Team, 2021.09) using the brms package (Bürkner, 2018), and using so-called weakly informative priors. Both models had the normalized duration of the nouns at the critical locations as DV, and included main effects of ambiguity condition of the prime, boundary location in the prime, prime/target position, noun phrase (NP1 vs. NP2), as well as all the interactions between them. A fixed effect of presentation order, as well as by-item and by-subject random intercepts were added to the models.

The effect of the boundary location in the primes on the duration of the noun phrases for the relative clause sentences is plotted in Figure 2. For these sentences, the estimated probability of an

interaction effect of boundary location and noun phrase on the critical noun duration is 0.13 with a 95% CI of [0.11,0.15], suggesting a rather clear effect of boundary location on the NP durations. Post-hoc analyses showed that this priming effect pattern is present in the repetition of the listened sentences ($\hat{b} = 0.20$, with a 95% CI of [0.17, 0.22]) and, more interestingly, it is carried over, to a lower extent, in the production of the target sentence as well ($\hat{b} = 0.06$, with a 95% CI of [0.03, 0.09]). So, the results of the model for the relative clause sentences revealed that speakers produced longer NP1 than the corresponding NP2 after hearing a sentence with an early boundary; whereas speakers produced longer NP2 than NP1 after hearing a sentence with a late boundary; and this was true for both the repeated sentences and the production of the new ambiguous read sentences. In line with previous production studies, and in line with our predictions, the model for the transfer-of-location sentences showed a repetition priming effect in the listened sentences ($\hat{b} = 0.10$, with a 95% CI of [0.07, 0.13]) but no prosodic priming effect in the production of the new sentences ($\hat{b} = 0.00$, with a 95% CI of [-0.03, 0.04])

Figure 2. Duration of NPs at critical locations in relative clauses after hearing sentences with different boundaries



3.3 Discussion

The results of the priming study showed that priming different intonational boundaries promotes the production of boundaries at the corresponding location, when repeating the primed sentence (repetition) as well as when reading a new target ambiguous sentence (production), when there is not a strong syntactic preference boosting one interpretation of the target sentence over the other and thus reducing the saliency of prosodic cues. These preliminary analyses extend previous work in perception and provide the first experimental evidence that, in proper conditions, intonational phrase structure can be primed in production as well. Our findings suggest that intonational phrase structure is not merely the result of semantic and syntactic processes, but indeed it affects the processing of new sentences. All this points out the relevant role of prosody in planning and facilitating language processing and production, and the need to further investigate how and what exactly are the mechanisms underlying these processes.

Selected references

- Hemforth, B., Fernandez, S., Clifton, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, 166, 43–64. <https://doi.org/10.1016/j.lingua.2015.08.010>
- Jun, S.-A., & Bishop, J. (2015). Priming Implicit Prosody: Prosodic Boundaries and Individual Differences. *Language and Speech*, 58(4), 459–473. <https://doi.org/10.1177/0023830914563368>
- Tooley, K. M., Konopka, A. E., & Watson, D. G. (2014). Can Intonational Phrase Structure be Primed (like Syntactic Structure)? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 348–363. <https://doi.org/10.1037/a0034900>
- Tooley, K. M., Konopka, A. E., & Watson, D. G. (2018). Assessing priming for prosodic representations: Speaking rate, intonational phrase boundaries, and pitch accenting. *Memory & Cognition*, 46(4), 625–641. <https://doi.org/10.3758/s13421-018-0789-5>

Gapping in Romanian comparatives: an experimental perspective on case marking and animacy

Gabriela Bîlbîie, University of Bucharest & LLF

In this paper, we concentrate on the ways in which animacy affects case marking in Romanian elliptical comparatives that involve gapping (Ross 1967), cf. (1).

In order to tease apart contradictory data obtained from informal introspective judgments ((2a), cf. Van Peteghem 2009 vs. (2b), cf. Zafiu 2013), we present an experimental study on the alternation between nominative and accusative case for the pronominal subject remnants in Romanian, showing the relevance of the semantic factor of animacy: there seems to be a preference for accusative marking of the subject remnant when it is not followed by an animate remnant (3a), whereas the nominative seems to be the only strategy which is available when both remnants are animate (3b).

We ran an acceptability judgment task, by using a 2x2 factorial design (experimental factors: CASE and ELLIPSIS). We built 20 experimental items, 10 with animate and 10 with inanimate nouns. For each item, there are four conditions, as illustrated in (4) for inanimate remnants. The conditions (4c) and (4d) served us as control items: ungrammatical control in (4c) vs. grammatical control in (4d). The rating (Likert) scale was 1-7. We had 56 participants who did the task on the IbexFarm platform.

Our results confirm our hypothesis: there is indeed a case alternation (*pace* Van Peteghem 2009), which is not free (*pace* Zafiu 2013), but rather conditioned by a more general semantic constraint of animacy. Both linear mixed-effect and cumulative link models on our participants' acceptability ratings reveal a significant interaction between ellipsis, case, and animacy ($p < .001$, see also Figure 1).

In order to account for the preferences we observe with respect to the case marking of the subject remnant in the presence of a second remnant in gapped comparatives, we propose an explanation in terms of processing (cf. Sag et al. 1985): gapping is more acceptable if both remnants are clearly dissociated by a linguistic mean, e.g. case marking or different semantic type. The strong preference for nominative case when the second remnant is animate could be explained by the explicit case marking dissociation we observe in these contexts (the accusative animate object usually bearing the differential object marker *pe*), whereas the strong dispreference for accusative case could be explained by the redundancy of case marking (both the subjects and the object remnants bearing the accusative case). If both remnants have the same semantic type in terms of animacy, one should have different case marking in order to disambiguate. On the other hand, if remnants do not have the same semantic type, there is no significant acceptability difference in terms of case marking.

Overall, our study shows the importance of experimental methods, which provide more reliable and richer data (Wasow & Arnold 2005, Gibson & Fedorenko 2013, Sprouse et al. 2013).

Linguistic examples from Romanian:

(1) *Ion o iubește pe Maria mai mult decât ea pe el.*

‘Ion loves Maria more than she him.’

(2) a. *Ea lucrează mai mult acasă decât {tu_{NOM}/*tine_{ACC}} la serviciu.* (Van Peteghem 2009)

‘She works more at home than you at the office.’

b. *Eu sunt mai bucuros azi decât {tu_{NOM}/tine_{ACC}} ieri.* (Zafiu 2013)

‘I am happier today than you yesterday.’

(3) a. *Ana iubește geografia mai mult decât {tine_{ACC}/tu_{NOM}} istoria.*

‘Ana likes geography more than you history.’

b. *Ana îl iubește pe Ion mai mult decât {tu/??tine} pe Dan.*

‘Ana loves Ion more than you Dan.’

(4) a. [accusative, ellipsis]

Ana iubește geografia mai mult decât tine_{ACC} istoria.

‘Ana likes geography more than you history.’

b. [nominative, ellipsis]

Ana iubește geografia mai mult decât tu_{NOM} istoria.

‘Ana likes geography more than you history.’

c. [accusative, verb]

Ana iubește geografia mai mult decât iubești tine_{ACC} istoria.

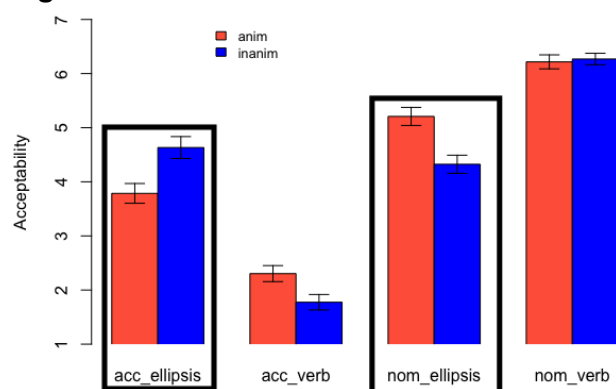
‘Ana likes geography more than you like history.’

d. [nominative, verb]

Ana iubește geografia mai mult decât iubești tu_{NOM} istoria.

‘Ana likes geography more than you like history.’

Figure 1. Animate vs. inanimate distinction in the 4 experimental conditions



References

- Gibson, E. & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1-2), pp. 88-124.
- Ross, J. R. (1967). *Constraints on Variables in Syntax*. PhD thesis. Indiana University.
- Sprouse, J., Schütze, C. & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001-2010. *Lingua* 134, pp. 219–248.
- Van Peteghem, M. (2009). Sur le subordonnant comparatif dans les langues romanes. *Langages* 174, pp. 99-112.
- Wasow, T. & Arnold, A. (2005). Intuitions in linguistic argumentation. *Lingua* 115(11), pp. 1481-1496.
- Zafiu, R. (2013). Comparative constructions. In G. Pană-Dindelegan (ed.), *The Grammar of Romanian*, pp. 503-510. Oxford University Press.

Interdiction as a primary reading of negated deontic necessity in child Romanian

Adina Camelia Bleotu^{a,b}, Anton Benz^a & Gabriela Broz**ă**^b
ZAS Berlin, University of Bucharest

We investigate the interaction between negation and deontic necessity in child Romanian by looking at unambiguous forms such as *nu e nevoie s**ă*** ‘not is need S**Ă**’, expressing lack of necessity, and *trebuie s**ă** nu* ‘must S**Ă** not’, expressing interdiction, as well as at the ambiguous *nu trebuie s**ă*** ‘not must S**Ă**’, a negated modal with two intonationally differentiated meanings (lack of necessity and interdiction). Experimental evidence from a ternary reward task shows that, unlike adults, Romanian 5-year-olds interpret both interdiction and lack of necessity forms as interdiction, regardless of surface scope or intonation. Children’s answers may reflect an initial preference for strong scope and/or a tendency to choose one single alternative out of several.

1 Motivation for the study

It has been shown that children prefer strong (interdiction) readings of negated modals both in production and comprehension. An investigation of the spontaneous speech of French and Spanish children (Jereti**č** 2018) reveals that children use weak (lack of necessity) negated modals much less than predicted by the input, using strong modals instead. Experimental work by Gualmini & Moscati (2009) shows that children tend to interpret *pu**ò** non* (‘may not’) as ‘non pu**ò**’ (‘cannot’). Moscati & Crain (2014) and Koring et al. (2018) further reveal a similar preference for strong interpretations of negated epistemic modals in Italian and Dutch.

2 Theoretical problem and contribution

Our contribution is to investigate children’s scopal preferences in a different environment and language: deontic necessity (rather than epistemic necessity) and negation in Romanian. We explicitly address the question which reading comes first: the weak (lack of necessity) reading or the strong (interdiction) reading. Several proposals have been put forth in the literature. One possibility is that the weak reading is primary, and the strong reading is derived from it either via *negative strengthening* (in accounts which treat *must* as a PPI and *need* as an NPI, such as Israel 1996, Homer 2010, 2015, Iatridou & Zeijlstra 2013) or as a *scaleless implicature* (Jereti**č** 2021). Another possibility is that children interpret negation and modals based on *surface scope* (Musolino 1998). A third possibility is that the strong reading is primary, and, consequently, acquired first. Children may start out with *strong scope* preferences (Semantic Subset Principle-see Crain et al. 1994) and/or they may show *premature closure* (Acredolo & Horobin 1987, Ozturk & Papafragou 2015, Leahy & Carey 2020, a.o.), committing to only one alternative out of several when handling lack of necessity, a modal notion involving multiple alternatives. To test the predictions of these accounts, we conducted an experiment testing Romanian children’s understanding of negated modals expressing lack of necessity or interdiction. Unlike adults, Romanian 5-year-olds mostly interpret all negated modals as interdiction, a finding which supports the *Semantic Subset Principle* and *premature closure*.

3 Investigating deontic necessity and negation in child Romanian experimentally

3.1 Aim

We look at *nu e nevoie s**ă*** ‘not is need S**Ă**’, which unambiguously expresses lack of necessity, *trebuie s**ă** nu* ‘must S**Ă** not’, which unambiguously expresses interdiction, and *nu trebuie s**ă*** ‘not

must SĂ', a negated modal with two readings (lack of necessity and interdiction), which, apart from context, are disambiguated through prosody (https://osf.io/tas6k/?view_only=941c5bc7ec664e159434fbe9ce0dcb5b): for interdiction, F0 goes from 230 Hz to 370 Hz (*nu*) and then to 230 Hz (*trebuie*), while for lack of necessity, F0 stays around 400 Hz for *nu* and the first syllable of *trebuie* and then drops to 250 Hz. We argue *nu* has a contrastive L->H* accent for interdiction, and an L accent for lack of necessity (Estebas-Vilaplana & Prieto (2010)). Given the important role of prosody for interpretation, we are also interested in whether children and adults are equally sensitive to it, in line with previous studies by Armstrong (2014) and Stoddard & de Villiers (2021). In a previous forced choice task we conducted, Romanian 5-year-olds could accurately identify interdiction based on the different intonations of *nu trebuie să*. We here investigate experimentally Romanian children's scopal and prosodic preferences for both unambiguous and ambiguous negated necessity modals.





3.2 Participants

We tested 25 Romanian monolingual children (Mean age: 5;27; Age range: 5-5;11, 12 M, 13 F) and 37 adults.

3.3 Methodology

We employed a ternary reward task, inspired by Katsos & Bishop (2011). Participants are familiarized with contexts where a (grand)parent and their child are looking at two different fruits/drinks/toys/pieces of clothing together. The (grand)parent tells the child that he/she must not/need not do a certain action X. The child then performs action X or action Y. Participants have to reward the child with a sad face if he/she did something forbidden by the (grand)parent, a blue star if what he/she did was so-so, but it was allowed by the (grand)parent, and two blue stars if what he/she did was the best thing, exactly what the (grand)parent said (Table 1). Participants were presented with 32 sentences addressed by the (grand)parent to the child character: 16 sentences with an ambiguous modal (*nu trebuie să* 'not must SĂ' with a *Necessary-Not* or *Not-Necessary* intonation) and 16 with unambiguous modals (*trebuie să nu* 'must SĂ not', expressing interdiction, and *nu e nevoie să* 'not is need SĂ', expressing lack of necessity). The child performed the forbidden/unnecessary action X in half of the sentences and the action Y in the other half. The materials were recorded and analyzed in Praat.

Table 1. Example of an experimental item for *nu trebuie X* 'not must X' with a *Not-Necessary* Intonation, where the child performs action X

<p>Mama și fata se uită la două fructe: o prună și un ananas. Mama îi spune fetei: Mother and daughter are looking at two fruits: a plum and a pineapple. Mother tells the daughter:</p>	
<p>"Nu trebuie să mănânci pruna" (<i>Not-Necessary</i> Intonation). not must SĂ eat plum-the 'You need not eat the plum'</p>	
<p>Fata mănâncă pruna. The girl eats the plum.</p>	
<p>How would you reward the daughter?</p>	
	
	

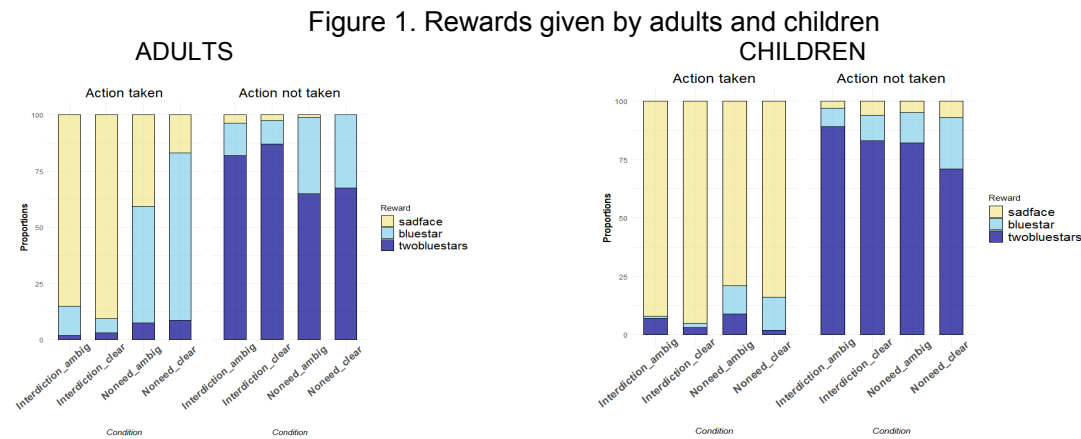
3.4 Expectations

If the child character performs action X, we expect adults to give more *one blue star* rewards for *nu e nevoie să* and for *nu trebuie să* with a *Not-Necessary* intonation than for *trebuie să nu* and

for *nu trebuie să* with a *Necessary-Not* intonation, where the expected reward is clearly a *sad face*. If the character performs action Y, we expect more *one blue star* rewards for lack of necessity modals and more *two blue stars* rewards for interdiction modals. If children interpret lack of necessity as interdiction, we generally expect them to give fewer *one blue star* rewards than adults in the same contexts, for all negated modals.

3.5 Results

The differences between children and adults appear only when the character performs action X (Figure 1). Here, adults reward the character with significantly more *one blue star* rewards and fewer *sad face* rewards after lack of necessity statements (with *nu e nevoie să* or *nu trebuie să* with a *Not-Necessary* intonation) than after interdiction statements (with *trebuie să nu* and *nu trebuie să* with a *Necessary-Not* intonation). In contrast, while children also give slightly more *one blue star* reward, they give significantly fewer *one blue star* rewards than adults. Children's interdiction preference is confirmed by logistic regressions with Reward type/Interpretation as a DV, Modal, Group as fixed effects, and random slopes per Item, Participant.



Legend: Interdiction_ambig = *nu trebuie să* 'not must SĂ' with interdiction intonation, Interdiction_clear = *trebuie să nu* 'must SĂ not', Noneed_ambig = *nu trebuie să* 'not must SĂ' with lack of necessity intonation, Noneed_clear = *nu e nevoie să* 'not is need SĂ'

4 Account

We find that interdiction is the primary reading of children: they interpret weak lack of necessity negated modals as expressing interdiction, while never interpreting interdiction as lack of necessity. Interestingly, in the current task, children's prosodic sensitivity is obscured by their interpretation of (un)ambiguous lack of necessity as interdiction. To capture the results, *negative strengthening* would have to assume children obligatorily move necessity above negation at LF—which might be costly. A *scaleless implicature account* would have to assume children obligatorily compute scaleless implicatures. However, children are known to generally derive implicatures to a lower extent than adults (Noveck 2001, even though they are more adult-like with free choice inferences—e.g., Tieu et al. 2016). It is thus unlikely they would strengthen lack of necessity to interdiction as a default. Our findings also cast doubt on a *surface scope explanation*. Given that, in Romanian, the negative marker *nu* 'not' occurs before the modal, except for *trebuie să nu*, a surface scope account (as in Lidz & Musolino 2002) would predict lack of necessity readings for all necessity modals except for *trebuie să nu*, contrary to our findings. Instead, our results are more compatible with a *strong scopal preference account*, where children initially prefer to assign unique strong scope. This account is supported by similar findings for ambiguous sentences with negation and modality/quantifiers (Musolino & Lidz 2006, Gualmini & Moscati 2009, a.o.). Another (additional) explanation for children's dispreference for lack of necessity might be that lack of

necessity involves multiple alternatives, and children are known to show *premature closure* (Acredolo & Horobin 1987, Ozturk & Papafragou 2015, Leah & Carey 2020, a.o.), a cognitive tendency to commit to only one alternative when faced with several.

References

- Acredolo, C., & Horobin, K. (1987). Development of relational reasoning and avoidance of premature closure. *Developmental Psychology*, 23(1), 13–21.
- Armstrong, M. E. (2014). Child comprehension of intonationally-encoded disbelief. In Orman W., Valteau M.J. (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development*, vol.2, pp. 25–38. Somerville, MA: Cascadilla Press.
- Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Lawrence Erlbaum
- Estebas, E. & Prieto, P. (2010). Castilian Spanish intonation. *Transcription of Intonation of the Spanish Language*, 17–48.
- Gualmini, A., & Moscati, V. (2009). The early steps of modal and negation interactions: Evidence from child Italian. In E. Aboh, E. van der Linden, J. Quer, & P. Sleeman (Eds.), *Romance languages and linguistic theory 2007*. Amsterdam: John Benjamins
- Homer, V. (2011). Polarity and modality. Los Angeles, CA: University of California dissertation.
- Homer, V. (2015). Neg-raising and positive polarity: The View from modals. *Semantics and Pragmatics* 8.
- Iatridou, S., & Zeijlstra, H. (2013). Negation, polarity, and deontic modals. *Linguistic Inquiry* 44(4), 529–568.
- Israel, M. (1996). Polarity sensitivity as lexical semantics. *Linguistics and Philosophy* 19, 619–666.
- Jeretič, P. (2018). Evidence for children’s dispreference for modal weakness. Second Qualifying Paper. New York: New York University.
- Jeretič, P. (2021). *Neg-raising modals and scaleless implicatures*. New York: New York University dissertation. lingbuzz/006390.
- Leahy B.P., & Carey, S.E. (2020). The Acquisition of Modal Concepts. *Trends in Cognitive Sciences*, 24(1), 65-78.
- Lidz, J., & Musolino, J. (2002). Children’s command of quantification. *Cognition*, 84(2), 113–154.
- Katsos, N., & Bishop, D. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature, *Cognition* 120 (1), 67–81.
- Koring, L., Meroni, L. & Moscati, V. (2018). Strong and weak readings in the domain of worlds: A Negative polar modal and children’s scope assignment. *Journal of Psycholinguistic Research*, 47(6), 1193–1217.
- Moscati, V., & Crain, S. (2014). When negation and epistemic modality combine: the role of information strength in child language. *Language Learning and Development*, 10(4), 345–380.
- Musolino, J. (1998). *Universal grammar and the acquisition of semantic knowledge*. University of Maryland: University of Maryland dissertation.
- Musolino, J., & Lidz, J. (2006). Why children aren’t universally successful with quantification. *Linguistics*, 44(4), 817–852.
- Noveck, I. (2001). When children are more logical than adults. *Cognition*, 78(2), 165-188.
- Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, 11(3), 191–214.
- Stoddard, B., & de Villiers, J. (2021). Children’s sensitivity to prosody and ostension in answers to *wh*-questions. Philosophy: Faculty Publications, Smith College, Northampton, MA.
- Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2016). Children’s knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, 33(2), 269–298.

French evaluative adjectives: position and interpretation

Lisa Brunetti & Gabriel Thiberge (Université Paris Cité, LLF, CNRS)

We present three experiments on French multidimensional evaluative adjectives that partly replicate Kaiser & Wang (2021)'s study on English. We tested whether the subjectivity of these adjectives varies depending on their syntactic position, and by consequence their pragmatic status along the given-new and the (non-)at-issueness dimensions. Since in French evaluative adjectives can both be pre- and post-nominal, both positions were tested. Results showed the same tendency as in English for different positions to affect the perception of subjectivity. Nevertheless, no significant difference was found between the pre- and the post-nominal position, contrary to previous claims.

1 Introduction

In both French and English, adjectives can either be noun modifiers (*the amazing orchestra*) or they occupy a predicative position. An adjective in a predicative position can either occur in a matrix clause (*The orchestra was amazing*) or in a subordinate clause, specifically, in an appositive relative clause (*The orchestra, which was amazing, ...*). As pointed out by Kaiser & Wang (2021) (from now on, K&W), these three positions have an impact on the information-structural and discourse functions of the adjective. From the point of view of their at-issue/non-at-issue status, the predicate of a matrix clause is clearly at issue content; on the contrary, if we follow Potts (2005)'s analysis of appositive relative clauses (from now on, ARC) as conventional implicatures, then the content of an ARC is by definition *not* at issue. Finally, the status of the adjective as a noun modifier depends on the (in)definiteness of the noun phrase: if the noun phrase is definite, it triggers an existential presupposition: the modifying adjective, as part of the presupposition, is therefore non-at-issue content. From an information-structural point of view, adjectives whose content is part of a presupposition will also be considered as given, while in a predicative position they will, in most contexts, convey new information (focus),¹.

K&W studied how these differences affect the subjectivity of evaluative adjectives such as *amazing* or *good* in English. By means of four experiments, they showed that such adjectives in the three positions were perceived as expressing significantly different degrees of subjectivity: a noun-modifying adjective in a subject definite noun phrase (*The amazing orchestra played...*) was perceived as less subjective than a predicative adjective, and a predicative adjective within an ARC as less subjective than one in a matrix clause. In other words, both being part of the at-issue content and expressing new information makes the interpretation of the adjective more subjective.

If we now look at how the same research question can be addressed for French, we see that the picture is slightly different. Unlike English, which almost exclusively allows for a prenominal position, French allows for certain classes of adjectives to alternate between a prenominal and a postnominal position (*cet incroyable orchestre / cet orchestre incroyable* 'this amazing orchestra') without any apparent change in meaning (Riegel et al., 1994). This option is available for *multidimensional* evaluative adjectives (such as 'amazing', 'boring', 'interesting', 'unfair'), namely adjectives which convey evaluative information that is based on multiple criteria (see McNally & Stojanovic (2017) and references quoted therein).

Several factors influence the choice of the position, among which the type of determiner and the length of the adjective, in itself and with respect to the modifying noun (Thuillier, 2013). There is also much discussion in the literature on the possible semantic/pragmatic effects of the two positions. Different and sometimes contradictory statements are made in this respect. In Riegel et al. (1994)'s descriptive grammar, it is claimed that with adjectives such as *merveilleux* 'wonderful' or *horrible* 'horrible', the prenominal position favors an interpretation where it is the

¹An exception is when the subject is focused (focus fronting).

speaker who is responsible for the attribution of the quality denoted by the adjective, while in the postnominal position such a subjective interpretation disappears (see also Jones (1996)). According to other scholars, a prenominal adjective conveys information that is taken for granted or is shared common knowledge (cf. Martin (2013) and quoted references). On the contrary, a postnominal adjective can be narrowly focused or contrasted (Nølke, 1996; Martin, 2013). If the latter observations are correct, then postnominal adjectives should be judged as behaving more similarly to ARCs, which are assumed to convey new information.

Before taking the two syntactic positions of the noun-modifying adjective into account and their potential effects on subjectivity, we wanted to see whether the results on French were comparable to those on English. To do so, we conducted an experiment that was as close as possible to the original one. That was the purpose of Experiment 1, which therefore only included noun-modifiers in a prenominal position.

2 Experiment 1

Experiment 1 is meant to replicate the first of K&W's experiments, namely the one testing *multidimensional* evaluative adjectives. With such adjectives, results were stronger in K&W's study than with unidimensional ones. As we said above, for a more pertinent comparison with the English study, the noun-modifier occupied a prenominal position. The choice was also determined by the idea that the difference between the prenominal adjective and the ARC would be greater, given the obligatorily postnominal position of the relative clause in French.

Half of the adjectives that we used expressed a positive evaluation, and half of them a negative evaluation. Following K&W, sentences contained objective information that justified the evaluation expressed by the adjective. French adjectives only partly overlapped the ones used for the English experiment. We wanted to control for certain factors that have an effect on the acceptability of a prenominal position, such as the frequency and the length of the adjective, as well as the choice of the determiner in the noun phrase (Thuilier, 2013). We chose adjectives that had a comparable frequency of use across items; we excluded monosyllabic adjectives (and nouns) and there was at most a one-syllable difference between the adjective and its noun. A demonstrative determiner (*ce/cette* 'this') was used for all items, which made the prenominal adjective more natural and sound less formal. Finally, because of the effect that tense might have on subjectivity, all verbs had imperfect tense in both the matrix clause and the ARC.

Table 1 illustrates an item in all conditions for Experiment 1, as well as for Experiments 2 and 3 (see Sections 3 and 4).

Table 1: An experimental item in all its conditions across experiments

Baseline (Exp 1/2)	Cet orchestre comprenait cinq violonistes avec l'oreille absolue. <i>This orchestra included five violinists with perfect pitch.</i>
Noun Modifier (Exp 1/3-prenom.)	Cet incroyable orchestre comprenait cinq violonistes avec l'oreille absolue. <i>This amazing orchestra included five violinists with perfect pitch.</i>
Noun Modifier (Exp 2/3-postnom.)	Cet orchestre incroyable comprenait cinq violonistes avec l'oreille absolue. <i>This orchestra amazing included five violinists with perfect pitch.</i>
Predicate (Exp 1/2)	Cet orchestre était incroyable , il comprenait cinq violonistes avec l'oreille absolue. <i>This orchestra was amazing, it included five violinists with perfect pitch.</i>
Relative clause (Exp 1/2)	Cet orchestre, qui était incroyable , comprenait cinq violonistes avec l'oreille absolue. <i>This orchestra, which was amazing, included five violinists with perfect pitch.</i>

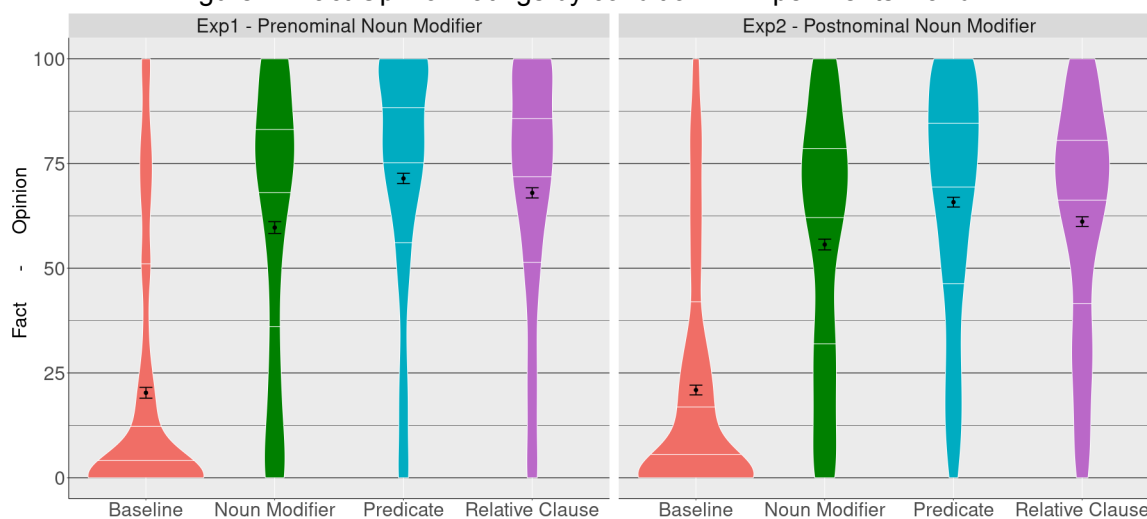
Boldface indicates the adjective.

The materials, design, and procedure were analogous to those used by K&W. On a computer, participants had to indicate on a scale whether written sentences contained an opinion or a fact. Participants were told that such sentences had been extracted from newspapers, magazines, blogs, etc. K&W used a 6-point numerical fact/opinion scale, but we presented participants with a 100-point scale that appeared as a slider with virtually no transition from one point to the next. A near-continuous scale has two benefits: it allows participants to better nuance their judgments, and it better qualifies for an analysis of the data with linear mixed mod-

els (dependent variable = number ranging from 0 to 100). Our study had 24 target items and 43 fillers, presented in a latin square design with randomization so that participants saw 6 target items in each of the four conditions (Baseline, Noun Modifier, Predicate and Relative Clause). The experiment was preceded by three trials. 96 French native speakers were recruited via Prolific (www.prolific.co) and received a compensation to answer the questionnaire, which was run via an instance of IbexFarm (Drummond, 2016) hosted on university servers.

Results showed a similar pattern to the one found for English adjectives: all three conditions differed significantly from the baseline, which was taken as the reference level in a linear mixed model, with items and participants as random effects (see below Figure 1, left, $p < .001$ for all 3 target conditions). A pairwise comparison with Bonferroni correction showed that the least subjective interpretation is the one where the adjective modifies the noun, like in K&W's study. Adjectives in a predicate position and inside an ARC were interpreted as significantly more subjective than adjectives modifying the noun ($p < .001$), but subjectivity ratings in the predicate and the relative-clause conditions were less distinguishable. While predicates received the highest mean rating overall, the difference between the two, unlike K&W's results, was not statistically significant.

Figure 1: Fact/Opinion ratings by condition in Experiments 1 and 2



Ratings range from 0 = purely factual sentence to 100 = purely subjective sentence.
White markings = quartiles, black dot = mean, error bars = standard error.

3 Experiment 2

As discussed in the Introduction, the position of the noun-modifying adjective in French is assumed to have an effect on subjectivity. In order to test this, we replicated the first experiment by replacing prenominal adjectives with postnominal ones. If the intuition of part of the literature that a prenominal position increases subjectivity is correct, then a *postnominal* adjective would be judged as even less subjective in this experiment; on the contrary, within the hypothesis that a postnominal adjective is focused, we would expect it to be interpreted more similarly as an adjective in an ARC. The procedure was identical to that of Experiment 1 and 98 other native speakers of French, recruited via Prolific again, took part.

Results showed the same overall pattern as those in Experiment 1: noun-modifying adjectives were rated as significantly less subjective than those inside an ARC and the latter as less subjective than predicative ones (Figure 1, right). However, this time the difference between ARC and predicate position was significant, even after Bonferroni correction ($p < .05$). The absolute numerical difference between the baseline and the noun-modifier condition was slightly smaller than in Experiment 1 (20.3 vs 59.7 in Experiment 1, 20.9 vs 55.7 in Experiment 2), but participants globally rated the three subjective conditions as less subjective in Experiment 2.

A linear mixed model was also run including the data from both experiments and an *Experiment* predictor (mean-center coded) in interaction with the adjective conditions. The model did not single out the noun-modifier condition from the ARC and the predicate conditions across experiments, thus ruling out a simple effect of the postnominal position of the adjective.

4 Experiment 3

The results of Experiment 2 did not confirm what was predicted in the literature concerning an effect of the position of the noun-modifying adjective. In order to verify these results, a third experiment was conducted, consisting of noun-modifying adjectives in a prenominal vs postnominal position only. The same 24 target items were presented to 53 French native speakers recruited from the RISC network (CNRS, UMR 3352), as well as 48 fillers, with no baseline condition this time. The same procedure and the same 100-point slider scale were used. Results were in line with previous experiments, and no significant difference was found in subjectivity judgments between the two positions.

5 Discussion and conclusions

Given the important difference between the three conditions and the baseline (which was constructed as conveying factual information), one may wonder whether the presence of the adjective does not have a pragmatic "subjectivizing" effect on the whole utterance, despite the fact that, strictly semantically, only the adjective is subjective.² Future experiments may test this hypothesis by giving more explicit instructions to participants in this respect.

Concerning the difference among the three conditions, in French, like in English, both pragmatic dimensions play a role in the interpretation of the sentence as conveying subjective information. Results in both languages, however, point out the major role played by the given-new distinction: adjectives conveying new information (ARCs and predicates) are evaluated as distinctly more subjective than those that are part of given information. From a syntactic point of view, the difference is not between being inside a noun phrase (the noun-modifier and the ARC, which also depends on the noun) or not; rather, it is between adjectives followed by a copula, therefore in a predicative position (in a matrix clause or in an ARC) and adjectives that directly modify the noun. In order to disentangle the predicative position from new information, future research might test adjectives within an *indefinite* noun phrase occupying a position inside the verb phrase (e.g. 'The concert was played by an amazing orchestra'), which is a typical position for new information (both in English and in French), though the adjective is not predicative.

Concerning the specificity of the French data, if it is indeed the given/new distinction that plays a role, one would expect a more subjective interpretation of postnominal adjectives, within the claim that they are in focus. Our results could not support this nor a more general hypothesis of an effect on subjectivity of the pre-/postnominal position. Our results could also not support the hypothesis of a difference between English and French due to the additional position available in French. A larger cross-linguistic analysis, targeting languages that have a yet different inventory of syntactic positions, might shed more light in this respect.

References

- Drummond, A. (2016). *Ibexfarm (version 0.3. 9)*.
- Jones, M. A. (1996). *Foundations of french syntax*. Cambridge University Press.
- Kaiser, E., & Wang, C. (2021). Packaging information as fact versus opinion: Consequences of the (information-) structural position of subjective adjectives. *Discourse Processes*, 58(7), 617–641.
- Martin, F. (2013). *Decomposing (non)-restrictivity. Evaluative modifiers in post-head positions*.
- McNally, L., & Stojanovic, I. (2017). Aesthetic adjectives. *Semantics of Aesthetic Judgements*, 17–37.
- Nølke, H. (1996). Où placer l'adjectif épithète? focalisation et modularité. *Langue française*, 38–58.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Riegel, M., Pellat, J.-C., & Rioul, R. (1994). *Grammaire méthodique du français*. Presses universitaires de France.
- Thuillier, J. (2013). Experimental approach: The case of the syntax of attributive adjectives. In *10th syntax and semantics conference in paris (cssp 2013)*.

²We thank a reviewer for pointing out to us such intriguing possibility.

Processing past time reference, but which one? An ERP study on the Mandarin Chinese aspect morphemes *guo*₁ and *guo*₂ with definite and indefinite time adverbs

Aymeric Collart, Academia Sinica

The processing in the brain of the temporal concord of aspect markers in Mandarin has attracted the attention of researchers, but the picture is not complete yet. This study focuses on *GUO*, the phonological realization of the two morphemes *guo*₁ and *guo*₂. They occur with time adverbs of different natures: *guo*₁ with indefinite time adverbs ('in the past'), *guo*₂ with definite time adverbs. The ERP results show that *guo*₁ temporally violated elicits a biphasic N400-P600, and *guo*₂ a P600, suggesting common cognitive processes ('localization') between the two, and specific cognitive process ('existential') for *guo*₁, in line with their linguistic analyses.

1 Introduction

1.1 The multidimension of the perception of an event in time

The time at which an event occurs can be perceived in different ways by the human cognition (Baggio et al., 2012). For instance, an event can be seen as located in time according to the time of speech ('temporal localization'). Conversely, it can be placed in time by taking another event as its reference, such that the event in question is located in a sequence of two events ('temporal sequencing'). Another facet of the cognitive perception of time concerns the existential status of the event: events placed in the past and in the present are more likely to be taken as *real*, and events in the future as *unreal* (Lyons, 1977). Such cognitive ways to perceive time have correlates in language. 'Tense' in its traditional definition expresses the temporal localization of an event (Comrie, 1985), but it can also be used to refer to events in a sequence (Klein, 1994). Similar claims can be made for 'aspect'. When it comes to the occurrence of an event in time, perfective aspect can exhibit a temporal sequencing meaning. Some aspect markers may also be used to express temporal localization as well (Iljic, 2010).

1.2 Expressing temporal reference in Mandarin with aspect markers and their neuronal processing in temporal concord violation paradigms

Mandarin is considered as a tenseless language, i.e., it does not have morphemes encoding tense distinctions, but several aspect markers can be used to express temporal relations. Among them, the postverbal morphemes *-le* and *-guo* are analyzed as expressing different cognitive facets of time: temporal sequencing for *-le* (Woo, 2018), temporal localization for *-guo* (Iljic, 2010). Interestingly, *-le* and *-guo* are grammatical when occurring with a deictic past time adverb, but not with a deictic future time adverbs. See (1).

- (1) Zhangsan zuotian/*mingtian jiao-le / jiao-guo yangtai de hua.
Zhangsan yesterday/*tomorrow pour-LE / jiao-GUO balcony DE flower
'Zhangsan (already) poured the flowers on the balcony yesterday/*tomorrow.'

The verb *you* 'to have' has grammaticalized into a verbal auxiliary in some Mandarin-speaking regions, expressing the temporal existence of the event (Collart & Su, 2022). Similar with *-le* and *-guo*, '*you* + VP' exhibits the same temporal concord violation. See (2).

- (2) Zhangsan zuotian/*mingtian you jiao yangtai de hua.
Zhangsan yesterday/*tomorrow have pour balcony DE flower
'Zhangsan poured the flowers on the balcony yesterday/*tomorrow.'

The processing in the brain of the temporal concord violations in (1-2) has been explored with the ERP technique in previous experiments. *-Le* elicited a LAN effect, analyzed as reflecting the violation of the temporal sequencing process, *-guo* a P600, reflecting the temporal localization process (Qiu & Zhou, 2012; Collart & Chan, 2021). The temporal concord violation

of 'you + VP' was reflected by a N400 component on the verb, taken as indicating the violation of the existence of the event in time (Collart, 2018). Overall, it appears that the aspect markers in Mandarin placed in temporal concord configurations can be used to express the different cognitive perceptions of the occurrence of the event in time, and this idea finds supportive evidence from electrophysiological data.

1.3 Distinguishing between *guo*₁ and *guo*₂: The present study

However, the picture is not complete. The linguistic literature points out that *-guo* is the phonological realization of two distinct morphemes: *guo*₁ ('experiential *guo*'), expressing the existence and the localization of the event in time, and *guo*₂ ('phasal *guo*'), only indicating the localization of the event in time (Ilic, 2010; Collart, 2022). Syntactic tests support this distinction: (a) *guo*₁ can appear with the negation *mei*, but not *guo*₂, (b) *guo*₂ can occur with *-le*, but not *guo*₁, (c) *guo*₁ appears after indefinite deictic time adverbs (e.g., *yiqian* 'in the past'), *guo*₂ after definite deictic time adverbs (e.g., *zuotian* 'yesterday'). Crucially, the temporal concord violation pattern is also found for *guo*₁ as well (see (3), to compare with (1)).

- (3) Zhangsan yiqian/*yihou jiao-guo (*-le) nüwang de hua.
Zhangsan in.the.past/*in.the.future jiao-GUO₁ (*-LE) queen DE flower
'Zhangsan poured the queen's flowers in the past/*in the future.'

The present study proposes to investigate the neuronal processing of the temporal concord violation of *guo*₁ and to compare with the one of *guo*₂ in order to determine whether the different linguistic analyses are reflected in their neurophysiological processing. To do so, the processing of sentences as in (1), involving *guo*₂ (for which the brain process has already been explored) is compared with sentences as in (2), involving *guo*₁. Different predictions can be made based on their linguistic analyses and previous ERP results: (a) if *guo*₁ is used to express both temporal existence and temporal localization, then its temporal concord violation is expected to elicit a biphasic N400 + P600 response, while only a P600 is expected for *guo*₂; (b) if the brain does not distinguish between *guo*₁ and *guo*₂ concerning the processing of their temporal concord violation, the same ERP component is expected: A P600 effect.

2 Methodology

2.1 Participants

The participants were 24 native speakers of Mandarin (14F, mean age: 26.2 y.o., range: 20–40 y.o.), right-handed and without brain disorder. They signed a consent form approved by the Center for Research Ethics of National Taiwan Normal University.

2.2 Materials

The material was the same as in Collart & Chan (2021), except for the deictic time adverbs: the definite time adverbs were *zuotian* 'yesterday' and *mingtian* 'tomorrow', the indefinite time adverbs were *guoqu* 'in the past' and *weilai* 'in the future', and not *yiqian/yihou* to avoid 'past in the past/future' interpretation. The 272 sentences, dispatched into four lists with a Latin square design (34 per condition + 136 fillers to counterbalance the grammaticality of the sentences) were tested for naturalness in an online pilot study (74 participants who did not take part in the ERP experiment, 7-point Likert scale).

2.3 Procedure and data acquisition

The procedure and data acquisition parameters replicated the ones in Collart & Chan (2021).

2.4 Preprocessing pipeline and statistical analyses

The preprocessing pipeline and statistical plan in Collart & Chan (2021) was adopted, except for the length of the ERP epochs. The epochs were computed starting from the onset of the deictic time adverbs and lasted until the end of the sentence (3200ms long). The ERP data

were analyzed with repeated-measure ANOVAs (DV = voltage, IV = Type (*Definite* vs. *Indefinite*, Grammaticality (*Grammatical* Vs. *Ingrammatical*), Anteriority (*Frontal* vs. *Central* vs. *Posterior*), Hemisphere (*Left* vs. *Midline* vs. *Right*) at two time windows.

3 ERP Results

The temporal concord violation involving definite adverbs only elicited a P600 (see Fig. 1), but a biphasic N400-P600 was found for the indefinite adverbs (see Fig. 2).

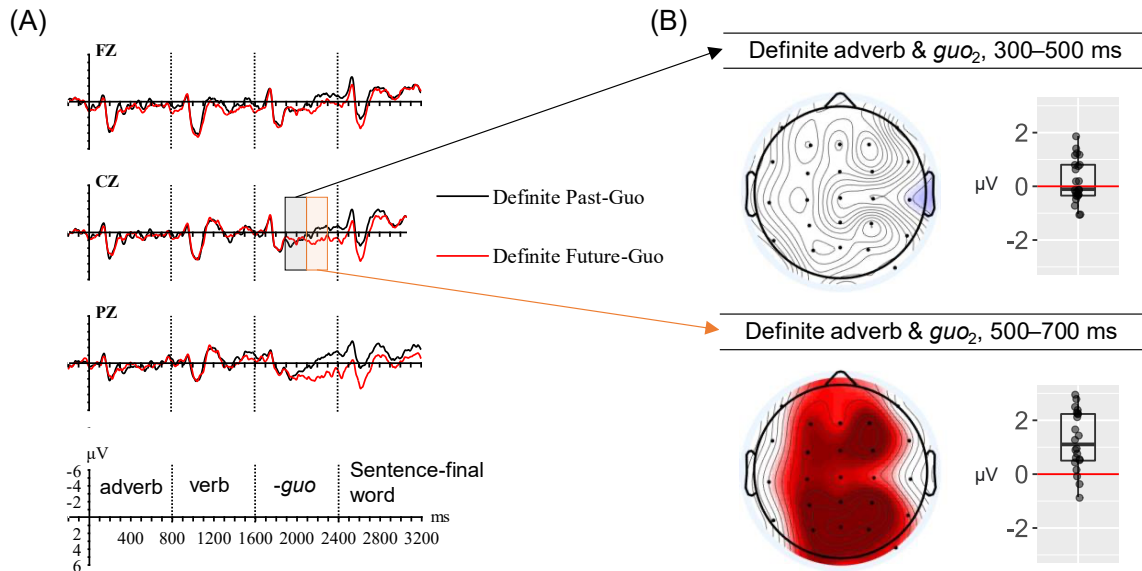


Fig. 1. Panel A: ERP waves of the definite temporal concord conditions (black = grammatical, red = ungrammatical, negative voltage plotted upward). **Panel B:** topographic maps and boxplots of the difference wave at guo_2 (ungrammatical *minus* grammatical; mean amplitude from $-1.5 \mu\text{V}$ (blue) to $+1.5 \mu\text{V}$ (red)) at the 300–500 and 500–700 ms time windows

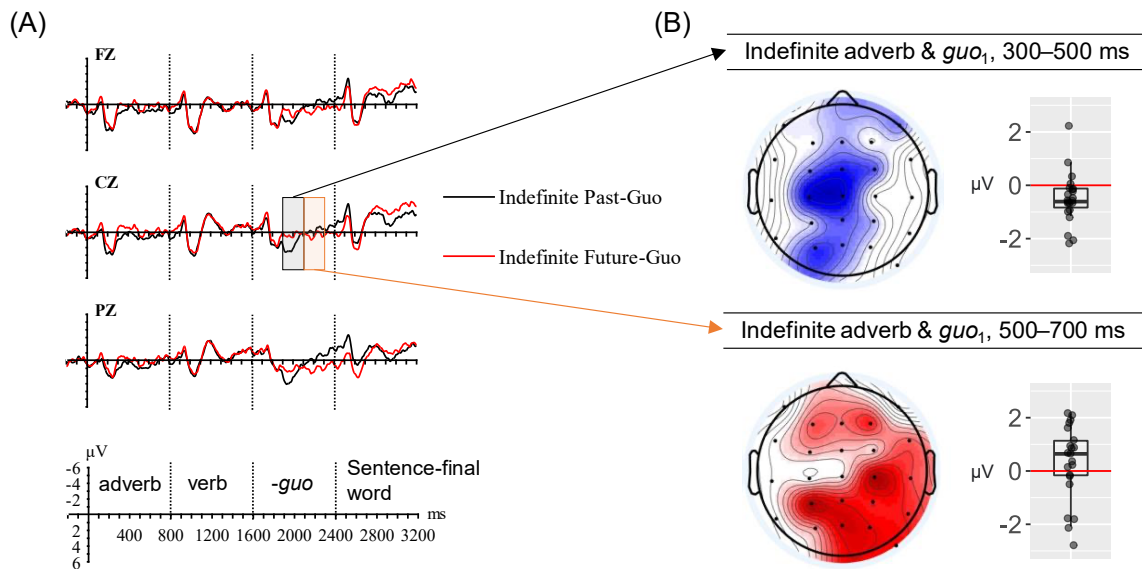


Figure 2. Panel A: ERP waves of the indefinite temporal concord conditions (black = grammatical, red = ungrammatical, negative voltage plotted upward). **Panel B:** topographic maps and boxplots of the difference wave at guo_1 (ungrammatical *minus* grammatical; mean amplitude from $-1.5 \mu\text{V}$ (blue) to $+1.5 \mu\text{V}$ (red)) at the 300–500 and 500–700 ms time windows

These observations were statistically verified: only the Type x Grammaticality interaction was significant at the 300–500ms time window ($F(1,23) = 9.62, p < .01$); only the main effect of Grammaticality was significant at the 500–700ms time window ($F(1,23) = 13.89, p < .01$).

4 Discussion

The ERP results indicate distinct and also overlapping processing patterns for *guo*₁ and *guo*₂. *Guo*₁ elicited a N400 effect followed by a P600. The N400 can be seen as indicating the difficulty to integrate an event which is asserted as existing (meaning of *guo*₁) in a non-existing context (meaning of future time adverb). The P600 reflects the difficulty to locate an event perceived as be past (meaning of *guo*₁) in a future time context. Conversely, only a P600 was found for *guo*₂, which can be analyzed the same way as for the P600 elicited by *guo*₁. These results, coherent with the linguistic analyses of *guo*₁ and *guo*₂, are modeled in Fig. 3.

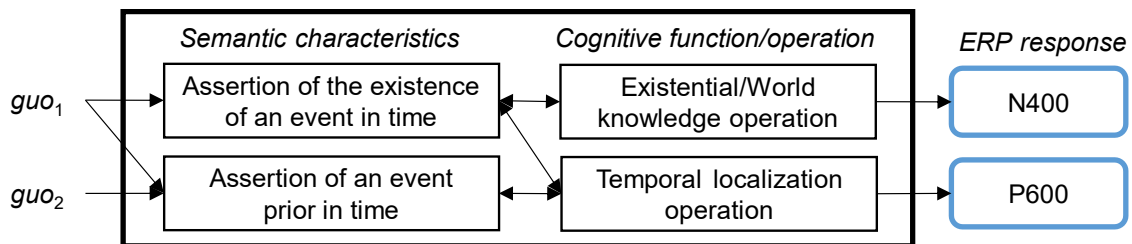


Fig. 3. Linking the linguistic and cognitive analyses of *guo*₁ and *guo*₂ with their ERP responses

Finally, these results may even suggest a cognitive continuum between the meaning of the aspect markers in Mandarin and their ERP signatures: ‘you + VP’ (N400, existential) – *guo*₁ (N400+P600, existential+localization) – *guo*₂ (P600, localization) – *-le* (LAN, sequencing).

5 Conclusion

This study bridges linguistic analyses of the aspect markers *guo*₁ and *guo*₂ in Mandarin with their ERP responses as well as a tentative model with other markers. More insight could be given by comparing with the processing of tense markers in Indo-European languages.

Selected references

- Baggio, G., van Lambalgen, M., & Hagoort P. (2012). Language, linguistics and cognition. In R. Kempson, T. Fernando, & N. Asher (Eds.). *Handbook of the philosophy of science* (pp. 325–355). Amsterdam: North Holland.
- Collart, A. (2018). Same interpretation, different process? The case of verbal *-le* and *You+VP* in Taiwan Mandarin. Paper presented at the joint ICPEAL 17-CLDC 9 conference, National Taiwan University, Taipei.
- Collart, A. (2022). *Bridging time in language and in the brain: Insights from Mandarin Chinese*. Doctoral dissertation, National Taiwan Normal University, Taipei, Taiwan. Retrieved from <https://hdl.handle.net/11296/w7mrm5>.
- Collart, A., & Chan, S.-H. (2021). Processing past time reference in a tenseless language: an ERP study on the Mandarin aspectual morphemes *-le* and *-guo*. *Journal of Neurolinguistics*, 59, 100998.
- Collart, A., and Su, H.-K. (2022). Encoding the existence of an event with ‘you (to have) + VP’ in Taiwan Mandarin: A corpus-based investigation. *Concentric: Studies in Linguistics*. Issue TBA.
- Comrie, B. (1985). *Tense*. Cambridge, New York: Cambridge University Press.
- Ilijic, R. (2010). Holisticity, discontinuity, class in Chinese aspect. *Language Research*, 46(2), 147–198.
- Klein, W. (1994). *Time in language*. London: Routledge.
- Qiu, Y., & Zhou, X. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain Research*, 1446, 91–108.
- Woo, I.-H. (2018). On perfective *-le* in Mandarin Chinese: theoretical and pedagogical issues. *Acta Linguistica Asiatica*, 8(2), 139–158.

Two measures of Backgroundedness predict island status of Wh-questions and RCs across 7 English constructions

Nicole Cuneo & Adele E Goldberg
 Department of Psychology, Princeton University
 Princeton, NJ 08544

Abstract

The extent to which a construction *backgrounds* information inversely predicts how available that information is for long-distance dependency (LDD) constructions ($N=680$). Acceptability ratings were collected on declarative sentences and two types of corresponding LDDs (*Wh*-questions, relative clauses) which correlate ($r = .84$). Two measures of backgroundedness were collected and found to predict island status: i.e., each predicts the acceptability of both *wh*-questions and relative clauses more than declaratives. We interpret this to mean that island effects arise from a clash of discourse properties where LDD constructions make an argument prominent in the discourse while “island” constructions background information to varying degrees.

We report results in support of the claim that constraints on islands arise because of a clash between the functions of grammatical constructions to be combined (Goldberg, 2006; Cuneo & Goldberg; 2022; Namboordiripad et al. 2022). In particular, we argue that long distance dependency (LDD) constructions, both questions and relative clauses, make the “moved” constituent prominent in the discourse, while island constructions, including clausal temporal adjuncts, non-bridge verb complements, relative clauses, ensure their content is “backgrounded” to varying degrees. Results indicate that it is infelicitous for a speaker to make a constituent prominent in the discourse (via a LDD) to the extent that the prominent constituent is backgrounded (via an island construction). For example, if a speaker wishes to request information about a later event it is infelicitous for the same speaker to position that information within a clausal temporal adjunct which is typically backgrounded in discourse (i.e., not “at-issue”, Potts, 2004).

(1) He researched the purchase [after comparing prices]_{backgrounded}.

(2) What_{prominent} did he research the purchase [after comparing _]_{backgrounded}?

We collected from separate groups, 1) acceptability judgments on 84 declarative sentences 2) corresponding *wh*-questions, 3) corresponding sentences containing relative clauses. Table 2 provides example stimuli for each of 8 construction types tested. Several of the construction types are generally considered islands (e.g., relative clauses, clausal adjuncts), others are not (main clauses, “bridge” verb complements). However, we do not presuppose which constructions are islands or to what extent. Sample stimuli are provided in Table 1.

Table 1. Example declarative response sentences (no island violations) and *wh*-questions (potential island violations)

Constructions	Example Declaratives and LDDs (<i>Wh</i> -Question & Relative Clauses)
<i>Main Clauses</i>	The door that leads to the basement was closed. Where does the door that was closed lead to _ ? She admired the stone that the door that led to the basement was made of _ .

<i>Relative Clauses</i>	The door that was closed leads to the basement. Where does the door that leads to _ was closed? She admired the stone that the door that was made of _led to the basement.
<i>Causal Adjuncts</i>	He researched it by comparing prices. What did he research the question by comparing _ ? He was aware of the prices that he researched the purchase by comparing _
<i>Temporal Adjuncts</i>	He researched it after comparing prices. What did he research the question after comparing _ ? He was aware of the prices that he researched the purchase after comparing _.
<i>DO Recipients</i>	She showed Sam the portrait. Who did she show _ the portrait? The artist knew the buyer who Gary showed the portrait ___.
<i>PO Recipients</i>	She showed the portrait to Sam. Who did she show the portrait to _? The artist knew the buyer who Gary showed the portrait to ___.
<i>Clausal complements</i>	Alicia believed he got hired in Hawaii. What did Alicia believe that he got _ in Hawaii? Alicia wanted the job she believed he got _in Hawaii.
<i>Clausal complements</i>	Alicia forgot he got hired in Hawaii. What did Alicia forget he got _ in Hawaii? Alicia wanted the job she forgot he got _in Hawaii.
<i>Parasitic Gaps</i>	She puts conditioner in her hair after washing it. What does Mara put conditioner in ___ after washing ___? Her fans were impressed with her hair which she puts conditioner in after washing her face.
<i>Nonparasitic Gaps</i>	She puts conditioner in her hair after washing her face. What does Mara put conditioner in her hair after washing ___? Her fans were impressed with her hair which she puts conditioner in after washing.

We separately collected judgments on two measures of Backgroundedness. The first is the Negation task used in previous work (e.g., Ambridge & Goldberg, 2008; Namboodiripad, et al., 2022). Constructions that background their information are presumed to be less affected by main clause negation; put differently, information that is more at-issue is more likely to be negated by main clause negation. In this task, the main clause of each of the declarative sentences was negated and participants were asked to what extent the information in the various target constructions was negated.

Finally, we used a Discourse task introduced in Cuneo & Goldberg (2022). Trials prompt participants for certain information (e.g., *Tell me why Nicole is so happy today*), and then asks which of two sentences is the “more direct and cooperative” response (see Table 2): Critically, both response options on each trial include the requested information and only differ in *how* the requested information is packaged. The sentence pairs are minimally different and neither contains any island violation. An acceptability survey confirmed there was no systematic difference in acceptability between the two types of responses. One response provides the requested information within a

construction that was hypothesized to be an “island” while the other provided the requested information within a construction that was hypothesized to be a non-island.

Table 1: Example stimuli in Discourse Task

Tell me why Ali got up so early.	
His rowing club that meets at the lake starts at 6:00. (<i>Main Clause</i>)	His rowing club that starts at 6:00 meets at the lake. (<i>Relative Clause</i>)
Tell me why that puppy is so happy.	
The owner got Fido outside by giving him treats. (<i>Causal adjunct</i>)	The owner got Fido outside while giving him treats. (<i>Temporal adjunct</i>)
Tell me what you did in the garden.	
I planted a tree without watering it. (<i>Parasitic coreference</i>)	I planted a tree without watering the flowers. (<i>Non-parasitic reference</i>)
Tell me why Iris took time off from school.	
Dan heard that she wasn't feeling well. (<i>potential "Bridge" Verb</i>)	Dan hated that she wasn't feeling well. (<i>"Non-bridge" Verb</i>)

Preregistration Each part of the experiment was preregistered before data collection, including number of participants, exclusion criteria, stopping rule and analyses Discourse:

<https://aspredicted.org/2bv9s.pdf>. Negation: https://aspredicted.org/see_one.php; Wh/declaratives: https://aspredicted.org/see_one.php; RC/dlinked: https://aspredicted.org/see_one.php

Participants For each measure, separate groups of 120 unique participants were recruited via the Cloud Research platform as a front end on Mechanical Turk (Litman et al. 2017).

Procedure For acceptability ratings on declaratives, wh-questions, relative clauses, and for judgments on the negation task and the discourse task, 72 stimuli were quasi-randomly assigned to one of 4 lists of 21 target sentences, with the stipulation that no participant judged more than one of any highly similar pair of sentences. Order of presentation of the stimuli was randomized for each participant. Acceptability ratings were based on a 7-point scale, negation scores, on a 5-point scale. Filler trials were included as catch trials on all lists.

Results.

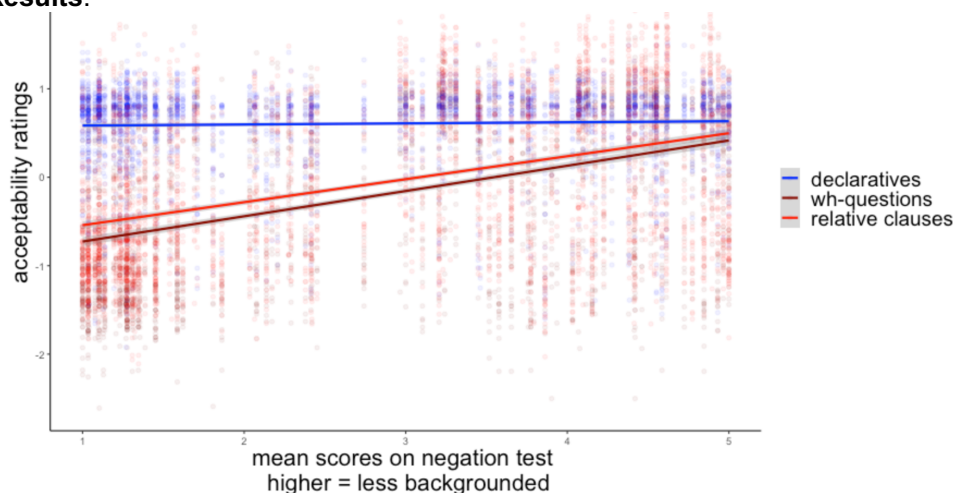


Figure 1. Responses to the **Negation task** predict the acceptability of *wh*-questions (dark red) and relative clauses (red) more than the acceptability of the declarative responses themselves (blue).

The Ordinal package was used in all analyses reported, with random effects for subjects, items and construction types.

Acceptability ratings on the two types of LDD constructions tested (questions and relative clause) strongly correlate ($r = .84$) (cf. Abeillé, et al. 2020; Sag 2010). As predicted, the interaction between Type (Declarative vs. LDD) and Discourse measure when predicting acceptability was significant for both discourse measures (negation: $\beta = 0.53218$, $p < .00001$; discourse: $\beta = 1.42$, $p < .0001$) Figure 1 shows the degree to which main clause negates the target construction (x-axis) predicts acceptability ratings on *wh*-questions (y-axis, in red), and RCs (y-axis, brown), and not declaratives (blue). Figure 2 shows the same is true for the same is true of the independent measure of Backgroundedness: the discourse task.

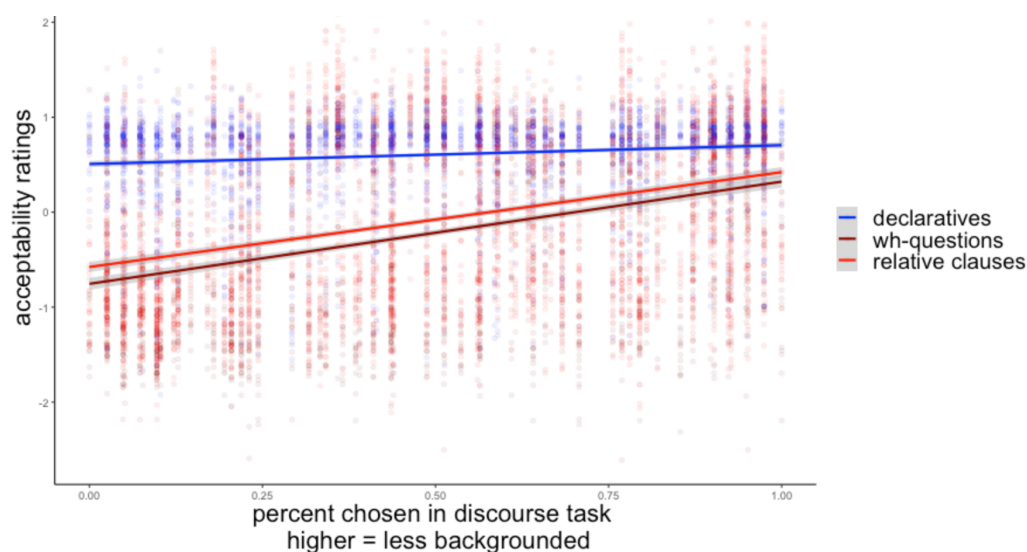


Figure 2. Responses to the **Discourse task** predict the acceptability of *wh*-questions (dark red) and relative clauses (red) more than the acceptability of the declarative responses themselves (blue)

References

- Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from Subjects: Differences in Acceptability Depend on the Discourse Function of the Construction. *Cognition*, 204, 104293.
- Ambridge, B. & Goldberg, A.E. (2008). The island status of clausal complements: evidence in favor of an information structure explanation. *Cognitive Linguistics* 19(3) 349-381.
- Chaves, R. P, and Putnam, M. T. (2021). *Unbounded Dependency Constructions : Theoretical and Experimental Perspectives*. Oxford University Press.
- Cuneo, N & Goldberg, AE. (2022) Islands effects without extraction: the discourse function of constructions predicts island status. *Cognitive Science Society Proceedings*.
- Goldberg, A. E. (2006). *Constructions at Work: the nature of generalization in language*. Oxford University Press.
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433-442.
- Namboordiripad, S., Cuneo, N. et al. (2022). Backgroundedness predicts island status of non-finite adjuncts in English. *Cognitive Science Society Proceedings*.
- Sag, I A. (2010). English Filler-gap constructions. *Language* 86.486-545.
- Sprouse, J., Wagers, M., & Phillips, C. (2012). Working-memory capacity and island effects: A reminder of the issues and the facts. *Language*, 88(2), 401-407.

Embedded questions: Evidence in a decision-theoretic paradigm for ‘surprise’ & ‘agree’

Lea Fricke¹, Emilie Destruel,² Edgar Onea¹ & Malte Zimmermann³
¹Graz University, ²University of Iowa, ³Potsdam University

In this paper, we present new cross-linguistic empirical evidence on the interpretation of questions, specifically those embedded under the two verbs *surprise* and *agree*, in German and English. Our study contributes to the long-standing debate between introspection and experimental findings on what are the (exhaustive) readings that speakers actually derive in such embeddings. In our experiment, participants were confronted with a decision problem involving all potential exhaustive readings, for which we created probabilistic models of the participants’ beliefs in a Bayesian analysis. Our results align with prior empirical evidence: embedded questions under *surprise* and *agree* are accepted under weakest readings.

1 Introduction

The present paper aims at answering the following question: What (exhaustive) readings associate with embedded questions and to what extent do speakers derive them? Here, we focus specifically on two embedding verbs, *surprise* and *agree*, illustrated in (1) and (2),

- (1) Ali was **surprised** who danced at the party.
- (2) Kim and Ali **agree** on who of the dancers lack talent.

Past theoretical literature discusses different readings that arguably exist for these two verbs, which are summarized in Table 1 and 2. Readings are ordered based on their logical strength, the top reading being the strongest. For *surprise*, the general assumption is that (1) is true under a weakly exhaustive reading (**WE**). This WE reading has two manifestations following Lahiri (2002), who claims this verb is only optionally distributive. Others have argued for strong exhaustivity (**SE**). Under *agree*, the question meaning characterizes the extent to which the attitude holders’ (AH) beliefs must align. Incompletely aligned beliefs (**IA**) served as a negative baseline in our study.

Table 1
überraschen/surprise Q (A+ = A has Q-property; A- = A does not have Q-property)

Reading	Mental state of Attitude Holder	Facts in the World
WE _{distributive} (Berman 1991)	Expectation: A-, B-, C-	A+, B+, C+, D-, E-
WE _{non-distributive} (Lahiri 2002)	Expectation: A-, B-	
SE (Klinedinst & Rothschild 1999)	Expectation: A+, B+, C+, D+, E+	

Recent experimental evidence is partially at odds with some of these theoretical claims, suggesting that sentences like (1) and (2) are indeed true under the weakest interpretations, i.e., **SE** and **CA** (Cremers & Chemla 2017, Chemla & George 2016). However, the status of the readings detected in these experiments is unclear: Since the experiments used truth-value judgment tasks, the communicative reliability and robustness of the accepted readings remains unknown. Our study aims at extending this empirical line of research from a cross-linguistic perspective by testing the availability of pragmatically robust and reliable readings only.

Table 2
sich einig sein/agree Q (A+ = A has Q-property, A- = A lacks Q-property, A? = uncertain whether A has Q-property.)

Reading	Mental states of Attitude Holder 1 & Attitude Holder 2
CA+ (complete alignment) (Kratzer 2006)	Both AHs believe: A +, B+, C+, D-, E-
CASU (complete alignment, same uncertainties) (Beck & Rullmann 1999)	Both AHs believe: A +, B+, C+, D?, E?
CA (complete alignment of positive belief) (Lahiri 2002)	Both AHs believe: A+, B+, C+, D- AH1 believes: E-; AH2 believes: E?
IA (incomplete alignments)	Both AHs believe: A+, B+, C+, D- AH1 believes: E+; AH2 believes: E-

2 Experiment

2.1 Participants

A total of 24 native speakers of German (mostly Austrian German) were tested, namely 17 females and 7 males between the ages of 20 and 31 ($M = 24.37$ years). 20 of these were university students. Participants were recruited via postings on university-related Facebook groups and via email messages and printed posters on campus. The financial compensation varied between 9.40 and 10.40 euros. For the English version of the experiment, we tested 26 monolingual native speakers of American English, including 12 males and 14 females who were aged between 19 and 24 years old ($M = 20.65$ years). All were either undergraduates ($n = 20$) or graduate students ($n = 6$) in a Midwestern university, and were recruited via email messages. All undergraduates were enrolled in a first-year language class, and they were compensated for their time with extra points toward their final course grade, which corresponded to the amount of the financial gain earned in the experiment (varied between 8.80 and 11.40 dollars).

2.2 Materials

Before starting the experiment, participants had to read a context presenting them with an enacted betting scenario of a TV show and ensured all experimental stimuli were considered as part of a context rather than in isolation. Moreover, the context made very clear what the domain under the discussion was – the five contestants – and that all of them are relevant for the interpretation of the stimuli. The concrete task was for participants to judge bets as won or lost; judging a bet as “won” corresponded to accepting a target sentence. On each trial, participants saw experimental materials presented on slips with two sides. The front side of the slip included a bet concerning the happenings in the show, which they had to evaluate. The bet appeared in the form of a sentence containing an embedded question under *to surprise* or *to agree*, and it also contained a monologue/dialogue that expressed the beliefs of the attitude holder in question. In the case of *surprise*, the back side displayed a table summarizing what actually happened, i.e., the facts in the world. For the verb *agree*, this was not the case since there is no objective factual base against which to measure the attitude holders’ subjective agreement. To create our experimental materials, we manipulated two factors. First, we manipulated **READING**, tested by changing the contents of the attitude holder’s statement and the reported actual facts in the world. The readings tested depended on the embedding verb and appear in Table 1 & 2. As the semantic analysis of *surprise* is controversial (Roelofsen et al. 2019), we tested the same readings in both languages for this verb (Table 1). For *agree*, we tested partly different readings in English and German, since we had no reason to believe that there could be cross-linguistic differences. However, as *agree+Q* involves two AHs, there are a number of possible belief configurations that we wanted to cover in

the experiments. Second, we tested the factor **ROLE**: In Role 1, participants had to redeem bets and would profit from bets that are won. In Role 2, they had to review betting slips and decide whether to pay out a reward. In this role, one profits from lost bets. To harness in the biases of the two roles, fees for redeeming bets, and fines for not paying out rewards for won bets were part of the rules. Thus, participants had a real financial incentive for answering correctly.

Our experiment had four crucial features: a) Correctness of the answers was evaluated post-experiment, such that no training artefact could emerge; b) the roles induced a different financial bias to judge bets as lost/won; c) participants had to reckon with negative financial consequences in case the bet was incorrectly judged as won/lost, thereby boosting reliability and robustness as a design feature; d) the use of direct financial incentive is known to increase effort (Camerer & Hogarth 1999). The linking hypothesis between participants' responses and readings is based on utility maximization in simple decision problems. Expected utility is measured in terms of direct financial payoff. The facts in the world and the financial gains/losses were correlated with the readings we tested (Table 1 & 2) such that the payoff, illustrated in Table 3, emerges for a person who decides to redeem/pay out a bet. Expected utility was calculated based on situation and on the probabilities of the three readings.

Table 3

Situation \ Reading		Role 1			Role 2		
		WE _{distr.} / CA+	WE _{nondistr.} / CASU	SE/CA/ IA	WE _{distr.} / CA+	WE _{nondistr.} / CASU	SE/CA/IA
WE _{distributive} /CA		20	-10	-10	10	-20	-20
WE _{nondistrib} /CASU		20	20	-10	10	10	-20
SE/CA-/IA		20	20	20	10	10	10

2.3 Analysis

We created two Bayesian statistical models for the experimental data. The first model, which we call the standard model, assumed that there is a fixed value of subjective probability for the three readings in the entire population. The second model, the variable-value model, considers the possibility that different persons have different probabilities for the different readings. We ran the models using a Hamiltonian Monte Carlo simulation with the no-U-turn sampling (NUTS) algorithm (Hoffman & Gelman, 2014) and performed model comparison using the Bayes factor (BF) and the bridgesampling package (Gronau, Singmann, & Wagenmakers, 2020). For the verb *surprise* the best model was the variable value-model, see Figure 1 for the posteriors. In both languages, the SE reading is the dominant interpretation, but there is also a group of people that has a non-distributive weak exhaustive interpretation of questions embedded under this verb. For German *agree*, the best model was the standard model with the CASU reading as the dominant interpretation. For English *agree*, the CA reading was dominant followed at a big distance by the CASU interpretation. The best model in this case was the variable-value model. Figure 2 shows the posterior of the best model for each language. Note that for German, we tested a negative baseline reading instead of the CA reading.

2.4 Conclusion

Our findings align with prior empirical evidence. They suggest reliability and robustness for the previously found readings. For *surprise*, the weakest reading (SE) is indeed dominant, cf. Cremers & Chemla (2017). For *agree*, we found that it is not required for the AHs to have the same opinions on the entire answer space, cf. Chemala & George (2016). Future research should take into account more the cross-linguistic landscape and investigate the effect of varying gains and losses in terms of prospect theory (Kahneman & Tversky 1979).

Figure 1

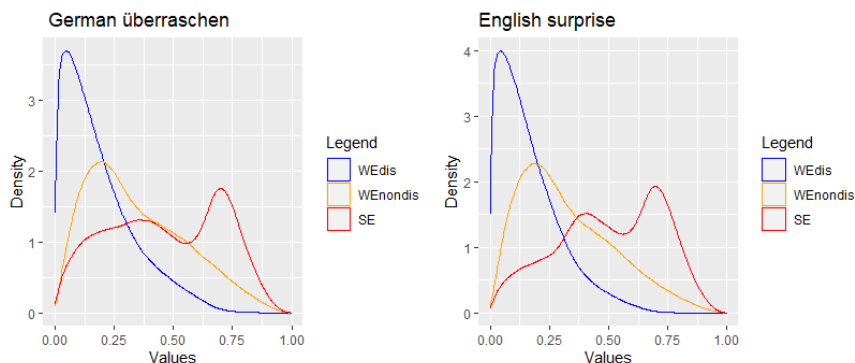
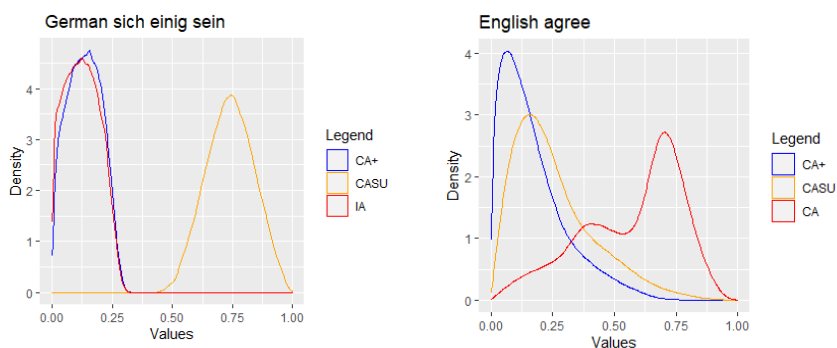


Figure 2



References

- Beck, S. & Rullmann, H. (1999). A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7, pp. 249-298.
- Berman, S. (1991). On the semantics and logical form of Wh-clauses. PhD thesis. Amherst, MA: University of Massachusetts.
- Camerer, C. F. & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19(1), pp. 7-42.
- Cremers, A. & Chemla, E. (2017). Experiments on the acceptability and possible readings of questions embedded under emotive-factives. *Natural Language Semantics* 25, pp. 223-236.
- Chemla, E. & George, B. R. (2016). Can we agree about 'agree'? *Review of Philosophy and Psychology* 7(1), pp. 243-264.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), pp. 457-472.
- Gronau, Q. F., H. Singmann, and E.J. Wagenmakers (2020). "bridgesampling: An R Package for Estimating Normalizing Constants". In: *Journal of Statistical Software* 92.10, pp. 1-29.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), pp. 263-291.
- Lahiri, U. (2002). Questions and answers in embedded contexts. Oxford: Oxford University Press.
- Klinedinst, N. & Rothschild, D. (2011). Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4(2), pp.1-23.
- Kratzer, A. (2006). *Exhaustive Questions*. Linguistics Colloquium, MIT
- Roelofsen et al. 2019. The *whether puzzle. In v. Heusinger et al. (eds.). *Questions in Discourse*, 172-197. Leiden: Brill.

Je peux ou je dois ? Faudrait savoir ! **Acquiring modals' force: evidence from French**

Anouk Dieuleveut, Laboratoire de Linguistique Formelle, Université Paris Cité, CNRS

Learning the “force” of modals raises a *subset problem* for children. Given that necessary entails possible, what prevents learners from assuming possibility meanings for necessity modals like *must*? Acquisition studies show that children struggle with necessity modals. But they tend to focus on English, where necessity modals are rare in the input. By looking at French, this study first shows that frequency alone can't explain English results: despite being more exposed to necessity modals, French children still display a ‘Necessity Gap’. Second, we discuss *how* children figure out force, focusing on one logical way to solve *subset problems*: negation. We show that learners need other cues: given irregularities of scope, informative cases are almost absent from their input.

1 Introduction

Modals can express different forces: possibility (e.g. *can*), or necessity (e.g. *must*). Learning modals' force raises a semantic *subset problem* for children (Berwick 1985; Wexler & Manzini 1987). Given that *necessary* entails *possible*, possibility modals (e.g., *can*) are used in contexts where necessity modals (e.g., *must*) are also true: What prevents children from hypothesizing possibility meanings for necessity modals? One logical solution is that children rely on negative (Downward-entailing) environments, since these reverse the direction of logical relationships (Gualmini & Schwarz 2009).

The acquisition literature reports an asymmetry in children's early mastery of modals' force. Studies on English show that by age 2, children use possibility modals (e.g. *can*) frequently, productively, and in an adult-like way. But they use necessity modals (e.g. *have to*) later on, less frequently, and in a non-adult-like way: they use them when adults prefer possibility modals (Dieuleveut et al. 2019). Comprehension experiments also report difficulties: 4 to 5-year-old children tend to both accept possibility modals in necessity situations—for instance, they consider as appropriate “There might be a bear in the box” in a situation where it is certain that there is one—; and necessity modals in possibility situations (e.g. “There must be a bear in the box” when it is just a possibility that there is one) (Noveck 2001; Ozturk & Papafragou 2013; Cournane et al. 2021). The source of children's difficulties, sometimes called the ‘Necessity Gap’, is debated. One possibility is that children would have issues with the *meaning* of necessity modals: they might not have figured out their underlying force yet.

This literature mostly focuses on English, where necessity modals are actually quite rare in parents' speech (children's ‘input’) (Dieuleveut et al. 2019). Previous corpus studies on French tend to focus on how children learn the other dimension of modals, flavor (the type of modality modals express: possibility and necessities given some rules, some goals, some capacities, or given what we know) (Bassano 1996; Cournane & Tailleux 2020), or how they interact with negation (Jeretič 2018). By looking at French, where necessity modals are more frequent in the input, this study assesses how differences in exposure might affect children's proficiency. Based on a detailed assessment of French young children's modal input and of their own modal production, we show that the delay for necessity modals reported for English-speaking children extends to French. While hearing more necessity modals, French children still produce them later on and less frequently. Second, we focus on the role negation might play in figuring out force, and argue that children need other cues: given irregularities of scope, informative cases are almost absent from their input.

2 Methods

We use the Lyon Corpus of French (Demuth & Tremblay 2008) (5 children; 3 females; age range: 1;00-3;00), and the Paris corpus (Morgenstern & Parrisé 2007); 6 children; 3 females; age range: 0;7-6;03), on CHILDES (MacWhinney 2000). Children are recorded during

spontaneous interactions with their parents at home. Both corpora contain audio/video data. 14,596 utterances (adults: 11,082; child: 2,939; other children (excluded): 575; excluding repetitions: adult (2.4%): 10,813; child (8.3%): 2,695) containing modal auxiliaries (*pouvoir/devoir/falloir/avoir-à*) were extracted and coded for force ((1)), negation ((2)) and flavor ((3)).

- (1) **Force** POSSIBILITY (P): *pouvoir*
NECESSITY (N): *falloir, devoir, avoir-à*
- (2) **Negation** *absent*: Il doit partir. ('He must leave.')
- present*: Il (**ne**) doit {**pas**/...} partir ('He (NEG) must NEG leave.')
- other DE-environments*: **Personne** (**ne**) doit partir ('**Nobody** must leave.')
- (3) **Flavor** *Root*: MOTHER: y a plein d'habits sales ! ('There are many dirty clothes!')
- MOTHER: elle **doit** laver tout le linge. ('She **must** do all the laundry')
- (Lyon corpus, Marie, 3;06,19)
- Epistemic*: CHILD: je trouve pas la grosse. ('I can't find the big one')
- MOT: elle **doit** être restée dans la voiture. ('It **must**'ve stayed in the car')
- (Lyon corpus, Marie, 2;05,16)

3 Results

Utterances containing modal auxiliaries represent 3.8% of all French adults' utterances (vs 5.8% in English), and 1.9% of children's utterances between age 2 and 3 (1.6% in English).

Force. **Table 1** summarizes adults' and children's modal productions in French (**1a**), with English as comparison point (**1b**). Data for English are taken from Dieuleveut et al. (2019). We see that while hearing more necessity modals in their input (French: N=61.9%; P=38.1%, vs English: N=28.4%; P=71.6%), French children produce more possibility modals (N=39.9%; P=60.1%) (**Fig1a/1b**).

Flavor. As reported in other languages (Kuczaj & Maratsos 1975; Papafragou 1998, a.o.), French children produce few epistemic modals, and tend to produce them later on (the so-called 'Epistemic Gap', Cournane 2014). We see that the asymmetry is reflected in their input: there is a huge bias towards root uses in adult speech (5.9% of epistemic modal uses in adults' speech, and 0.4% in children's).

Negation. Children don't produce necessity modals with negation frequently (14.3% of necessity modals with negation; 17.4% for adults). We find no negated *devoir* between age 2 and 3. A noteworthy difference between French and English is that, while English children produce many negated *possibility* modals, much more than adults (51.0%, vs 20.9% for adults), French children produce them less often than adults (children=16.1%; adults=22%).

Figure 1a/b Distribution of possibility and necessity modals with and without negation, by force and speaker (**1a**: French; **1b**: English).

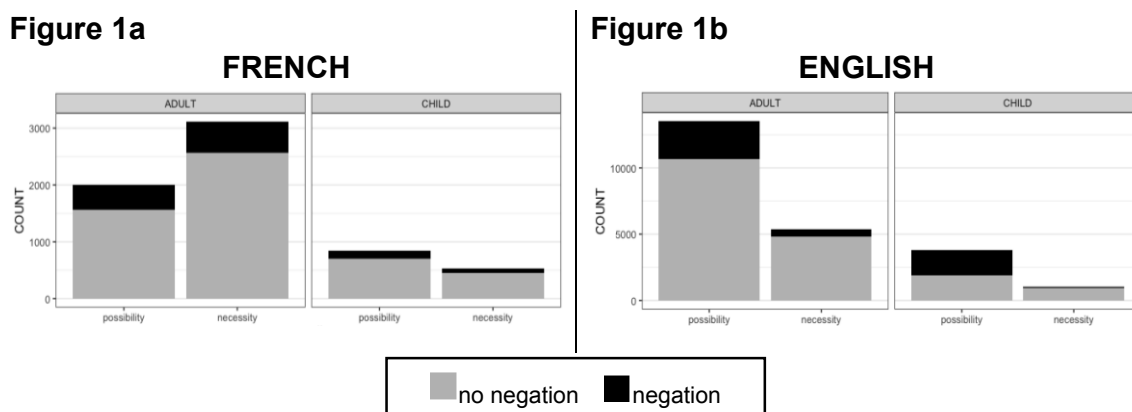


Table 1a/b Counts and percentages of modal uses by force and lemma for adults and children, ordered by lemma frequency in adult speech (repetitions excluded), with usage frequency with negation and proportion of root/epistemic uses. Instances taking NP complements (e.g. “il faut du pain”, ‘we need some bread’), are excluded (*falloir*: 6.5%; *devoir*: 0.4%; *avoir-à*: 64.7%).

Table 1a

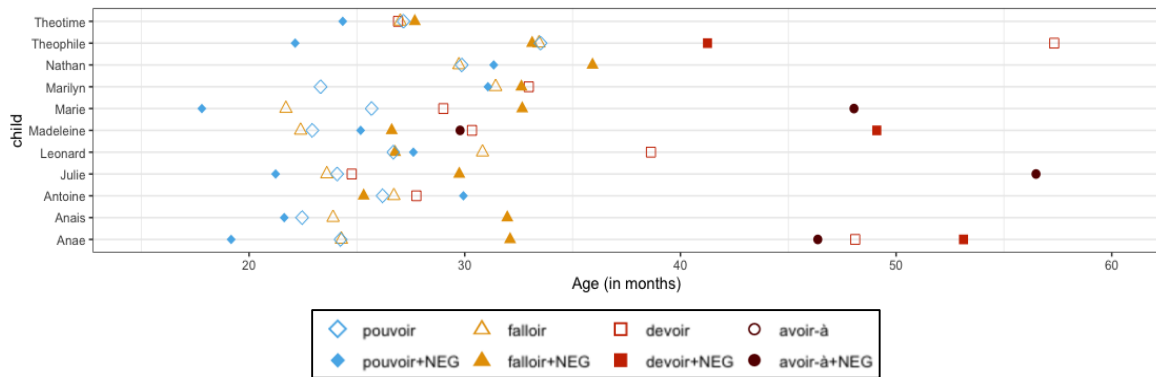
	ADULTS (n=5116)				CHILDREN (n=1378)			
	all	%mod	%Neg	%Epi	all	%mod	%Neg	%Epi
POSS	2008	38.1%	22.0%	2.9%	849	60.1%	16.1%	0.4%
<i>pouvoir</i>								
NECE	3108	61.9%	17.4%	7.8%	529	39.9%	14.3%	0.6%
<i>falloir</i>	2659	53.2%	17.6%	0.0%	492	36.8%	12.0%	0.0%
<i>devoir</i>	403	7.7%	6.6%	60.0%	21	1.8%	0.0%	14.3%
<i>avoir-à</i>	46	1.0%	93.3%	0.0%	16	1.3%	100.0%	0.0%
ALL	5116	100.0%	19.2%	5.9%	1378	100.0%	15.4%	0.4%

Table 1b

	ADULTS (n=18853)				CHILDREN (n=4800)			
	all	%mod	%Neg	%Epi	all	%mod	%Neg	%Epi
POSS	13500	71.6%	20.9%	9.8%	3798	79.1%	51.0%	2.4%
<i>can</i>	11934	63.3%	20.4%	1.3%	3700	77.1%	50.9%	0.2%
<i>might</i>	1213	6.4%	17.1%	95.2%	86	1.8%	2.3%	93.0%
<i>able</i>	315	1.7%	57.5%	0.0%	3	0.1%	66.7%	0.0%
<i>may</i>	38	0.2%	10.5%	42.1%	9	0.2%	0.0%	55.6%
NECE	5353	28.4%	10.1%	6.3%	1002	20.9%	5.2%	2.1%
<i>have to</i>	2398	12.7%	4.5%	0.3%	352	7.3%	2.0%	0.3%
<i>got to</i>	937	5.0%	1.2%	0.7%	288	6.0%	1.7%	0.0%
<i>should</i>	696	3.7%	22.8%	7.5%	21	0.4%	19.0%	9.5%
<i>need to</i>	493	2.6%	17.0%	0.0%	217	4.5%	6.0%	0.0%
<i>must</i>	411	2.2%	15.8%	64.2%	114	2.4%	17.5%	15.8%
<i>supposed</i>	335	1.8%	31.3%	2.7%	9	0.2%	33.3%	0.0%
<i>ought to</i>	83	0.4%	8.4%	0.0%	1	0.0%	0.0%	0.0%
ALL	18853	100%	18%	8.8%	4800	100%	41%	2.4%

Age of first production. As reported for English, children tend to produce possibility modals earlier. The mean age of 1st production for *pouvoir* is 1:11; for *falloir*, 2:03; for *devoir*, 2:11; and for *avoir-à*, 5:06 (**Figure 2**).

Figure 2. Age (in months) of first productions of *pouvoir*, *falloir*, *devoir* and *avoir-à* with and without negation (n = 11 children).



4 Discussion

While more exposed to *necessity* modals, French children use *possibility* modals earlier and more frequently. This suggests that the delay for English necessity modals are not just due to their low frequency.

Why are necessity modals delayed? Several hypotheses have been proposed, and several factors might be involved. It does not have to reflect deep semantic or conceptual issues: the production asymmetry could simply come from differences between children and adults’ conversational goals: children might be less prone to giving orders, or to expressing certainty, than their parents. Moreover, children’s first modals tend to express ability flavor, which lack clear necessity counterparts (Horn 1972; Hackl 1998). Another factor, specific to the syntax of French *falloir*, could participate in the delay: *falloir* only takes expletive subjects, and might therefore be acquired later. But this could reflect deeper *semantic* issues, with the meaning of necessity modals: If children are unsure about the force of necessity modals, they might produce them less often. Future research will probe further whether French children use them in an adult-like way, using the same variant of the Human Simulation Paradigm (Gillette et al. 1999) used to assess English-speaking children productions (Dieuleveut et al., 2019).

The second question this study addresses is *how* children might solve the *subset problem*. Our results suggest that as previously argued for English (Dieuleveut et al. 2022), as well as for the acquisition of *every* (Rasin & Aravind 2021), it is unlikely that French children can rely on negative environments to figure out the force of necessity modals: given irregularities of scope, informative cases are almost absent from their input. But the problem in French is even more acute than in English: both *falloir* and *devoir* outscope negation (Iatridou & Zelijstra 2013; Homer 2015), cases in which using negation might be even confusing. The one French necessity modal that scopes under negation, *avoir-à* ('have-to'), is extremely rare and almost only occurs in exceptive constructions (*n'avoir-qu'à*, '~ only have-to'), where it means possible (von Fintel & Iatridou 2007) (adults: 40/43 cases of negated *avoir-à*; children: 16/16 cases). Children thus need other cues to solve the subset problem, which we will probe in future research.

References

- Bassano, D. (1996). Functional and formal constraints on the emergence of epistemic modality. *First Language*, 16(46), 077-113.
- Berwick, R. C. (1985). *The acquisition of syntactic knowledge* (Vol. 16). MIT press.
- Cournane, A., & Tailleux, S. (2020). La production épistémique chez l'enfant francophone. *Arborescences: revue d'études françaises*, (10), 47-72.
- Cournane, A. (2015). *Modal development: input-divergent L1 acquisition in the direction of diachronic reanalysis*. University of Toronto (Canada).
- Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of child language*, 35(1), 99-127.
- Dieuleveut, A., van Dooren, A., Cournane, A., & Hacquard, V. (2022). Finding the force: How children discern possibility and necessity modals. *Natural Language Semantics*, 1-42.
- Dieuleveut, A., van Dooren, A., Cournane, A., & Hacquard, V. (2019). Learning modal force: evidence from children's production and input. In *Proceedings of the 2019 Amsterdam Colloquium* (pp. 111-122).
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2), 135-176.
- Gualmini, A., & Schwarz, B. (2009). Solving learnability problems in the acquisition of semantics. *Journal of Semantics*, 26(2), 185-215.
- Hackl, M. (1998). On the Semantics of "Ability Attributions".
- Homer, V. (2015). Neg-raising and positive polarity: The view from modals. *Semantics and Pragmatics*, 8, 4-1.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California, Los Angeles.
- Iatridou, S., & Zelijstra, H. (2013). Negation, polarity, and deontic modals. *Linguistic inquiry*, 44(4), 529-568.
- Jeretič, P. (2018). Evidence for children's dispreference for weakness: A corpus study. *Manuscript, New York University*.
- MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.
- Manzini, M. R., & Wexler, K. (1987). Parameters, binding theory, and learnability. *Linguistic inquiry*, 413-444.
- Maratsos. 1975. What children can say before they will. *Merrill-Palmer Quarterly* 21(2): 89–111.
- Morgenstern, A., & Parisse, C. (2012). The Paris Corpus. *Journal of French language studies*, 22, 7-12.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165-188.
- Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, 11(3), 191-214.
- Papafragou, A. (1998). The acquisition of modality: Implications for theories of semantic representation. *Mind & language*, 13(3), 370-399.
- Rasin, E., & Aravind, A. (2021). The nature of the semantic stimulus: the acquisition of every as a case study. *Natural Language Semantics*, 29(2), 339-375.
- Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of child language*, 28(1), 127-152.
- Von Fintel, K., & Iatridou, S. (2007). Anatomy of a modal construction. *Linguistic Inquiry*, 38(3), 445-483.

“Spinnt sie” or “Spinnt die”?

Empirical studies on d- and p-pronouns in German

Luise Ehrmantraut, Saarland University

D-pronouns (*der/die/das*) are traditionally classified as demonstrative pronouns, because they often occur in the same contexts of use as *dieser*. However, they also compete with p-pronouns (*er/sie/es*), so they stand between two pronominal subtypes. I therefore propose a new partial classification of the German pronoun system that distinguishes between two d-pronoun types. Based on corpus data, I show that the use of one d-pronoun type is reserved for conceptually oral text types. Furthermore, I present experimental data on prosodic properties of the different d-pronoun types. Finally, I showcase two experiments highlighting valuation and gender as relevant factors influencing pronoun choice.

1 Reclassification

Based on existing classification attempts (Bethke 1990, Weinrich 2007), I propose a new partial classification of the German pronoun system, which considers the observed usage and function spectrum of d-pronouns. It provides for a summary and subdivision of the pronouns of the third person under the term *reference pronouns* (In the following: RPs). D-pronouns are listed in it as demonstrative RPs (besides *dieser, jener*, etc.) and as rhematic RPs. In the latter function, they stand in opposition to the p-pronouns classified as thematic RPs.

Table 1: New partial classification of German pronouns

reference pronouns (third person)	thematic reference pronoun	<i>er, sie, es</i>
	rhematic reference pronoun	<i>der, die, das</i>
	demonstrative reference pronoun	<i>dieser, jener, etc.</i> <i>der, die, das</i>

Recent research on d-pronouns is predominantly concerned with the use of d-pronouns as demonstrative RPs in relation to p-pronouns. The focus is on cases of referential ambiguity as in (1), in which p- and d-pronouns show different reference preferences (Bosch & Umbach 2007, Ellert 2013, Bosch & Hinterwimmer 2016, Bader & Portele 2016/2019).

(1) Paul war heute mit Hans verabredet. Er / Der hat aber kurzfristig abgesagt.
Paul was today with Hans appointed. He_{them/dem} has but short term canceled.

My work, conversely, focuses on contexts of use of d-pronouns as rhematic RPs as in (2), where d-pronouns behave as p-pronouns and pronoun choice is guided by pragmatic factors.

(2) Peter wollte mich heute besuchen. Er / Der hat aber kurzfristig abgesagt.
Peter wanted me today to visit. He_{them/rhem} has but short term canceled.

According to Bethke (1990) and Weinrich (2007), thematic and rhematic RPs differ in their attention value. While thematic RPs indicate low-attentional continuity, rhematic RPs maintain attention with respect to the referent.

2 Corpus Study

Starting point was the hypothesis that d-pronouns are used more frequently relative to p-pronouns the more conceptually oral the language use is (Bethke 1990).

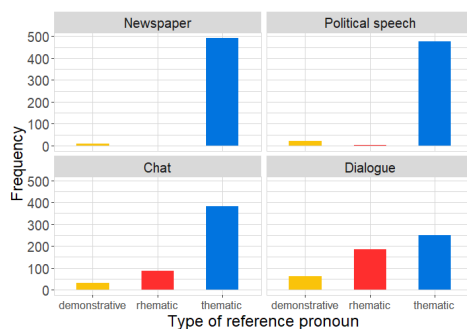
2.1 Procedure

Hence, four text types were examined for occurrences of p- and d-pronouns that cover the range of conceptual orality and writtenness: *newspaper, political speech, chat* and *dialogue*. Suitable corpora were the sub-corpus *TAGGED-T2* of the DeReKo, the sub-corpora *Politische Reden* and *Dortmunder Chatkorpus* of the DWDS and the corpus *FOLK* of the DGD. For each text type, 500 random search hits were considered. Every d-pronoun was analyzed in terms of the d-pronoun type (rhematic or demonstrative).

2.1 Results

A simple multinomial model was calculated using the GAMLj module of the program Jamovi (The jamovi project 2021). For both d-pronoun types (rhematic and demonstrative), the frequency ratio to thematic RPs was analyzed as a function of text type. For the frequency ratio of demonstrative and thematic RPs, the paired comparison of all text types revealed significant differences. For the frequency ratio of rhematic and thematic RPs, the text type *newspaper* was excluded from the analysis, because no rhematic d-pronoun was found. In the paired comparison of the other text types, significant differences were found in each case.

Figure 1: Frequency ratio of demonstrative, rhematic and thematic RPs in different text types



2.2 Discussion

For both d-pronoun types, the results confirm the hypothesis that the more conceptually oral a text or a speech is, the more frequently they occur. One possible explanation is that conceptually oral speech, because of its spontaneity, and spoken speech, because of its volatility, more likely or frequently require attention signals to the listener, for example, in form of rhematic or demonstrative RPs, which fulfill different attention-controlling functions (attention-maintaining vs. attention-diverting). Furthermore, conceptually oral speech is generally more characterized by emotion and expressivity and less by polite restraint, which favors the use of rhematic RPs (Bethke 1990, see chapter 4). For the demonstrative RP, there also exist alternative expressions such as *diese/r*, which are primarily used in conceptually written language, where they represent competitors for the demonstrative d-pronoun.

3 Rating Studies

Since contradictory generalizations about prosodic properties of d-pronouns can be found in the research literature – partly described as basically strongly accented (Weinrich 2007, Engel 2009), partly as usually not conspicuously accented (Ahrenholz 2007) – I conducted two experimental rating studies investigating the accentuability of p-pronouns and d-pronouns in their different functions (thematic, rhematic and demonstrative RPs).

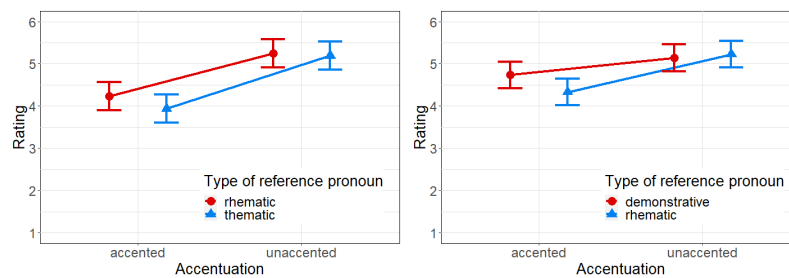
3.1 Design

The first experiment contrasted thematic and rhematic RPs in terms of their accentuability, the second experiment contrasted rhematic and demonstrative RPs. Thus, both experiments crossed the factors pronoun type and pronoun accent, and the dependent variable was the subjects' rating on a seven-point scale (2x2 design). Acoustic stimuli were presented to the subjects (16 items, 32 filler items) embedded in a written description of the utterance situation. The different conditions were constructed as minimal pairs similar to examples (1) and (2), but without referential ambiguity. 60 subjects participated in each of the experiments. The hypothesis was that the acceptability of an accent on the pronoun depends on the pronoun type: demonstrative RPs like the d-pronoun in (1) referring to *Hans* show a higher acceptability of the accented variant in relation to the unaccented variant (default) than thematic and rhematic RPs like the p- and d-pronoun in (2).

3.2 Results

The data was analyzed by calculating a linear mixed effects model using lme4 for R (Bates et al. 2015). The result confirms the hypothesis: Accented demonstrative RPs are rated significantly better than accented rhematic or thematic RPs compared to non-accented ones.

Figure 2: Acceptability rating in dependence of pronoun type and pronoun accent



3.3 Discussion

While d-pronouns as rhematic RPs maintain attention with respect to a salient referent (Bethke 1990, Weinrich 2007), the function of the demonstrative RP is to emphasize a less salient referent (Bader & Portele 2016/2019). Since accentuation serves to emphasize sentence elements on which no special attention was previously focused, it is more compatible with the function of the demonstrative RP than with that of the rhematic RP. The result thus additionally supports the differentiation of the two d-pronoun types.

4 Forced Choice Studies

That thematic and rhematic RPs differ in their attention value becomes clear, for example, in the context of evaluative utterances. The initial hypothesis was that evaluative utterances about personal referents, whether positive or negative, increase the willingness to use rhematic RPs instead of thematic RPs, regardless of the gender of the referent.

4.1 Design

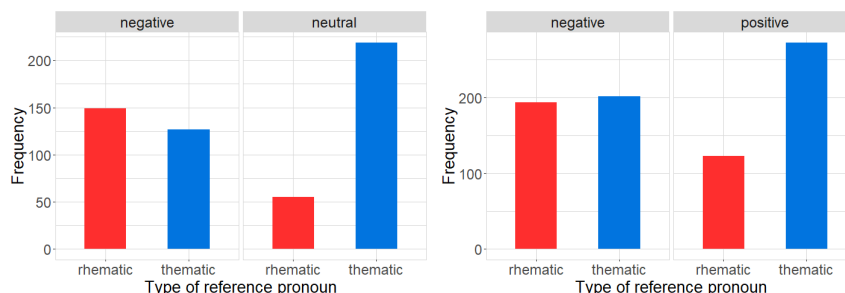
Two experimental studies were conducted. Both crossed the factor valuation with the factor gender. The dependent variable, pronoun type, was determined by the subjects themselves. In the first study, negative (3b) and neutral (3a) utterances were contrasted; in the second, negative and positive (3c) utterances. Apart from that, the experiments were identically structured (2x2 design). In each of the two experiments, 36 subjects were presented with 24 items and 48 filler items.

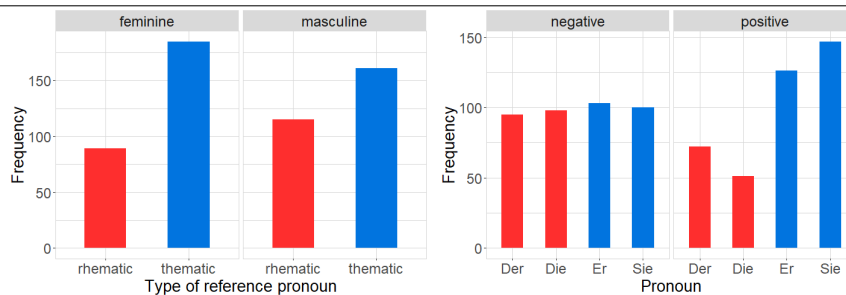
- (3) Gestern war ich bei Kerstin zum Essen eingeladen.
 Yesterday was I at Kerstin for dinner invited.
- a. Sie/Die hat Nudeln gekocht. b. Sie/Die kocht echt furchtbar.
 She_{them/rhem} has noodles cooked. She_{them/rhem} cooks really terrible.
- c. Sie/Die kocht echt super.
 She_{them/rhem} cooks really great.

4.2 Results

For statistical analysis, the program Jamovi was used. A generalized mixed model calculated using the GAMLj module revealed a significant main effect for the factor valuation for both experiments, a significant main effect for the factor gender for experiment 1, and an interaction effect between the factors valuation and gender for experiment 2 (Figure 3).

Figure 3: Results of experimental studies on the influence of valuation and gender on pronoun choice





4.3 Discussion

The results confirmed that valuation increases the likelihood of rhematic d-pronoun occurrence, but also revealed unexpected effects and interaction concerning positive valuation and gender. One possible explanation for those effects is based on research results according to which d-pronouns are classified as more impolite for referring to personal referents (Bellmann 1990). Thus, in positively evaluated utterances, the tendencies to keep attention on the referent by using a d-pronoun and to behave politely towards the person by refraining from using a d-pronoun compete. The fact that rhematic d-pronouns occur much less frequently in positively evaluative utterances with reference to feminine referents could be related to a stronger unconsciously anchored command of politeness towards women. Since Bellmann's study only investigates the reference to present persons by d-pronouns, it must be examined whether the same applies to the reference to absent persons in order to substantiate this explanatory approach. An alternative explanation for the more frequent use of rhematic RPs in negatively evaluative utterances could be that the inhibition threshold for negative utterances about a referent is basically higher than for positive ones and that for this reason negative valuation is basically perceived as more intense compared to positive valuation. Accordingly, the rhematic RP as an attention-controlling intensity marker would be more likely and more frequently required in negative utterances than in positive ones.

References

- Ahrenholz, Bernt. (2007). *Verweise mit Demonstrativa im gesprochenen Deutsch*. Berlin/New York: Walter de Gruyter.
- Bader, Markus & Portele, Yvonne. (2016). Accessibility and Referential Choice: Personal Pronouns and D-pronouns in Written German. *Discours*, 18. DOI: 10.4000/discours.9188.
- Bader, Markus & Portele, Yvonne. (2019). The interpretation of German personal pronouns and d-pronouns. *Zeitschrift für Sprachwissenschaft*, 38(2), 155-190. DOI: 10.1515/zfs-2019-2002.
- Bates et al. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. DOI: 10.18637/jss.v067.i01.
- Bellmann, Günter. (1990). *Pronomen und Korrektur*. Berlin/New York: Walter de Gruyter.
- Bethke, Inge. (1990). *der, die, das als Pronomen*. München: Iudicum.
- Bosch, Peter & Umbach, Carla. (2007). Reference Determination for Demonstrative Pronouns. *ZAS Papers in Linguistics*, 48, 39-51. DOI: 10.21248/zaspil.48.2007.353.
- Bosch, Peter & Hinterwimmer, Stefan. (2016). Anaphoric Reference by demonstrative pronouns in German. In search of the relevant parameters. *Linguistische Arbeiten*, 563, 193-212.
- Ellert, Miriam. (2013). Information Structure Affects the Resolution of the Subject Pronouns *Er* and *Der* in Spoken German Discourse. *Discours*, 12, 3-24. DOI: 10.4000/discours.8756.
- Engel, Ulrich. (2009). *Deutsche Grammatik* (2nd ed.). München: Iudicum.
- Gallucci, Marcello. (2019). *GAMLj: General analyses for linear models*. Retrieved from <https://gamlj.github.io/>.
- R Core Team. (2021). *R: A Language and environment for statistical computing*. Retrieved from <https://cran.r-project.org/>.
- The jamovi project. (2021). *jamovi* (Version 1.6). Retrieved from <https://www.jamovi.org>.
- Weinrich, Harald. (2007). *Textgrammatik der deutschen Sprache* (4th ed.). Hildesheim: Georg Olms.

Investigating lexical bias in Persian light verb constructions: What can we learn from priming experiments?

Pegah Faghiri, Monique Flecken and Eva van Lier (University of Amsterdam)

The verbal lexicon in Persian is overwhelmingly formed by multiword expressions (MWE) including a verb and a non-verbal element, mainly a noun, such as *bāzi kardan* ‘play’ (game do) or *qadam zadan* ‘walk’ (step hit), known as complex predicates (CPs) or light verb constructions (LVCs). There is an ongoing theoretical debate about the nature of these LVCs: The compositional view, formulated in various studies in generative syntax, is mainstream. In contrast, numerous studies, taking a lexicalist approach, have stressed the idiomatic nature and word-like properties of (a part of) Persian LVCs (see Samvelian 2018 for an overview). While some recent contributions present experimental evidence (e.g. Faghiri & Samvelian 2021), the type of data involved remains fundamentally the same: acceptability ratings vs. grammatical judgments. In particular, a language processing perspective is largely lacking (Shabani-Jadidi 2012 is a notable exception). In this study, we use a (structural) priming paradigm to investigate to what extent Persian LVCs display lexicalized behaviour. The alternation we are focusing on concerns the choice of LV involved in forming the LVC, given that a part of Persian LVCs display the possibility to alternate between two or more verbs, without affecting the meaning of the combination (see e.g. Faghiri & Samvelian 2013).

Our priming study builds on related research on the well-known dative alternation in Germanic languages. Corpus studies reveal that alternating ditransitive verbs show a statistical preference for (or occur more often in) one of the two alternating constructions. Importantly, such lexical preferences or biases manifest themselves in priming production experiments (e.g. Melinger & Dobel 2005, Bernolet & Hartsuiker 2010, Segaert et al. 2014). Specifically, presenting a verb in its dispreferred construction is shown to exert a stronger priming effect compared to that verb’s preferred construction; this is called “inverse priming”. However, the overwhelming majority of studies on lexical bias in priming concerns Germanic ditransitives. Therefore, our study not only offers a new, processing-based perspective on Persian LVCs, but also broadens the scope of verb-bias studies in terms of the languages and constructions covered.

As mentioned, a number of verbs can participate in forming Persian LVCs. *Kardan* ‘do’ is the most frequent LV and is used in many newly coined LVCs, e.g. *imeyl kardan* ‘email’, *čat kardan* ‘chat’, *fālo kardan* ‘follow (in social networks)’, *še(y)r kardan* ‘share (in social networks)’. But there are other productive LVs as well, such as *zadan* ‘hit’, which is also used in newly coined LVCs, e.g. *payāmak zadan* ‘send a text (on a mobile phone)’, *lāyk zadan* ‘like (a post on social networks)’ (see Samvelian 2012). Some of these can occur with both LVs, e.g. *lāyk kardan/zadan*, *imeyl kardan/zadan*, whereas others cannot, e.g. *fālo kardan/*zadan*. As such, the LV alternation in Persian LVCs is lexically restricted, like the dative alternation.

As a pretest for our priming study, we first collected baseline data for a set of 52 LVCs potentially displaying *kardan/zadan* (‘do’/‘hit’) alternation. We were able to use picture stimuli for 24 of these and hence included them in a constrained picture-description experiment (Fig. 1). For the remaining 28, we ran a Cloze task experiment (Fig. 2). In both experiments, participants completed sentences with a verb, i.e. the verbal element of the LVC, while the nominal element was specified – either in the picture (Exp. 1) or as part of the sentence (Exp. 2). Both experiments included 3 practice items at the beginning and 2 filler items between critical items. They were carried out online via PCIbex Farm. We collected a total number 2085 datapoints from respectively 71 (Exp. 1) and 56 (Exp. 2) native speakers of (Iranian) Persian. We excluded 15 LVCs which exclusively occurred with ‘hit’ or ‘do’ and for which we did not have corpus evidence for the possibility to occur with the other LV, as well as 7 for which the rate of ‘do’ and/or ‘hit’ were lower than the rate of a third option, and classified the 34 remaining into three “bias” types based on the rate of ‘do’/‘hit’ in the sentences produced:

1. Do-bias LVCs: LVCs with more than two-third (66.7%) of do- responses

-
2. Hit-bias LVCs: LVCs with less than one-third (33.4%) of do- responses
 3. No-bias LVCs: other/in-between LVCs

11 LVCs were labelled as having a do-bias (e.g. *atse* ‘sneeze’, *mangane* ‘staple’, *imeyl* ‘email’, *emzā* ‘sign’), 14 as having a hit-bias (e.g. *telefon* ‘phone’, *baxiye* ‘stich’, *susu* ‘twinkle’, *taxmin* ‘estimate’) and 9 as having no-bias (e.g. *rang* ‘paint’, *vasle* ‘patch’, *mohr* ‘stamp’, *jamz* ‘sum’).

Thus, not only do our pretest data confirm LV alternations but they also show that LVCs can display lexical biases and that the choice of the LV is not necessarily motivated by syntactic and/or semantic constraints. If Persian LVCs display different lexical preferences and these preferences are indeed part of speakers knowledge, then we would expect to see priming effects similar to what has been observed for dative alternations. In particular, we expected a difference in the strength of priming effects, depending on the prime LVC bias type (i.e. do-bias, hit-bias and no-bias) and the priming condition: prime LVCs with a lexical bias were predicted to exert a stronger priming effect when used with their dispreferred LV (i.e. the inverse priming effect); Also, we expected target LVCs to show resistant to priming depending on the strength of their lexical bias, in particular do-/hit- bias target LVC types should be in general more resistant to priming than no-bias target LVCs. We expected to see these difference in the distribution of LVs produced by the participants and/or in their response latencies (see e.g. Segaert et al. 2014).

To test these predictions, we carried out a priming experiment via PCIBex Farm using a task similar to Ziegler et al. (2018). In each trial participants read a sentence (prime item) and then saw a picture (target item) that they had to describe (by typing) in one sentence using the given noun. They also answered a recognition memory question (cover task) after each item. We included three priming conditions (baseline do-prime, hit-prime) and used a 3x3x3 mixed design with priming condition and prime LVC bias type as a within-items factor and target LVC bias type as a between-items factor – across 9 experimental lists. For example:

1. Do-prime: *mesvāk kardan* ‘Ali brush-**did** his teeth after dinner.’ (simplified)
2. Hit-prime: *mesvāk zadan* ‘Ali brush-**hit** his teeth after dinner.’
3. Baseline: ‘Ali had a toothache after dinner’

Accordingly, with these combinations we get three different priming types depending on the prime LV expectedness: “expected” (when the prime bias type and the prime LV match, e.g. a do-bias LVC used with ‘do’), “unexpected” (when the prime bias type and the prime LV mismatch, e.g. a do-bias LVC used with ‘hit’), “neutral” (for all no-bias primes). Our stimuli consisted of 12 target pictures corresponding to 4 no-bias, 4 hit-bias and 4 do-bias LVCs from Experiment 1. Each target was paired with three different prime sentences, one of each type. In total 17 different prime LVCs were used. Each list contained all the 12 target items and across all lists each target item occurred an equal number of times with each prime of the three prime types as well as with each of the three priming conditions. Each list contained a different subset of 12 prime items appearing only once in one of the three conditions. In addition, each list included a set of 39 filler trials, which sums to 51 total trials.

129 (Iranian) Persian speakers participated in the experiment, of which results from 4 were excluded. We used the target sentence completion time span between two consecutive items as a measure of response latency (RL) for each item. We normalized RLs for participants across all the items only excluding the first one. We coded participants responses (1500 in total) according to the verb used: *kardan* (776), *zadan* (610), other (96) and incomplete (18), and filtered out irrelevant responses (114=7.6% of the data). For each target LVC, we calculated its baseline bias as the log-odds for *kardan* ‘do’ responses in the baseline condition (see Bernolet & Hartsuiker, 2010:457). We analysed the data using logistic mixed models for LV (with *kardan* = 1 and *zadan* = 0) and linear mixed models for RL.

With respect to the choice of LV, the results show a clear main effect of target bias, but we do not see any priming effects including the expected inverse priming effect: prime sentences with ‘do’ or ‘hit’ LVs trigger more ‘do’ responses (which is the overall more frequent choice) than the baseline for target LVCs that do not (already) have a bias for ‘do’, regardless of the LV used in the prime (Fig. 3) or the prime bias type (Fig. 4). RLs, on the other hand, are sensitive to both priming conditions (Fig. 5) and prime bias types (Fig. 6). While in the baseline condition, RLs are comparable for *kardan* ‘do’ and *zadan* ‘hit’ responses i) with ‘do’ primes RLs are lower with ‘do’ responses and with ‘hit’ primes RLs are lower with ‘hit’ responses and ii) with do-bias primes RLs are lower with ‘do’ responses, whereas with hit-bias primes RLs are lower with ‘hit’ responses. The results of our model fitted to a subset of the data excluding the no-bias prime type and the baseline condition (641 datapoints) showed i) a 3-way significant interaction between LV choice, priming condition and priming type and ii) a 4-way significant interaction including target bias. These results show that response latencies are differentially affected by the priming condition and the prime expectedness type: when participants choose ‘do’ with hit-bias targets they are significantly faster when primed with an unexpected ‘hit’ than when primed with an expected ‘hit’ (while with ‘do’ primes, RLs are comparable). Similarly, when participants choose ‘hit’ with hit-bias targets they are much faster when primed with an unexpected ‘do’ than when primed with an expected ‘do’ (while with ‘hit’ primes, RLs are comparable).


Our study confirms that some Persian LVCs show LV alternations that cannot be explained based on syntactic and/or semantic rules and that these LVCs display lexical biases in terms of which verb they preferably appear with. Moreover, the results of our priming experiment suggest that speakers have knowledge of these biases, which, when primed, manifests not only in their usage, i.e. significant difference in the rate of ‘do’ depending on the target LVC’s bias, but also in their response latencies, which are clearly affected by inverse priming.

References

- Bernolet, S. & R. J. Hartsuiker. 2010. Does verb bias modulate syntactic priming? *Cognition* 114: 455–461.
- Faghiri, P. & P. Samvelian. 2021. The issue of “separability” in Persian complex predicates. In *One-to-many relations in morphology, syntax, and semantics*, pp.117–149. Language Science Press.
- Melinger, A. & Ch. Dobel. 2005. Lexically driven syntactic priming. *Cognition* 98: B1–B20.
- Samvelian, P. 2012. *Grammaire des prédicats complexes. Les constructions nom-verbe*. Paris: Lavoisier.
- Samvelian, P. 2018. Specific features of Persian syntax. The Ezāfe Construction, Differential Object Marking and Complex Predicates. In A. Sedighi & P. Shabani-Jadid (eds.), *The Oxford handbook of Persian linguistics*, pp 226–269.
- Samvelian, P. & P. Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian Complex Predicates, In *Proceedings of The 9th Workshop on Multiword Expressions (MWE2013)*, pp 11-20.
- Segaert, K., Weber, K., Cladder-Micus, M. & Hagoort, P., 2014. The influence of verb-bound syntactic preferences on the processing of syntactic structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5): 1448.
- Shabani-Jadidi, P., 2012. Processing Compound Verbs in Persian. Ph.D. thesis, University of Ottawa.
- Ziegler, J., Snedeker, J., & Wittenberg, E. 2018. Event structures drive semantic structural priming, not thematic roles: Evidence from idioms and light verbs. *Cognitive Science* 42(8): 2918-2949.

Fig. 1 Screenshot of an Item from Experiment 1 (English translations are added for illustration)

منگنه



یک نفر دارد چند برگ کاغذ را بهم

Lit.: ‘Someone is a few sheets of paper together **[staple_N....]**.’
 Intended ‘Someone is stapling a few sheets of paper together.’
 NB. Persian is an SOV language.

Fig. 2 Screenshot of an item from Experiment 2 (English translations are added for illustration)

چه روزی احسان گزارش پیشرفت کار را ایمیل

Lit.: ‘What day Ehsan the progress report [email_N ...]?’
 Intended ‘When does Ehsan email the progress report?’

Fig. 3 Mean rate of ‘do’ by target bias type and prime conditions (Baseline, Do and Hit)

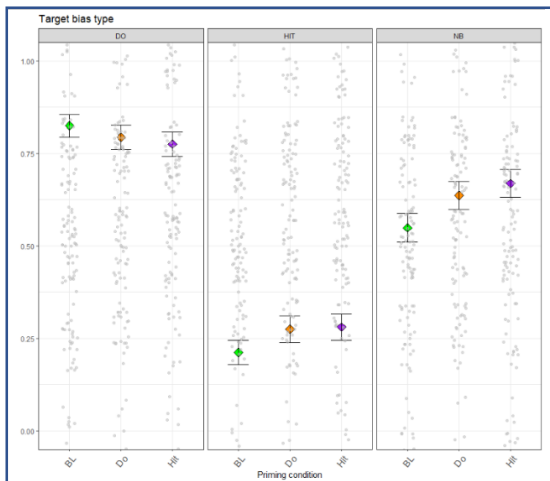


Fig. 4 Mean rate of ‘do’ by prime bias type and prime conditions (Baseline, Do and Hit)

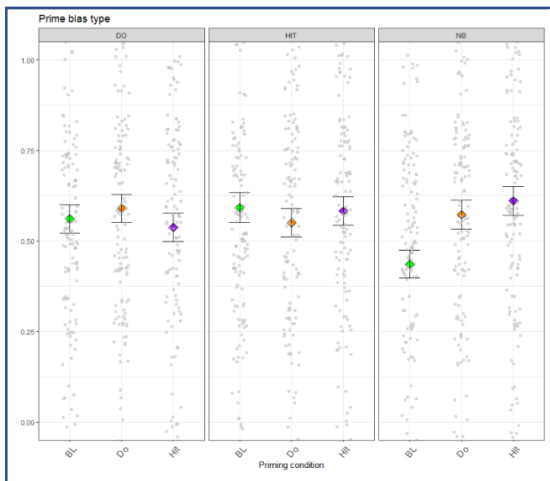


Fig. 5 Mean RIs by prime condition and response (kardan ‘do’ vs. zadan ‘hit’)

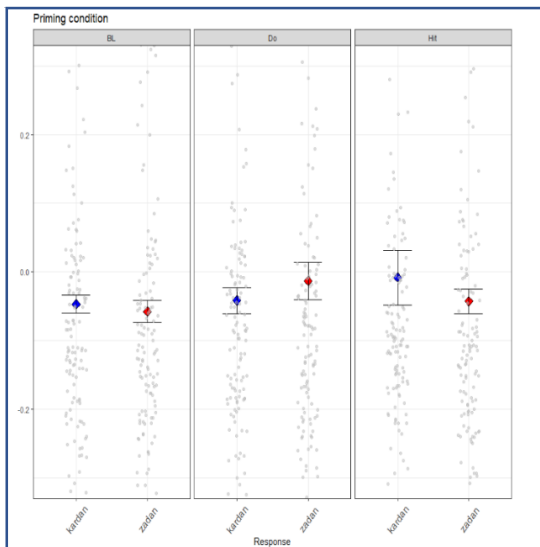
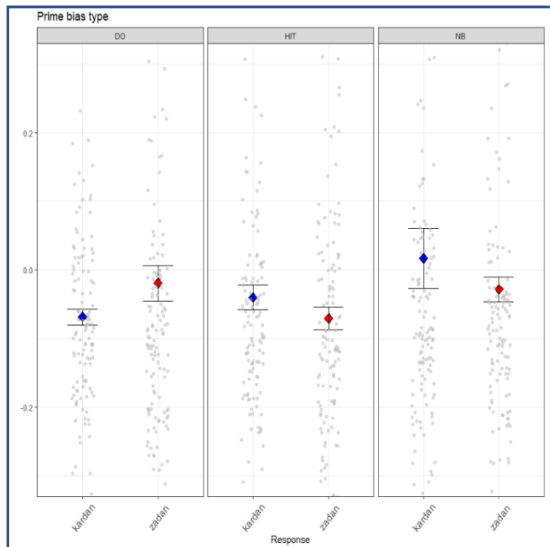


Fig. 6 Mean RIs by prime bias type and response (kardan ‘do’ vs. zadan ‘hit’)



Preposition omission in French sluicing: An empirical approach

Amal Hassen, Anne Abeillé

amal.hassen@etu.u-paris.fr, anne.abeille@u-paris.fr

LLF, Université Paris Cité & CNRS

Abstract

We show that omitting prepositions in French sluicing, a non-Preposition-stranding language, is acceptable, contrary to syntactic analyses based on syntactic reconstruction and deletion (Merchant, 2001). We tested Nykiel (2013&2017)'s theory, which predicts preposition omission under sluicing to be favored by nominal remnants/correlates, which are more informative than pronominal ones. We conducted two acceptability judgment experiments using pronominal correlates and remnants in the first one (*qui* 'who') and nominal ones in the second (*quel* 'which' +noun), and a corpus study (Frantext). Preposition omission under sluicing is acceptable in French and is rated higher/more frequent with nominal correlates and remnants than with pronominal ones. The effect of informativeness of the remnants and correlates confirms cognitive and information-based hypotheses on preposition omission under sluicing.

1 Introduction

Sluicing is a type of ellipsis where only a wh-word (i.e. remnant) is left (Ross, 1969). There are two main syntactic approaches: a fragment-based analysis with direct interpretation (Ginzburg & Sag, 2000; Culicover & Jackendoff, 2005) and syntactic reconstruction with deletion under identity (Ross, 1969; Merchant, 2001). According to the former, preposition omission comes from the choice between a PP or an NP correlate (1a). According to the latter, preposition omission is derived from preposition stranding (1b), and should not be allowed in non-preposition-stranding languages like French (2).

- (1) a. Kim talked [to [someone]], but I don't know [**who**] / [**to who**]
b. Kim talked to someone, but I don't know to who ~~she talked~~/who ~~she talked to~~
- (2) a. Anne l'a offert à quelqu'un, mais je ne sais pas à **qui** /*qui (Merchant 2001:115)
'Anne offered it to someone, but I don't know to who/*who'
b. *Anne l'a offert à quelqu'un mais je ne sais pas qui ~~elle l'a offert à~~
'Anne offered it to someone, but I don't know who she offered it to'

Nykiel (2013, 2017) suggests that preposition is optional in most languages and relies on a cognitive hypothesis for preposition omission, based on Hawkins (2004, 2014)'s minimize Form (MiF). MiF is the possibility of minimization or reduction of an element if its syntactic and semantic features are rendered obvious by the surrounding context. Nykiel found less omission with pronominal correlates (*someone*) and more with nominal correlates (*a friend*) in both English and Polish (a non P-stranding language), and henceforth the preposition has more chances to be omitted with a more informative remnant/correlate. In the same line, we consider Uniform Information Density (UID) (Levy & Jaeger, 2006), initially proposed for *that* omission, since a fragment with a preposition is more informative than a fragment without.

2 Experiments

In a corpus study on *-ça* (*that*) questions, Smirnova & Abeillé (2021) found two cases of preposition omission in matrix sluices (3), therefore, we tested the acceptability of preposition omission in French sluicing.

- (3) a. – L'adresse de Rosine Portinari, tu l'as pas? – Qui ça? (Thérame, 1985)

- the address of Rosine Portinari you it have NEG ? - who ça
 ‘You don’t have the address of Rosine Portinari? – Who?’
 b. « La ville de Jaufré Rudel ! » « Qui ça? » (Garat, 1984)
 the city of Jaufré Rudel ! who ça?
 « The city of Jaufré Rudel! » « Who? » (Smirnova & Abeillé, 2021, p. 249)

We tested preposition omission in two on-line acceptability rating tasks, with a 2x2 design: ± ellipsis, ± preposition, with half *à* (‘to’) and half *de* (‘of’), which are the most frequent prepositions in French.

2.1 Experiment 1

We tested *qui* (animate) remnants and PP nominal correlates (4). We had 20 target items and 20 fillers from an unrelated experiment. The participants read sentence pairs and rated the acceptability of the second sentence on a 1-5 Likert scale. 40 native speakers were recruited on Prolific and paid 1.8 GBP. The results are shown in Figures 1 and 2, and table 1 below. As expected, (4d), which is ungrammatical in French, was rated lowest. Contrary to Merchant’s theory, (4c) was rated much higher. There was no difference between ellipsis and no ellipsis in the +prep conditions but an interaction between prep and ellipsis so that -prep was rated higher with ellipsis (mean: 3.2) than without (mean: 2.1). We also found a difference between *à* and *de*, where *de* was rated higher than *à* with ellipsis, with or without preposition (Figure 2). No effect of preposition type is expected if the -prep condition is supposed to be ungrammatical.

(4) Context sentence : *J’ai parlé à un ami.* (‘I talked to a friend’)

- +prep+ellipsis: À qui? (‘To who?’)
- +prep-ellipsis: À qui as-tu parlé? (‘To who did you talk?’)
- prep+ellipsis: Qui? (‘Who?’)
- prep-ellipsis: Qui as-tu parlé? (‘Who did you talk?’)

Figure 1: Acceptability judgment results

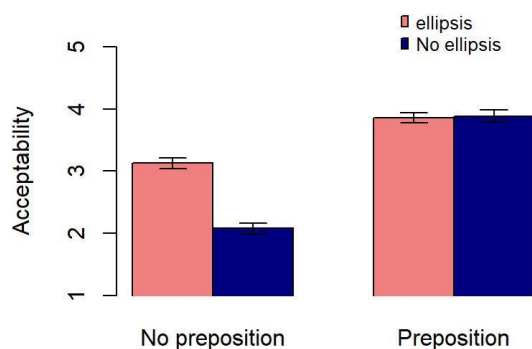


Figure 2: Rating differences between *à* and *de*

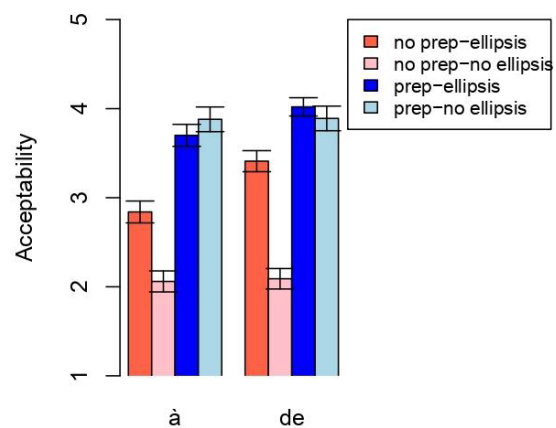


Table 1. Cumulative link mixed model results of Experiment 1

Conditions	Rating results (<i>p</i> -values)
prep / no prep	6.62e-16 ***
ellipsis / no ellipsis	0.0148 *
Interaction: prep: ellipsis	1.52e-09 ***

2.2 Experiment 2

The design is similar to experiment 1. In Experiment 2, we tested nominal remnants *quel+noun* (5). 47 native speakers were recruited on Prolific and paid 1.8 GBP. Results in Figure 3 and Table 2 show that examples like (5d), which is ungrammatical in French, were rated lowest. (5c) was rated much higher than in Experiment 1, and as high as (5b) (+prep-ellipsis), as predicted by Nykiel's hypothesis. No rating differences between *à* and *de* were detected (figure 4).

(5) Context sentence : *J'ai parlé à un ami.* ('I talked to a friend')

- a. +prep+ellipsis: *À quel ami?* ('To which friend?')
- b. +prep-ellipsis: *À quel ami as-tu parlé?* ('To which friend did you talk?')
- c. -prep+ellipsis: *Quel ami ?* ('Which friend?')
- d. -prep-ellipsis: *Quel ami as-tu parlé?* ('Which friend did you talk?')

Figure 3: Acceptability judgment results

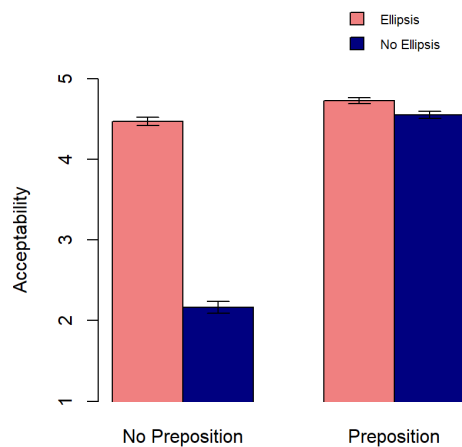


Figure 4: Rating differences between *à* and *de*

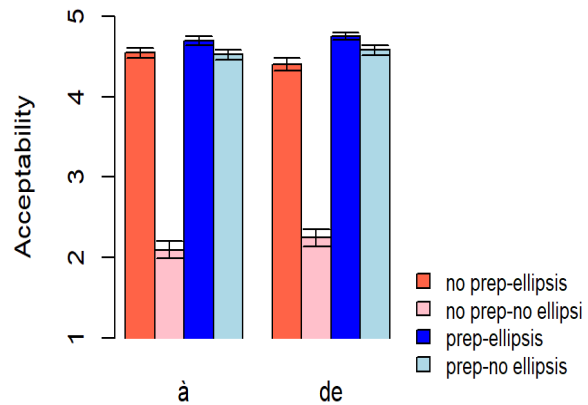


Table 2. Cumulative link mixed model results of Experiment 2

Conditions	Rating results (p -values)
prep / no prep	4.05e-12 ***
ellipsis / no ellipsis	2.67e-15 ***
Interaction: prep: ellipsis	1.24e-10 ***

3. Corpus study

In a corpus study (Frantext database, texts after 1980), we extracted 245 sluices with a PP correlate (6) (table 3). We found an overall 44.9% prep omission rate, which stands between English (67%) and Polish (18.3%). We annotated various factors: nominal/pronominal remnant/correlate, definite/indefinite correlates, preposition type and matrix/embedded sluices. We found that the remnant type plays a role: 74.8% omission with nominal remnant (*quel+noun*) vs 10.5% with a pronominal correlate (*qui, quoi* 'who, what'). We also found that matrix sluices favor prep omission with 51.3% vs. 24.1% for embedded sluices. The other factors that favored prep omission are reprise sluices, nominal correlates and stand-alone prepositions (e.g. *pour, contre, avec*, etc.).

- (6) a.— Je venais d' apprendre la mort de quelqu'un . — **De qui ?** (Gavalda, 2008)
 'I just learnt the death of someone. - Of whom ?'

- b. – Je ne sais pas. Demande à ton amie. – **Quelle amie** ? (Labruffe, 2019)
 ‘I don’t know. Ask to your friend. - Which friend ?’

Table 3. Extracted sluices from Frantext (with a PP correlate)

Sluices	Matrix	Embedded	Total
Prep + <i>qui/quoi</i>	67	9	76
∅ <i>qui/quoi</i> ?	7	9	16
Prep + <i>quel</i> + noun	24	35	59
∅ <i>quel</i> + noun	89	5	94
Total	187	58	245

4. Conclusion

We conclude that preposition omission under sluicing is acceptable in French, contrary to Merchant’s theory, and that the same factors play a role as in English and Polish (Nykiel 2014 & 2017). Preposition omission in French sluices is preferred with a nominal correlate and a nominal remnant (*quel* ‘which’), which are more informative than their pronominal counterparts. Preposition alternation in French sluicing is thus compatible with Nykiel’s theory and also in line with the Uniform Information Density hypothesis.

This is problematic for deletion based analyses based on a full syntactic interrogatives (Merchant 2001). An alternative deletion based analysis has been proposed by Rodrigues et al 2009, which suggests a it-cleft source (7a). However, such a source is problematic for additive sluicing (what else?). We do find preposition omission with additive sluices in French (7b)

- (7) a. Anne a parlé à quelqu’un mais je ne sais pas qui ~~e’est~~
 ‘Anne talked to someone, but I don’t know who it is’
 b. Cela retombe sur vous — **qui d’autre**, puisque vous êtes le seul [...] ? (Debray, 1966)
 ‘This falls back on you - who else, since you are the only one?’

Our results thus speak in favor of a fragment-based analysis with direct interpretation (Ginzburg & Sag 2000, Culicover & Jackendoff 2005, Ginzburg 2012).

Selected references

- Ginzburg, J., & Sag, I. A. (2000). *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. CSLI Publications.
 Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
 Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
 Levy, R. & Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. *NIPS’06: Advances in Neural Information Processing Systems*. Vol(19). 849-856.
 Merchant J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. OUP.
 Nykiel, J. 2017. Preposition stranding and ellipsis alternation. *English Language and Linguistics* 21(1): 27–45.
 Nykiel, J. (2013). Clefts and preposition omission under sluicing. *Lingua* 123. 74–117.
 Rodrigues, C., Nevins, A. & Vicente, L. (2009). Cleaving the interactions between sluicing and preposition stranding. In Wetzels, L., Weijer, J. van der (eds.), *Romance Languages and Linguistic Theory* (2006), pp. 175-198. John Benjamins, Amsterdam.
 Smirnova, A., Abeillé, A. (2021). The French question particles *ça* and *donc*: a corpus study, *Linguistic Research*, 38(2): 239-269.

Romanian subject alternation: the position of antecedent hypothesis

Fabian Istrate, Anne Abeillé, Barbara Hemforth

istratemariusfabian@gmail.com

Paris Graduate School of Linguistics,

Laboratoire de Linguistique Formelle, Université Paris Cité-CNRS

The *Position of Antecedent Hypothesis* (Carminati 2002, 2005) is one of the main factors accounting for preferences in languages with null / pronominal subject alternation. According to this hypothesis, null subjects favor a subject antecedent and pronominal subjects, a non-subject antecedent. We ran a first experiment in which we found that null subjects are preferred over pronominal ones in Romanian. Additionally, negative polarity slightly increased the number of pronominal subjects. Our second experiment revealed that preferences for a subject or an object antecedent for null vs pronominal subject are less categorical than in Italian, and this tendency might be linked to a less widespread use of null subjects in Romanian, compared to Italian.

1. Introduction

Like other Romance languages (except French), Romanian is a pro-drop language, allowing pronominal subject omission with any verbal form. The choice between null and pronominal subjects has been shown to be sensitive to several factors in Romance languages: word order (Mayol 2010, Catalan), person and animacy (Correa Soares et al., 2019, 2020; Fernandes et al 2018, Brazilian Portuguese), discourse status (Runner & Ibarra 2016, Spanish), function of the antecedent (Carminati 2005, Italian), etc. According to the *Position of Antecedent Hypothesis* (Carminati 2002, 2005), subject preferences come in two parts: (i) null subjects favor a subject antecedent, (ii) overt (pronominal) subjects favor a non-subject antecedent. These tendencies have been shown to play a role both in production and comprehension (anaphora resolution). Experiments carried out on other Romance languages (Spanish, Catalan, Portuguese) have shown that, whereas the first part of the hypothesis seems to be rather robust cross-linguistically (e.g. Filiaci et al. 2013, de la Fuente et al. 2016), the second part is subject to variation: e.g. in Spanish, the object preference associated with pronominal subjects is often weaker than in Italian (Sorace & Filiaci 2006, Sorace et al. 2009, Filiaci et al. 2014, Chamorro 2018, Contemori et al. 2019). In our research project, we want to find out which factors favor the use of null vs pronominal subjects in Romanian, starting with the role of contrast (realized as polarity in Experiment 1) and in how far the PAH applies to Romanian (Experiment 2).

2. Empirical evidence

2.1 Null subjects in Romanian

Following earlier work on the role of discourse status (Correa Soares et al. 2020), we looked at the role of polarity in this experiment. We assume a contrastive role of negation as for example in Farkas (2010). Thus, in an example like (1), with a negative answer, the pronominal subject may be interpreted as a contrastive topic (Krifka 2006). We consider that there may be a presupposition that the person already mentioned (Maria) represents one of several possible alternatives in the communication situation.

We ran a first binary forced-choice experiment on IlexFarm (20 experimental items, 20 distractors and 10 comprehension questions) with one experimental factor Polarity, with two values (*affirmative*, *negative*) cf. (1). 49 Romanian native speakers (6 male, 43 female, mean age=21), students from the University of Bucharest, read 20 items consisting of a polar question and a choice of two answers with either a null or pronominal subject. As prescribed by the Romanian norm, participants showed a clear preference for null subjects, across

Polarity conditions (Figure 1, Table 1). However, we also found a main effect of polarity: for a negative answer participants chose more frequently the pronominal subject than for a positive answer.

- (1) Question: *A ascultat Maria sfaturile părinților săi?*
 AUX.3SG listen.PST Maria advices.DEF parents.GEN.DEF POSS.3SG
 'Did Mary listen to her parents' advice?'
 Answer 1: *Da, (ea) ascultă mereu sfaturile părinților săi.*
 'Yes, (she) always listens to her parents' advice.'
 Answer 2: *Nu, (ea) a luat decizia pe cont propriu.*
 'No, (she) made the decision on her own.'

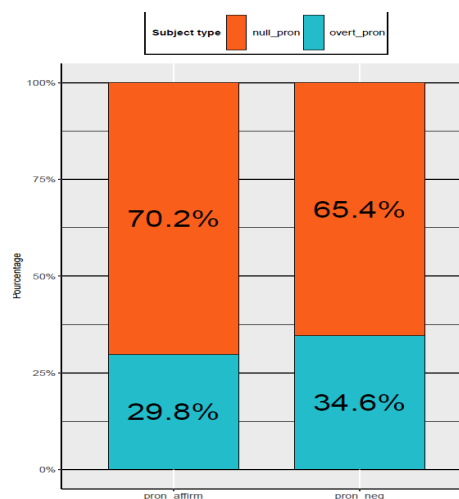


Figure 1. Results of Experiment 1

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.5240	0.4414	3.452	0.000555 ***
conditionpron_neg	-0.4327	0.1902	-2.276	0.022860 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
 (Intr)
 cndtnprn_ng -0.234

Table 1. Statistical analysis for Exp. 1

2.2 Experiment 2: The PAH in Romanian

We ran a second binary forced-choice task on IlexFarm (20 experimental items, 30 distractors, comprehension questions after 20% of the trials), by comparing antecedent choices in two conditions: *null*, *overt*, cf. (2,3). Each experimental item contained a transitive verb with two human arguments (subject and object), followed by a temporal embedded clause containing a null or pronominal subject. The task of the participants was to determine the antecedent of the embedded subject (either matrix subject or object). Half of our items contained female arguments, cf. (2), and another half male arguments, cf. (3) (no effect of gender was found in this experiment). We tested 64 Romanian native speakers (25 male, 39 female, mean age=24), all students from the University of Bucharest. As shown in Figure 2, participants mostly prefer to associate the embedded null subject with the matrix subject antecedent (64.2%), and the embedded pronominal subject with the matrix object antecedent (58.2%), confirming Carminati's hypothesis. For the statistical analysis, we used a general linear mixed model as shown in Table 2. The factor type of subject (null vs. overt) is statistically significant ($p < .001$).

- (3) *Ioana a văzut-o pe Alexandra când (ea) s-a urcat în autobuz.*
 Ioana AUX.3SG see.PST-CL.3SG.F.ACC DOM Alexandra when (she) REFL-AUX.3SG
 get.PST in bus
 'Ioana saw Alexandra when she got on the bus.'
 Question: *Cine s-a urcat în autobuz?*
 who REFL.AUX.3SG get.PST on bus
 'Who get on the bus?'
 Answer 1: *Ioana.*

Answer 2: *Alexandra*.

(3) *Matei* *l-a* *întâlnit* *pe* *Cătălin* *când* *(el)* *a* *ajuns*
 Matei CL.3SG.M.ACC-AUX.3SG meet.PST DOM Cătălin when he AUX.3SG arrive.PST
la Bușteni.

at Bușteni

'Matei met Cătălin when he arrived in Bușteni.'

Question: *Cine a ajuns în Bușteni?*

'Who arrived in Bușteni?'

Answer 1: *Matei*.

Answer 2: *Cătălin*.

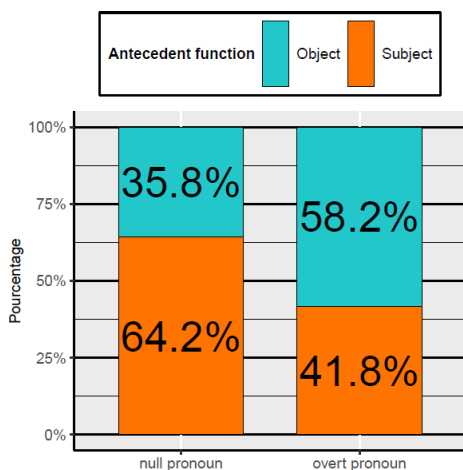


Figure 2. Results of Experiment 2

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.14638	0.14606	1.002	0.316
condition	-0.53639	0.06417	-8.358	<2e-16 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
 (Intr)
 condition -0.016

Table 2. Statistical analysis for Exp. 2

3. The Position of Antecedent Hypothesis in Romance languages

In a Romance perspective (cf. Table 3), our experiments show that (i) Romanian has a preference for null subjects (70%), (ii) a null subject favors a subject antecedent (64%), as in Spanish and Catalan (though not categorically as in Italian or European Portuguese), and (iii) a pronominal subject favors an object antecedent (58%), as in Spanish (Filiaci et al. 2013, de la Fuente et al. 2013) and Catalan (Mayol 2010). In contrast to this group of languages (Romanian, Spanish, Catalan), Italian and European Portuguese are characterized by a stronger tendency to interpret null subjects as referring to a subject antecedent and a stronger tendency to link the pronominal subjects to an object antecedent (about 80%). We thus observe some variation across Romance languages in the preference for subject/object, since Italian is characterized by a stronger tendency to link the pronominal subject to an object antecedent (about 80%), while in Romanian and Catalan, for instance, preferences seem to be less categorical.

Language	Subject/ Object + null subject	Subject/ Object + overt subject
Italian	76.8 % / 23.1%	17% / 82.9%
European Portuguese	79 % / 21%	24 % / 76%
Spanish	66.3 % / 36.8%	33.6 % / 63.1%
Catalan	59.1 % / 40.9%	35.2 % / 64.8%
Romanian	64.2 % / 35.8%	41.8 % / 58.2 %

Table 3. Romanian & other Romance languages

4. Conclusion

We conclude that subject alternation in Romanian is sensitive to the antecedent function as in other Romance languages. As in Spanish and Catalan, the first part of Carminati's hypothesis is more robust than the second part. Although our results will have to be tested in a larger scale fully parallel crosslinguistic study, we assume that the less categorical choice between an object or a subject antecedent for a pronominal subject might be related to the generally weaker preference for null subjects in Romanian compared to Italian or European Portuguese (Fernandes et al. 2018). As suggested by the first experiment, the choice of a null or pronominal subject is also sensitive to semantic features like polarity, an issue which has to be further investigated. The baseline of null subjects in Romanian (70%), compared to Italian (77%, see Torregrossa et al. 2020) and Catalan (62%, see Casanova 1999) is likely to trigger gradual preferences when associating the type of subject with its antecedent (subject or object): the higher the percentage of the null subject baseline, the stronger we expect the preferences to link a null subject to subject antecedent and the pronominal subject to a non-subject antecedent to be. Ongoing research (corpus analyses and experiments) extends our work on the constraints for the usage of null vs pronominal subjects and their role in pronoun resolution.

Selected references

- Carminati, M. N., 2005, Processing reflexes of the Feature Hierarchy (Person > Number > Gender) and implications for linguistic theory, *Lingua*, 115, 259-285.
- Correa Soares, E., Miller, P., & Hemforth, B. (2019). The effect of verbal agreement marking on the use of null and overt subjects: a quantitative study of first person singular in Brazilian Portuguese. *Fórum Linguístico* 16, no. 1 (2019): 3579-3600. <https://periodicos.ufsc.br/index.php/forum/article/view/1984-8412.2019v16n1p3579>.
- Correa Soares, E., Miller, P., & Hemforth, B. (2020). The Effect of Semantic and Discourse Features on the Use of Null and Overt Subjects - A Quantitative Study of Third Person Subjects in Brazilian Portuguese. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 36(1), 2020360107. Epub July 06, 2020. <https://dx.doi.org/10.1590/1678-460x2020360107>
- de la Fuente, I. & Hemforth, B., 2013, *Effects of Clefting and Left-Dislocation on Subject and Object Pronoun Resolution in Spanish*, in *Proceedings of the 16th Hispanic Linguistics Symposium*, 27-45. Cascadilla Proceedings Project.
- de la Fuente, I., Hemforth, B., Colonna, S., & Schimke, S. (2016). The role of syntax, semantics, and pragmatics in pronoun resolution: A cross-linguistic overview. In *Empirical perspectives on anaphor resolution*, 11-32. Anke Holler, Katja Suckow (eds). Berlin/Boston: De Gruyter.
- Farkas, D., 2010, The grammar of polarity particles in Romanian, in A. M. Di Sciullo & V. Hill (eds.), *Edges, Heads, and Projections: Interface properties*, John Benjamins, 87-124.
- Fernandes, E. G., Luegi, P., Correa Soares, E., de la Fuente, I. & Hemforth, B., 2018, *Adaptation in pronoun resolution: Evidence from Brazilian and European Portuguese. Learning, memory, and cognition*, 44(12), 1986–2008.
- Filiaci, F., Sorace, A. & Carreiras, M., 2013, *Anaphoric biases of null and overt subjects in Italian and Spanish : a cross-linguistic comparison*, Language and Cognitive Processes.
- Mayol, L., 2010, *Refining Saliency and the Position of Antecedent Hypothesis: a Study of Catalan Pronouns*, Proceedings of the 33rd Annual Penn Linguistics Colloquium.

Dependency formation during real-time processing: Evidence from webcam-based eye-tracking with subjective and objective adjectives

Elsi Kaiser & Jesse Storbeck, University of Southern California

We report an eye-gaze-tracking study showing that the idea of certain expressions triggering dependency formation/retrieval of previously-mentioned referents can, at least in certain contexts, be extended to a new domain: adjectives. Our study, testing subjective and objective adjectives, suggests under certain communicative circumstances, even adjectives can trigger looks to previously-mentioned referents. In addition to providing novel evidence that predicates of personal taste can trigger activation of a judge referent, our findings suggest that referential dependencies, broadly construed, may be more widespread than previously thought. Moreover, source-of-information and adjective-type (subjective / objective) effects occur rapidly during comprehension, indicating that these dependency-building are constrained in semantically-principled ways.

1 Introduction

We report a webcam eye-gaze-tracking study showing that the idea of certain linguistic expressions triggering dependency formation/retrieval of previously-mentioned referents can – at least in certain communicative contexts – be extended to a new domain: adjectives.

In a sentence like “*Ben told Jon that the meal was {delicious/vegetarian}*,” who thinks that the meal was delicious/vegetarian? You would probably say Ben. In this communicative event, the **source-of-information** (the agent of *told*) is naturally interpreted as being the one who thinks the meal was delicious/vegetarian (see also Kaiser 2020, 2022). This intuition brings up the possibility of **source retrieval effects** during real-time processing; could comprehenders mentally retrieve/reactivate the source-of-information referent (Ben) when they process the embedded clause about the meal being delicious or vegetarian, even though Ben is not explicitly mentioned in it?

Furthermore, it’s important to acknowledge that the kind of information communicated by source referents can vary: Saying that a meal is delicious expresses the **source’s own subjective taste/opinion**, using a predicate of personal taste (PPT, e.g. Lasersohn 2005). But in saying that a meal is vegetarian, the source is stating a more **objective fact**. (Although one could disagree about what counts as *vegetarian*, this kind of debate involves matters of definition. In contrast, debates about what is *delicious* are matters of subjective taste.)

1.1 Subjective vs. objective

The distinction between subjective adjectives (e.g. *delicious*, *scary*) and objective adjectives (e.g. *vegetarian*, *wooden*) is well-established in theoretical semantics and philosophy (e.g. Lasersohn 2005, McNally & Stojanovic 1997, Bylinina 2014 and many, many others). Semanticists have proposed that subjective adjectives have as part of their meaning a special ‘**judge**’ argument/parameter that identifies *the person whose opinion/judgment the adjective expresses* (Ben in (1), e.g. Lasersohn 2005). On this view, subjective adjectives are always linked to a judge, unlike objective adjectives. If this view is on the right track, and if the notion of the judge is at play during real-time processing, it raises the possibility that, when a person encounters a subjective adjective, they mentally retrieve the judge referent linked to that adjective (**judge retrieval effects**).

(1a) Ben: “The meal is delicious” (1b) Ben: “The meal is delicious_{BEN}” (*delicious to Ben*)

So far, to the best of our knowledge, there have been no systematic experimental investigations regarding the questions of (a) whether source retrieval effects occur during real-time processing and, more specifically relevant for semantic theories of subjective adjectives,

(b) whether judge retrieval effects can be detected during the real-time processing of predicates of personal taste. Our experiment aims to shed light on these issues.

2 Experiment

We used webcam-based gaze-tracking to test for evidence of comprehenders retrieving previously-mentioned source referents or judge referents during real-time processing of adjectival expressions that do not explicitly mention the source or the judge.

2.1 Participants, procedure

In a visual-world set-up with auditory stimuli presented alongside images, people (117 native U.S.-English speakers) heard sentences like ex.(2) while seeing displays like Fig.1a, as eye-gaze was tracked (with PCibex, Zehr & Schwarz 2018, Webgazer.js, Papoutsaki et al. 2016).

2.2 Design

We manipulated adjective type (subjective vs. objective). We also used voice (active vs. passive) to manipulate whether the subject or the object is the source-of-information (2x2, 20 targets), as illustrated in the examples in (2). Participants' eyegaze was tracked during the critical display (Fig.1a), as they heard the critical sentences (ex.2) with a variety of subjective and objective adjectives.

(2) Example (auditory, presented as eye-gaze was tracked. Items were 50/50 male/female):

(a) **Ben**_{SOURCE} told **Jon**_{PERCEIVER} that the meal at the event was {delicious/vegetarian}

(b) **Ben**_{PERCEIVER} was told by **Jon**_{SOURCE} that the meal at the event was {delicious/vegetarian}

(3) Question (written): Who thought that the meal was {delicious/vegetarian}?

In the critical displays, the L/R positions of the subject and the object, as well as the pairing of names and pictures, were randomized. The central image of the thing being talked about, e.g. *meal*, appeared on the screen when the determiner (*the*) started in the audio and remained on-screen until adjective onset, e.g. during the time when people heard *the meal at the event was*. After this time, the central image disappeared. Thus, image of the thing being talked about, in this example the meal, was not on the screen when participants heard the critical adjective. The appearing-and-disappearing was done in order to attract participants' eyegaze to a central, neutral location, equidistant from the two characters.

Fig.1a. Example of critical display

(The central object was no longer visible when the adjective started, i.e., it was not visible during the time window when people's eyegaze was analyzed).

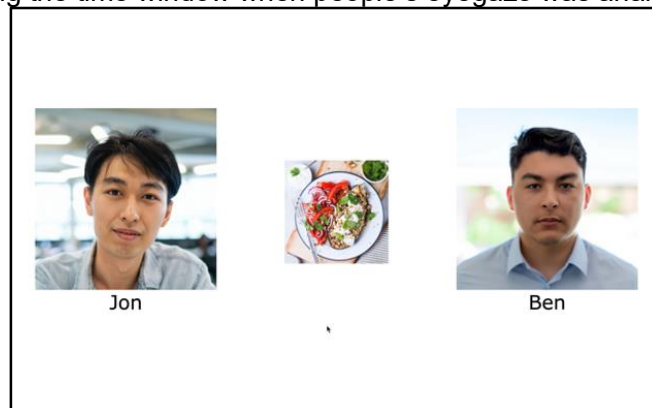


Fig.1b Example ‘who thought’- question display. (Participants responded by clicking on the relevant person; the question was shown in writing on the screen)



After each trial, people saw a ‘who thought’ question on the next screen (ex.3, Fig.1b) and clicked the relevant person. These questions ensure participants pay attention to the task, and also render the source-of-information relevant for the task at hand. There were included to ensure that participants do not default to shallow processing. It’s important to note that the subjective and objective adjective conditions do not differ in this regard: Both were followed by a ‘who thought’ question. Thus, any differences between subjective and objective adjectives cannot be attributed to the question, as it is present for both.

As expected based on Kaiser (2020, 2022), participants’ click responses to the ‘who thought’ questions yield the expected source bias, >90% in all conditions. This also serves as a sanity check to confirm participants are paying attention.

2.3 Predictions

First, if **source retrieval effects** occur during online processing of (subjective and objective) adjectives, especially when source retrieval is encouraged by the ‘who thought’-questions, both subjective and objective conditions should elicit *more looks to the source than the perceiver* when either a subjective or objective adjective is encountered. Second, if **judge retrieval effects** occur in real-time (if presence of a judge argument triggers retrieval of the judge referent), subjective adjectives should elicit *even more looks to the source* than objective adjectives, since it is only with subjective adjectives that the source is also a judge. (Note that semantically, objective adjectives are analyzed as lacking a judge argument, so semantic theories do not predict objective adjectives to trigger judge retrieval effects.)

3 Results and discussion

Overall, the results provide evidence for both source retrieval and judge retrieval effects, providing novel data for online processing of these kinds of semantic dependencies:

Source-advantage scores (source minus perceiver looks) are plotted in Fig.2. The turquoise lines show source-advantage scores for the subjective adjective conditions, the reddish orange lines show the source-advantage scores for the objective adjective conditions. The y-axis shows *source-advantage scores* (looks to source minus looks to perceiver). The x-axis shows *time*: 0 ms is the onset of the adjective; the source and perceiver images remained on-screen for 3 s after sentence offset (which is when the trial ended).

As can be seen in Fig.2, soon after adjective onset (0ms, dotted line), the source advantage scores increase steeply with both subjective and objective adjectives, in both active and passive conditions (more looks to source than perceiver; positive numbers on the y-axis). This suggests that adjectives can trigger **retrieval of the source**. Given the presence of the ‘who thought’ questions, we view these source-retrieval effects as expected: They provide a sanity check that participants are attending to the task at hand and show that source activation can occur rapidly after participants hear the relevant adjective.

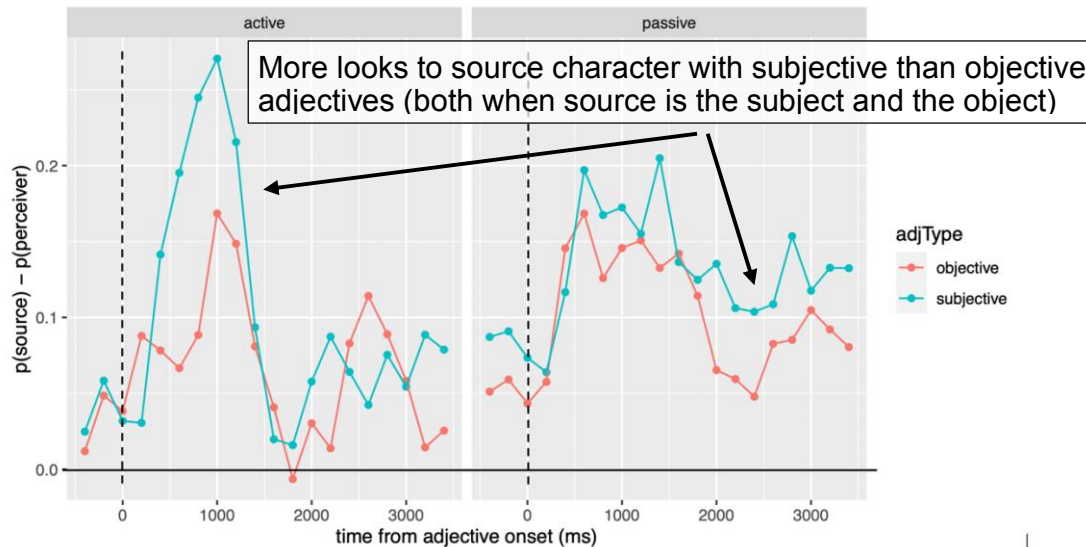
Crucially, we also find an effect of *adjective type*: From adjective onset until the end of the

trial, subjective adjectives trigger more looks to the source character than objective adjectives ($t=2.15$, $p=0.033$). This is exactly what is predicted under the view that only subjective adjectives trigger **retrieval of the judge**, due to their special semantic status (judge argument/parameter) that is not shared by objective adjectives.

We also observe a marginal effect of voice ($t=1.86$, $p=0.068$), suggesting subjects may be more easily retrieved than objects, a finding that fits with prior work indicating that subjects are more prominent/salient – but no interaction between voice and adjective type ($p>.9$).

Time permitting, we will also discuss results from another study that did not include ‘who thought’ questions and thus explores how automatic the source and judge retrieval effects are.

Fig.2: Eye-gaze patterns during the critical sentence
(left panel: active voice; right panel: passive voice)



4 Conclusions

As a whole, our results suggest that in the right context, even adjectives (not typically viewed as ‘anaphoric’) can trigger looks to previously-mentioned referents. This suggests referential dependencies, very broadly construed, may be more widespread than often thought. Moreover, gaze patterns reveal that effects of source-of-information and adjective type occur rapidly during comprehension, and show that the dependency-building/retrieval processes triggered by adjectives are incrementally constrained in semantically principled ways.

References

- Bylinina, L. (2014) *The grammar of standards: Judge-dependence, purpose-relativity, and comparison classes in degree constructions*. LOT Dissertation Series 347. Utrecht: LOT.
- Kaiser, E. (2020). Shifty behavior: Investigating predicates of personal taste and perspectival anaphors. In *Semantics and Linguistic Theory 30 (SALT 30)*, 821–842. Ithaca, NY: CLC. DOI: <https://doi.org/10.3765/salt.v30i0.4850>
- Kaiser, E., (2022) Do perspective-sensitive anaphors and subjective adjectives exhibit perspectival uniformity? An experimental investigation. *Glossa: a journal of general linguistics* 7(1). doi: <https://doi.org/10.16995/glossa.5768>
- Lasersohn, P. (2005) Context dependence, disagreement, and predicates of personal taste. *Linguistics & Philosophy*, 28, 643-686.
- McNally, L. & Stojanovic, I. (2017). Aesthetic adjectives. In James O. Young (ed.), *The Semantics of Aesthetic Judgment*, 17-37. Oxford University Press.
- Papoutsaki, A. et al. (2016). *Webgazer: Scalable webcam eye tracking using user interactions*. In Proceedings of the 25th IJCAI, 3839-3845. ACM.
- Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832>

Referring to someone using only their last name: Insights from gender-marked pronouns

Elsi Kaiser, Deborah Ho, Haley Hsu, Claire Post & Madeline Rouse
University of Southern California

In some contexts, last-name-only format is used for people (e.g. *Hedberg came in*). At least in U.S. English, men are more likely to be referred to by last-name-only format than women (male bias, e.g. McConnell-Ginet 2003). Moreover, researchers referred to with last-name-only are judged more famous/ eminent (eminence bias, Atir & Ferguson 2018). However, the robustness of these biases is not yet well-understood, nor how they interact with other semantic biases. We show that these biases persist in informationally-impoverished contexts, the male bias persists when even pitted against verbs' implicit-causality biases, and these biases persist even when use of the last-name-only format for women is primed.

1 Introduction

How we refer to someone matters. Whether a speaker chooses refer to someone as *Sofiana*, *Sofiana Romano*, *Professor Romano* or *Romano* is influenced by numerous factors. The speaker's choices also influence others' impressions. The present work focuses on the phenomenon of referring to people with last-name only (e.g. *Hedberg came in*, *Ramirez was promoted*). Further examples are in (1a-c).

- (1a) I would go so far as to say that had Watson and Crick not come into Rosalind's photograph -- by hook or crook; whichever way it was -- they would have lost the race entirely (*from a podcast by the Scientific American on Dr. Rosalind Franklin*)
- (1b) Johnson is a great professor. He is funny (*from ratemyprofessor.com*)
- (1c) Welsh is my favorite professor. She's just amazing (*from reddit.com*)

Although last-name-only does not carry an explicit marker of male gender (and does not have male *phi*-features, in formal terms) and can also be used for women (ex.1c), it has been shown that, at least in U.S. English, in multiple contexts (e.g. politics, academia, sports, science, even informal conversation), men are more likely to be referred to by last-name-only than women (**male bias**, e.g. McConnell-Ginet 2003, Atir & Ferguson 2018, Gardner & Brown-Schmidt 2019, 2020). Moreover, referring to a researcher by last-name only results in them being judged more famous, more eminent, higher status and more deserving of awards (**eminence bias**, Atir & Ferguson 2018). Thus, using last-name-only more for men than for women is not without consequences.

However, Atir & Ferguson (2018) focused on naturalistic communication and other rich contexts where people knew a lot about each referent. For example, they used names of famous people or provided participants with information about scientists and their research. Furthermore, they did not specifically control the linguistic properties of the sentences mentioning the referents, which leaves open the possibility of factors such as topicality or salience playing a role. Thus, it is not yet clear whether the last-name only format on its own is robust enough to elicit a male bias or an eminence bias in more linguistically-rigorously controlled settings, and how these biases interact with other semantic interpretation biases.

1.1 Aims of this work

Our work has three main aims. The first aim is to test whether these two semantic biases – the male bias and the eminence bias – associated with the last-name-only form are sufficiently robust and sufficiently strongly linked to this particular linguistic form that they emerge even in linguistically-controlled, informationally-impoverished contexts where speakers lack rich

mental representations of the referents. I.e., do we still make assumptions about gender when the only cue is the use of the last-name only format?

The second aim is to test we also test whether the male bias of the last-namely-only format is strong enough to persist even when *pitted against a different and well-established semantic bias* (verbs' implicit causality (IC) biases).

Finally, we test whether the male bias persists even if participants are primed beforehand with text using last-name-only format for female referents.

2 Experiment 1

Experiment 1 (20 targets, 22 fillers, 91 native U.S.-English speakers) used a sentence-completion task. Participants read sentence fragments ending in 'because + pronoun' and wrote continuations. We manipulated three factors, as illustrated in ex.(2-3):

The **first factor** is the verb's implicit causality (IC) bias: When followed by an explanation relation signaled by *because*, does the verb elicit mostly subject (IC1) or object (IC2) continuations? The existence of implicit causality biases are well-established in prior work. The verbs were selected based on norms reported by Hartshorne & Snedeker (2013). The **second factor** is whether the pronoun prompt is *he* or *she*. The **third factor** is whether the verb is *eminent* (presents the IC biased referent in a positive light, e.g. IC1: *impressed*, IC2: *promoted*) or *noneminent* (presents the referent in a negative light, e.g. IC1: *disappointed*, IC2: *despised*).

Targets contained one first-name referent and one last-name referent. The last name was always in the position favored by verbs' IC bias (subject/IC1 verbs, object/IC2 verbs). This was done to pit verb bias and male-bias against each other. Examples are in (2-3) below.

(2) IC1 verbs (subject-biased)

- | | |
|---|--------------------------|
| (a) Smith impressed Eric because he | [he + eminent verb] |
| (b) Smith impressed Amanda because she | [she + eminent verb] |
| (c) Smith disappointed Eric because he | [he + non-eminent verb] |
| (d) Smith disappointed Amanda because she | [she + non-eminent verb] |

(3) IC2 verbs (object-biased)

- | | |
|--|--------------------------|
| (a) Frank promoted Mayfield because he | [he + eminent verb] |
| (b) Claire promoted Mayfield because she | [she + eminent verb] |
| (c) Frank despised Mayfield because he | [he + non-eminent verb] |
| (d) Claire despised Mayfield because she | [she + non-eminent verb] |

3 Results of Experiment 1

Data were double-coded by coders blind to condition and with gender cues removed, to ensure any potential biases that the coders might have do not distort the date.

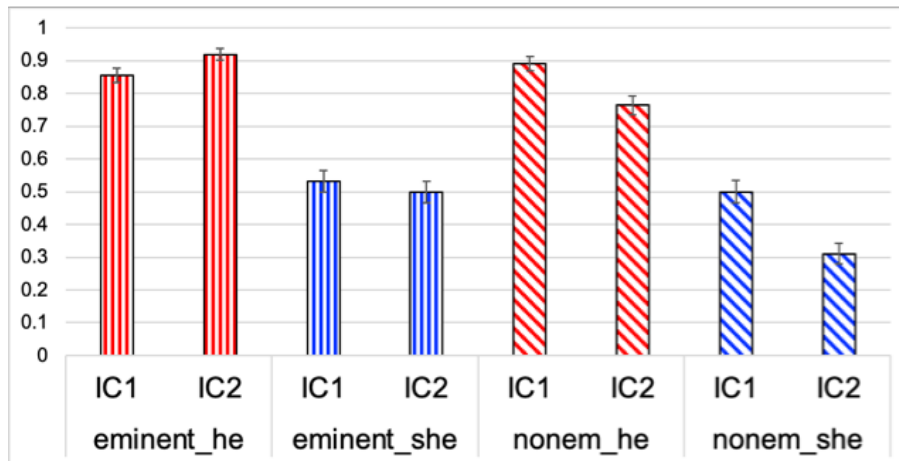
Fig.1 shows the *proportion of verb-bias-compatible continuations* – i.e., how often participants use the pronoun to refer to the referent favored by the verb's implicit causality (IC) bias (subject of IC1 verbs, object of IC2 verbs). Recall that this referent is realized using last-name-only. Thus, if verb bias is all that matters, all conditions should show very high bars, i.e. high rates of verb-bias compatible continuations.

Indeed, this is what we see in the **he conditions**: When the prompt pronoun is *he*, all conditions show clear verb bias effects. Both eminent and non-eminent IC1 verbs elicit mostly subject continuations ($p's < .001$), and eminent and non-eminent IC2 verbs elicit mostly object continuations ($p's < .001$). In sum, in the *he* conditions we observe the expected IC patterns familiar from prior work on implicit causality.

Strikingly, none of the **she conditions** show a rate of verb-bias-compatible continuations above chance. The rate of verb-bias-compatible continuations does not differ from chance with eminent IC1 and IC2 verbs or with noneminent IC1 verbs, and is in fact below chance with noneminent IC2 verbs ($p < .05$).

In other words, with *she*, even when the verb's semantic IC bias pushes towards the last-name-only referent (e.g. 'Smith' in *Smith impressed Amanda because she...*), participants are 'reluctant' to interpret that referent as the antecedent of the pronoun *she* – despite the verb biasing it. This reveals a *dispreference for interpreting a last-name-only form as referring to a female referent*. These patterns show up in the responses of male and female participants.

Fig.1 Experiment 1: Proportion of continuations that are compatible with the verb's implicit causality bias (subject with IC1 verbs, object with IC2 verbs).



Further analyses show that with IC2 verbs, verb-bias-compatible continuations are less frequent with non-eminent objects with both *he* and *she* ($p < .002$, $p < .02$). We attribute this to the **eminence bias**: If last-name-only style is associated with eminence, participants could be reluctant to provide explanations of why a last-name-referent would be being *criticized*, *despised*, *distrusted* etc. This pattern obtains with both *he* and *she*, indicating that the eminence effect is at play also when the pronoun signals the referent is female. In contrast to IC2 verbs, IC1 verbs show no effects of (non)eminence; perhaps they are masked by the greater overall prominence of subjects. This an important direction for future work.

4 Experiment 2

Experiment 2 was the same as Experiment 1 (92 new participants), but now we primed people beforehand with five paragraphs about famous female scientists that used last-name-only reference for female referents (ex.4), to see if priming/exposure could boost likelihood to interpret last-name-only referents as female. Participants were asked to read the five female scientist paragraphs at the start of the study and to answer comprehension questions about them, to ensure they paid attention. Then, participants did the same sentence-completion task, with the same stimuli, as in Experiment 1.

(4) Example of female scientist paragraph from Experiment 2

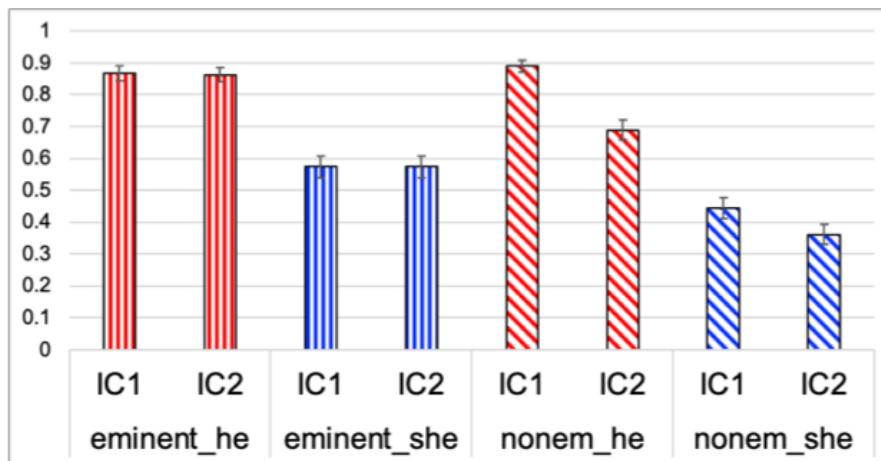
Dr. Rosalind Franklin was an English chemist who lived from 1920 to 1958. An expert in x-ray crystallography, she made groundbreaking contributions to the study of genetics, in particular the molecular structure of DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). Today, Franklin is perhaps best known for the x-ray diffraction photographs that she and a graduate student, Raymond Gosling, took of DNA fibers. These include the famous "photo 51," showing the three-dimensional structure of DNA. However, Franklin is sometimes called the "dark lady of DNA" because her important role in the discovery of a second type of DNA and its remarkable double helix structure went largely unrecognized for decades. Nevertheless, Franklin's pioneering work on the molecular structure of coal and viruses was already appreciated during her lifetime. In fact, Franklin's discoveries helped the Allies use more fuel-

efficient coal during World War II. Franklin died of ovarian cancer at age 37, but her team continued her research which eventually won the Nobel Prize in Chemistry in 1982.

5 Results of Experiment 2

As shown in Fig.2, the results largely replicate Experiment 1. There is no strong indications of the male bias (or the eminence bias) weakening in the presence of the priming manipulation.

Fig.2 Experiment 2 with priming: Proportion of continuations that are compatible with the verb's implicit causality bias (subject with IC1 verbs, object with IC2 verbs).



6 Conclusions

The dependency-building elicited by gender-marked pronouns shows that referring to someone by last-name only triggers strong semantic inferences in comprehenders' minds, at least in the U.S. English context. Our results show that (i) the male and eminence biases persist even in informationally-impooverished contexts where the prior linguistic context is controlled to avoid potential confounds, (ii) the male bias persists even when pitted against established verb IC biases, and (iii) the male bias and the eminence bias do not seem to be affected by a priming manipulation seeking to boost the likelihood of last-name only being used for women. These results suggest that even when participants have no other information about someone, simply seeing that person referred to by means of (the in-principle gender-ambiguous) last-name-only format is still enough to trigger an assumption that the referent is male and eminent. Put together, these results highlight an under-researched aspect of pronoun interpretation that can have societal consequences.

References

- Atir, S., & Ferguson, M. J. (2018). How gender determines the way we speak about professionals. *PNAS Proceedings* 115(28), 7278-7283
- Gardner, B. & Brown-Schmidt, S. (2019). "Call me by my last name": Gender bias through production about and memory for names. Poster presented at the 60th Annual Meeting of the Psychonomic Society. Montreal, Canada.
- Gardner, B. & Brown-Schmidt, S. (2020). "No gender mentioned but I'd say male": Gender bias through production about news stories. Poster presented at the 33rd Annual CUNY Conference on Human Sentence Processing.
- Hartshorne, J. & Snedeker, J. (2013). Verb Argument Structure Predicts Implicit Causality, *Language Cognition and Neuroscience* 28, 1474-1508.
- McConnell-Ginet, S. (2003). "What's in a Name?" Social Labeling and Gender Practices. *Handbook of Language and Gender*, Wiley. <https://doi.org/10.1002/9780470756942.ch3>

Extraction asymmetries in complex participle adjuncts

Andreas Kehl, University of Tübingen

In this talk, I will argue that extraction asymmetries observed in the literature for participle adjuncts do not need to be explained by additional syntactic or semantic licensing principles. The factors identified in the literature apply equally to the corresponding declarative structures. I will present experimental evidence indicating that factors like aspectual class and verb type show effects which are independent of whether the sentence appears in declarative or interrogative form. These results are in line with other approaches to islands which call into question the relevance of grammatical licensing principles.

1 Extraction from participle adjuncts

Long-distance dependencies such as *wh*-extraction into adjunct constituents are traditionally considered to be impossible in the generative framework, as stipulated in the *Condition on Extraction Domain* (CED, Huang, 1982). The existence of apparently grammatical extractions from some adjunct constituents, also in attested examples, has given rise to a rich literature on how these extractions can exceptionally be licensed in the grammar and what distinguishes good extraction candidates from bad ones.

I will argue that the grammatical licensing mechanisms suggested in the literature, whether they are formulated in syntactic or semantic terms, add complexity to the grammar that are unwarranted in the light of experimental evidence which suggests that the distinction between good and bad extraction candidates is independent of the application of extraction and also leads to acceptability differences in the declarative counterparts. This line of research ties in with the more general impetus in the discussion of island phenomena which has moved away from grammatical licensing principles towards accounts based on processing complexity and discourse factors that are independent of extraction (Abeillé et al., 2020; Chaves & King, 2019; Chaves & Putnam, 2020; Culicover, Varaschin, & Winkler, 2022; Culicover & Winkler, 2018, 2022; Hofmeister & Sag, 2010; Liu et al., 2022).

It has to be pointed out first that adjuncts come in a variety of shapes that seem to react differently to extraction: tensed adverbial clauses appear to resist extraction stronger than untensed adjunct clauses and are thus typically given in the literature as evidence for the validity of the CED. However, other data patterns observed in the literature suggest that there are additional extraction asymmetries within certain adjunct types that further complicate the empirical landscape because some extractions are considered to show a relatively high degree of acceptability whereas minimally different sentences do not. Truswell (2007), for example, observes that participial adjuncts modifying atelic main verbs like *work* in (1a) result in ungrammatical extractions, whereas extraction is fine if the main verb is telic like *arrive* in (1b):

- (1) a. *What_i did John work [whistling _i]?
b. What_i did John arrive [whistling _i]?

The declarative counterparts of these sentences in (2) are both considered grammatical, so that the grammaticality difference in interrogatives is a result of extraction:

- (2) a. John worked whistling a funny song.
b. John arrived whistling a funny song.

There is a considerable number of approaches which explain the judgment differences in (1) by additional licensing principles that permit extraction in some cases but not others. In the case of Truswell (2007), extraction is only licensed if event-semantic conditions are fulfilled.

Depending on whether these event-semantic properties are encoded structurally (as for example in Borer, 2005 or Ramchand, 2008) or not, the application of locality operations needs to be enriched to include sensitivity to such properties.

There is a growing body of research into islands and related phenomena that seeks to provide alternative explanations for the observed judgment patterns. Particularly for the adjunct types discussed here, Brown (2017) proposes that acceptability differences in interrogative and declarative structures are related, meaning that the effects of telicity and transitivity determine acceptability in the presence and, crucially, also in the absence of extraction. I will argue that the observation in Brown (2017) about the relation between acceptability in declarative and interrogative participle adjuncts is on the right track and that additional licensing principles as in Truswell (2007) are not necessary; instead, the same acceptability contrasts reported for interrogatives appear in equal strength in the corresponding declaratives.

2 Experimental evidence

2.1 Method and design

To test the claim in Truswell (2007) that the aspectual class of the matrix predicate has an influence on the strength of extraction, an acceptability judgment experiment was conducted. The 2x2 factorial design, shown in (3), crosses the two factors TELICITY and STRUCTURE.

- | | | |
|-----|---|--------------|
| (3) | a. Dennis came home eating ice cream. | [telic/-wh] |
| | b. What did Dennis come home eating _ ? | [telic/+wh] |
| | c. Dennis walked around eating ice cream. | [atelic/-wh] |
| | d. What did Dennis walk around eating _ ? | [atelic/+wh] |

This allows to determine whether the aspectual class of the matrix predicate influences how strongly the application of extraction degrades the two declarative conditions. The research question was whether conditions with atelic matrix predicates are affected to a stronger degree by extraction than conditions with telic matrix predicates.

2.2 Participants and materials

Four lexicalizations of the four conditions were created, resulting in a total of 16 target items. They were distributed across four lists according to the Latin square design. 48 participants were recruited via Mechanical Turk and judged the 16 target items on a 7-point Likert scale. The experiment included 32 distractors with varying degrees of naturalness, including the 15 cardinal well-formedness sentences from Gerbrich, Schreier, and Featherston (2019) to anchor the scale. All items were randomized for each individual participant to avoid list position effects. Four participants were excluded from the statistical analysis because they indicated a native language other than English, another participant because of a technical problem with the experimental platform, and another four participants due to poor performance on the cardinal well-formedness sentences. This left 39 participants that were included in the statistical analysis.

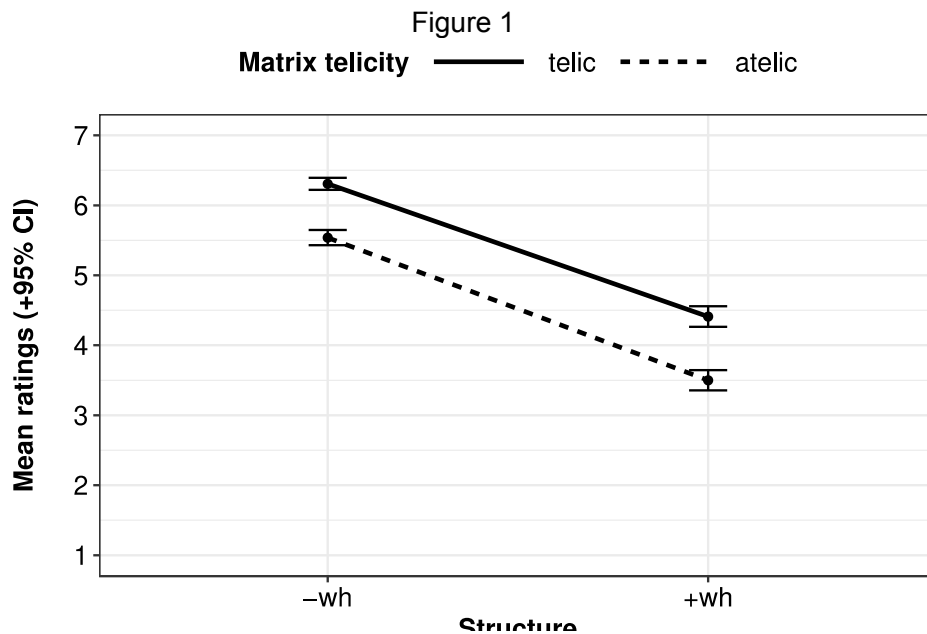
2.3 Results and analysis

The results are shown graphically in Figure 1 with untransformed mean ratings. It is clearly visible that interrogative conditions are judged less acceptable than declaratives, and that telic matrix predicates have an advantage over atelic ones; this advantage exists in both declarative and interrogative conditions.

A linear mixed effects model was fit over the data using the afex package (Singmann et al., 2020) in the R programming language (R Core Team, 2020, version 4.0.2); the model includes TELICITY and STRUCTURE plus their interaction as fixed effects as well as participants and items as random effects. Following Barr et al. (2013), the model includes the maximal random effect structure that led to model convergence in order to include as many sources of variability as possible (see Singmann & Kellen, 2020; Winter, 2020). The results show significant effects for TELICITY ($\beta = .413$, $t = 3.712$, $p < .01$) and STRUCTURE ($\beta =$

.995, $t = 9.578$, $p < .001$), but the interaction is not significant ($\beta = .037$, $t = .327$, $p = .71567$). This means that the effect of TELICITY exists independently of extraction. This is not as predicted in Truswell (2007); Truswell is right about the advantage of telic matrix predicates over atelic ones, but this effect is not related to whether extraction has taken place or not.

Similar results are also obtained for the distinction between unaccusative, unergative, and transitive matrix predicates discussed in Borgonovo and Neeleman (2000). Again, the effect resulting from the verb type of the matrix predicate does not interact with that of extraction.



3 Discussion and conclusion

The experimental results confirm the hypothesis that the aspectual class of the matrix predicate has an effect that is independent of extraction. Truswell's (2007) judgment pattern in (1) and (2) may be the result of translating the gradient acceptability seen in the experiment to a binary distinction between grammatical and ungrammatical structures. Since both declarative conditions in the experiment show a fairly high degree of acceptability, they are both mapped to a grammatical judgment; as far as the interrogative conditions are concerned, a different mapping of the two conditions to grammatical for telic matrix predicates and to ungrammatical for atelic ones depends on where the cut-off line for the binary distinction is drawn.

The effect of telicity can be explained in terms of differences in the semantic compatibility between the two predicates, which is observable regardless of whether extraction has taken place or not. Relative acceptability in declaratives serves as a reliable predictor of acceptability differences in interrogatives. There is thus no need for a licensing principle that permits extraction if the matrix predicate is telic but not if it is atelic. These experimental results show that the individual differences between types of matrix predicate deserve closer attention independently of extraction, which can be captured by complexity differences as proposed in Culicover, Varaschin, and Winkler (2022).

References

- Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, 204, Article 104293. <https://doi.org/10.1016/j.cognition.2020.104293>

- 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.
- Borer, H. (2005). *The normal course of events*. Oxford University Press.
- Borgonovo, C., & Neeleman, A. (2000). Transparent adjuncts. *Canadian Journal of Linguistics/La Revue Canadienne de Linguistique*, 45(3/4), 199–224.
- Brown, J. M. M. (2017). *Heads and adjuncts: An experimental study of subextraction from participials and coordination in English, German and Norwegian* [Doctoral dissertation, University of Cambridge].
- Chaves, R. P., & King, A. (2019). A usage-based account of subextraction effects. *Cognitive Linguistics*, 30(4), 719–750.
- Chaves, R. P., & Putnam, M. T. (2020). *Unbounded dependency constructions: Theoretical and experimental perspectives*. Oxford University Press.
- Culicover, P. W., Varaschin, G., & Winkler, S. (2022). The radical unacceptability hypothesis: Accounting for unacceptability without universal constraints. *Languages*, 7(2), Article 96. <https://doi.org/10.3390/languages7020096>
- Culicover, P. W., & Winkler, S. (2018). Freezing: Between grammar and processing. In J. M. Hartmann, M. Jäger, A. Kehl, A. Konietzko, & S. Winkler (Eds.), *Freezing: Theoretical approaches and empirical domains* (pp. 353–386). Mouton de Gruyter.
- Culicover, P. W., & Winkler, S. (2022). Parasitic gaps aren't parasitic, or, the case of the uninvited guest. *The Linguistic Review*, 39(1), 1–35.
- Gerbrich, H., Schreier, V., & Featherston, S. (2019). Standard items for English judgment studies: Syntax and semantics. In S. Featherston, R. Hörnig, S. von Wietersheim, & S. Winkler (Eds.), *Experiments in focus: Information structure and semantic processing* (pp. 305–328). de Gruyter.
- Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, 86(2), 366–415.
- Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Doctoral dissertation, Massachusetts Institute of Technology].
- Liu, Y., Winckel, E., Abeillé, A., Hemforth, B., & Gibson, E. (2022). Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics*, 8, 495–525.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. www.r-project.org/
- Ramchand, G. C. (2008). *Verb meaning and the lexicon: A first phase syntax*. Cambridge University Press.
- Singmann, H., Bolker, B. M., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *afex: Analysis of factorial experiments* [R package version 0.22-2]. <https://cran.r-project.org/package=afex>
- Singmann, H., & Kellen, D. (2020). An introduction to mixed models for experimental psychology. In D. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Routledge.
- Truswell, R. (2007). Extraction from adjuncts and the structure of events. *Lingua*, 117, 1355–1377.
- Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.

Dutch long passive was tried to reject

Iva Kovač (University of Vienna) & Gert-Jan Schoenmakers (Radboud University)

Long passive is a complementation configuration with control, matrix passive, and promotion of the embedded object to matrix subject. It has long been regarded as unavailable in Dutch, but the results of our judgment experiment provide evidence that its acceptability depends on grammatical factors, in particular the matrix verb class. A rather high degree of inter-speaker variation further shows that long passive is acceptable to some participants. We propose a syntactic interpretation of our findings, showing how a marginal phenomenon can open up new perspectives on the syntax of a language.

1 Purpose of the study

In this paper, we provide first experimental evidence for the existence of long passive in Dutch. Long passive is a complementation configuration involving control in which an embedded object is promoted to matrix subject due to the passivization of the matrix verb (Wurmbrand 2001, 2014). The prevailing view in the linguistic literature is that long passive is absent from Dutch (Broekhuis 1992). However, recent claims about speaker variation (Tavenier 2020) and many examples found on the internet call for reconsideration. Two examples found in a local newspaper and on the internet are given in (1). That we are dealing with long passive is evident from plural agreement on the matrix auxiliary (1a) and nominative case on the pronominal object (1b).

- (1) a. [De beide personen]_i werd-en geprobeerd t_i te reanimeren ...
the both persons AUX-PL tried t to resuscitate ...
lit. 'Both people were tried to resuscitate ...' (De Gelderlander; August 11, 2019)
- b. Hij_i werd nog geprobeerd t_i over te halen door Eugene Reaper ...
he.NOM was still tried t PRT to convince by Eugene Reaper ...
lit. 'He was still tried to convince by Eugene Reaper ...' (GTA IV Wikipedia page)

We report on a judgment experiment further investigating the status of long passive in Dutch, having isolated two parameters from the literature that may impact its acceptability. Both factors concern the matrix verb class.

2 Rationale of the study

In their investigation of complementation configurations in a collection of typologically diverse languages, Wurmbrand and Lohninger (2019) make a distinction between EVENT verbs, such as *try* or *dare*, and SITUATION verbs, such as *decide* or *promise*. They show that complements of Event verbs are cross-linguistically more transparent for various dependencies, including long passive (see also Wurmbrand 2014). We expect in line with the cross-linguistic empirical landscape that configurations with a matrix Event verb are more amenable to long passive than those with a matrix Situation verb in Dutch as well.

Secondly, Pitteroff and Schäfer (2019) report that implicit control with *beginnen* 'begin' (2a) is less acceptable than with *proberen* 'try' (2b). Both verbs belong to Wurmbrand and Lohninger's (2019) Event class, but *beginnen* 'begin' is an aspectual and *proberen* 'try' a non-aspectual verb. Unlike long passive, these examples do not involve promotion of the embedded object. However, they do share two crucial properties with it: matrix passive and control of the embedded understood subject by the matrix implicit argument (hence *implicit control*). We expect Pitteroff and Schäfer's observation to extend to long passive as well.

- (2) a. Er werd begonnen de woonkamer op te ruimen.
there AUX begun the living.room PRT to clean
lit. 'It was begun to clean the living room.'

- b. Er werd geprobeerd de analyse te begrijpen.
 there AUX tried the analysis to understand
 lit. 'It was tried to understand the analysis.'

In all, we hypothesize that:

- H1** implicit control configurations (without long passive) are better than long passive;
H2 long passive is better with Event verbs than with Situation verbs;
H3 both configurations are better with non-aspectual verbs than with aspectual verbs.

3 The experiment

80 native speakers of Dutch rated how natural (0%-100%) sentences (24 target items, 48 fillers) would sound if uttered by a native speaker. Each target sentence was preceded by a single context sentence that licenses the use of a passive construction. The target sentences contained an implicit control or a long passive configuration and one of three matrix verb types: Event_{aspectual} (e.g. *begin*), Event_{non-aspectual} (e.g. *try*), or Situation (e.g. *decide*).

The raw judgment data were transformed into z-scores and entered into an LMER (Table 1). Statistical analysis revealed that long passive items were rated significantly lower than implicit control ones [**H1**] and items with an aspectual verb significantly lower than items with a non-aspectual verb [**H3**]. Post-hoc pairwise comparisons revealed significant differences between Event and Situation verbs in long passive items [**H2**], their direction depending on the sub-class (aspectual vs. non-aspectual) of the Event verb (aspectual: $t = -10.34$, $p < .001$; non-aspectual: $t = 3.46$, $p = .007$; Tukey-adjusted). The results are visually presented in Figure 1. Furthermore, we find considerable inter-speaker variation in items with long passive and non-aspectual Event matrix verbs: long passive is in fact perfectly available to at least some participants, see Figure 2.

Table 1: Specifications of the statistical analysis
 ($z \sim \text{verb class} * \text{construction} + (1 + \text{construction} | \text{participant}) + (1 + \text{construction} | \text{item})$)

	β	SE	t	p	
(<i>intercept</i>)	-0.12	0.05	-2.52	.018	*
construction type	-0.80	0.09	-8.93	< .001	***
verb class	0.22	0.02	11.47	< .001	***
aspectuality	0.75	0.05	13.77	< .001	***
construction * verb class	-0.18	0.04	-4.75	< .001	***
construction * aspectuality	0.04	0.11	0.37	.712	

4 Theoretical implications

Our findings indicate that the acceptability of long passive in Dutch is dependent on the class of the matrix verb. This is an important finding, because it may very well be where the grammatical core of the acceptability judgment shines through (cf. Schütze 1996) despite the marginal status of the configuration. Since the verb classes we studied behave differently with regard to a range of grammatical phenomena in various languages (Wurmbrand & Lohninger 2019), we conclude that the observed contrasts in the compatibility of these classes with long passive in Dutch (for those speakers who accept it) are connected to grammatical properties of the verbs in question.

Furthermore, the marginal status of long passive is a result of strong inter-speaker variation, which implies that our participants have distinct grammars. We believe that Dutch long passive therefore deserves follow-up research, especially in light of recent findings that an investigation of individual grammars may be required to arrive at, or render obsolete, certain generalizations (Lyskawa & Ranero 2022). Schütze (1996: 37) makes the case most eloquently: “[i]t has come to be generally acknowledged that not all speakers of ‘the same language’ might have the same competence, but that does not justify basing the theory only on sentences for which

Figure 1: Mean judgment scores (left) and z-scores (right) per condition (error bars indicate within-subject standard errors from the mean)

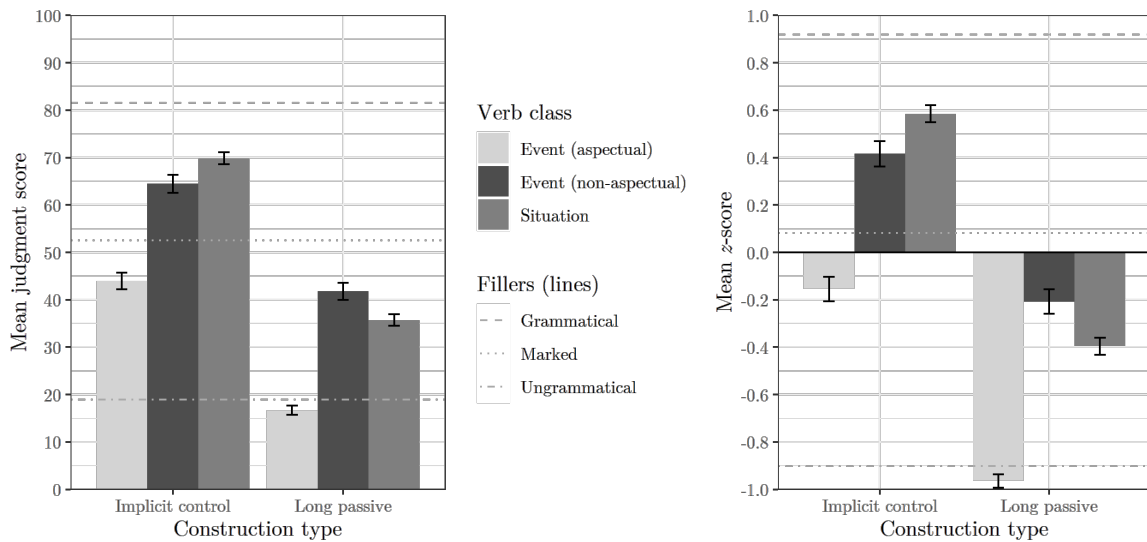
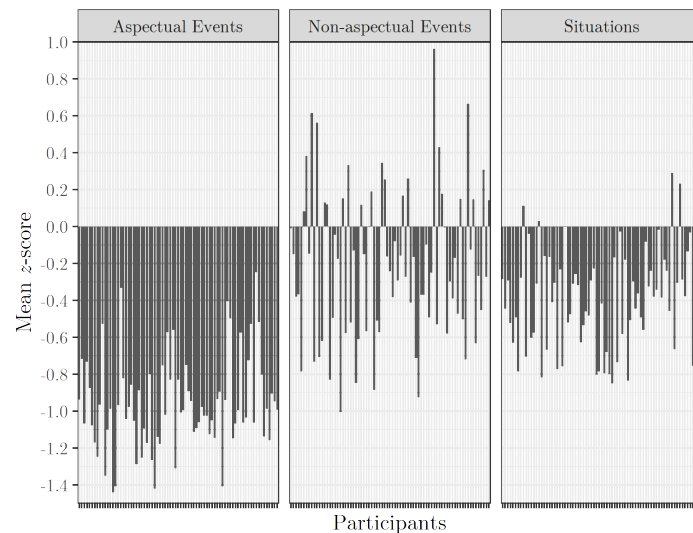


Figure 2: Mean z-scores for three different verb classes in long passive, per participant



there is universal agreement, and extrapolating by some means to dictate the status of the remainder. In cases where people disagree, that fact cannot be ignored; the theory must be able to describe *every* speaker’s competence, and thus must allow for variation wherever it occurs.”

5 Deriving the distribution

The judgment patterns we find can be accounted for under [Wurmbrand and Shimamura’s \(2017\)](#) analysis of long passive. These authors argue that long passive involves a syntactic dependency between an embedded underspecified Voice_R head and the matrix passive Voice head encoding the passive implicit argument (e.g. [Legate 2014](#)). This relation serves as a basis for semantic argument sharing, and a control-like relation can thus be established without PRO, allowing promotion of the embedded object to matrix subject (otherwise, PRO would intervene).

We propose that [H1] is due to markedness of Voice_R in Dutch and a preference for control via PRO. Specifically, Dutch speakers whose grammars are able to generate long passive have

Voice_R in their lexical inventories, while speakers who reject long passive do not and can only establish control via PRO, which makes promotion of the embedded object impossible. Concerning [H2], Wurmbrand and Lohninger (2019) argue that Event verbs, but not Situation verbs, can take radically reduced complement clauses such as VoiceP_R (as opposed to e.g. TP or CP). The relatively high ratings of Situation verbs in long passive may be due to a coercion of these verbs into a simultaneous interpretation (see Wurmbrand 2001), making them compatible with VoiceP_R, but also explaining their degraded status compared to non-aspectual Event verbs. Regarding [H3], Dutch aspectual verbs may combine with an infinitive or with a prepositional phrase (e.g. *beginnen met zingen* lit. ‘begin with singing’). We propose that the low ratings for long passive with aspectual verbs are the result of the infinitive not being their complement, but an obligatory control adjunct. Following Landau (2021), obligatory control adjuncts are CPs embedded into a PP-layer. This analysis for infinitival dependents of Dutch aspectual verbs is corroborated by the observation that the preposition may also be overt, as illustrated in (3) for the verb *ophouden* ‘cease’. Long passive is then unacceptable because such an elaborate structure blocks both promotion of the embedded object and the dependency between the embedded and matrix Voice heads which is needed for control to be established in long passive.

- (3) ... voldoende om op te houden **met** onze tijd te verknoeien.
 ... enough for PRT to cease **with** our time to waste
 ‘... enough to cease wasting our time.’ (J. van de Wetering: De zaak Ijsbreker, p. 70)

Our results show considerable contrasts between verb classes, reflected both in the general acceptability of long passive and the extent of individual variation, providing new perspectives on Dutch verbal syntax. We conclude that devoting attention to marginal phenomena as well as inter-speaker variation may provide valuable insights into broader aspects of grammar.

References

- Broekhuis, H. (1992). *Chain-government: Issues in Dutch syntax*. ICG Printing: Holland Institute of Generative Linguistics.
- Landau, I. (2021). *A selectional theory of adjunct control*. Cambridge, MA: MIT Press.
- Legate, J. A. (2014). *Voice and v: Lessons from Acehnese*. Cambridge, MA: MIT Press.
- Lyskawa, P., & Ranero, R. (2022). Optional agreement as successful/failed Agree: Evidence from Santiago Tz’utujil (Mayan). *Linguistic Variation*, 22(2), 209–267.
- Pitteroff, M., & Schäfer, F. (2019). Implicit control crosslinguistically. *Language*, 95(1), 136–184.
- Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press. (Reprinted in 2016 by Language Science Press, Berlin)
- Tavenier, L. (2020). *On the nature of θ -roles: Dutch evidence for a feature-based account* (Unpublished master’s thesis). Utrecht University.
- Wurmbrand, S. (2001). *Infinitives: Restructuring and clause structure*. Berlin/New York: Mouton de Gruyter.
- Wurmbrand, S. (2014). Restructuring across the world. In L. Veselovská & M. Janebová (Eds.), *Complex visibles out there. Proceedings of the Olomouc Linguistics Colloquium 2014: Language use and linguistic structure* (pp. 275–294). Palacký University.
- Wurmbrand, S., & Lohninger, M. (2019). An implicational universal in complementation: Theoretical insights and empirical progress. In J. M. Hartmann & A. Wöllstein (Eds.), *Propositional arguments in cross-linguistic research: Theoretical and empirical issues*. Tübingen: Gunter Narr Verlag.
- Wurmbrand, S., & Shimamura, K. (2017). The features of the voice domain: Actives, passives, and restructuring. In R. D’Alessandro, I. Franco, & Á. Gallego (Eds.), *The verbal domain* (pp. 179–204). Oxford: Oxford University Press.

Predictability effects on the acceptability of antecedent-target mismatches under verb phrase ellipsis

Robin Lemke, Lisa Schäfer, Heiner Drenhaus, Ingo Reich (Saarland University)

We present an information-theoretic account of acceptability contrasts between antecedent-target mismatches under verb phrase ellipsis (VPE) and three experiments which support its prediction: The acceptability of mismatches (and VPE in general) increases as a function of the context-based likelihood of the target of ellipsis. Our approach relies on general processing mechanisms rather than specific assumptions about the syntax and processing of VPE mismatches, which have been proposed in previous research.

Since Sag (1976); Williams (1977), it has been assumed that verb phrase ellipsis (VPE) requires a syntactically identical antecedent, as in (1a), whereas structural differences like the voice mismatch in (1b) are not acceptable. However, (1c) shows that voice mismatches are not unacceptable across the board.

- (1) a. Ann invited Bill and Sue did ⟨invite Bill⟩, too.
b. *Ann invited Bill and Sue was ⟨invited by Ann⟩, too.
c. This problem was to have been looked into, but obviously nobody did ⟨look into the problem⟩. (Kehler, 2002, 548)

1 Background

Mismatches have been used as diagnostics in the ongoing debate on the syntactic analysis of VPE. On the one hand, *syntactic* accounts claim that the ellipsis site contains rich syntactic structure, which is deleted under identity with the antecedent (e.g. Sag, 1976). On the other hand, *pragmatic* accounts (e.g. Hardt, 1993) argue that VPE involves a null anaphor resolved through context. Unacceptable mismatches support syntactic accounts, because VPE without identity is degraded, whereas acceptable mismatches favor pragmatic accounts by suggesting that full identity is not necessary. From the syntactic perspective, which in principle predicts identity, acceptable mismatches have been addressed in two ways: They could be grammatical, but sometimes unacceptable (Kim et al., 2011), or they could be ungrammatical, but sometimes acceptable (Arregui et al., 2006; Parker, 2018). The mismatches between acceptability and grammaticality have been accounted for by different processing mechanisms: Kim et al. (2011) argue that unacceptable mismatches violate the parser's expectations. Arregui et al. (2006) propose a VPE-specific reconstruction mechanism, which allows hearers to use syntactic derivation to construct a matching antecedent if none is present. Parker (2018) proposes a more general noisy memory retrieval mechanism. While Kim et al. (2011) argues with specific parsing rules, the latter two accounts predict that, in general, mismatches are more acceptable when antecedent and target are syntactically similar to each other. A problem for these accounts is that their predictions rely on syntactic properties of the antecedent, the target, or the degree of syntactic similarity between these expressions. However, the acceptability of mismatches is also driven by nonsyntactic factors like discourse relation (Kehler, 2002), information structure (Kertz, 2013), implicit Questions under Discussion (QuD) (Miller and Hemforth, 2014) and even extralinguistic context (Geiger and Xiang, 2021). In sum, these studies indicate that VPE mismatches improve when the target is likely in context, for instance because it refers to a salient QuD or its information structure is aligned with the antecedent. If structurally similar conjuncts are relatively likely, likelihood might also play a role in the similarity-driven effects observed by Arregui et al. (2006).

2 Information-theoretic account

The reduction of predictable material is a crucial prediction of information-theoretic accounts of linguistic encoding, which derive it from a general tendency toward distributing the hearer's processing effort uniformly across the utterance. Since predictable expressions are easier to process (Hale, 2001), reducing them avoids underutilizing processing resources (Fenk-Oczlon, 1989; Levy and Jaeger, 2007). In turn, using longer forms for unpredictable expressions avoids exceeding the processing resources. Applied to VPE, we therefore expect that mismatches (but in principle any instance of VPE) is more strongly preferred when the target is more likely. This likelihood might be modulated by diverse factors, such as linguistic context, the form of the antecedent, pragmatic inferences, and extralinguistic context. Therefore, the information-theoretic account is potentially capable of explaining some of the previous empirical findings with a single and independently evidenced processing mechanism, which has been shown to constrain the distribution of other omission and reduction phenomena (Jaeger and Buz, 2017).

3 Experimental rationale

We test the predictions of the information-theoretic account at the case of the stimuli used by Arregui et al. (2006), which they provide in their appendix. Across the four conditions in (2), Arregui et al. (2006) find a gradual acceptability cline ($a > b > c > d$), which they interpret as indicating the effort of constructing a parallel antecedent by syntactic derivation: When antecedent and target are similar, this effort is low and the mismatch is relatively acceptable, but the more derivation steps are required to build a matching antecedent, the more unacceptable is the mismatch. We hypothesize that the acceptability cline might result from the decreasing likelihood of the target from (2a) through (2d), which our information-theoretic account predicts to result in higher processing effort and degraded acceptability of VPE. We use an acceptability rating experiment to replicate the pattern in Arregui et al. (2006), a production task to measure the likelihood of the target VP and corresponding production preferences and a self-paced reading study to test for effects on processing effort.

- (2) a. None of the astronomers saw the comet, but John did. (Available VP)
 b. Seeing the comet was nearly impossible, but John did. (Embedded VP)
 c. The comet was nearly impossible to see, but John did. (VP with trace)
 d. The comet was nearly unseeable, but John did. (Negative adjective)

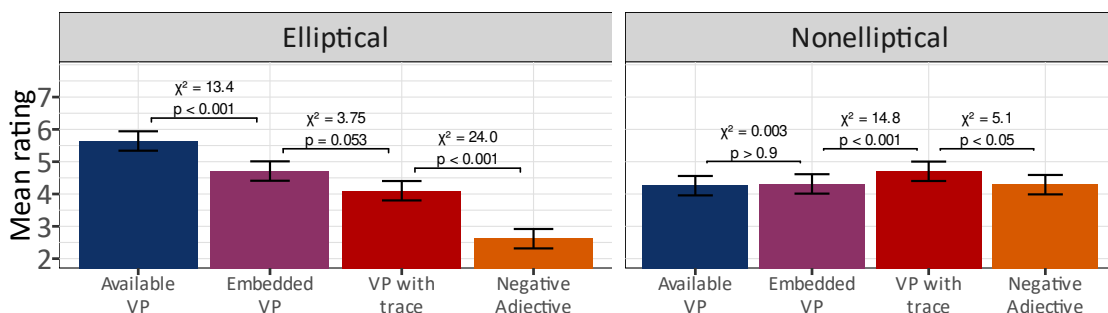


Figure 1: Mean acceptability ratings for VPE (left) and nonelliptical controls (right).

4 Experiment 1 – Acceptability rating

We first replicated the data by Arregui et al. (2006) in a web-based rating study conducted with LimeSurvey (LimeSurvey GmbH, 2012). Unlike Arregui et al. (2006), we also included the corresponding nonelliptical utterances (...but John saw the comet.) to investigate whether the effect is specific to ellipsis. Arregui et al. (2006) tested the first conjuncts as controls, but this does not rule out the possibility of an ellipsis-independent mismatch penalty (Kim et al., 2011).

The 16 items were mixed with 60 fillers and rated on a 7-point Likert scale (7 = fully acceptable) by 96 subjects recruited on the *Clickworker* crowdsourcing platform. COMPLETENESS was tested between subjects to ensure that the results are comparable to Arregui et al. (2006). The data (see Fig 1) were analyzed separately for the elliptical and nonelliptical conditions with cumulative link mixed models for ordinal data (Christensen, 2019). The predictor CONSTRUCTION was forward-coded, so that each of the three contrasts compared a level to the subsequent one(s). For the elliptical conditions, this replicates the pattern reported by Arregui et al. (2006). The nonelliptical sentences show that the effect is specific to ellipsis: The contrasts are either not significant, or their direction is inverted.

5 Experiment 2 – Production

We used a web-based written production task implemented in LimeSurvey to investigate two predictions of the information-theoretic account: (i) The potentially omitted VP is more likely in conditions where it is more acceptable and (ii) subjects produce more instances of VPE in that case. 120 participants recruited on *Clickworker* were asked to provide the most natural continuation of the materials, which were cut off after the subject in the 2nd conjunct (*John*, in (2)). The data were annotated for whether subjects produced a VP identical to the one that would be omitted under VPE (*saw the comet*) and, if so, whether this VP was reduced by VPE. The data (Fig. 2) were analyzed with logistic mixed effect regressions (Bates et al., 2015) predicting one of the binary dependent variables IDENTITY or ELLIPSIS from the forward coded CONSTRUCTION predictors. The analyses show that in two out of three contrasts, the VP is significantly more often produced in conditions where VPE is judged as more acceptable in the rating study, and that – among the VPs that can be reduced – VPE is more frequent the more likely the VP is. Taken together, this shows that the acceptability differences in experiment 1 are related to a gradual decrease in likelihood from (2a) through (2d), which is also reflected in a stronger preference for omitting the VP in a production task.

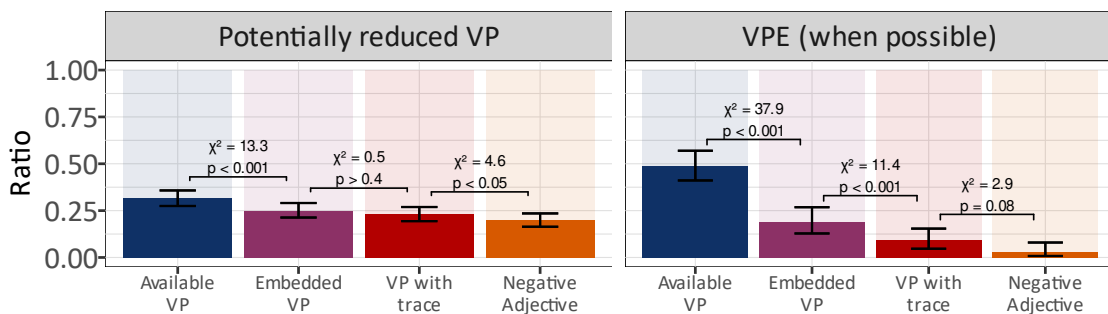


Figure 2: Ratio of potentially reduced VPs (left) and VPE ratio among these VPs (right).

6 Experiment 3 – Self-paced reading

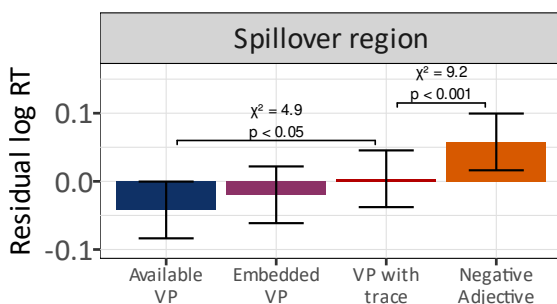


Figure 3: Residual log RTs at the SOR onset.

We then used a web-based self-paced reading study to investigate whether the link between the target VP likelihood and the preference for VPE is mediated by processing effort. 48 subjects recruited on *Clickworker* read 16 items like (2), which were extended with a spillover region (SOR, a causal or temporal clause like *because he had a special telescope* for (2)) and mixed with 60 fillers. The stimuli were presented word-by-word in

a centered self-paced reading paradigm using PCIBex (Zehr and Schwarz, 2018). We analyzed the mean residual log reading times of the first three words of the spillover region which followed the auxiliary *did* (Fig. 3) with linear mixed effects regressions (Bates et al., 2015) using the same predictors as in the previous experiments. The analysis reveals a similar gradual pattern as experiments 1 and 2: Ellipsis seems to be more difficult to process when the VP is less likely. In the main analysis (forward coding), only one of the CONSTRUCTION contrasts is significant (VP with trace vs. negative adjective). However, pairwise comparisons show that the difference between the available VP and VP with trace conditions is also significant (due to the gradual difference between the first three conditions, forward coding cannot test this).

7 Discussion

Our experiments show that the gradual acceptability pattern reported by Arregui et al. (2006) is in line with production preferences and processing effort: As the information-theoretic account predicts, VPE is more acceptable when the omitted VP is more likely (in this case, subjects are also more likely to produce VPE) and it is more easily processed. The data do not speak against Kim et al. (2011) and Kertz (2013), but the predictors they investigate do not explain the complete pattern in our data. Our findings are partially in line with previous accounts of mismatches, in particular with Arregui et al. (2006) and Parker (2018). However, Arregui et al. (2006) assume VPE-specific repair mechanism, whereas our account operates on more general processing principles. Furthermore, Arregui et al. (2006) and Parker (2018) operate on the similarity between antecedent and target, so that they cannot take into account pragmatic or extralinguistic factors, which have been shown to modulate the acceptability of mismatches in other studies. The information-theoretic account predicts effects of all factors which modulate the likelihood of the target. A further implication is that some of the identity conditions proposed in the literature for VPE and other ellipses might be traced back to differences in predictability. To what extent this is possible must remain open for future research, and since information-theoretic optimization is limited to a choice between grammatical structures (Jaeger, 2010), some mismatches might still be ruled out by grammar. Taken together, the information-theoretic account is supported by the data, it provides a uniform explanation of VPE mismatches and other omissions, and it is based upon independently motivated processing principles.

Selected references

- Arregui, A., Clifton, C., Frazier, L., and Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, 55(2): 232–246.
- Geiger, J. and Xiang, M. (2021). At the syntax-discourse interface: Verb phrase ellipsis interpretation in context. *Language*, 97(1):e89–e110.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL (Vol. 2)*, pages 159–166.
- Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing*. Number 182 in IRCS Technical Reports Series.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI, Stanford, California.
- Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, 89(3):390–428.
- Kim, C. S., Kobele, G. M., Runner, J. T., and Hale, J. T. (2011). The Acceptability Cline in VP Ellipsis. *Syntax*, 14(4):318–354.
- Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in neural information processing*, volume 19, pages 849–856. MIT Press, Cambridge, MA.
- Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry*, 44(1):77–108.
- Miller, P. and Hemforth, B. (2014). *VP Ellipsis beyond syntactic identity: the case of nominal antecedents*.
- Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa: a journal of general linguistics*, 3(1):129.
- Sag, I. A. (1976). Deletion and logical form. PhD thesis, MIT, Cambridge, MA.
- Williams, E. (1977). Discourse and logical form. *Linguistic Inquiry*, 8:101–139.

Testing extraction of and out of subjects and objects in Mandarin Chinese relative clauses

Ruihua MAO, Université Paris Cité

ruihua.mao@etu.u-paris.fr

This study probes into the subject relative clause preference and subject islands in Mandarin through two contextualized acceptability judgment experiments. Specifically, we manipulate grammatical function (subject vs. object) and resumptive pronouns (with vs. without RP) in simple subject/object relative clauses and sub-extraction from subject and object where the resumptive pronoun is possessive. The results showed a lack of the subject preference for simple relative clauses, but existence of preference for extraction out of subjects in RCs, which is consistent with discourse-based theory (Abeillé et al., 2020).

1 Introduction

Unlike simple extraction, which typically favors subjects, sub-extraction out of subjects is often penalized (so-called subject islands) even if there is cross-linguistic and cross-construction variation (Sprouse et al 2016). For sub-extraction, Abeillé et al (2020) found a subject advantage for French and English (with pied-piping) when extracting out of NPs in relative clauses but not in *wh*-questions (or *it*-clefts). In order to obtain the cross-linguistic pattern, we tested extraction preferences in Mandarin Chinese relative clauses.

Mandarin Chinese relative clauses (RC), which combine the basic SVO word order and head-final RC structures, appear to be inconsistent for a subject relative advantage. These different studies may be hard to compare because they look at different kinds of contexts and constructions (Yun et al., 2015), sometimes suffering from a variety of insufficiencies (Vasishth 2015). Hsiao and Gibson (2003), Gibson and Wu (2013), and Sung et al. (2016) found an Object preference, whereas Li et al. (2010), Wu et al. (2012), and Jäger et al. (2015) found a (slight) subject preference when the NP with RC starts with a classifier. In a meta-analysis, Vasishth et al. (2013) concluded that the general evidence points towards an advantage for subject relative clauses. The question of extraction out of subjects and objects in Mandarin Chinese RCs has been studied much less than simple relative clauses. We will provide new data on this question in our Experiment 2.

Also, resumptive pronouns (RP) are often thought of as a last resort strategy ameliorating long-distance dependencies (Zenker & Schwartz, 2021). In our first experiment, we will test extraction of the subject or the object in Mandarin relative clauses (with classifiers) with or without resumptive pronouns. If resumptive pronouns are a last resort strategy, they should be more acceptable in the more complex construction, i.e. the object relative clauses in our Experiment 1 and possibly extractions out of both subjects and objects in Experiment 2.

In sum, Mandarin RCs play a crucial role in addressing two research questions: (1) Is there a subject or object preference, for extraction and/or sub-extraction? (2) Does RP in Mandarin ameliorate extraction acceptability?

2 Experiment

2.1 Methods

Two contextualized acceptability judgment experiments were conducted on IBEX, where we crossed grammatical function (subject vs. object) and Resumptive pronouns (with vs. without RP) in a 2x2 Latin-square design. Materials for Experiment 1 comprised Mandarin simple subject/object RCs, with a classifier to avoid temporal ambiguity, and an RP in subject or object position (1), consisting of 20 items and 32 fillers. Materials for Experiment 2

are adapted from Exp1 for sub-extraction out of subject or object, and in which the RP is possessive (2), consisting of 20 test items and 25 fillers (from an unrelated experiment).

(1) Exp 1 example set

Context:

Wajue gongzuo yijing jinxingle shunian le.
Excavation work already going on several years.

“Excavation work has been going on for years.”

SRC_0RP

Kaoguxuejia faxian-le yige zhenhanle gudai suoyou laifang guowang de jinzita
Archaeologist found one-CL astound-PFV in-ancient-times all visiting kings REL pyramid
“The archaeologist found a pyramid which __ astounded all visiting kings in ancient times.”

SRC_RP

Kaoguxuejia faxian-le yige ta zhenhanle gudai suoyou laifang guowang de jinzita
Archeologist found one-CL it astound-PFV in-ancient-times all visiting king REL
pyramid
“The archeologist found a pyramid, which it astounded all visiting kings in ancient times.”

ORC_0RP

Kaoguxuejia faxian-le yige gudai suoyou laifang guowang dou zantan de jinzita
Archeologist found one-CL in-ancient-times all visiting kings DOU admire REL pyramid
“The archeologist found a pyramid, which all visiting kings admire_ in ancient times.”

ORC_RP

Kaoguxuejia faxian-le yige gudai suoyou laifang guowang dou zantan ta de jinzita
Archeologist found one-CL in-ancient-times all visiting kings DOU admire it REL pyramid
“The archeologist found a pyramid, which all visiting kings admire it in ancient times.”

(2) Exp 2 example set

Context:

Wajue gongzuo yijing jinxing-le shunian le.
Excavation work already going on several years.
Excavation work has been going on for years

RC_S_0RP

Kaoguxuejia faxian-le yige gaodu zhenhanle gudai suoyou laifang guowang de jinzita
Archaeologist found one-CL height astound-PFV in-ancient-times all visiting kings REL pyramid
“The archaeologist found a pyramid of which the height astounded all visiting kings in ancient times.”

RC_S_RP

Kaoguxuejia faxian-le yige qi gaodu zhenhanle gudai suoyou laifang guowang de
jinzita
Archeologist found one-CL its height astound-PFV in-ancient-times all visiting king REL pyramid
“The archeologist found a pyramid of which its height astounded all visiting kings in ancient times.”

RC_O_0RP

Kaoguxuejia faxian-le yige gudai suoyou laifang guowang dou zantan gaodu de
jinzita
Archeologist found one-CL in-ancient-times all visiting king DOU admire height REL
pyramid
“The archeologist found a pyramid of which all visiting kings admire the height in ancient times.”

RC_O_RP

Kaoguxuejia faxian-le yige gudai suoyou laifang guowang dou zantan qi gaodu de jinzita
Archeologist found one-CL in-ancient-times all visiting king DOU admire its height REL pyramid

“The archeologist found a pyramid of which all visiting kings admire its height in ancient times.”

2.2 Participants

Native Mandarin speakers residing in Mainland China completed the experiments online. Participants were presented with sentence pairs (context and experimental sentence) and asked to rate the second sentence on a 1-7 Likert scale, and answer simple yes/no comprehension questions. Only data from participants with an accuracy rate above 80% were analyzed, and the effective data is composed of judgment from 60 participants (Exp1) and 50 participants (Exp2) respectively.

3 Predictions

Most syntax-based theories predict a subject preference in Exp1 (simple RC) and a subject penalty in Exp2 (subject island). However, some syntactic approaches suggest that Mandarin Chinese in general has very few or no island constraints (see Cheng 2009 for wh-questions). No penalty for extraction out of the subject would be predicted in this case. If RP makes complex structures easier, this would predict an advantage for RP in case of island violations (if they exist), hence an interaction in Exp2. Resumptive pronouns may also make simple object relatives easier (Exp1). A discourse-based theory predicts no subject penalty in Exp2 because RCs are not focalizing constructions (Abeillé et al 2020). A distance-based processing theory predicts an object preference for both experiments because of the shorter linear distance between the object and the head noun. A frequency-based processing theory predicts a subject advantage in Exp1 because subject RCs are more frequent in corpora (Yun et al 2015).

4 Results and Discussion

4.1 Experiment 1: Subject and object relative clauses

Bayesian analyses (brms package by Bürkner, 2018, 8 chains, 9000 iterations, weakly informative priors, maximal model, cumulative) show a high probability for a main effect of RPs (higher ratings on RCs without RPs) and a high probability for an interaction (object relatives are judged more acceptable with RPs than subject relatives), but a low probability for a main effect of syntactic function, possibly because subject relatives are less acceptable with RP. It also supports Aoun & Li (2003), who argue that the syntax of RCs with RP is more complex (with an empty operator) than RCs with a gap. This might mean that RPs are only used to help with a very complex construction.

Figure 1a : Results Exp1

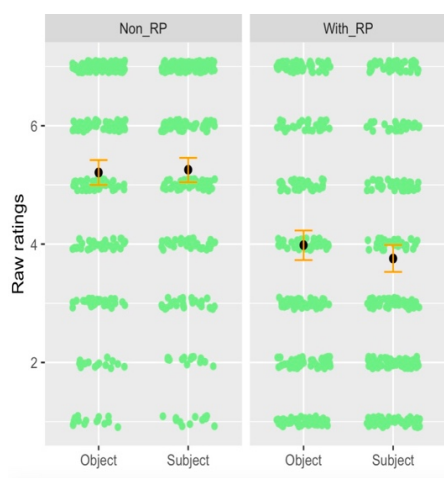
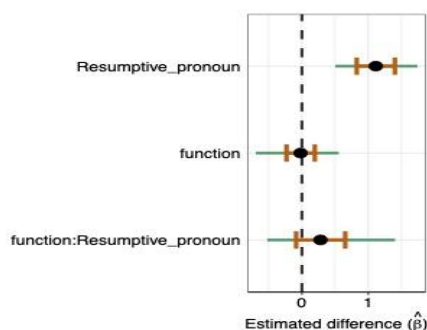


Figure 1b: Analysis Exp 1

Resumptive_pronoun: $p(\beta > 0) = 1$
 Function: $p(\beta < 0) = .58$
 Function:Resumptive_pronoun: $p(\beta < 0) = .94$



4.2 Experiment 2

With a maximal model in Bayesian analysis (all parameters as in Exp1), we found a tendency for a subject advantage with the probability of .86, but low probability for a main effect of RPs and interaction effect. This subject preference is similar to what was found for Italian, French and English RCs (with pied-piping). No penalty for RP was found for the extractions in contrast to the simple RCs. The lack of subject penalty is predicted by Abeillé et al 2020's discourse based theory but would also be compatible with a general lack of islands in Mandarin Chinese. However, it is not consistent with a universal subject island constraint and is not predicted by a linear distance-based processing theory either. Fukuda et al (2020) found that Japanese also lacked a subject-object asymmetry in subextraction in RC, although the distance for extraction out of objects is shorter. Our data may also be compatible with a frequency-based processing theory if subject sub-extraction is more frequent in corpora. This will be tested in future work.

Figure 2a: Results Exp 2

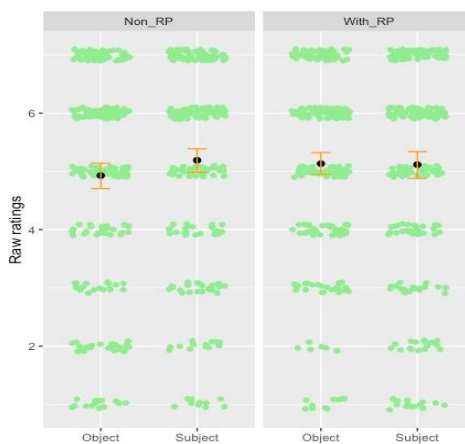
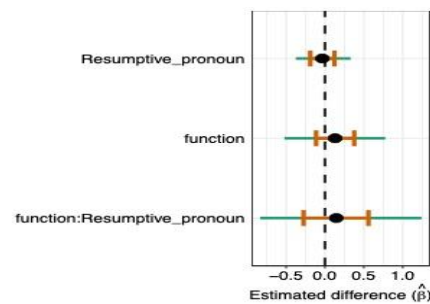


Figure 2b: Analysis Exp 2

Resumptive pronoun: $p(\beta < 0) = .70$
Function: $p(\beta < 0) = .86$
Function: Resumptive_pronoun: $p(\beta < 0) = .76$



Selected references

- Abeillé, A., Winckel, E., Hemforth, B., Gibson, E. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204. <https://doi.org/10.1016/j.cognition.2020.104293>
- Aoun, J., & Li, Y. H. A. (2003). Essays on the representational and derivational nature of grammar: The diversity of wh-constructions. MIT press.
- Cheng, L. L. S. (2009). Wh-in-situ, from the 1980s to Now. *Language and Linguistics Compass*, 3(3), 767-791. <https://doi.org/10.1111/j.1749-818X.2009.00133.x>
- Li, Q., Zhang, J., & Yue, W. (2010). Chinese relative clauses processing in supportive context removing ambiguity. *Studies in literature and language*, 1(4), 12-19. <http://dx.doi.org/10.3968/n>
- Jäger, L., Chen, Z., Li, Q., Lin, C. J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, 79, 97-120. <https://doi.org/10.1016/j.jml.2014.10.005>
- Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, 8(10). <https://doi.org/10.1371/journal.pone.0077006>
- Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, 27(10), 1489-1524. <https://doi.org/10.1080/01690965.2011.614423>
- Yun, J., Chen, Z., Hunter, T., Whitman, J., & Hale, J. (2015). Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics*, 24(2), 113-148. <https://doi.org/10.1007/s10831-014-9126-6>
- Zenker, F., & Schwartz, B. D. (2021). Resumptive pronouns facilitate processing of long-distance relative clause dependencies in second language English. *Proceedings of the Linguistic Society of America*, 6(1), 325-339. <https://doi.org/10.3765/plsa.v6i1.4972>

Anaphoric Potential of Weak Definites contrasted with Implicit Entities and Indefinites in German

Fereshteh Modarresi & Manfred Krifka
Leibniz-Zentrum Allgemeine Sprachwissenschaft Berlin (ZAS)

Weak definites (WDs) as in *take the train* (Carlson & Sussman 2005) are formally definite but differ from anaphoric definites or definites that refer to a unique familiar entity. They have a number-neutral reading (cf. *Mary took the train to Paris* – could be one or more trains), they show sloppy readings under conjunctions (as in *Peter took the train and Mary too* – could be different trains) and they take narrow scope under quantifiers (as in *everybody took the train* – could be different trains). WDs appear to be similar to non-specific, narrow-scope indefinites (IDs) as *take a train*.

However, WDs differ from IDs in their potential to be taken up by anaphora. Schwarz (2014) argues that WD interpretations disappear when anaphoric uptake is present, as in *Every accident victim was taken to the hospital [both WD and SD interpretation] and discovered that it had a roof garden [only SD interpretation survives]*. Scholten & Aguilar (2010) investigate what they call the “questionable ability to set up discourse referents” in an experiment for Dutch. Assuming that pronouns and anaphoric DPs differ insofar as they prefer more vs. less salient discourse referents (DRs), they show that pronouns are more often selected after IDs (91%), but less so after WDs (yet still 59%).

We report on a series of experiments (some presented first in Modarresi, Fortmann & Krifka 2019) that investigate the anaphoric potential of WDs. Some of them make use of the fact that in German, WDs as objects of certain prepositions are formally distinct from anaphoric definites, as in *ins Kino* vs. *in das Kino* ‘to the cinema’. They are not distinct from familiar definites but this interpretation can be made unlikely by choice of examples.

Exp. 1 followed the procedure of Scholten & Aguilar (2010) but with parallel sentence frames. In English translation, participants were presented with one condition of a text like ‘Luise likes to go out to a pub with friends. Last week she drank a bit more and took {a / the} bus home. Luckily, [it | the bus] waited in front of the pub when she came out’. They had to select the pronoun ‘it’ or the full DP ‘the bus’ (32 participants, 14 sentences). Although WD antecedents elicited slightly more full DP anaphors, the difference was not significant. As the choice might be affected by the length of the anaphor, in **Exp 2** we investigated uptake with slightly longer demonstrative pronouns vs. full DPs (e.g. German *dieser* vs. *der Bus*), again with no significant difference.

We also conducted **Exp 3**, a free sentence completion experiment (online, 15 items, 30 participants). Participants should complete

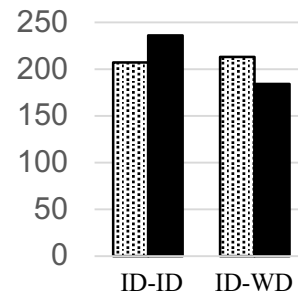
- (1) *Sophie ist wegen starker Bauchschmerzen {zum / zu einem} Arzt gegangen.*
Als erstes fragte...
‘Sophie went because to the.WD / to a.ID doctor because of belly ache.
At first she asked...’

WD antecedents failed to be taken up slightly more often than ID antecedents (no uptake 48 vs. 40 times). Also, WDs were more often taken up by full DPs than by pronouns (94 vs. 81 times). However, these differences are not significant.

In **Exp 4** we used a novel technique that did not involve production but interpretation. We presented participants with antecedent sentences that contained an ID antecedent followed by a WD or an ID antecedent. The second sentence contained a pronoun compatible with either antecedent (in its gender and its plausible interpretation):

- (2) *Nora hat sich gestern ein Museum angeschaut, bevor sie {ins Kino / in ein Kino} gegangen ist. Es war gerade neu eröffnet worden.*
 ‘Nora went to a museum yesterday before going {to the.WD cinema / to a.ID cinema}. It had just opened.’

The pronoun has two possible antecedents. We predicted that the selection of the second antecedent should be modulated by its form, with WDs being chosen less often than IDs. In an online survey (60 participants, 14 + 7 filler items), the participants read one version of the sample item, and then were asked, on a separate screen, to decide whether the pronoun refers to the first or the second antecedent (e.g., *Was ist gerade neu eröffnet worden?* ‘What was newly opened?’, followed by a selection between *das Museum* und *das Kino*). In the ID-ID case, we found a preference for the second antecedent, which is to be predicted as it is more recent, hence more salient (Ariel 1991). In the ID-WD case, the second antecedent was slightly disfavored. The difference between uptake of IDs and WDs in second position was significant (Chi-square $p = 0,01$). But there were many cases in which the pronoun was interpreted as referring to the WD antecedent.



We take the results of Exp 4, and tentatively of Exp 1, 2 and 3, as evidence that WDs do introduce DRs but that they are less salient than DRs introduced by IDs in the same position. (These results are in line with a study by Brocher et al. 2020 which use a different technique, eye tracking).

There are a number of theoretical models that do not assume that WDs introduce DRs, hence appear in conflict with all the experimental results reported so far. If WDs are seen as instances of pseudo-incorporation, which is expressed by bare nominals in certain languages, this includes theories where they are interpreted as kind-referring (Aguilar-Guevara & Zwarts 2010, Schwarz 2014), as property denoting (McNally 1995), as involving predicate restriction instead of argument saturation (Ladusaw & Chung 2003, Dayal 2015), or as not involving any FRs in other ways (Farkas & de Swart 2003).

Such theories can resort to an explanation of anaphoric reference to WDs as an instance of bridging inferences or associative anaphora. However, if uptake were via bridging, we expect that WDs should prefer definite DPs over pronouns (cf. Garrod & Sanford 1982). We have seen in Exp 1 that WD antecedents do not differ significantly from ID antecedents in this respect. In **Exp 5** we directly contrasted associative anaphora with anaphora to WDs with experimental items as the following.

- (3) *Susanne ist Journalistin bei einem Nachrichtensender. Gestern ist sie {mit dem Flugzeug / \emptyset } nach Costa Rica geflogen.* ‘Susanne is a journalist working with a news agency. Yesterday she flew to Costa Rica {by airplane / \emptyset }.’

Participants should select a pronoun vs. a full DP as the best option to continue:

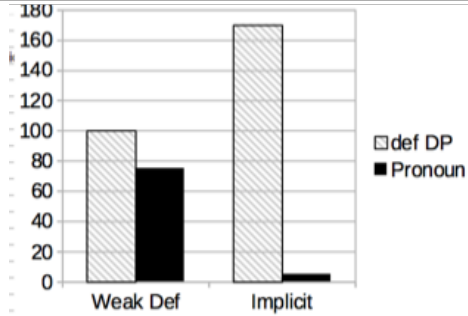
Da über dem Atlantik starke Stürme herrschten, geriet [es | das Flugzeug] öfters in Turbulenzen. ‘Since there were strong storms over the Atlantic Ocean, [it / the airplane] often ran into turbulence.’

The experiment (36 participants, 25 items, each participant saw one version of each item), revealed a clear difference: While participants continued implicit antecedents overwhelmingly with definite DPs as predicted, they picked up WD antecedents nearly

as often with pronouns. The difference is highly significant. We conclude that WDs must allow for anaphoric uptakes distinct from associative anaphora.

Our results show that WDs do introduce DRs, but that these DRs are less salient than those introduced by IDs. One modelling option would be to assume saliency values to antecedents, and assign higher values for IDs than to WDs.

However, this would neither relate to the definiteness feature of WDs, nor to their potentially non-unique and narrow-scope interpretation observed above.



Krifka & Modarresi (2016) propose a model for pseudo-incorporation in Persian that can be adapted to WDs, and explain their semantic and anaphoric properties (cf. also Yanovich 2008). It is formulated within Discourse Representation Theory (DRT) of Kamp & Reyle (1994). (3) represents the standard uptake of DRs introduced by IDs; here, d_2 is introduced by *ein Kino* ‘a cinema’, and the DR of *it*, d_3 , is identified with d_2 . We assume a novel way of introducing event DRs via existential sub-DRSs.

- (4) *Nora ist in ein Kino gegangen.* ‘Nora went to a cinema’
 $[d_1 d_2 \mid d_1 = \text{Nora, cinema}(d_2), \exists[e_1 \mid e_1 < \text{now, go-to}(e_1, d_1, d_2)]]$

Es ist neu geöffnet worden. ‘
 $[d_1 d_2 d_3 \mid d_1 = \text{Nora, cinema}(d_2), \exists[e_1 \mid e_1 < \text{now, go-to}(e_1, d_1, d_2)],$
 $d_3 = d_2, \exists[e_2 \mid e_2 < \text{now, recent}(e_2, \text{now}), \text{open}(e_2, d_3)]]$

In contrast, WDs introduce a DR that is dependent on the event quantifier; *ins Kino* identifies the unique cinema of e_1 . As a consequence, this DR is not directly accessible but can be recovered by a process of abstraction and summation, as in (4). This explains both the narrow-scope effects and the reduced anaphoric potential.

- (5) *Nora ist ins Kino gegangen.* ‘Nora went to the.WD cinema’
 $[d_1 \mid d_1 = \text{Nora}, \exists[e_1 d_2 \mid d_2 = \text{cinema}(e_1), e_1 < \text{now, go-to}(e_1, d_1, d_2)]]$

Es ist neu geöffnet worden.
 $[d_1 d_3 d_3 \mid d_1 = \text{Nora}, \exists[e_1 d_2 \mid d_2 = \text{cinema}(e_1), e_1 < \text{now, go-to}(e_1, d_1, d_2)]$
 $d_3 = \Sigma d_2 \exists[e_1 d_2 \mid d_2 = \text{cinema}(e_1), e_1 < \text{now, go-to}(e_1, d_1, d_2)],$
 $\exists[e_2 \mid e_2 < \text{now, recent}(e_2, \text{now}), \text{open}(e_2, d_3)]]$

Kamp & Reyle (1994) have introduced abstraction and summation for anaphoric uptake of IDs in the scope of quantifiers. This predicts that the anaphoric potential of such IDs is similar to WDs, as it relies on the same processes. However, there are also differences – in particular the nature of the quantifier involved, in particular their plurality compared with the existential quantifier. To investigate this, we designed **Exp 6**, a free text completion experiment, with prompts like (5).

- (6) *{Every hiker / The hiker} prepared a sandwich for the picnic. ...*

This was an online experiment on English with 15 items and 60 participants. We classified the various anaphoric uptakes. We found slightly more uptakes after singular subjects (*the hiker*) than after quantified subjects (*every, each, all* and definite plural DPs in sentences suggesting distributive interpretations, as in *The hikers prepared a sandwich*), but the difference was not significant (e.g. 48% uptake under *every*-DPs, 52% uptake under singular DPs). We take this as evidence that anaphoric uptake of indefinites under the scope of quantifiers is indeed easily possible.

In the same experiment, we also investigated the uptake under indefinites and weak definites, as in (6).

(7) *The woman went to {the hospital / a hospital} for treatment. ...*

We found slightly more uptakes of IDs (24%) than of WDs (21%), again not significant. We take this as further evidence that anaphoric uptake of WDs is easily possible. The fact that overall we found fewer uptakes in cases like (7) than in cases like (6) is probably a consequence from the fact that both cases of (7) describe more stereotypical situations, where the object does not matter as much as in (6). Also, while the differences between *every / the* cases in (6), and WD/ID cases in (7) were not significant with the number of cases and participants tested, these differences showed the expected tendencies.

We conclude that the results of the experiments lend support to the proposed representation of weak definites.

References

- Aguilar-Guevara, A & J Zwarts. 2010. Weak definites and reference to kinds. *SALT*. 20. 1-15.
- Ariel, M.. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Brocher, A., Weeber, F., Hoek, J., Heusinger, K. von, 2020. Referent management in discourse. The accessibility of weak definites. 42nd Annual Conference of the Cognitive Science Society, pp. 2829-2835.
- Carlson, G & R S Sussman. 2005. Seemingly indefinite definites. In: Kepser, S. & M. Reis, (eds), *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Berlin: Mouton, 26-30.
- Dayal, V. 2015. Incorporation: Morpho-syntactic vs. semantic considerations. In: Borik, O. & B. Gehrke, (eds), *The syntax and semantics of pseudo-incorporation*. Leiden: Brill, 189-221.
- Farkas, D. & H. de Swart. 2003. *Semantics of incorporation. From argument structure to discourse transparency*. Stanford: CSLI Publications.
- Garrod, S. C. & A. J. Sanford. 1982. The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* 1: 21-41.
- Kamp, H & U Reyle. 1993. *From discourse to logic. Introduction to model theoretic semantics of natural language, formal logic, and Discourse Representation Theory*. Dordrecht: Kluwer.
- Krifka, M & F Modarresi. 2016. Number neutrality and anaphoric uptake of pseudo-incorporated nominals in Persian (and weak definites in English). *SALT* 26. 874-891.
- Ladusaw, W & S Chung. 2003. *Restriction and saturation*. Cambridge, Mass.: MIT Press.
- McNally, L. 1995. Bare plurals in Spanish are interpreted as properties. In: Morrill, G. & D. Oehrle, (eds), *Formal grammar*. Barcelona: Universitat Politècnica de Catalunya, 197-122.
- Modarresi F., Fortmann, J. & M. Krifka 2019. Weak definites vs. implicit entities vs. indefinites in German. Presented at the 41st Annual meeting of the DGfS, Bremen.
- Scholten, J., Aguilar-Guevara, A., 2010. Assessing the discourse referential properties of weak definites. *Ling. Neth.* 27, 115e128.
- Schwarz, F 2014. How weak and how definite are weak indefinites? In: Aguilar-Guevara, A, B LeBruyn & J Zwarts, (eds), *Weak Referentiality*. John Benjamins.
- Yanovich, I. 2008. Incorporated nominals as antecedents for anaphora, or How to save the thematic arguments theory. *University of Pennsylvania Working Papers in Linguistics* 14: 367-379.

Islands with bridges? Extraction from adjunct clauses in Danish

Anne Mette Nyvad, Ken Ramshøj Christensen, Christiane Müller
Aarhus University

Abstract

We present data from a series of acceptability judgement experiments on extraction from three types of finite adjunct clauses in Danish: relativization with and without supporting context, and topicalization with context. The results revealed a strikingly stable pattern: Extractions from temporal and causal clauses are significantly less acceptable than extractions from conditionals, which are significantly less acceptable than extractions from complement clauses. We shall argue that adjunct clauses are not strong islands in Danish, and that the variation in the acceptability ratings across constructions and languages depends on extra-syntactic factors.

1 Introduction

Adjunct clauses are traditionally assumed to be universal, strong syntactic islands blocking extraction. This characterization has traditionally been captured by the *Condition on Extraction Domains* (CED, Huang 1982, 505), which bans extraction from domains that are not properly governed. However, several studies have challenged this assumption. Extraction from finite adjunct clauses has been shown to be acceptable to varying degrees in a number of languages, including Danish and the other Mainland Scandinavian languages, based on examples like (1):

- (1) *Den vase* får du ballade [hvis du taber ____].
that vase get you trouble if you drop
'You are in trouble if you drop that vase.'
(Hansen and Heltoft 2011, 1814)

In (1), the DP *den vase* 'that vase' has been felicitously topicalized from a finite adjunct clause introduced by *hvis* 'if'. Examples like this appear to violate the CED and thus raise the question whether island constraints are subject to cross-linguistic variation.

However, recent studies on Norwegian (Bondevik, Kush, and Lohndal 2020) and Swedish (Müller 2017) suggest that extraction from adjunct clauses is not unconstrained, given that these languages exhibit an acceptability pattern that varies as a function of the type of adjunct clause in question. In other studies, the relative acceptability of extraction has also been suggested to depend on dependency type (Sprouse et al. 2016) and the presence of contextual facilitation (e.g. Kush, Lohndal, and Sprouse 2019). Although it has previously been suggested that neither *wh*-islands (Christensen, Kizach, and Nyvad 2013) nor relative clauses (Christensen and Nyvad 2014) are strong islands in Danish, it has not yet been explored in an acceptability study on adjunct clause extraction.

2 Present study

The purpose of the present study was twofold: One, to find out whether the acceptability of extraction out of different types of adjunct clauses varies in Danish, as has been found in the other Mainland Scandinavian languages Swedish and Norwegian. Second, to compare adjunct island sensitivity in Danish and English, based on the same basic set of stimuli, differing minimally across conditions.

2.1 Methods

Based on a recent study on English (Nyvad, Müller, and Christensen 2022), we conducted a series of acceptability judgment experiments (off-line judgments on a 7-point Likert scale) on extraction from three types of finite adjunct clauses in Danish. For the design of this experiment, we used a 2x2 factorial design with the two factors \pm Island and \pm Extraction. In addition, there was an additional Complementizer factor with four different levels: *at* 'that' (introducing a complement clause), *hvis* 'if', *da* 'when' and *fordi* 'because' (all introducing adjunct clauses). The *that*-clauses and the three different types of adjunct clauses were embedded under adjectival psych-predicates, and the target sentences were all preceded by a supporting context, e.g. as follows:

(2) **Context:**

I det sidste træningsprogram jeg udarbejdede for Emma, ville jeg gerne gøre det så godt som umuligt for hende og inkluderede derfor endnu et sæt virkelig brutale pull-ups.

'In the latest workout routine I designed for Emma, I really wanted to make it impossible for her and included another set of particularly brutal pull-ups.'

Target:

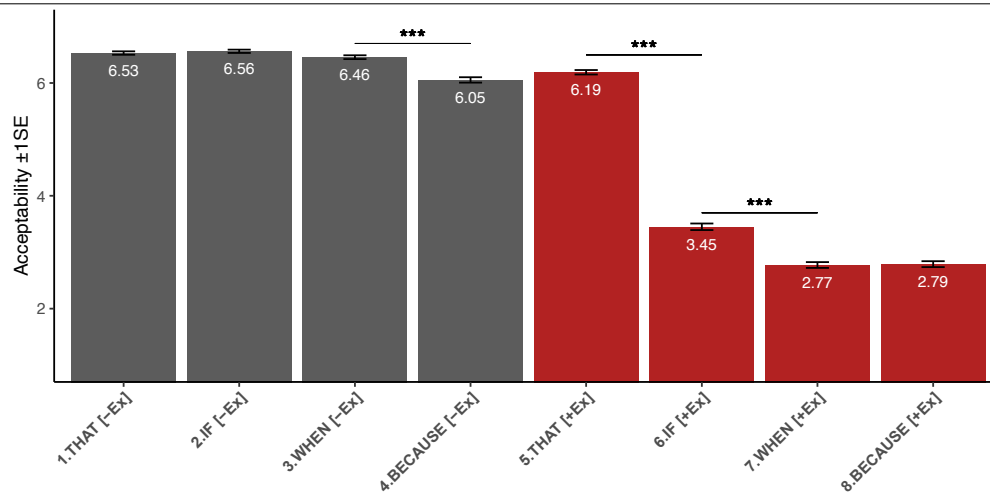
Det er dét program som jeg ville blive overrasket hvis hun faktisk gennemførte ____ .
it is this exercise that I would become surprised if she actually completed
'This is the exercise that I would be surprised if she actually completed.'

Two types of fillers were added, one involving extraction from NP subjects and another involving extraction from coordinate structures. These structures served additionally as points of comparison, as both subjects and coordinate structures are also assumed to be strong islands in both English and Danish. The acceptability survey was disseminated in the form of an online questionnaire using Google Forms and participants were recruited through social media platforms.

2.2 Results

A total of 335 native speakers of Danish (330 female, 5 male) volunteered to participate in the experiment. The results and sliding pairwise comparisons for the eight target sentence types are shown in Figure 1 below. The acceptability ratings for the four baseline conditions [-Ex] all had a rating above 6 on the 7-point scale. There was a significant drop in acceptability between the [-Ex] and [+Ex] types, indicating a significant negative main effect of extraction, [+Ex]>[-Ex], as shown by the contrast between *that* [+Ex] and *if* [+Ex]:

Figure 1



Furthermore, we tested whether the acceptability level of extraction out of adjunct clauses in Danish would be affected by (a) removing the supporting context, and (b) employing topicalization out of the adjunct clauses instead of relativization. The overall results revealed a strikingly stable pattern: Extraction from *when*- and *because*-clauses is significantly less acceptable than extraction from *if*-clauses, which was significantly less acceptable than extraction from complement clauses headed by *that*. While we find that Danish is similar to Swedish (Müller 2017) and Norwegian (Bondevik, Kush, and Lohndal 2020) in showing variability in the acceptability pattern of extraction from conditional, temporal and causal adjunct clauses, the results relating to the raw scores of extractions from adjunct clauses are surprisingly low. In addition, the DD scores (difference-in-differences) suggest that all three types of adjunct clause extraction are island violations in Danish, according to the 0.75 threshold suggested by Kush et al. (2019, 401), while the results for English (Nyvad, Müller, and Christensen 2022) indicated that relativization out of conditional, temporal, and causal clauses does not appear to constitute strong island violations. In addition, there seems to be very little (if any) positive or negative effect of supporting context, and the overall acceptability pattern was the same for relativization and topicalization.

3 Conclusions

We shall, however, argue that though the pattern in Danish could be taken to support the assumption that adjunct clauses are strong syntactic islands, a closer analysis of the results suggests otherwise. First of all, the acceptability ratings are all much higher than for ungrammatical controls (in particular, coordinate structure violations). Secondly, the *if*-type shows more inter-participant variation, reflected in a less uniform (less skewed) response distribution than the other types.

Based on the variability observed in the data, we argue that adjunct clauses are not strong syntactic islands in Danish. Moreover, while extraction out of adjunct clauses must be licensed by the grammar, the variation in the acceptability of extraction depends on extra-syntactic factors, including processing and discourse function, both cross-linguistically and cross-construction.

References

- Bondevik, I., Kush, D. & Lohndal, T. (2020). Variation in Adjunct Islands: The Case of Norwegian. *Nordic Journal of Linguistics*, December, 1–32. <https://doi.org/10.1017/S0332586520000207>.

-
- Christensen, K.R., Kizach, J. & Nyvad, A.M. (2013). Escape from the Island: Grammaticality and (Reduced) Acceptability of Wh-Island Violations in Danish. *Journal of Psycholinguistic Research* 42 (1), 51–70. <https://doi.org/10.1007/s10936-012-9210-x>.
- Christensen, K.R. & Nyvad, A.M. (2014). On the Nature of Escapable Relative Islands. *Nordic Journal of Linguistics* 37 (1), 29–45. <https://doi.org/10.1017/S0332586514000055>.
- Hansen, E., & Heltoft, L. (2011). *Grammatik over det Danske Sprog. bind I-III*. København: Det Danske Sprog- og Litteraturselskab.
- Huang, C.J. (1982). Logical Relations in Chinese and the Theory of Grammar. PhD dissertation, Cambridge, MA: MIT. http://pubman.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf.
- Kush, D, Lohndal, T. & Sprouse, J. (2019). On the Island Sensitivity of Topicalization in Norwegian: An Experimental Investigation. *Language* 95 (3), 393–420. <https://doi.org/10.1353/lan.2019.0051>.
- Müller, C. (2017). Extraction from Adjunct Islands in Swedish. *Norsk Lingvistisk Tidsskrift* 35 (1). <http://ojs.novus.no/index.php/NLT/article/view/1399>.
- Nyvad, Anne Mette, Christiane Müller, and Ken Ramshøj Christensen. 2022. “The Nonuniformity of Adjunct Clause Extraction in English.” *Languages*.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. “Experimental Syntax and the Variation of Island Effects in English and Italian.” *Natural Language & Linguistic Theory* 34 (1): 307–44. <https://doi.org/10.1007/s11049-015-9286-8>.

How specific are linguistic structures? Mathematical priming on relative clause attachment in French

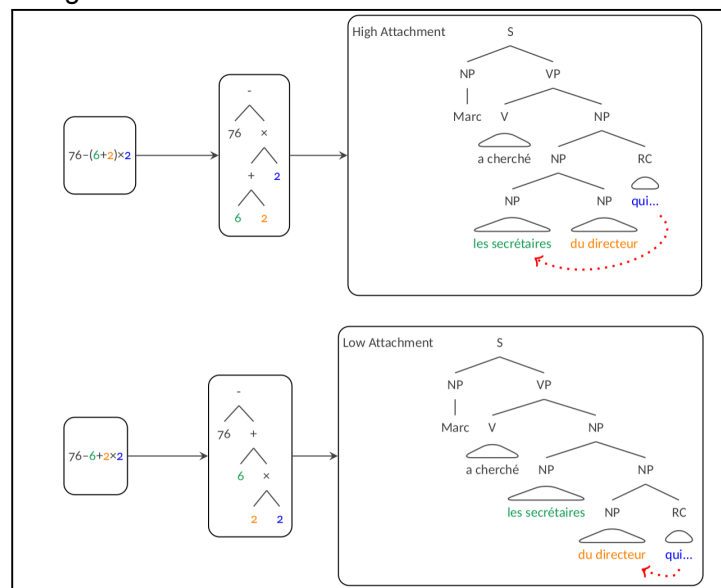
Céline Pozniak (SFL, U. Paris 8), Mireille Copin (LNPL, U. Toulouse Jean Jaurès),
Giuseppina Turco (LLF, U. Paris Cité) & Barbara Hemforth (LLF, U. Paris Cité)

The specificity of linguistic structures has been a central question in research related to questions on language universals and learnability. Scheepers et al. (2011) found that calculations like $76-(6+2)\times 2$ vs. $76-6+2\times 2$ respectively correspond to high and low attachment relative clause constructions. We report results from two production experiments further exploring the nature of mathematical priming on language. Contrary to Scheepers et al. (2011) and other previous studies, our results rather support relative clause attachment as association to thematic domains, as suggested by construal theory.

1 Introduction

The specificity of linguistic structures has been a central question in research related to questions on language universals and learnability. This question is still under debate (see Scheepers et al., 2019 for a review). Scheepers et al. (2011) suggested that calculations like $76-(6+2)\times 2$ structurally resemble a high-attachment RC construction (*[[the friend of a colleague] who lived in Spain]*) whereas calculations like $76-6+2\times 2$ are more similar to a low-attachment RC construction (*[the friend of [a colleague who lived in Spain]]*). This also applies for French as illustrated in Figure 1.

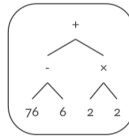
Figure 1. Calculations and relative clause attachment



Using a syntactic priming task, Scheepers et al. (2011) found that solving calculations like $76-(6+2)\times 2$ vs. $76-6+2\times 2$ (Figure 1) influenced high versus low relative clause (RC) attachment preferences in subsequently presented English sentences that participants had to complete (e.g, *The tourist guide mentioned the bells of the church that...*).

Although mathematical priming effects have been found based on this method, replicability issues are regularly raised. Hedier et al. (2020) and Hedier (2020) found that the calculation corresponding to low attachment was not correct. Indeed, for the calculation $76-6+2\times 2$, participants would rather start with $76-6$, calculate 2×2 , and then subtract the latter result from the first, leading to a different representation, as illustrated in Figure 2.

Figure 2. Real representation for 76-6+2×2 (Hedier, 2020)



According to the authors, the correct calculation should be $76-(6+2\times 2)$. Therefore, we used the new calculation for low attachment from Hedier et al. (2020) in experiment 1 in French. The results found suggested that priming may not be structural but associated to thematic domain. That's why we used other calculations in experiment 2 for high and low attachment primes.

2 Experiment 1 One variable was manipulated: prime (low vs. high). Primes consisted in calculations whose structure was equivalent to either that of a low attachment or of a high attachment of a RC. Items (N=12) were preceded by two structurally similar calculations, and consisted of structurally ambiguous target sentences containing complex NPs of the form *NP1 of NP2*, half of them singular-plural, half plural-singular, followed by the relativizer “qui” (Table 1). We used non-perceptual verbs in the main clause so as to avoid pseudo-RCs (Grillo & Costa, 2014), which have a different syntactic structure (Hedier et al. 2020).

Table 1. Example of two items in all conditions

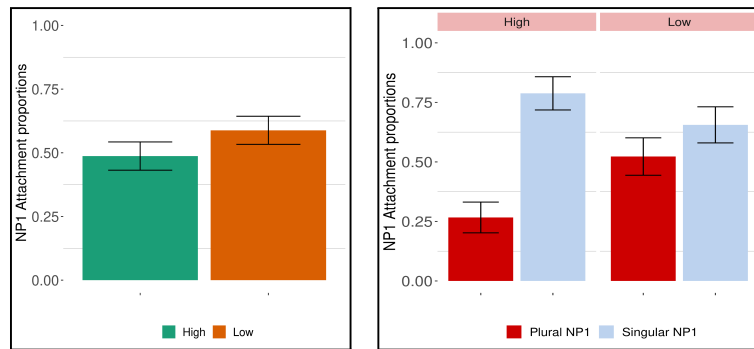
Item	NP1	Prime	Sentence
1	Plural	High	$76-(6+2)\times 2=$
		Low	$76-(6+2\times 2)=$
2	Singular	High	$8 - (6 - 2) / 2 =$
		Low	$8 - (6 - 2 / 2) =$

2.1 Prediction If shared structural representations between mathematics and language exist that trigger priming (Scheepers et al., 2011), high attachment calculations should prime RC high attachment, and low attachment calculations should prime RC low attachment.

2.2 Procedure Inspired by Scheepers et al. (2019), participants were reminded before the experiment of the arithmetic operator-precedence rules. Then, they answered calculations, read and had to continue sentences in writing (see Table 1). Eighty native speakers of French (recruited on Prolific) participated. We present the results from monolingual participants who correctly answered the prime calculations in the critical trials (N=618 observations).

2.3 Results As seen in Figure 3 (left), contrary to our expectations, Bayesian analyses showed that after a low calculation prime, continuations referred more to high attachment than low attachment ($\hat{\beta}=-0.53$, CrI[-1.47,0.36], $P(\hat{\beta})>0=0.88$). In post hoc analyses, taking number marking into account (Figure 3, right) we found an interaction between NP1 number and prime ($\hat{\beta}=2.69$, CrI[0.30,5.19], $P(\hat{\beta})>0=0.99$), meaning that there was a higher proportion of high attachment continuations with a singular NP1 than with a plural NP1, in the high prime condition.

Figure 3. Results from Experiment 1



2.4 Discussion Calculations did not prime the expected structure of RC attachment and, moreover, we found a strong interaction with number marking. Reanalyses of previous experiments in French showed the same pattern regarding NP1 number which seems to be a robust effect. We suggest that the calculations primed RC construal to thematic domains (TD) as it has been suggested in *Construal Theory* (Frazier & Clifton, 1997, see also Keller, 1995, or Crysmann, 2005, for corresponding analyses in HPSG). According to construal theory, RCs (as other non-primary relations) are associated to a thematic domain (Figure 4). The antecedent of the RC is then chosen based on a variety of factors (e.g. number). Priming of association to thematic domains could lead to the observed effects with our “high attachment” calculations priming a thematic domain above the complex NP (with a free choice of antecedents based on non-structural factors) and the “low attachment” calculations priming a thematic domain defined by NP1 (Table 2).

We tested the robustness of thematic domain priming with the adequate calculations experiment 2. If shared representations triggering thematic domain priming do exist, the high attachment calculations should prime RC high attachment, and the low attachment calculations should prime RC low attachment (Table 2, Primes 2 and 3).

Figure 4. Association to thematic domains

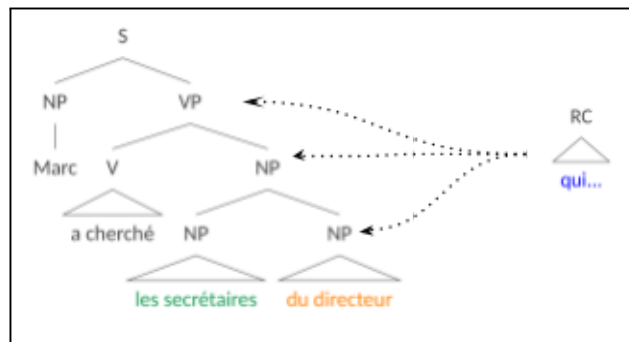


Table 2. Nature of priming for the calculations

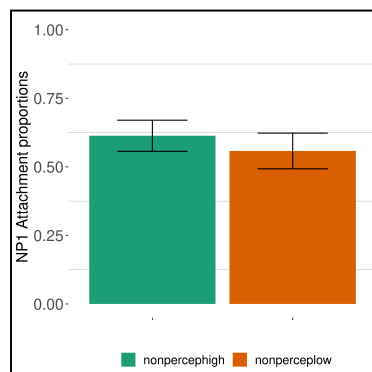
Prime	Calculation	Structural Priming	Thematic Domain Priming
1	76-(6+2)×2	High attachment (NP1)	Thematic domain defined by the main verb (NP1 or NP2 equally accessible)
2	76-(6+2×2)	Low attachment (NP2)	Thematic domain: Complex NP, defined by NP1, so high attachment (NP1)
3	76-6+(2×2)	Low attachment (NP2)	Thematic domain: Last NP, so low attachment (NP2)

3 Experiment 2

3.1 Design & procedure Design and procedure are the same as Experiment 1. We present the results from participants who correctly answered the prime calculations (N=506 observations).

3.2 Results & Conclusion As seen in Figure 5, continuations referred more to high attachment than low attachment after a high attachment prime ($\beta=0.57$, CrI[-0.6,1.7], $P(\beta)>0=0.84$). Our results support the hypothesis of shared representations between mathematics and language. However, it seems that structural priming may not be at stake here, but rather thematic domain priming, suggesting that relative clauses as non-primary relations are associated to a thematic domain. This would explain replicability issues from previous studies, especially in French since the calculations were rather structural and not linked to thematic domain (see Hedier et al., 2020 for the calculations). A better analysis of the number effect, especially regarding the semantic particularity of singulars, is also necessary to understand the nature of priming. Finally, our results support relative clause attachment as association to thematic domains, as suggested by construal theory.

Figure 5. Results from Experiment 2



Selected references

- Crysmann, B. (2005). Relative clause extraposition in German: An efficient and portable implementation. *Research on Language and Computation*, 3(1), 61-82.
- Frazier, L., & Clifton, C. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research*, 26(3), 277-295.
- Grillo, N., & Costa, J. (2014). A novel argument for the universality of parsing principles. *Cognition*, 133(1), 156-187.
- Hedier, A. (2020) Cross-domain priming: quantifying the mathematical priming effect. *Master's thesis*, Université de Paris.
- Hedier, A., Pozniak, C., & Hemforth, B. (2020, September). Mathematical to language priming: New evidence from French. In *Architectures and Mechanisms of Language Processing*.
- Keller, F. (1995). Towards an Account of Extraposition in HPSG. *arXiv preprint cmp-lg/9502027*.
- Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science*, 22(10), 1319-1326.
- Scheepers, C., Galkina, A., Shtyrov, Y., & Myachykov, A. (2019). Hierarchical structure priming from mathematics to two-and three-site relative clause attachment. *Cognition*, 189, 155-166.

Reading non-canonical sentences in context: Identity vs. Poset

Yvonne Portele & Markus Bader (Goethe University Frankfurt)

Isolated non-canonical object-before-subject (OS) sentences usually cause comprehension difficulties compared to canonical sentences with a subject-before-object (SO) structure. As theories of discourse structure have shown, non-canonical structures are only licensed under certain discourse conditions. We investigated whether given objects - given either by identity or set-membership - ease the processing of German OS sentences. Results from a self-paced-reading study show that both discourse relations eliminate processing difficulties for non-canonical sentences. Their overall processing pattern does not differ from their canonical counterparts (apart from influences caused by divergent lexical items).

1 Introduction

Non-canonical sentences, such as active object-before-subject (OS) clauses in German, usually cause processing difficulties during online comprehension compared to canonical sentences with an active subject-before-object (SO) structure. This pattern has been detected in several languages for isolated sentences (e.g., Bader & Meng, 1999 for locally ambiguous structures in German; Hyönä & Hujanen, 1997 for unambiguous sentences in Finnish). It is well-known, however, that the occurrence of non-canonical structures must be motivated by discourse-pragmatic conditions. Theories of discourse structure (e.g., Birner & Ward, 1998; Lambrecht, 1996; Prince, 1981) highlight *givenness* – already being introduced via the previous discourse – and *partially ordered set relations (poset)* – standing in a poset relation to a referent already mentioned in the previous discourse (see (1) for a set/subset relation) – as some of the central factors in the licensing of non-canonical word order.

- (1) We don't get involved in *all* murders, *but this one we thought we 'd take a look at*.
["ABC World News Tonight"; taken from Birner & Ward, 1998, p. 220]

In line with these theories, there is psycho- and neurolinguistic evidence that supportive contexts, when contrasted with unsupportive ones, indeed attenuate processing difficulties for non-canonical sentences (e.g., Burmester, Spalek, & Wartenburger, 2014; Gattei, París, & Shalom, 2021; Kaiser & Trueswell, 2004; Schumacher & Hung, 2012).

For German, Speyer (2005) found that a general preference to fill the sentence-initial position in German with the topic (the backward looking center) is usually overridden when the sentence contains an element standing in a poset relation to the preceding context, because this element then occupies the initial position. Furthermore, Weskott, Hörnig, Fanselow, and Kliegl (2011) found in a self-paced reading study that possible comprehension disadvantages for non-canonical object-subject (OS) in comparison to canonical subject-object (SO) sentences disappear and can even be turned into processing advantages when the object stands in a poset relation to the context. The question whether an identity relation also alleviates possible disadvantages for German OS sentences has received far less attention (but see Experiment 3 in Weskott, 2003 for referential identity in a more complex discourse configuration). In an offline acceptability study, Bader and Portele (2021) found that OS sentences in which the object referent was given and realized as an NP with a demonstrative determiner were rated equally high as their SO counterpart. To our knowledge, corresponding online data are missing.

2 Experiment

The current experiment investigated the question whether appropriate discourse structure, manipulated via contexts, facilitates the online comprehension of non-canonical sentences in Ger-

Table 1: Example stimulus used in the experiment

Identity	
Context	Letzte Woche am Donnerstag hat der Sponsor mit einem Torhüter telefoniert. <i>Thursday last week, the sponsor talked to a goalkeeper on the phone.</i>
SO target	Er hat diesen Torhüter zu einer Verlängerung seines Vertrages bewegt. <i>He persuaded this goalkeeper to extend his contract.</i>
OS target	Diesen Torhüter hat er zu einer Verlängerung seines Vertrages bewegt. <i>This goalkeeper, he persuaded him to extend his contract.</i>
Membership	
Context	Letzte Woche am Donnerstag hat der Sponsor mit der Fußballmannschaft telefoniert. <i>Thursday last week, the sponsor talked to the soccer team on the phone.</i>
SO target	Er hat den Torhüter zu einer Verlängerung seines Vertrages bewegt. <i>He persuaded the goalkeeper to extend his contract.</i>
OS target	Den Torhüter hat er zu einer Verlängerung seines Vertrages bewegt. <i>The goalkeeper, he persuaded him to extend his contract.</i>
Question:	Hat der Sponsor mit einem Torhüter telefoniert? <i>Did the sponsor talk to a goalkeeper on the phone?</i>

man. More specifically, we asked whether an identity relation (not investigated so far) eases the processing of non-canonical sentences in a similar way as poset relations or whether poset relations (found to alleviate and even override disadvantages) constitute licensing conditions that are more helpful for the human parser than an identity relation. A poset relation was established by creating a set-membership constellation, i.e., by including a collective noun (e.g., *soccer team*; *employment agency*) and mentioning a member of this collective (e.g., *goalkeeper*; *clerk*).¹ We used this relation instead of the more frequently investigated *part-whole* relation (e.g., Weskott et al., 2011) to extend the investigation of different poset relations.

2.1 Methods

2.1.1 Participants

45 native speakers of German were recruited via Prolific or participated for course credit.

2.1.2 Materials

Twenty experimental item sets consisting of a context sentence and a target sentence were created (see Table 1). We manipulated the two factors *Discourse Relation* (Identity vs. Membership) and *Word Order* (SO vs. OS). In the Identity conditions, a context sentence introduced two male referents. The subject of the context sentence was a definite NP (e.g., *the sponsor*). The object was introduced by using an indefinite NP (e.g., *a goalkeeper*). The subject of the context sentence was rementioned with a personal pronoun (*he*) as subject of the target sentences – either in sentence-initial (SO) or sentence-medial (OS) position. The object was rementioned in the target sentences by using a demonstrative NP including a lexical repetition of the co-referent noun (e.g., *this goalkeeper*) and consequently also appeared in sentence-initial (OS) or sentence-medial (SO) position. In the membership conditions, the same definite NP subject as in the identity conditions was used. Instead of a male character referent, the object was a set-denoting definite NP (e.g., *the soccer team*). The subject of the target sentences was again the personal pronoun (*he*) referring back to the previous subject and also appeared in sentence-initial (SO) or sentence-medial (OS) position. The object of the target sentences

¹ Although the set-membership relation is not a poset relation in a technical sense, it is usually subsumed under poset relations in the relevant literature, e.g., IS-A-MEMBER-OF in Ward and Prince (1991). Under the mereological approach to plurals and collectives (Link, 1983), there is a poset relation also in the technical sense in our membership condition.

was a definite NP denoting a member, a human character, of the set introduced in the previous sentence. It could appear either sentence-initially (OS) or in mid-sentence position (SO). Fifty-two filler items were included in the experiment.

2.1.3 Procedure

The 72 items were distributed onto 4 lists according to a Latin Square design. Materials were presented in a self-paced reading experiment (word-by-word moving-window presentation) conducted via Ixby farm (Drummond, Von Der Malsburg, Erlewine, Yoshida, & Vafaie, 2016). Eight of the experimental sentences were followed by a yes-no comprehension question.

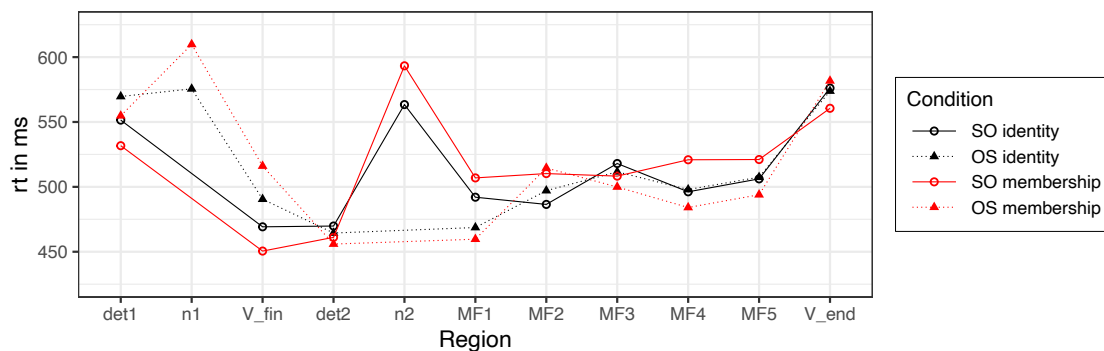
2.2 Results

Raw reading times are shown in Figure 1. Linear mixed-effects models showed significant effects at the positions n1 (first noun), V_fin (finite verb), det2 (second determiner) and MF 1 (first position of the midfield). At n1, there is a significant effect of Discourse Relation. Identity relations were processed faster than membership relations. In the identity relation, this is where the lexical item from the previous sentence is repeated. In the membership relation, this is where the noun denoting a singled-out member is mentioned for the first time. This difference is also visible as a spill-over effect in the form of a significant interaction at the following position V_fin. At det2, there was an effect of Word Order as well as Discourse Relation. OS sentences were processed faster. This difference stems from processing the lexical items for the personal pronoun *he* (er) in OS conditions vs. a determiner in SO conditions. The processing advantage for membership over identity relations at this point reflects the lexical difference between the articles *this* (dieser) in identity relations vs. *the* (den) in membership relations. This difference is still visible at the position of MF1, where there was an effect of Word Order with SO sentences being processed faster than their OS counterpart. No significant differences were detected in later regions within the sentence. Mean accuracy in the question comprehension task was 89.7%. There were no significant differences between conditions.

3 Discussion

We found several effects associated with different lexical items. We see two non-exclusive reasons for the effect for OS sentences found at the position of the first noun (n1), which was not significant for SO sentences at the position of n2. This effect can represent a processing advantage due to the lexical repetition of the noun in identity conditions or a processing disadvantage for having to infer the noun in membership conditions. We are currently running further self-paced reading studies investigating this issue by contrasting a poset relation with an identity relation established without lexical repetition, i.e., by using synonym expressions. We will compare this study with the current results to disentangle shortcomings of the current

Figure 1: Raw reading times in the different conditions. Note: There are no n1 and n2 data points in the respective SO or OS conditions since the subject was the one word personal pronoun *he* (er) whereas the object was a two word NP (determiner + noun).



experiment (such as the confound of discourse relation and article of the noun phrase). Overall, our results suggest that both discourse relations investigated in this study – identity and membership – alleviate potential processing difficulties of non-canonical OS sentences compared to their canonical SO counterpart in German. Self-paced reading patterns did not differ apart from lexical influences. Furthermore, there was no distinct advantage for a poset relation over an identity relation. We leave it to future work to investigate further discourse relations and their effect on the processing of non-canonical sentences. In addition, investigating different manifestations of *givenness* (e.g., situationally inferable vs. standing in a poset relation) will reveal important insight on the role of discourse licensing of non-canonical sentences.

References

- Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: an across-the-board comparison. *Journal of Psycholinguistic Research*, 28, 121-143. doi: 10.1023/A:10232062081
- Bader, M., & Portele, Y. (2021). Discourse and form constraints on licensing object-first sentences in German. *Languages*, 6(2), 82. doi: 10.3390/languages6020082
- Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam/Philadelphia: John Benjamins.
- Burmester, J., Spalek, K., & Wartenburger, I. (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and language*, 137, 62–76.
- Drummond, A., Von Der Malsburg, T., Erlewine, M. Y., Yoshida, F., & Vafaie, M. (2016). *Ibex Farm*. Retrieved from <https://github.com/addrummond/ibex>
- Gattei, C. A., París, L. A., & Shalom, D. E. (2021). Information structure and word order canonicity in the comprehension of spanish texts: An eye-tracking study. *Frontiers in Psychology*, 12.
- Hyönä, J., & Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in finnish: An eye fixation analysis. *The Quarterly Journal of Experimental Psychology: Section A*, 50(4), 841–858.
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94, 113–147. doi: 10.1016/j.cognition.2004.01.002
- Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press. doi: <https://doi.org/10.1017/CBO9780511620607>
- Link, G. (1983). The logical analysis of plurals and mass terms: a lattice-theoretical approach. In R. Bäuerle, S. Christoph, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (p. 302-323). Berlin & New York: Walter de Gruyter.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (p. 223-255). New York etc.: Academic Press.
- Schumacher, P. B., & Hung, Y.-C. (2012). Positional influences on information packaging: Insights from topological fields in german. *Journal of Memory and Language*, 67(2), 295–310.
- Speyer, A. (2005). Competing constraints on Vorfeldbesetzung in German. In *Proceedings of the Constraints in Discourse Workshop* (pp. 79–87).
- Ward, G. L., & Prince, E. F. (1991). On the topicalization of indefinite NPs. *Journal of Pragmatics*, 16(2), 167–177.
- Weskott, T. (2003). *Information structure as a processing guide: The left periphery of German verb-second sentences and its interpretation in context* (Unpublished doctoral dissertation). Universität Leipzig.
- Weskott, T., Hörnig, R., Fanselow, G., & Kliegl, R. (2011). Contextual licensing of marked OVS word order in German. *Linguistische Berichte*, 225, 3–18.

The interplay between quotation and referentiality: An empirical investigation into name-mentioning constructions

Natascha Raue & Álvaro Cortés Rodríguez
U. Kassel / U. Kassel & U. Tübingen

This paper presents a large-scale corpus investigation into quotational constructions involving the predicate *call* as in the sentence *This natural phenomenon is called a “moonbow”*. The nominal mentioned in the quotation, i.e. a “moonbow”, adopts a referring interpretation. Importantly, the determiner preceding the quoted nominal is optional. The study puts evidence forward that names used in name-mentioning constructions that are accompanied by a determiner differ referentially from nominals without a determiner. Our corpus data provides evidence that there are significantly more valid name-mentioning constructions with a determiner which we interpret as indicating a difference in referential salience.

1 Introduction

Name-mentioning constructions (NMC), as in (1a), are a type of quotational construction used to point to linguistic shapes and inform the addressee about the name of a lexical concept.

- (1) a. This piece of writing is commonly called (a) “short story”.
b. *A “short story” has three syllables.

NMCs involve naming verbs such as *call* or *refer to as* that function as name-selecting predicates. Name-selecting predicates typically involve three arguments, as shown in (2a): an (implicit) agent *x* (*one*), a theme *y* (*phenomenon*), and the name *z* (*moonbow*) of the theme *y*.

- (2) a. This natural phenomenon is called “moonbow”.
b. This natural phenomenon is called a “moonbow”.
c. *This natural phenomenon is called a “moonbow” but this phenomenon is not a “moonbow”.

The event argument in the NMC in (2a) adopts a generic meaning. The semantics of the verb *call* entails a copular relation (Matushansky 2008; Härtl 2020), more specifically, an identificational copular relation in which the two nominals are referentially identified, which can be used to explain the referentiality of the quoted material.

Evidence for the assumption of an implicit copula comes from the fact that copular relations cannot be negated, as shown in (2c). Evidence for the assumption of an implicit copula comes from the fact that copular relations cannot be negated, as shown in (2c). Assuming that the speaker “veridically commits” (see Giannakidou & Mari 2019) to the truth of the utterance, i.e. that the phenomenon is a moonbow. The semantics of the nominal does not deviate from the conventionalized semantics, meaning that the nominal is used non-metaphorically. In sentence (2a-b), we are dealing with the same meaning of *moonbow*, and the negation in (2c) renders the sentence illogical and hence unacceptable. In other words, to call a phenomenon “moonbow” entails that the phenomenon is a moonbow.

1.2 NMCs and the use-mention distinction

In comparison to a sentence like *Kassel is a city in Hesse*, in which the word “Kassel” is used with its customary reference, the expression “Kassel” is mentioned in constructions like

“Kassel” has six letters, describing the linguistic setup of the word. As a matter of fact, mentioned expressions create reference to the word itself and are frequently accompanied by quotation marks (Cappelen & Lepore 1997; Saka 1998). Crucially, quotational constructions, as represented in (1a), may be accompanied by a determiner. In contrast, a metalinguistic quotation blocks the occurrence of a determiner as shown in (1b). We would like to emphasize that demonstrations of linguistic shapes usually do not contain an article. Further, metalinguistic quotations have a long-lasting tradition in philosophical debates and are commonly regarded as acceptable sentential constructions (see e.g. Washington 1992; Saka 1998; Brendel et al. 2011).

1.3 NMCs as instances of pure quotation

In recent debates, quotational constructions have been separated into at least four different kinds of quotation. A stereotypical example of pure quotation as in *“Table” has two syllables* is characterized by its metalinguistic operation, i.e. by creating reference to a linguistic dimension of the quoted expression *table* (see, e.g., Quine 1981). Here, the quoted expression is only mentioned and refers metalinguistically. Name-mentioning constructions of the type in (1a) and (2a/b) have been argued to be instances of pure quotation (see Schlechtweg & Härtl 2020). Pure quotations are distinct from further types of quotation, including direct quotation (*Lena said yesterday at the museum of modern art: “This piece of art is difficult to understand”*), indirect quotation (*Lena said that this piece of art is difficult to understand*), and mixed quotation (*Max believes that the Pope “has God on speed dial”*), see, e.g., Brendel et al. (2011); Cappelen & Lepore (2007).

1.4 State of the art and preliminary work

There is evidence from a corpus study conducted in German that in NMCs with the verb *nennen* (‘call’) quotes occur significantly more often when the mentioned nominals are preceded by a determiner (Härtl 2020). This can be interpreted as a pragmatic mechanism where the use of quotes serves to compensate for the denotational interpretation the determiner indicates by highlighting the mentioned expression’s metalinguistic status. In an acceptability judgment study, we further observed a significant preference in English for nouns contained in NMCs to occur with a determiner. This difference is present for both *call* and *refer to as* predicates.

2 Empirical investigation

The current study aims to provide a more robust empirical foundation for the individual differences that determiner use induces in name-mentioning constructions.

2.1 Methods

The empirical investigation presents a large-scale corpus study using detailed concordance queries. We systematically use the enTenTen20 corpus (Jakubíček et al. 2013), and sampled N=2000 NMCs, which contained the predicate *call*. The four queries contained double quotes around the nominal and followed the patterns given in (3).

- | | | | |
|-----|----|--------------------|----------|
| (3) | a. | W calls X “Y”. | [noDETa] |
| | b. | X is called “Y”. | [noDETp] |
| | c. | X is called a “Y”. | [DETp] |
| | d. | W calls X a “Y”. | [DETa] |

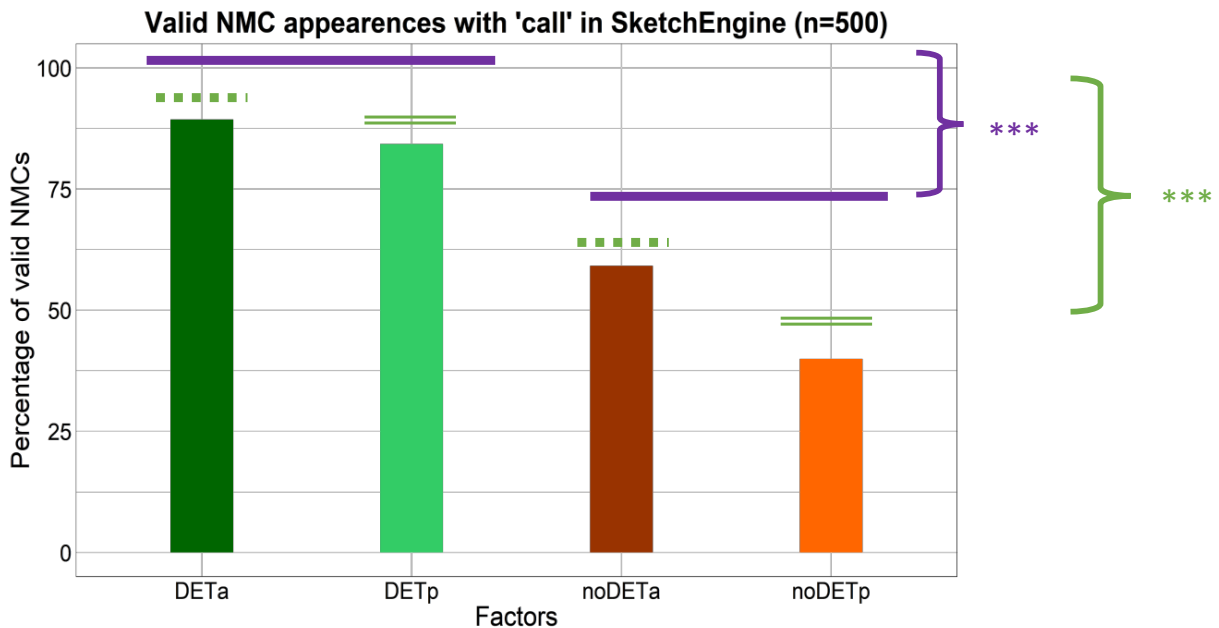
As a first step, the randomly selected constructions were labelled manually with regard to three criteria for valid NMCs. Valid NMCs have been defined as syntactical structures that (i) involve the naming predicate *call* functioning as a verb, (ii) involve exclusively nominal

tokens as the argument of the respective verb embedded in double quotes in the written mode, and (iii) constructions in which the postverbal nominal refers generically.

2.2 Results

For the statistical analysis, we ran a Generalized Linear Model (GLM) in R, which applies a logistic regression to a response parameter considered binomial (Winter 2020), see the descriptive statistics in Figure 1.

Figure 1: Percentage of valid NMCs per condition



The results showed a significant effect for the presence of a determiner as well as a main effect for VOICE where significantly more valid *active* NMCs were found.

2.3 Discussion

We argue that this significant preference for the presence of a determiner indicates a difference in referentiality levels, revealing that nouns preceded by a determiner are referentially more salient. We follow a definition of referential salience claiming that it is a function activating the referent’s conceptual representation in the discourse model (cf. Arnold & Griffin 2007). The degree of salience influences the assessability of an entity targeting a referent, meaning that highly salient entities are easier to assess in discourse operations. Our understanding of the notion of referential salience assumes this to be a graded notion (cf. Giora & Fein 1999).

Based on our corpus investigation, we conclude that NMCs containing a determiner introduce a more salient referent than those without a determiner. In other words, we argue that the name in NMCs accompanied by a determiner is perceived as referentially different from uses without a determiner. In future research, we will address the question of why referential salience should be considered relevant for naming, given the fact that the referent in NMCs is already introduced by another argument of the sentence, i.e., the theme argument.

Given that the statistical analysis also revealed that there are significantly more valid NMCs in active voice as opposed to passive constructions, it supports studies revealing that active structures generally appear more frequently (cf. Bada 2018). Naming constructions can be analysed as a small clause (cf. Matushansky 2008, Fara 2015). We follow the

underspecified copular approach for name-mentioning constructions as presented in Härtl (2020), arguing that NMCs are small clauses involving either a subject or an object. When considering the grammatical specifications of NMCs in active and passive, the nominals are assigned nominative case in passive voice (4b) as compared to a double accusative case in active constructions (4a).

- (4) a. Man nennt das Phänomen einen Mondregenbogen
 one calls this phenomenon.ACC a moonbow.ACC
 b. Dieses Phänomen wird ein Mondregenbogen genannt.
 this phenomenon.NOM is a moonbow.NOM called

In our discussion, will argue that NMCs should not be subsumed as a type of mixed quotation, i.e., a combination of direct and indirect quotation. Instead, we argue that NMCs should be interpreted as representing an instance of pure quotation. In our conclusion, we aim at implementing NMCs in the taxonomy of quotation types.

References

- Arnold, J. & Z. M. Griffin. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* 56(4), pp. 521–536.
- Bada, E. & Ulum, Ö. (2018). Utilization of active and passive constructions in English academic writing. *International Journal of Human Sciences* 15(1), pp. 413–421.
- Brendel, E. & Meibauer, J. & Steinbach, M. (2011). Exploring the Meaning of Quotation. In Elke Brendel (ed.), *Understanding quotation*, pp. 1-34. Berlin: Mouton de Gruyter.
- Cappelen, H. & Lepore, E. (1997). Varieties of quotation. *Mind* 106(423), pp. 429–450.
- Cappelen, H. & Lepore, E. (2007). *Language Turned on Itself: The Semantics and Pragmatics of Metalinguistic Discourse*. Oxford/New York: Oxford University Press.
- Fara, D. G. (2015). Names are predicates. *Philosophical Review* 124(1), pp. 59–117.
- Giannakidou, A. & Mari, A. (2019). Veridicality, subjectivity and grammatical mood. The view from Greek and Italian. Talk at *Sinn und Bedeutung* 24. Universität Osnabrück.
- Giora, R. & Fein, O. (1999). Irony Comprehension: The Graded Salience Hypothesis. *Humor* 12(4), pp. 425–436.
- Härtl, H. (2020). Referring nouns in name-informing quotation: A copula-based approach. In Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab (eds.), *Proceedings of Sinn und Bedeutung* 24(1), pp. 291–304. U Osnabrück / HU Berlin.
- Jakubiček, M. & Kovář, V. & Pavel Rychlý & Suchomel, V. (2013). The TenTen corpus family, *7th International Corpus Linguistics Conference CL*, pp. 125–127.
- Matushansky, O. (2008). On the linguistic complexity of proper names. *Linguistics and Philosophy* 31(5), pp. 573–627.
- Matushansky, O. (2008). On the linguistic complexity of proper names. *Linguistics and Philosophy* 31(5), pp.573–627
- Saka, P. (1998). Quotation and the use-mention distinction. *Mind* 107 (425), pp. 113–135.
- Schlechtweg, M. & H. Härtl (2020). Do we pronounce quotation? An analysis of name-informing and non-name-informing contexts. *Language and Speech* 63(4), pp. 769-798.
- Washington, C. (1992). The identity theory of quotation. *Journal of Philosophy* 89 (11), pp. 582–605.
- Winter, B. (2020). *Statistics for linguists: An introduction using R*. New York/London: Routledge.

What the use of the German focus particle *auch* can tell us about the influence of structural properties of the context

Laura Reimer (University of Münster), Bettina Braun (University of Konstanz) and Christine Dimroth (University of Münster)

In four experiments using a sentence fragment arrangement task, we investigated to what extent speakers are influenced by structural properties of the context, and to what extent they rely on default mechanisms (e.g., subjects are prototypical topics, objects are likely to be the focus of an utterance) when they integrate unstressed or stressed German additive *auch*. Results show that speakers strongly rely on default mechanisms, but that they are also primed by the structure of the context. Word order has a greater impact than prosodic properties of the context, and the underlying structure does not seem to have an influence.

1 Introduction

Additive particles such as English *also*, French *aussi*, Italian *anche* and German *auch* associate with a constituent (the associated constituent, AC, see *Maria* in 1a and 1b), and this AC is related to contextually relevant discourse alternatives (see *Peter* in 1). However, compared to English, French or Italian, German distinguishes between two variants of the additive particle *auch*, namely an unstressed (see 1a) and a stressed variant (see 1b).

- | | |
|---|---------------------------------|
| (1) Peter hat Pflirsiche gegessen. | ‘Peter has eaten peaches.’ |
| a. Auch [Maria] hat Pflirsiche gegessen. | ‘Maria has eaten peaches, too.’ |
| b. [Maria] hat AUCH Pflirsiche gegessen. | ‘Maria has eaten peaches, too.’ |

There are several differences between these two variants. From an information structural perspective, the AC of stressed *AUCH* has the status of a contrastive topic (AC_T , see Krifka 1999), and the AC of unstressed *auch* is the focus of the utterance (AC_F). Intonationally, AC_T and stressed *AUCH* are said to build a bridge contour which combines two pitch accents: a rising accent on the contrastive topic (Braun 2012), and a falling accent on the corresponding focus part, i.e. the particle (e.g., Büring 1997). In the case of unstressed *auch*, the particle is unstressed and the AC_F carries a pitch accent. With respect to the syntactic surface structure, stressed *AUCH* follows its AC, while unstressed *auch* precedes its AC. Finally, stressed *AUCH* is often related to an AC being the subject, and unstressed *auch* to an AC being the object of the utterance (e.g., Höhle et al. 2009). However, several of these factors are intermingled, and speakers seem to rely on default mechanisms when choosing between the two variants of *auch*: If the AC of *auch* is the subject, it most likely has the information structural status of a topic (AC_T) (Reinhardt 1981), which preferably is the AC of stressed *AUCH*.

The data of a sentence fragment arrangement task (Reimer & Dimroth, in press) show that speakers indeed have a strong preference for stressed *AUCH* when the AC is the subject and for unstressed *auch* when the AC is the object, indicating that speakers strongly rely on these default mechanisms. However, the results further show that a speaker’s choice can be affected by (information) structural properties of a context sentence: While the preference for unstressed *auch* for objects was 93%, the preference for stressed *AUCH* for subjects was only 69%. This reduced preference for stressed *AUCH* in the subject condition can be attributed to the specific context that was used in the experiment: The context sentence contained the focus particle *nur* (see 2).

- | | |
|--|---|
| (2) Peter und Maria haben Appetit auf Obst. Ich wette, nur [Peter] hat Pflirsiche gegessen. | ‘Peter and Maria want to eat fruits. I bet only Peter has eaten peaches.’ |
| a) Nein! [Maria] hat AUCH Pflirsiche gegessen. | ‘No! Maria has eaten peaches, too.’ |
| b) Nein! Auch [Maria] hat Pflirsiche gegessen. | ‘No! Maria has eaten peaches, too.’ |

A continuation of the dialogue with 2a, where *auch* follows its AC, was expected to be the default option. However, participants often chose the continuation 2b. Thus, the presence of *nur* in the context sentence – a particle that precedes its AC, which in turn is a focus – may lead participants to integrate the particle *auch* in a way that results in a similar structure. Thus, speakers may have been primed by the (information) structure of the context sentence when they constructed their utterance (see Bock 1986, Branigan 2007, Ziegler et al. 2019, Ziegler and Snedeker 2019 for structural priming; see Fleischer et al. 2012 for information structural priming).

In order to examine this instance of priming more closely, we conducted four experiments using a sentence fragment arrangement task. Specifically, we asked whether the strong preference for unstressed *auch* in the object condition found in Reimer & Dimroth (in press) can be mitigated by structural properties of the context as well.

2 Experiment 1

2.1 Methods

As in the study by Reimer & Dimroth (in press), we used context sentences containing the particle *nur* (**nur** *Birnen*, ‘only pears’). However, since *nur* always precedes its AC, we further included two conditions with negations in different positions: *keine* (**keine** *Birnen*, ‘no pears’) which precedes the AC, hence similar to *nur*, and *nicht* (*Birnen* **nicht**, ‘pears not’), which follows the AC. If the surface structure of the context sentence [**X** + AC/ AC + **X**] influences the choice of the speaker to use stressed or unstressed *auch* in the target sentence, we expect speakers to use more stressed **AUCH** when the context comprises the negation *nicht* than if it comprises the focus particle *nur* or the negation *kein*. If, however, participants solely rely on default mechanisms (objects are likely to be the focus of the utterance; unstressed *auch* associates with a focused constituent), we expect speakers to use unstressed *auch*, independent of the elements in the context. Twenty-one native speakers of German took part in the web-based experiment (SoSci-Survey, Leiner 2014; all participants were recruited with the software hroot, Bock et al. 2014). After silently reading the context sentences, participants had to arrange a target sentence by dragging and dropping given words (presented in boxes in randomized order, see Figure 1).

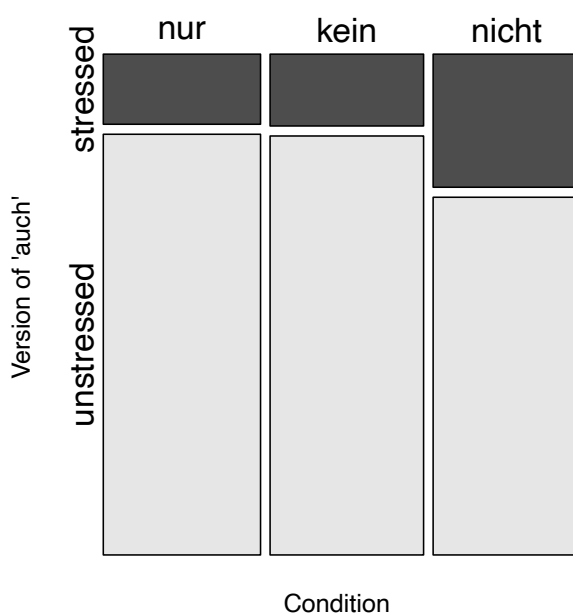
Figure 1



2.2 Results

We analysed the position of the particle *auch* relative to its object AC in the resulting target sentence. We performed a generalized linear model on the unstressed realizations in R (R core Team 2017, package lme4, Bates et al. 2015). The fixed-effects factor was the element in the context (*nur*/ *kein*/ *nicht*), and the random effects were items and participants. The results show that speakers have a general preference for unstressed *auch* (81%) if the AC is the object. Crucially, while the preference for unstressed *auch* does not differ between condition 1 (*nur*) and 2 (*kein*) ($\beta = -0.01$, $SE = 0.38$, $t = -0.04$), it differs significantly between condition 1 (*nur*) and 3 (*nicht*) ($\beta = -0.96$, $SE = 0.36$, $t = -2.68$, $p = .007$), and between condition 2 (*kein*) and 3 (*nicht*) ($\beta = -0.98$, $SE = 0.35$, $t = -2.72$, $p = .007$) (see Figure 2).

Figure 2



That *nur* and *kein* show a similar pattern and that both differ from *nicht* indicates that the shared surface structure has an influence on the choice of the two variants of the particle *auch*, and that speakers are primed by structural properties of the context. Crucially, the presence of *nicht* in the context with the word order [AC + X] leads speakers to use more stressed *AUCH* (against their actual preference to use unstressed *auch* if the AC is the object).

3 Experiment 2

3.1 Methods

To make sure that it is the structure [AC+ X] that lead speakers to use stressed *AUCH* and not the typical intonational contour that is related to *AUCH* and that might have been silently activated while reading the sentences, participants were not presented with the context sentences in written form, but auditorily. We included the conditions *kein* [X + AC] and *nicht* [AC + X] and manipulated the pitch accent of the AC and the negated element in the context (see 3 and 4). 44 native speakers participated in the experiment.

- (3) A: Peter hat Äpfel gegessen. 'Peter ate apples.'
 B: Ich wette, Peter hat /keiNE BIRnen\ gegessen. → L*+H H+L* (hat pattern)
 B': Ich wette, Peter hat KEIne BIRnen gegessen. → H* H* (double-peak)
 'I bet Peter didn't eat any pears.'

- (4) A: Peter hat die Äpfel gegessen. 'Peter ate apples.'
 B: Ich wette, Peter hat die /BirNEN NICHT\ gegessen. → L*+H H+L* (hat pattern)
 B': Ich wette, Peter hat die BIRnen NICHT gegessen. → H* H* (double-peak)
 'I bet Peter didn't eat any pears.'

3.2 Results

The results show a main effect of negated element in the context: As in Experiment 1, the presence of *kein* led to more uses of unstressed *auch* than the presence of *nicht* ($\beta = -1.09$, $SE = 0.18$, $t = -5.91$, $p < .001$). Crucially, there was no effect of intonational contour and no interaction ($p > .05$), indicating that the prosodic information of the context did not influence speakers when choosing between the two variants of *auch*.

4 Experiments 3 and 4

4.1 Methods

In Experiments 3 and 4, we investigated whether the syntactic priming effect found in Experiments 1 and 2 was driven by the linear ordering of the constituents on the surface, or whether the underlying structure influenced processing. To that end, we presented sentences in spoken form and participants had to arrange the fragments in written (Experiment 3) and spoken form (Experiment 4). The sentences contained the negated elements *kein* and *nicht*, which are related to different underlying structures (*kein* + AC/ AC + *nicht*), but which were presented with an identical surface structure in Experiments 3 and 4, (AC + *kein*/ AC + *nicht*; see 5). 31 native speakers participated in Experiment 3, and 29 native speakers participated in Experiment 4.

- (5) A: Peter hat Äpfel gegessen. `Peter ate apples.'
B: Ich wette, [Birnen] hat Peter **keine** gegessen. `I bet Peter didn't eat any pears.'
B': Ich wette, [Birnen] hat Peter **nicht** gegessen. `I bet Peter didn't eat any pears.'

4.2 Results

The results show a general preference for unstressed *auch*, indicating that speakers strongly rely on default mechanisms. In both experiments, there was a tendency of more uses of unstressed *auch* when the context contained *kein* compared to *nicht*, although this difference was not significant.

5 Discussion

The results of all four experiments indicate that speakers strongly rely on default mechanisms when constructing their utterances. They know that objects are likely to be integrated as the focus of the utterance, and this is reflected in their choice of unstressed vs. stressed *auch*. However, speakers are influenced by structural properties of the context to some extent, as shown by the results of Experiment 1 and 2. Thereby, word order has a greater impact than prosodic properties of the context. The results of Experiment 3 and 4 speak against an influence of the underlying structure.

References

- Bates, D., Mächler, M., Bolker, B. & Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, pp. 1-48.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), pp. 355-387.
- Bock, O., Baetge, I. & Nicklisch, A. (2014). hroot – Hamburg registration and organization online tool. *European Economic Review* 71, pp. 117-120.
- Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass* 1, pp. 1-16.
- Braun, B. (2012). Where to associate stressed additive particles? Evidence from speech prosody. *13th Annual Conference of the International Speech Communication Association*. Portland, Oregon.
- Büring, D. & Hartmann, K. (2001). The syntax and pragmatics of focus-sensitive particles in German. *Natural Language & Linguistic Theory* 19, pp. 229-281.
- Fleischer, Z., Pickering, M. J. & McLean, J. F. (2012). Shared information structure: Evidence from cross-linguistic priming. *Bilingualism: Language and Cognition*, pp. 1-12.
- Höhle, B., Berger, F., Müller, A., Schmitz, M. & Weissenborn, J. (2009). Focus particles in children's language: Production and comprehension of *auch* 'also' in German learners from 1 year to 4 years of age. *Language Acquisition* 16, pp. 36-66.
- Krifka, M. (1999). Additive particles under stress. In Deven Strolovitch & Aaron Lawson (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 8*, pp. 111-128. Cornell U: CLC Publications.
- Leiner, D. J. (2014). SoSci Survey (Version 2.6.00-i).
- Reimer, L. & Dimroth, C. (in press). Unstressed versus stressed German additive 'auch' – What determines a speaker's choice? *Linguistics Vanguard*.
- Ziegler, J., Bencini, G., Goldberg, A. & Snedeker, S. (2019). How abstract is syntax? Evidence from structural priming. *Cognition* 193, pp. 104045.
- Ziegler, J. & Snedeker, J. (2019). The use of syntax and information structure during language comprehension: Evidence from structural priming. *Language, Cognition and Neuroscience* 34, pp. 365-84.

Dimensions of judgment in stigmatized and non-stigmatized variation

Gert-Jan Schoenmakers (Radboud University)

Linguistic judgment experiments typically elicit responses in terms of the acceptability or surface probability of a sentence. Evidence that the conceptual dimension of the judgment scale influences the outcome of the experiment exists, but is only limited. This study investigates whether the scale dimension affects judgments for stigmatized (prescriptive norm violations) and for non-stigmatized (scrambling) variation in Dutch. Sentences are judged in one of three dimensions, *viz.* acceptability, surface probability, or aesthetic quality. The results indicate that participants take into account the scale dimension, but the effects depend on the type of variation. The findings are related to Schütze's (1996) theory of judgments.

1 Introduction

Linguistic judgments are increasingly considered the basis for inferences about linguistic representation (Goodall 2021, Schindler, Drożdżowicz, and Brøcker 2020, Schütze 1996). However, as Featherston (2021) points out, it is not exactly clear what these data quantify. Judgments are commonly interpreted as some sort of window into “grammaticalness” (Chomsky 1965), although it is well known that many other underlying factors contribute to a reported judgment as well. Sprouse (2020), for example, argues on the basis of Schütze's (1996) seminal work that linguistic judgments are the conscious reports of automatic responses to a stimulus. Such responses result from a composite of various considerations, yet they also contain an acceptability core that can sometimes serve as a proxy for grammaticality.

That it is difficult to determine what linguistic judgments are judgments of raises the question whether the instructions in judgment experiments can affect their outcome. In most cases, experiments prompt responses in a particular dimension through the experimental instructions, e.g. the *acceptability* or *naturalness* of a stimulus sentence. The present paper addresses the question to what extent linguistically naïve participants take into consideration the dimension of the judgment scale, and to what extent the scale dimension contributes to the acceptability core of linguistic judgments. Furthermore, the present study contrasts item sets of stigmatized and non-stigmatized variation so as to look for putative differences in the influence of the experimental instructions as a consequence of the semi-conscious application of linguistic rules in the former case (cf. Schütze 1996).

2 Background

Evidence that a manipulation of the scale dimension affects the output of linguistic judgment experiments is limited. Cowart (1997) presents an experiment in which two participant groups took part in the same task but under different instructions: ‘intuitive instructions’ based on personal judgment criteria and ‘prescriptive instructions’ eliciting judgments of well-formedness. This experiment did not yield any differences which are particularly relevant to linguistic theory. Langsford et al. (2019) investigate potential differences between judgments of acceptability and (confidence of) grammaticality for various grammatical illusion phenomena, as well as a set of judgment contrasts from *Linguistic Inquiry* (2001–2010). They find that reported judgments may differ somewhat between the two dimensions, but the instructions at least do not impinge on the relative acceptability between conditions. Turning now to cases of stigmatized variation, Bennis and Hinskens (2014) investigate judgments about ten different prescriptive norm violations in Dutch using a large scale questionnaire. Participants rated the norm violations on four scales (*good–bad*, *ugly–beautiful*, *sloppy–diligent*, *dialect–standard language*). The judgments were remarkably similar across the board, with only the last scale as a notable exception. Participants were thus at least sufficiently invested in the experiment to engage with its instructions, yet the scale manipulation did not yield linguistically relevant differences. Vogel (2019) investigates three German norm violations, eliciting judgments in terms of their

normativity, possibility, and aesthetic quality. Although the judgment scores of aesthetic quality were only slightly lower than those of normativity (18.7% and 24.5% respectively), the scores of probability were much higher (36.1%). Thus, prescriptive norm violations are not particularly good or pretty, but they do exist in the linguistic reality and speakers are aware of this. Crucially, here the instructions do influence the relative judgments between conditions.

It thus seems as though the dimension of the scale may impact the outcome of a linguistic judgment experiment, yet evidence is scarce and the reported effects so far are limited to cases of stigmatized variation. The present study replicates Vogel's (2019) experiment of prescriptive norm violations for Dutch, and expands on it by including an item set of non-stigmatized variation, *viz.* scrambling (see Schoenmakers 2022). More specifically, definite objects in the Dutch middle-field may appear on the left or right side of a clause adverb, a type of word order variation which has been argued to be driven by topic-focus structure and, crucially, which is not associated with sociolinguistic stigmatization.

3 The experiment

The experiment was an online questionnaire and contained two distinct item sets: one item set with prescriptive norm violations (stigmatized variation) and another with scrambling constructions (non-stigmatized variation). Three types of prescriptive norm violations were included in the experiment: subject *hun* 'them', comparative *als* 'as', and auxiliary *doen* 'do'. Items from this set either did or did not contain a norm violation and were additionally grammatical or ungrammatical (following Vogel 2019). The scrambling items contained a definite object on either side of a clause adverb. A sample target sentence from the scrambling item set is given in (1).

- (1) *Nora gaat (het museum) absoluut (het museum) bezoeken.*
Nora goes the museum absolutely the museum visit

All experimental items were preceded by three-sentence preambles, which in the scrambling items licensed the object as the topic or focus, in order to test for the 'discourse template' (see Schoenmakers 2020). Both item sets thus had a 2×2 design (\pm violation \times \pm grammatical in the stigmatized item set; *object position* \times \pm topicality in the scrambling set). 153 participants (M_{age} 48.51, range 18–91, $SD = 20.89$) rated 108 sentences (36 norm violations, 24 scrambling, and 48 fillers) in one of three dimensions, illustrated in (2) (i.e. *dimension* was a between-subjects factor). Judgments were given on a slider scale from 0–100%.

- (2) a. **Aesthetic judgment:**
Hoe mooi vind je de formulering van de bovenstaande zin?
How pretty do you find the wording of the above sentence?
- b. **Acceptability judgment:**
Hoe goed vind je de bovenstaande zin als Nederlandse constructie?
How good do you find the above sentence as a construction of Dutch?
- c. **Probability judgment:**
Hoe waarschijnlijk vind je het dat de bovenstaande zin is uitgesproken door een moedertaalspreker van het Nederlands?
How likely do you think it is that the above sentence has been uttered by a native speaker of Dutch?

4 Results

The judgment patterns of the raw scores are visually presented in Figure 1 for both item sets. The standardized scores were entered into two LMER models, with the above-mentioned factors and the scale dimension as fixed effects (with *acceptability* set as the reference category). The random structure of the models contained by-participant and by-item intercepts and slopes for the effects of both fixed factors and the by-participant (stigmatized) or by-item (scrambling) interaction. The full model specifications are given in Tables 1 and 2.

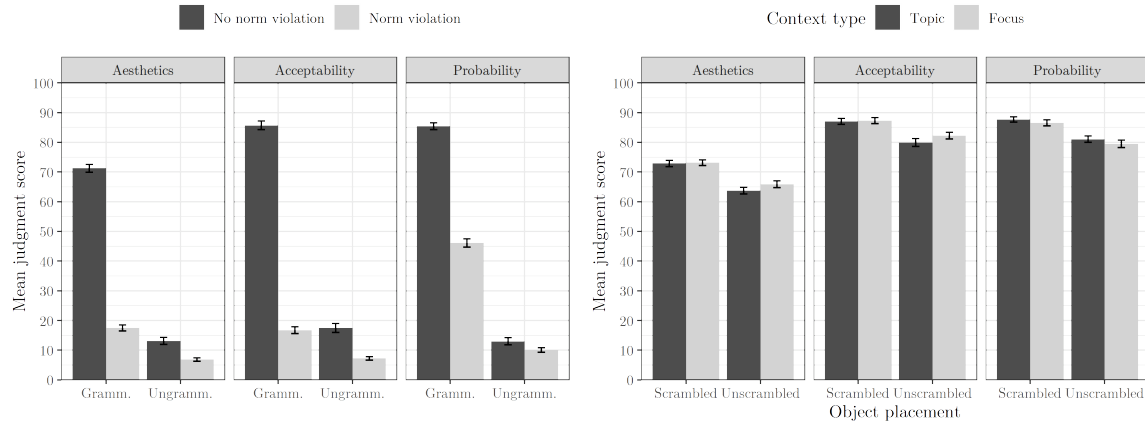


Figure 1: Mean judgment scores per condition for the two item set in three dimensions (error bars indicate within-subject standard errors from the mean)

Parameters	Fixed effects				Random effects (SDs)	
	β	Std. Error	t-value	p	by-participant	by-item
(Intercept)	-0.475	0.023	-20.452	< .001	0.082	0.095
norm violation	-0.996	0.055	-18.248	< .001	0.210	0.230
grammaticality	0.994	0.071	14.029	< .001	0.220	0.351
dimension (aesthetics)	-0.011	0.023	-0.492	.623	-	-
dimension (probability)	0.086	0.023	3.719	< .001	-	-
dimension (aesthetics) * norm violation	0.097	0.053	1.823	.068	-	-
dimension (probability) * norm violation	0.439	0.053	8.328	< .001	-	-
dimension (aesthetics) * grammaticality	0.066	0.055	1.203	.229	-	-
dimension (probability) * grammaticality	0.440	0.054	8.077	< .001	-	-
norm violation * grammaticality	-1.467	0.074	-19.752	< .001	0.388	-
dimension (aesthetics) * norm violation * grammaticality	0.009	0.102	0.086	.931	-	-
dimension (probability) * norm violation * grammaticality	0.544	0.101	5.392	< .001	-	-

Table 1: Model specifications of the linear mixed-effects model for the stigmatized item set (number of observations: 5506, groups: participant, 153; item, 36)

Parameters	Fixed effects				Random effects (SDs)	
	β	Std. Error	t-value	p	by-participant	by-item
(Intercept)	0.842	0.036	23.281	< .001	0.126	0.139
topicality	0.026	0.029	0.888	.374	0.054	0.058
object position	0.164	0.048	3.429	< .001	0.227	0.113
dimension (aesthetics)	-0.079	0.031	-2.564	.010	-	-
dimension (probability)	-0.046	0.030	-1.519	.129	-	-
dimension (aesthetics) * topicality	-0.015	0.036	-0.416	.677	-	-
dimension (probability) * topicality	-0.048	0.036	-1.327	.185	-	-
dimension (aesthetics) * object position	0.089	0.058	1.549	.121	-	-
dimension (probability) * object position	0.004	0.057	0.064	.949	-	-
topicality * object position	-0.038	0.061	-0.624	.533	-	0.168
dimension (aesthetics) * topicality * object position	-0.009	0.069	-0.132	.895	-	-
dimension (probability) * topicality * object position	0.027	0.068	0.393	.694	-	-

Table 2: Model specifications of the linear mixed-effects model for the scrambling item set (number of observations: 3671, groups: participant, 153; item, 24)

Statistical analysis led to the following conclusions:

- i. Prescriptive norm violations were rated higher on the scale of probability (46.1%) than on the scales of acceptability (16.7%) and aesthetic quality (17.5%), but lower than unmarked sentences on all three scales (71–86%). The difference between judgments of probability and acceptability was significant, and the effect was moderated by \pm violation and \pm grammatical. Further, the three-way interaction was also significant. The experimental instructions thus influenced judgment behavior in the case of stigmatized variation.

- ii. The manipulation of scale dimension did not impinge on the relative acceptability between conditions in items with a scrambling configuration. Thus, the present experiment does not provide evidence for an effect of the experimental instructions in non-stigmatized variation or for the idea that scrambling adheres to a ‘discourse template’, which is commonly assumed in the literature (see Schoenmakers 2020).
- iii. Unmarked items (grammatical fillers, non-violations, scrambling configurations) were rated considerably lower on the scale of aesthetic quality than on the two other scales (by at least a 10-point margin numerically). The difference between the dimensions of acceptability and aesthetic quality was significant in the scrambling set but not in the stigmatized item set.
- Taken together, the results indicate that participants take into account the scale dimension, in both stigmatized and non-stigmatized variation, but the effects depend on the type of variation.

5 Theoretical implications

A crucial difference between the norm violations and the scrambling items is that participants have conscious access to the prescriptive rules of their language. Schütze (1996: 83) notes that “we could imagine that expected judgment causes people to revert to conscious reasoning *about* sentences, rather than processing *of* them.” In case the norm is violated, the judgments of acceptability and aesthetic quality may thus reflect a binary opposition, in that the sentence either does or does not match a prescriptively correct form, whereas frequency estimations are much more open-ended (cf. Featherston 2021). Although work on norm violations is currently only limited, this type of rationale can explain findings of previous experimental studies as well (e.g. Hubers et al. 2020). The new results therefore call for much needed future research on norm violations and the way in which they are judged and processed.

Regarding the non-stigmatized (scrambling) item set, the findings do not provide much new evidence, but they do not reject Schütze’s (1996) theory of judgments either. Only main effects of the instructions were found (cf. Cowart 1997). One could argue that the new findings imply that the reported reactions do not reflect technical introspection in the Wundtian sense; rather, they consist of an acceptability core with additional effects from other cognitive processes that influence judgments of aesthetic quality and presumably other dimensions of judgment.

That is to say, judgments may reflect an amalgamation of (partially non-instructed) considerations on the part of the participant, and these can be vastly different conceptually. The data may thus serve as a proxy for grammaticality, but the judgment scale(s) used may have put an additional coat of paint on them. This must be considered when an attempt is made to answer the question what linguistic judgments quantify.

References

- Bennis, H., & Hinskens, F. (2014). Goed of fout: Niet-standaard inflectie in het hedendaags Standaardnederlands. *Nederlandse Taalkunde*, 19(2), 131–184.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. SAGE.
- Featherston, S. (2021). Response methods in acceptability experiments. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 39–61). Cambridge University Press.
- Goodall, G. (2021). Sentence acceptability experiments: What, how, and why. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 7–38). Cambridge University Press.
- Hubers, F., Redl, T., de Vos, H., Reinartz, L., & de Hoop, H. (2020). Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking [Article 186]. *Frontiers in Psychology*, 11.
- Langsford, S., Stephens, R., Dunn, J., & Lewis, R. (2019). In search of the factors behind naive sentence judgments: A State Trace Analysis of grammaticality and acceptability ratings [Article 2886]. *Frontiers in Psychology*, 10.
- Schindler, S., Drożdżowicz, A., & Bröcker, K. (2020). *Linguistic intuitions: Evidence and method*. Oxford University Press.
- Schoenmakers, G. (2020). Freedom in the Dutch middle-field: Deriving discourse structure at the syntax-pragmatics interface [Article 114]. *Glossa: A journal of general linguistics*, 5(1).
- Schoenmakers, G. (2022). *Definite objects in the wild: A converging evidence approach to scrambling in the Dutch middle-field* (Doctoral dissertation). Radboud University, Nijmegen.
- Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology* [Reprinted in 2016 by Language Science Press]. University of Chicago Press.
- Sprouse, J. (2020). A user’s view of the validity of acceptability judgments as evidence for syntactic theories. In S. Schindler, A. Drożdżowicz, & K. Bröcker (Eds.), *Linguistic intuitions: Evidence and method* (pp. 215–232). Oxford University Press.
- Vogel, R. (2019). Grammatical taboos. *Zeitschrift für Sprachwissenschaft*, 38(1), 37–79.

Topic position or prefield? – Disentangling the positional restriction of topic drop in German based on acceptability rating data

Lisa Schäfer (Saarland University)

This paper is concerned with the restriction of topic drop in German to the preverbal prefield position. Using acceptability rating data from four experiments, I argue (i) that topic drop is not necessarily the omission of a topic, and (ii) that topic drop is restricted to the prefield, (iii) in particular to the highest prefield position of an independent clause or to a prefield at the left edge of the utterance, but that (iv) topic drop is not bound to an utterance-initial positioning.

1 Overview

Germanic V2 languages like German systematically allow for a phenomenon called topic drop (TD), i.e. the omission of the preverbal constituent from a declarative sentence in spoken language or conceptually spoken text types (Fries 1988), see (1). As the term implies, it is often taken for granted in the literature that TD is the omission of the sentence topic (e.g. Helmer 2016). In the following, I will show that this view is unjustified in German (section 2). Instead, I will argue for a structural restriction of TD to the preverbal position of V2 clauses, the prefield in terms of the topological field model, and that it can potentially be captured as a restriction to the highest syntactic position, but not to the first element of an utterance (section 3).

(1) Δ Kann heute leider nicht kommen.
can today unfortunately not come

2 Topic drop as the omission of the (sentence) topic

Equating TD with dropping the sentence topic, i.e., the entity under which the comment information should be stored in the common ground (Krifka 2007), from the prefield implies a close connection between topicality and TD, e.g., topicality could be (i) a sufficient, i.e., any topical prefield constituent is omissible, and/or (ii) a necessary condition for TD, i.e., any omissible prefield constituent is topical. I will show that there are counterexamples for both predictions.

2.1 Topicality as sufficient or necessary condition for topic drop

(i) **Sufficiency** is questioned by non-droppable topics such as contrastive topics or topics which cannot be recovered like *Hans* in example (2). A and B both know Hans, but he is not present in the current discourse situation so that *Hans* cannot be omitted because a hearer would not be able to recover the reference of the TD. Topicality, then, is at least not strictly sufficient for TD; the lack of givenness in context can block the omission, as can the combination with focus in contrastive topics.

(2) A to B: Übrigens: *(Hans) hat letzte Woche geheiratet. (Krifka 2007: 43, adapted)
by.the.way Hans has last week married

(ii) **Necessity** can be refuted by corpus examples with dropped expletive subjects similar to (3a) in Ruppenhofer (2018). Expletives are not referential and hence cannot be topical (e.g. Lambrecht 1994). In defense of the necessity of topicality, Trutkowski (2016) suggests that the expletive subjects of weather verbs can be topical and thus droppable, but only when they refer to the current situation that must be present in the utterance context (3a vs. 3b).

(3) a. Δ Regnet grad. [uttered while looking out of the window]
rains right.now
b.* Δ Regnet bestimmt, wenn wir in Urlaub fahren. (Trutkowski 2011: 120, her judgments)
rains definitely when we in vacation go

2.2 Experiment 1: Topic drop of expletives

I test Trutkowski's (2016) account with an acceptability rating study on TD of the expletive subjects of weather verbs. I varied whether the current situation is present or not by presenting the target utterance in the context of a question that does or does not ask about the weather (4a vs. 4b). This results in a 2×2 design (COMPLETENESS (full form vs. TD) \times QUESTION TYPE (weather vs. other)). If topicality is a necessary condition, TD of non-topical expletives should

generally be degraded, while Trutkowski predicts an interaction, i.e., that TD is only degraded after a question not asking about the weather, i.e., not mentioning the current situation.

- (4) a. Was macht das Wetter bei dir? (weather question)
what makes the weather at you
 b. Wolltest du nicht joggen gehen? (other question)
wanted you not jog go
 (Es) regnet leider schon wieder ziemlich heftig
it rains alas already again pretty heavily

38 native speakers of German¹ recruited from the crowdsourcing platform Clickworker rated 24 items like (4) presented as instant messages with one of six weather verbs on a 7-point Likert scale (7 = completely natural), along with 80 fillers. I analyzed the data in R with CLMMs (Christensen 2019). The final model contained only a significant main effect of QUESTION TYPE ($\chi^2 = 12.05$, $p < .001$): Utterances after weather questions were rated as more acceptable (Fig. 1). This can be explained by pragmatics: An answer with a weather verb is more coherent after a weather question. With respect to the necessity of topicality, the result questions the general prediction and Trutkowski's (2016) account: TD is as acceptable as the full forms regardless of the question type. In conclusion, TD of non-topical expletives is possible, and topicality is neither a (strictly) sufficient nor a necessary condition for TD.

3 Topic drop as the omission from the prefield

Thus, TD should not be considered a topic omission, but an omission from the prefield. While the majority of research takes TD's prefield restriction for granted, Helmer (2016) argues that TD is also possible in the middle field. This is the motivation to test the prefield restriction.

3.1 Experiment 2: Topic drop in prefield vs. middle field position

In a COMPLETENESS \times TOPOLOGICAL POSITION (prefield vs. middle field) acceptability rating study 45 German Clickworkers rated 24 items like (5) and 72 fillers on a 7-point Likert scale.

- (5) *Context (in German): 'A: What do you have planned for tonight? B: We want to watch the new Matrix movie in the theater 😊 I totally like the first three movies'*
- a. (Ich) bin jetzt richtig gespannt auf den neuen Teil (prefield)
I am now rightly keen on the new part
 b. Jetzt bin (ich) richtig gespannt auf den neuen Teil (middle field)
now am I rightly keen on the new part

I analyzed the data with CLMMs. The main result is a significant COMPLETENESS \times TOPOLOGICAL POSITION interaction ($\chi^2 = 28.18$, $p < .001$): TD in the middle field is degraded compared to TD in the prefield (Fig. 2). (Future research needs to clarify whether the fact that TD in the middle field is not so bad in absolute terms could be due to some participants unconsciously inserting the missing pronoun in the middle field.) The result of experiment 2 supports the prefield restriction of TD, which I specify below.

Freywald (2020: 167), following Rizzi (1994), argues that TD is not restricted to any prefield position, but "dass Topik-drop nur in der höchsten Position eines selbstständigen Satzes stattfinden kann" ('that TD can only occur in the highest position of an independent clause'), i.e. in the highest [Spec, CP] of V2 clauses. This reasoning predicts that TD is impossible in embedded V2 clauses (Rizzi 1994; contra Trutkowski 2016), which I test in experiment 3.

3.2 Experiment 3: Topic drop in embedded prefields

In a 2 \times 3 acceptability rating study (COMPLETENESS \times EMBEDDING (initial vs. final vs. none)) 45 German Clickworkers rated 24 items like (6) on a 7-point Likert scale along with 72 fillers. If TD is possible in any prefield position, embedded TD should be acceptable. If TD is restricted to the syntactically highest prefield position, embedded TD should be degraded.

- (6) *Context (in German): 'A: What's new from Tim?'*

¹Participant numbers in this abstract are numbers after exclusions based on ungrammatical catch trials.

- a. (Er) hat seine neue Freundin betrogen, hat er mir am Freitag gebeichtet. (initial)
he has his new girlfriend cheated has he me on Friday confessed
- b. Am Freitag hat er mir gebeichtet, (er) hat seine neue Freundin betrogen. (final)
on Friday has he me confessed he has his new girlfriend cheated
- c. (Er) hat seine neue Freundin betrogen. (none)
he has his new girlfriend cheated

I analyzed the data with CLMMs using forward coding for the predictor EMBEDDING, which results in two variables (NONE VS. INITIAL/FINAL, NONE/INITIAL VS. FINAL). The main result is a significant COMPLETENESS × NONE/INITIAL VS. FINAL interaction ($\chi^2 = 9.96$, $p < .001$) indicating that final embedded TD is degraded (see Fig. 3). This suggests that TD is not possible in every prefield position. The interaction COMPLETENESS × NONE VS. INITIAL/FINAL is not significant, indicating that initial embedded TD is acceptable. There are two potential explanations for this pattern: 1) Embedded TD is acceptable, but only when it is positioned at the beginning of the utterance, potentially allowing for a better linking to the discourse (Trutkowski 2016). 2) The initial conditions are not really embeddings but independent V2 clauses with V1 parentheticals (Reis 1997; cf. Pauly 2013), so that in this condition TD would occur in the highest [Spec, CP] in line with Freywald's (2020) account. In sum, exp. 3 shows that a positioning in the prefield is not sufficient for TD and that TD might be restricted to the highest clause position.

Related to explanation 1) and to characterizations of TD's position as sentence-initial (Huang 1984, Trutkowski 2016) I tested whether TD must be the very first element of an utterance (which I call utterance-initial), or whether conjunctions can precede it.

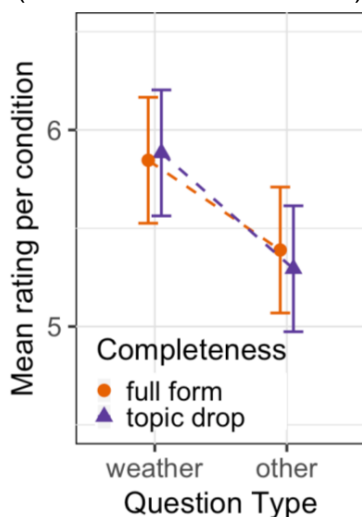


Fig. 1: Mean rating and 95% CIs for Exp. 1.

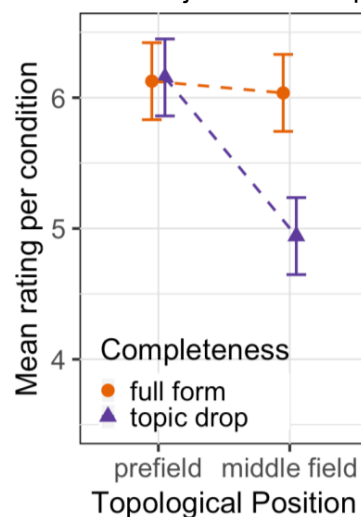


Fig. 2: Mean rating and 95% CIs for Exp. 2.

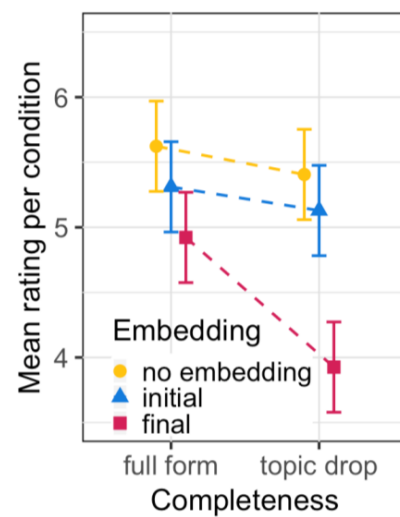


Fig. 3: Mean rating and 95% CIs for Exp. 3.

3.3 Experiment 4: Topic drop in non-initial position after conjunctions

In a COMPLETENESS × PRESENCE OF CONJUNCTION (PC) (present vs. absent) × SUBJECT GAP (licensed vs. blocked) acceptability rating study 58 German Clickworkers rated 24 items like (7-8) and 72 fillers on a 7-point Likert scale. The conjunction type was varied between items (8 × each *und* ('and'), *aber* ('but'), *denn* (parordinating 'because')). The control predictor SUBJECT GAP should ensure that TD after conjunctions is not in fact a cross-clausal subject gap construction where the subject *ich* ('I') is shared between clauses (Wilder 1997). Since such a reading of the target utterance (8) is only licensed when the speaker is the subject of both utterances, I manipulated exactly this: The speaker appears as subject in (7a), licensing a subject gap interpretation, and as object pronoun in (7b), blocking it. If TD after conjunctions is indeed a subject gap, only those TD conditions should be acceptable where such a reading is licensed.

- (7) Context (in German): 'A: What do you have planned for tonight? B: We want to watch the new Matrix movie in the theater 😊'

- | | |
|---|------------------------|
| a. Die ersten drei Filme mag ich total.
<i>the first three movies like I totally</i> | (subject gap licensed) |
| b. Die ersten drei Filme gefallen mir total gut.
<i>the first three movies please me totally well</i> | (subject gap blocked) |
| (8) a. Und (ich) bin jetzt richtig gespannt auf den neuen Teil.
<i>and I am now rightly keen on the new part</i> | (conjunction present) |
| b. (Ich) bin jetzt richtig gespannt auf den neuen Teil. | (conjunction absent) |

I analyzed the data first jointly and then post-hoc for each conjunction type separately with CLMMs. For *denn* I found a significant COMPLETENESS × PC interaction ($\chi^2 = 4.31$, $p < .05$) and significant main effects of PC ($\chi^2 = 17.4$, $p < .001$) and COMPLETENESS ($\chi^2 = 8.6$, $p < .01$): Utterances with TD and utterances with overt *denn* were degraded, TD with *denn* was particularly bad. The interaction was also present in the complete data ($\chi^2 = 6.02$, $p < .05$), but must have been caused there exclusively by *denn*, since for *und* and *aber* TD was rated as acceptable as the full forms after these conjunctions (interaction_{*und*}: $\chi^2 = 2.01$, $p > 0.1$; interaction_{*aber*}: $\chi^2 = 0.87$, $p > 0.3$, see Fig. 4). SUBJECTGAP did not have an impact on the acceptability of TD. In sum, exp. 4 suggests that TD does not require utterance-initial positioning, but that conjunctions like *und* and *aber* can precede it clause-internally or -externally regardless of a potential subject gap reading. *Denn* might deviate because it is syntactically parordinating, but semantically subordinating (Reich & Reis 2013), exhibiting properties similar to the final embeddings above.

4 Conclusion

I presented four experiments in which I systematically investigated theoretical claims concerning TD's positional restriction: I showed that TD in German is not restricted to topics – topicality is neither (strictly) sufficient nor necessary (exp. 1). TD occurs at least preferably if not obligatorily in the prefield (exp. 2). It is restricted to the highest prefield of independent clauses or at least to a prefield at the left edge (exp. 3). The latter does not entail utterance-initial positioning (exp. 4). In future research I will explore two potential explanations for these properties: (i) TD's left edge position might allow for an easy linking to the discourse. (ii) Syntactically, TD might involve an empty element that is restricted to positions where it must not be c-commanded sentence-internally (Rizzi 1994; see Stowell 1991 for a similar idea for null articles).

Selected references

- Christensen, R. H. B. (2019). *Ordinal. Version 2019.12-10*. • Freywald, U. (2020). Nach dem Öffnen rasch verbrauchen! Objektauslassung in direktiven Infinitiven – ein Fall von Topikdrop. In R. Külpmann, L. Neuhaus, & V. Symanczyk Joppe (Eds.), *Variation in der Argumentstruktur des Deutschen* (pp. 147–177). Buske. • Helmer, H. (2016). *Analepsen in der Interaktion: Semantische und sequenzielle Eigenschaften von Topik-Drop im gesprochenen Deutsch*. Universitätsverlag Winter. • Krifka, M. (2007). Basic Notions of Information Structure. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *The Notions of Information Structure* (pp. 13–55). Universitätsverlag Potsdam. • Rizzi, L. (1994). Early Null Subjects and Root Null Subjects. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (pp. 151–176). Benjamins. • Trutkowski, E. (2016). *Topic Drop and Null Subjects in German*. De Gruyter.

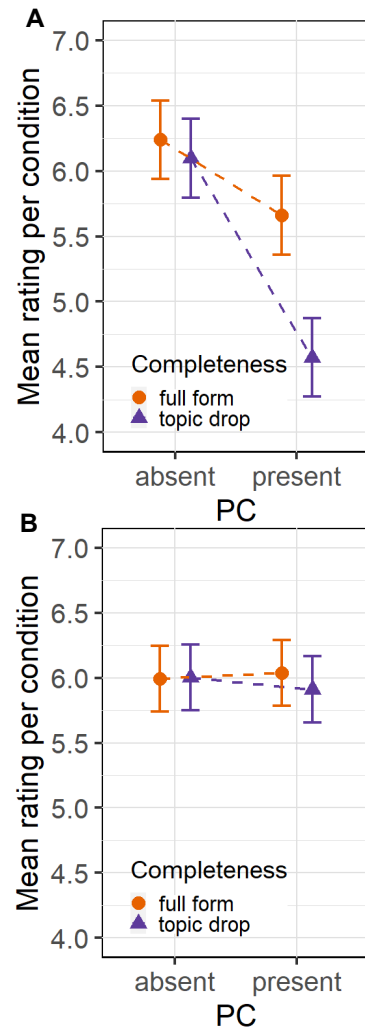


Fig. 4: Mean rating and 95% CIs for exp. 4 for *denn* (A) and *und* + *aber* (B)

The ‘human-like’ learnability of the *wh*- and coordination island constraints by artificial neural networks in Dutch

Michelle Suijkerbuijk, michelle.suijkerbuijk@ru.nl, Radboud University
Stefan Frank, stefan.frank@ru.nl, Radboud University
Peter de Swart, peter.deswart@ru.nl, Radboud University

The current study investigated whether an artificial neural network (ANN) can learn the *wh*- and coordination island constraints in Dutch, and directly compared its performance to that of native speakers. First, speakers’ sensitivity to these constraints was assessed with an acceptability judgement task, and subsequently the same sentences were presented to an ANN. While the Dutch speakers seem to show a sensitivity to *wh*- and coordination island violations, the ANN does not seem able to similarly learn to recognize these gap-resistant structures.

1 Introduction

Nearly all children acquire the syntax of their first language with ease, but how is that actually possible? While some have argued that humans need innate knowledge of language (Chomsky, 1971), recent research suggests that artificial neural networks (ANNs) can induce human-like grammatical knowledge without having linguistic abilities built in (Linzen & Baroni, 2021). However, this recent research almost all (1) use English as the input language, and (2) do not directly compare the ANN’s performance to that of actual native speakers, which makes it impossible to state that the ANN has acquired knowledge that is ‘human-like’. Therefore, we investigated whether an ANN can learn syntactic constraints **in Dutch**, and **directly compared the ANN’s performance to that of Dutch native speakers**. Specifically, we examined the learnability of syntactic island constraints.

2 Theoretical background

Syntactic island constraints are conditions on non-local dependency relations, prohibiting movement out of syntactic islands (Liu et al., 2022), such as *wh*-phrases or coordinations. These constraints have played an important role in the development of syntactic theories, and their predominant analysis in these theories relies heavily on the assumed innate language ability (Pearl & Sprouse, 2013). An example of a *wh*- and a coordination island violation can be found in (1), taken from Liu et al. (2022, p. 497).

- (1) a. *What did you wonder [_{wh-island} whether John bought <*what*>]?
a. *What did John buy [_{coordination island} a shirt and <*what*>]?

With regard to research on syntactic island constraints, either with ANNs or native speakers, Dutch is an underrepresented language. Previous studies with ANNs on island constraints have mainly been performed in English and show mixed results; some islands, such as *wh*- and coordination islands, are learned successfully, but others only partially or not at all (e.g., Wilcox et al., 2021). As it would be interesting to see whether these successfully learned constraints in English can also be learned in another language, typologically similar but also different from English, and underrepresented in the current state of affairs, the current study investigated whether these two island types could also be successfully learned by ANNs in Dutch. First, however, experimental data of native Dutch speakers on these island types had to be gathered as not much is known about whether these island constraints exist in Dutch, and if so, to what extent its speakers are sensitive to them. Beljon et al. (2021) is one of the few, if not the only, study that empirically investigated Dutch native speakers’ sensitivity to islands, specifically to *wh*-islands, and showed that they are strongly sensitive to the *wh*-island constraint. For the coordination island constraint, it has been argued that it is never possible, in any language, to extract an element out of (part of) a conjunct, although there is almost no

experimental evidence to support or oppose that claim (Liu et al., 2022). Due to this (near) absence of experimental evidence for *wh*- and coordination islands in Dutch, the current study gathered human experimental data to first establish whether the *wh*- and coordination island constraints exist in Dutch and if so, to what extent native speakers are sensitive to them. Next, to find out more about the cross-linguistic abilities of the ANNs, it was investigated whether a network demonstrates a human-like sensitivity to the island constraints in Dutch.

3 Methodology

3.1 Experimental design and materials

The interaction design used in the current study (based on Wilcox et al., 2021) builds on two predictions assumed to be made by the grammar: (1) gaps require fillers, and (2) fillers require gaps. Consequently, the independent variables PRESENCE OF GAP and PRESENCE OF FILLER were included in the design, resulting in four conditions, shown in a regular declarative sentence in Table 1.

Table 1

Conditions of PRESENCE OF FILLER and PRESENCE OF GAP in a regular declarative sentence.

Gap?	Filler?	Example sentence
No	No	<i>Ik weet dat jij zag dat de bakker <u>koekjes</u> maakte in de bakkerij.</i> I know that you saw that the baker cookies made in the bakery
No	Yes	<i>*Ik weet wat jij zag dat de bakker <u>koekjes</u> maakte in de bakkerij.</i> I know what you saw that the baker cookies made in de bakery 'I know (w/t)hat you saw that the baker made cookies in the bakery.'
Yes	No	<i>*Ik weet dat jij zag dat de bakker ___ maakte in de bakkerij.</i> I know that you saw that the baker GAP made in the bakery
Yes	Yes	<i>Ik weet wat jij zag dat de bakker ___ maakte in de bakkerij.</i> I know what you saw that the baker GAP made in the bakery 'I know (w/t)hat you saw that the baker made in the bakery.'

Furthermore, the independent variables ISLAND (non-island vs. island) and STRUCTURE (*wh*-island vs coordination island) were added to the design to compare non-islands to two island types. An example of the [-gap, -filler] condition within a *wh*-island can be found in (2).

- (2) *Ik weet dat jij je afvraagt of de bakker koekjes maakte in de bakkerij.*
I know that you REFL¹ wonder whether the baker cookies made in the bakery
'I know that you wonder whether the baker made cookies in the bakery.'

Additionally, as it has been suggested that native speakers and ANNs simply cannot thread information through the syntactically complex islands (Chowdhury & Zamparelli, 2018; Keshev & Meltzer-Asscher, 2019), control items were added to the experiment, in which participants had to maintain expectations for gendered pronouns either through a *wh*-/coordination island or a non-island configuration. An example of a *wh*-island control item can be found in (3).

- (3) *Ik weet dat de meester zich afvraagt of de leerlingen (zijn/?haar) uitleg begrijpen.*
I know that the male.teacher REFL wonders whether the students his/her explanation understand
'I know that the teacher wonders whether the students understand (his/?her) explanation.'

¹ REFL = reflexive pronoun

3.2 Participants, procedure and hypotheses

88 native Dutch speakers ($M_{age} = 19.61$, range 17-33, $SD_{age} = 2.04$) judged 160 sentences (64 experimental items, 32 control items, 64 fillers) on their acceptability in Dutch on a scale from 1 (*Erg slecht* ‘very bad’) to 7 (*Erg goed* ‘very good’). Moreover, a Long Short-Term Memory (LSTM) network, trained on 12 million sentences extracted from the Dutch *Corpora Of the Web* (NLCOW14, Schäfer, 2015), assigned surprisal values to the same experimental and control items, indicating the extent to which a word was unexpected by the network.

Before the creation of this experiment, the following hypotheses were made. First, within regular filler-gap dependencies, gaps should be less acceptable and more surprising when no *wh*-filler is present. Within island configurations, however, it should always be ungrammatical to create a gap within an island configuration, regardless of the presence of a filler. Second, filled argument positions should be less acceptable and more surprising when a *wh*-filler is present both within regular filler-gap dependencies and within island configurations; encountering a *wh*-filler should always give rise to the expectation of a gap, and not encountering this gap anywhere in the sentence should make the sentence less acceptable and more surprising.

4 Results

4.1 Acceptability judgement task

To start, the control items showed that Dutch native speakers can maintain gender expectancies through island configurations, as the gender matches were rated as more acceptable than the mismatches. This shows that they are capable of threading information through these complex structures.

The standardized acceptability judgements, illustrated in Figure 1A, were analysed in an LMER model, which revealed a significant interaction between PRESENCE OF FILLER, PRESENCE OF GAP and ISLAND ($\beta = -.04$, $SE = .02$, 95% CI of $\beta = [-.08, -.00]$, $p = .030$). First, for regular filler-gap dependencies, the figure shows that gaps are judged as less acceptable with no *wh*-filler present in the sentence, and that filled argument positions are perceived as less acceptable when there is a *wh*-filler in the sentence. Second, for island configurations, it can be seen that the presence of a *wh*-filler decreases the acceptability ratings of both sentences with and without gaps within the islands.

4.2 Long Short-Term Memory network

The surprisal values measured on the immediate post-gap verb, illustrated in Figure 1B, were analysed in an LMER model, which also revealed a significant interaction between PRESENCE OF FILLER, PRESENCE OF GAP and ISLAND ($\beta = 1.79$, $SE = .66$, 95% CI of $\beta = [.50, 3.10]$, $p = .007$). The figure shows exactly the same pattern for both regular filler-gap dependencies and island configurations; gaps are judged as more surprising with no *wh*-filler present in the sentence, but surprisal values do not seem affected by the presence of a *wh*-filler in sentences with filled argument positions. The interaction effect found thus seems to be solely driven by the difference in effect size as opposed to effect direction. Also important to note here is that the control items with gender expectations remained inconclusive, and can thus not provide the control that was aimed for.

5 Discussion and theoretical implications

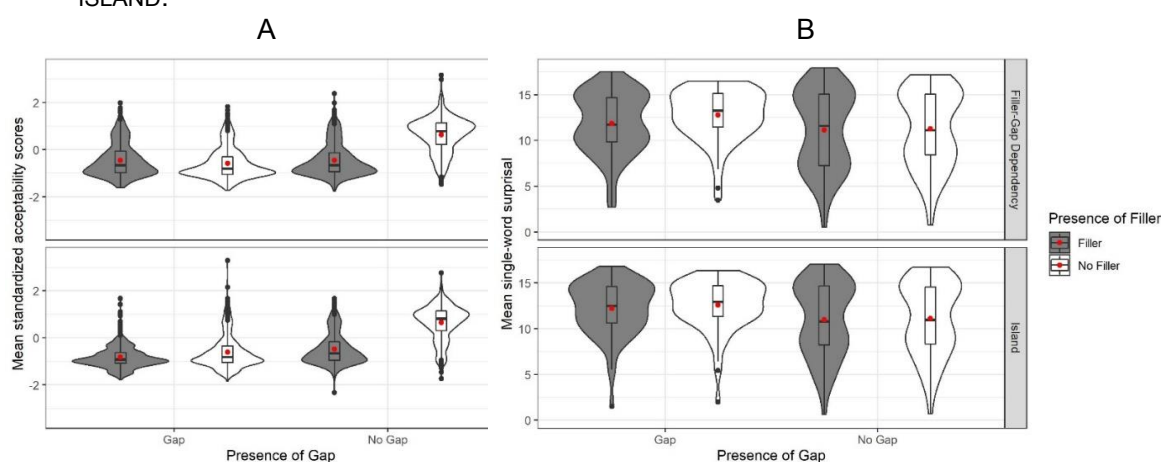
While the human native speakers of Dutch show sensitivity to *wh*- and coordination island violations, the LSTM network did not learn to recognize these gap-resistant structures. This could be due to various reasons either specific to this research project, such as the complexity of the items used, the analysis of the surprisal values and the training of the neural network architecture, or due to a more general reason, such as the structural properties of Dutch or the absence of an innate language ability in the network. While future research should first resolve the research-specific issues, it is still interesting to discuss the latter explanation.

While children can already recognize syntactic islands at the age of four, the network was unable to do the same with the current training data covering a lot more than four years of a human's life (Wilcox et al., 2021). This suggests that the information in the training data was not enough for the network to learn about syntactic island constraints, and that children might thus use something else than just external input to learn the syntactic island constraints not available to the network (e.g. internal language knowledge/abilities).

While this research thus provides relevant new insights for the debate about language acquisition, it also provides relevant new knowledge to the field of experimental syntax. The current results show that the *wh*- and coordinate structure island constraints exist in Dutch, which strengthens the results found by Beljon et al. (2021) and experimentally supports the theoretical claim that “it does not seem possible to extract one or more full conjuncts” (Liu et al., 2022, p. 503), at least for Dutch.

Figure 1

(A) Violin/boxplot of standardized acceptability ratings; and (B) mean single-word surprisal values; as a function of PRESENCE OF GAP, PRESENCE OF FILLER, and ISLAND.



References

- Beljon, M., Joosen, D., Koeneman, O., Ploum, B., Sommer, N., de Swart, P., & Wilms, V. (2021). The effect of filler complexity and context on the acceptability of the *wh*-island violations in Dutch. In M. Dingemanse, E. Van Lier, & J. Vogels (Eds.), *Linguistics in the Netherlands* (Vol. 38, pp. 4–20). John Benjamins Publishing Company.
- Chomsky, N. (1971). *Problems of Knowledge and Freedom*. Fontana.
- Chowdhury, S.A., & Zamparelli, R. (2018). RNN simulations of grammaticality judgments on long-distance dependencies. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 133–144). Association for Computational Linguistics.
- Keshev, M., & Meltzer-Asscher, A. (2018). A processing-based account of subliminal *wh*-island effects. *Natural Language & Linguistic Theory*, 37(2), 621–657.
- Linzen, T., & Baroni, M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, 7(1), 195–212.
- Liu, Y., Winckel, E., Abeillé, A., Hemforth, B., & Gibson, E. (2022). Structural, functional, and processing perspectives on linguistic island effects. *Annual Reviews*, 8, 495–525.
- Pearl, L., & Sprouse, J. (2013). Computational models of acquisition for islands. *Experimental Syntax and Island Effects*, 109–131.
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)* (pp. 28–34). Institut für Deutsche Sprache.
- Wilcox, E.G., Vani, P., & Levy, R.P. (2021). A Targeted Assessment of Incremental Processing in Neural Language Models and Humans. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 939–952). Association for Computational Linguistics.

No stone was left unturned: the passivizability of Dutch idioms.

Michelle Suijkerbuijk, michelle.suijkerbuijk@ru.nl, Radboud University
Ferdie Hubers, ferdy.hubers@ru.nl, Radboud University

The current study investigates the effect of the syntactic properties of the paraphrase of an idiom's literal and figurative meaning on the extent to which idioms can passivize in Dutch. Native speakers rated the literal interpretability of idioms of which both the figurative and literal paraphrase can passivize (Type 1) and of which only the latter can (Type 2) either in canonical or passive form, or a literal control. Results show that both idiom types can be interpreted figuratively in passive voice, suggesting that only the syntactic properties of an idiom's literal parts affect the passivizability of Dutch idioms.

1 Introduction

Idioms, such as *to kick the bucket* (meaning: to die) or *to spill the beans* (meaning: to reveal a secret), are fixed expressions whose meaning goes beyond the literal meaning of their parts and are an important part of everyday communication (Kyriacou et al., 2019). Over the years, many different hypotheses have been introduced about how these fixed expressions are stored in our mental lexicon and many have been able to explain numerous idiom phenomena (e.g., Everaert, 2010; Newmeyer, 1974; Tabossi et al., 2009). However, one phenomenon these hypotheses do not agree on is the idiom's syntactic flexibility, i.e. the specific syntactic transformations idioms can undergo without losing their figurative meaning, such as passivization (e.g., *the bucket was kicked by the old man*). The existing hypotheses have made theoretical predictions about the idiom's syntactic flexibility, hypothesizing an influential role of the syntactic properties of either the idiom's parts or the paraphrase of its figurative meaning (e.g. Everaert, 2010; Newmeyer, 1974). Most of these theoretical predictions, however, have never been empirically tested, and the existing empirical research, which is limited to English and Italian, has only investigated whether idioms can undergo certain syntactic transformations, but has not yet directly tested the reason why these idioms can or cannot (e.g. Kyriacou et al., 2019; Mancuso et al., 2020). Therefore, the current study aimed to empirically test the role of the syntactic properties of the paraphrase of an idiom's literal and figurative meaning on the extent to which they can be passivized in a language underrepresented in the previous empirical research, namely Dutch. Specifically, we directly tested the theoretical predictions put forward by Newmeyer (1974) and Everaert (2010), who explain the possible transformations of idioms differently.

2 Theoretical background

Over the years, many hypotheses have been introduced about how idioms are stored in our mental lexicon and what information about an idiom is available to us. Newmeyer (1974) argues that an idiom is stored as a single lexical unit, and that it is accompanied by two semantic components: M1 (the paraphrase of the figurative interpretation) and M2 (the paraphrase of the literal interpretation). This means, for example, that the idiom *kick the bucket* is stored as a whole in our mental lexicon, and that a paraphrase of its figurative (*to die*) and literal (*to kick the bucket*) meaning are stored with it. Newmeyer argues that a transformation can only take place if it possible on both the paraphrase of the literal and the figurative interpretation. Consequently, he predicts that *kick the bucket* cannot be interpreted figuratively in passive voice as only its literal paraphrase can passivize (*the bucket was kicked by the farmer*, but not **died by him*), but that *spill the beans* can (literal: *the beans were spilled over the table*; figurative: *the secret was revealed*).

Contrary to Newmeyer (1974), Everaert (2010) predicts that idioms are not stored as single units, but that our mental lexicon consists of so-called heads. Under these heads a phonological representation, various possible meanings (literal and potential figurative), and selection criteria are stored, such as C-selection and L-selection criteria to indicate which word

categories and which specific morphemes the head can select. The paraphrase of an idiom's figurative meaning is then stored under the heads of its literal components, each paired with specific L-selection criteria. For example, the figurative meaning of *kick the bucket* (i.e. *to die*) is stored under the V(erb)-head *kick* with the criteria that the V-head needs to be combined with the noun phrase *bucket* to receive the figurative meaning *to die*; the same is then indicated under the N(oun)-head *bucket*. As the literal meaning and the original syntactic properties of the literal parts of, for instance, the idiom *kick the bucket* remain available under this head, the idiom is predicted to be able to syntactically behave as the verb *kick* can when it is used in its literal sense. Therefore, in contrast to Newmeyer, Everaert predicts that the passive *the bucket was kicked* does retain its figurative meaning, because *kick* can passivize in its literal sense as well (as in *the ball was kicked by the boy*).

In sum, Newmeyer's (1974) hypothesis bases passivizability on the syntactic properties of the paraphrase of the literal and figurative meaning, but Everaert's (2010) only on the properties of the literal meaning. The current study aimed to test both of these predictions directly.

3 Methodology

3.1 Research design and materials

The current experiment had a 2×3 within-subjects design which included two independent variables, namely IDIOM TYPE and PRESENCE OF TRANSFORMATION. First, two types of Dutch idioms were used: (1) idioms of which both the literal and the figurative paraphrase can undergo passivization (e.g. 'iets op prijs stellen' fig. *to appreciate something*; lit. *to set something at a price*) and (2) idioms of which only the literal paraphrase can passivize (e.g. 'het loodje leggen' fig. *to die*; lit. *to put the lead down*). These are referred to as Type 1 and Type 2 idioms respectively. Second, these idioms were put in a sentence, which was either presented in its canonical form (e.g. 'ik stel dat op prijs' *I appreciate that*) or in passive voice (e.g. 'dat werd op prijs gesteld door mij' *that was appreciated by me*), or the paraphrase of the figurative meaning (i.e. the control condition) was presented which could only be interpreted literally (e.g. 'ik waardeer dat' *I appreciate that*). All of these sentences were preceded by a figuratively biasing context sentence. An example of an item set of the Type 1 idiom 'het ijs breken' (lit. *to break the ice*; fig. *to start a conversation*) as used in the experiment can be found in Table 1.

Table 1

Example sentences for every condition for the Type 1 idiom 'het ijs breken' (lit. to break the ice; fig. to start a conversation).

	Example sentences
Context sentence	<i>Freek had grappige weetjes opgezocht voor zijn date in de dierentuin</i> Freek had funny facts looked.up for his date in the zoo 'Freek had looked up funny facts for his date at the zoo.'
Canonical form	<i>Hij brak hiermee het ijs aan het begin van zijn date.</i> he broke with.this the ice at the beginning of his date 'He broke the ice with this at the start of his date.'
Passive form	<i>Het ijs werd hiermee gebroken door hem</i> the ice was with.this broken by him <i>aan het begin van zijn date</i> at the beginning of his date 'The ice was broken by him with this at the start of his date.'
Literal form	<i>Hij startte hiermee het gesprek aan het begin van zijn date.</i> he started with.this the conversation at the start of his date 'He started the conversation with this at the start of his date.'

According to Newmeyer's (1974) hypothesis, idioms are only able to undergo those syntactic transformations possible on both the paraphrase of an idiom's literal and figurative meaning. If

that is true, we expect only Type 1 idioms to be able to be interpreted figuratively in passive voice. On the other hand, Everaert (2010) predicts that idioms are able to undergo the same syntactic transformations when used in its literal sense. If this is true, we predict that both Type 1 and Type 2 idioms can be interpreted figuratively in passive voice.

3.2 Participants and procedure

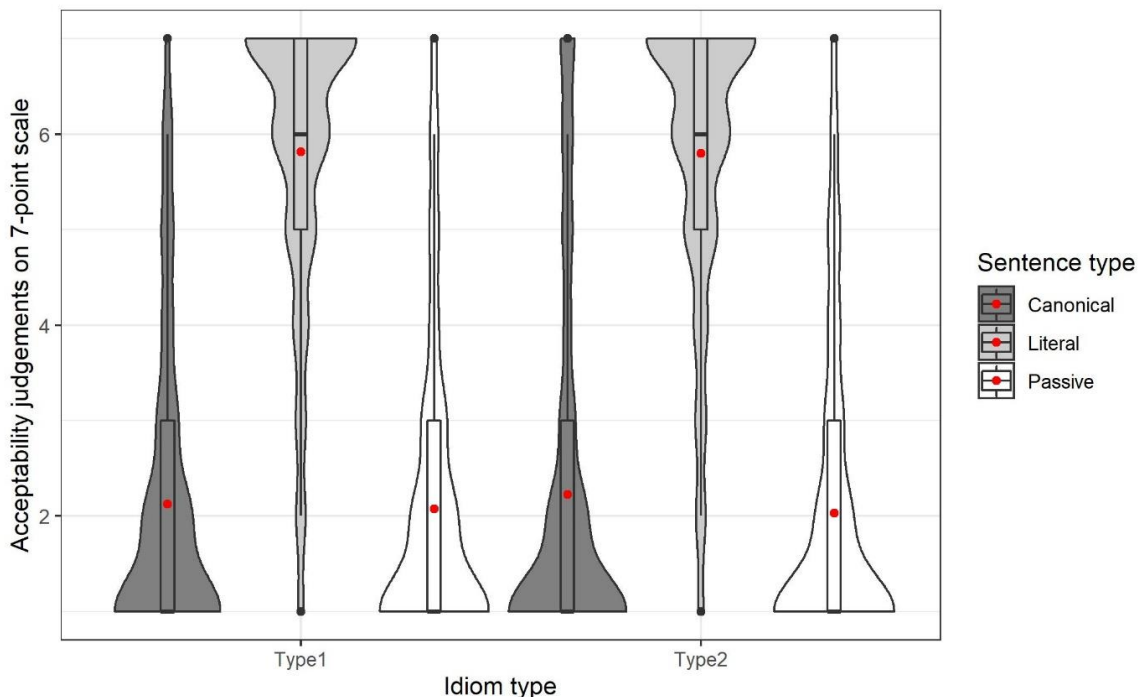
Sixty-two native Dutch speakers ($M_{age} = 19.76$, range 17-33, $SD_{age} = 2.92$) rated 64 stimulus items (8 canonical/Type1, 8 canonical/Type2, 9 passive/Type 1, 8 passive/Type2, 32 literal control) on the extent to which the target sentence following the context sentence could be interpreted literally on a scale from 1 ('Helemaal niet letterlijk' *totally not literally interpretable*) to 7 ('Helemaal letterlijk' *totally literally interpretable*).¹

4 Results

The standardized ratings were analysed using linear mixed effects regression and are visualized in Figure 1. The analysis revealed a significant main effect only of PRESENCE OF TRANSFORMATION ($F(2, 187.93) = 392.40, p < .001$); idioms in passive voice and in canonical form were rated as less literally interpretable than sentences that could only be interpreted literally, but there was no difference between idioms in passive or canonical form. No significant main effect of IDIOM TYPE ($\beta = .01, SE = .05, t = .14, p = .891$) nor any significant interaction effect between PRESENCE OF TRANSFORMATION and IDIOM TYPE was found (canonical vs. passive $\beta = .03, SE = .13, t = .26, p = .792$; passive vs. literal $\beta = .01, SE = .12, t = .10, p = .922$; canonical vs. literal $\beta = .02, SE = .12, t = .17, p = .863$). This means that Type 1 and Type 2 idioms were not rated differently, and that the ratings of the three sentence types is not significantly different for Type 2 idioms compared to Type 1 idioms, which is clearly illustrated in Figure 1.

Figure 1

Violin/boxplot with acceptability judgements on y-axis, the levels of IDIOM TYPE on the x-axis, and the levels of PRESENCE OF TRANSFORMATION representing the different colours.



¹ We decided to ask the participants to judge to what extent the sentence was literally interpretable instead of figuratively interpretable, as we thought that naïve participants would not necessarily be familiar with the term *figuratively*. However, we decided to interpret *not literally interpretable* as *figuratively interpretable*.

5 Theoretical implications

The results of the current empirical investigation show that the syntactic properties of the paraphrase of an idiom's figurative meaning do not seem to affect the extent to which Dutch idioms can be passivized; Dutch idioms can appear in passive voice if their literal parts can, even if the paraphrase of their figurative meaning cannot. Both Type 1 and Type 2 idioms were not rated differently from idioms in their canonical form when appearing in passive voice, strongly suggesting that passivized idioms can still retain their figurative meaning. This supports the hypothesis put forward by Everaert (2010), who predicts that idioms are able to undergo the same syntactic transformations as when they are used in their literal sense. However, the current results do not corroborate Newmeyer's (1974) hypothesis as he predicts that an idiom can only passivize if the paraphrase of both the literal and the figurative meaning can be passivized as well (i.e. only Type 1 idioms should be able to passivize).

Previous empirical studies already showed that English and Italian idioms, specifically those used in these previous studies, were able to passivize without losing their figurative interpretation (e.g. Kyriacou et al., 2019; Mancuso et al., 2020). However, these studies did not look at why passivization was possible for these specific idioms, for example because of their syntactic properties. The current study differentiated between two different idiom types, based on the syntactic properties of the paraphrase of the idiom's literal and figurative meaning, to investigate whether these properties influenced the extent to which an idiom could passivize. Therefore, the current study not only strengthens the findings of previous empirical investigations in English and Italian by showing that Dutch idioms are able to passivize while still retaining their figurative meaning, but it also shows that this is not constrained by the syntactic properties of the paraphrase of their figurative meaning. Idioms can passivize if their literal parts can, even if the paraphrase of their figurative meaning is not able to undergo this transformation.

In sum, the current study showed that Dutch idioms can be interpreted figuratively when appearing in passive voice, both when the idiom was Type 1 and Type 2. This strongly suggests that the syntactic properties of the paraphrase of the figurative meaning do not affect the extent to which idioms can be passivized in Dutch, but that the idiom can undergo passivization when its literal parts can as well. This is in accordance with Everaert's (2010) hypothesis.

References

- Everaert, M. (2010). The lexical encoding of idioms. In M. Rappaport Hovav, I. Sichel, & E. Doron (Eds.), *Syntax, Lexical Semantics and Event Structure* (pp. 76-99). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199544325.003.0005>
- Kyriacou, M., Conklin, K., & Thompson, D. (2019). Passivizability of Idioms: Has the Wrong Tree Been Barked Up? *Language and Speech*, 1-32. <https://doi.org/10.1177/0023830919847691>
- Mancuso, A., Elia, A., Laudanna, A., & Vietri, S. (2020). The Role of Syntactic Variability and Literal Interpretation Plausibility in Idiom Comprehension. *Journal of Psycholinguistic Research*, 49(1), 99–124. <https://doi.org/10.1007/s10936-019-09673-8>
- Newmeyer, F.J. (1974). The regularity of idiom behaviour. *Lingua*, 34, 327-342.
- Tabossi, P., Wolf, K., & Koterle, S. (2009). Idiom syntax: Idiosyncratic or principled? *Journal of Memory and Language*, 61(1), 77–96. <https://doi.org/10.1016/j.jml.2009.03.003>

Comprehending non-canonical and indirect speech acts in German

Andreas Trotzke (University of Konstanz) & Laura Reimer (University of Münster)

We compare the comprehension of non-canonical speech acts with the comprehension of indirect speech acts. Both speech act types are ‘mixed’ in the sense that they involve secondary and primary illocutionary forces. We conducted a speeded judgment experiment to test whether both types differ regarding how accurate the primary illocutionary force is identified and regarding how fast that force can be identified. Our results suggest that non-canonical speech acts are more accurately identified with their primary illocutionary force than indirect speech acts, but participants need more time to perform this identification for non-canonical speech acts than for indirect speech acts.

1 Introduction

In this paper, we present results from an experimental study on comprehending non-canonical versus indirect speech acts in German. To the best of our knowledge, comparing the comprehension of those two pragmatic categories enters new empirical territory and has not been done before in the psycholinguistic study of speech acts (see Holtgraves & Ashley 2001 for seminal work). For the category ‘non-canonical’ speech acts (NC-SA), we focused on two types of non-canonical questions: rhetorical questions and so-called surprise-disapproval questions (see Munaro & Obenauer 1999; Bayer & Obenauer 2011 for theoretical and Trotzke & Czipionka 2022 for recent experimental work). Both question types feature interrogative syntax, but at the speech-act level are interpreted either as assertions (rhetorical questions) or as complaints (surprise-disapproval questions). As for the category of indirect speech acts (I-SA), we used classic examples like *Can you pass me the salt?*, where the sentential force (aka ‘illocutionary mood’) is again a question (see Chierchia & McConnell-Ginet 1990; Hausser 1980 for fundamental discussion), but the utterance is interpreted as a request.

In our materials, we can clearly distinguish between NC-SA and I-SA: While the cases used for I-SA are ambiguous and comprehended by pragmatic inferencing in a particular context, our cases of NC-SA feature interrogative syntax (*wh*-V2 word order, see examples below), but are disambiguated by linguistic and non-contextual means: Rhetorical questions contained the German modal particle *schon* (which indicates the rhetorical-question interpretation) and surprise-disapproval questions were only formed with intransitive verbs where the *wh*-element was ‘what’ cannot refer to a syntactic argument (like in *What is he reading?* [*He is reading a book*]), but rather must express a meaning close to ‘why’, but with negative emotivity (like in *Why is he laughing so stupidly?* [cf. **He is laughing x*]). Accordingly, NC-SA and I-SA feature a different form-function mapping, and our study is the first to investigate whether they also differ in how they are comprehended by native speakers of German as compared to their canonical (C-SA) and direct (D-SA) speech-act counterparts.

2 Comprehending non-canonical and indirect speech acts: An experimental study

2.1 Methods

60 students from the University of Cologne took part in a web-based study (11 male/ 49 female, mean age 21.3). All were self-declared native speakers of German, except for five speakers. The data of these five speakers were excluded from the data analysis.

We created sentence pairs of two different non-canonical/canonical speech acts and one indirect/direct speech act involving interrogative sentences: a non-canonical rhetorical question vs. a canonical assertion (1), a non-canonical surprise-disapproval question vs. a canonical

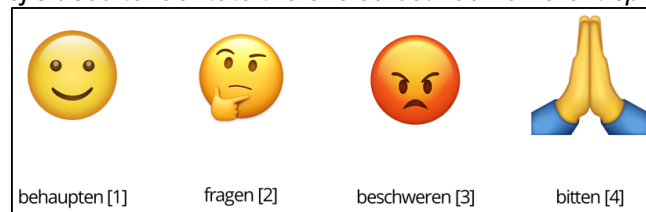
complaint (2), and an indirect request vs. a direct request (3). For each group, eight sentence pairs were created, leading to 24 sentence pairs.

- | | | | |
|-----|----|--|-----------------|
| (1) | a. | Wer mag schon gerne kalten Kaffee?
'After all, who likes cold coffee? (Nobody!) | [non-canonical] |
| | b. | Niemand mag gerne kalten Kaffee.
'Nobody likes cold coffee.' | [canonical] |
| (2) | a. | Was regnet das denn jetzt?
'Why is it raining now (of all times!)?' | [non-canonical] |
| | b. | Wie bescheuert, dass das jetzt regnet!
'How stupid that it's raining now!' | [canonical] |
| (3) | a. | Kannst du mir das Salz reichen?
'Can you pass me the salt?' | [indirect] |
| | b. | Gib mir bitte das Salz!
'Please pass me the salt!' | [direct] |

Each pair was embedded in a short story that ended with the respective speech act. The pairs were divided into two lists so that participants only saw one version of a sentence pair, leading to 24 critical items per list. We added 24 filler sentences that comprised of eight canonical assertions, eight canonical complaints and eight canonical requests—all differing in form from the canonical speech acts presented above.

The unsupervised web-based study was executed via the experiment software Gorilla (Anwyl-Irvine et al. 2020). Participants were presented with the context sentences and target sentences (performing a specific speech act) in written form in black letters on a white screen. Context and target sentences were presented separately. Participants first saw the context sentence and after pressing a button, the target sentence appeared. After reading the target sentence, participants had to decide what kind of action the speaker performs: 'behaupten' (asserting), 'fragen' (questioning), 'beschweren' (complaining), or 'bitten' (requesting). The judgment had to be executed with the help of four emojis (see Figure 1). The numbers [1] – [4] next to the speech act terms refer to the keys on the keyboard that had to be pressed in order to give the answer.

Figure 1. Emojis used to facilitate the choice between different speech act verbs.



The order of the four emojis and the respective keys were held constant to not overexert the participants. Before reading the critical items, participants were instructed by means of practice items involving explicit direct speech acts. Participants were asked to make their decision as fast and as precisely as possible. We measured the responses and their reaction times.

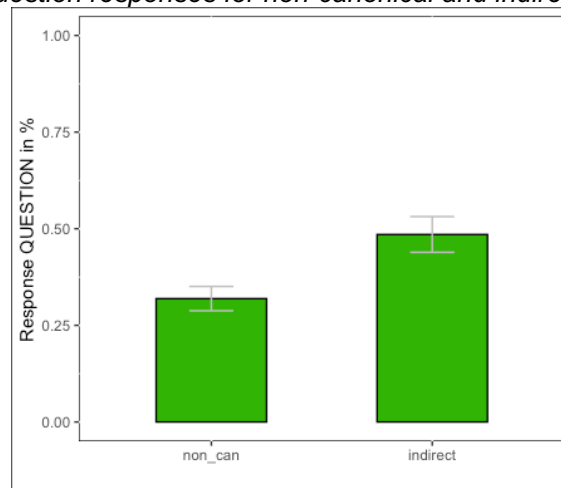
2.2 Results

In a first step, we coded whether a response was correct or incorrect according to our design of the materials. In the case of canonical and direct speech acts, each speech act had only one correct response (assert, complain, request). In the case of non-canonical and indirect speech acts, however, we coded question responses and the intended speech acts (assert, complain,

request) as correct responses. The overall accuracy was with 92% relatively high. The filler sentences had an overall accuracy of 90%, the canonical speech acts of 93%, the non-canonical speech acts of 96%, the direct speech act of 93%, and the indirect speech act of 99%. For the further analysis, we excluded all incorrect answers.

To find out whether the unintended interpretation as a question differs between non-canonical speech acts and indirect speech acts, we performed a binomial logistic mixed effects model on the question responses in R (R core Team 2017, package lme4, Bates et al. 2015). The fixed-effects factor was speech act (NONCAN/INDIRECT), and the model included by-participant intercepts as random factor. The significance of a factor was tested by comparing a model with this factor to a model that excluded it but had an otherwise identical structure. The results show that participants interpret the speech act more often as a question in the case of indirect speech acts (48.5%) than in the case of non-canonical speech acts (31.9%) ($\beta = -0.91$, $SE = 0.19$, $t = -4.72$, $p < .001$, see Figure 2). There is no difference between the two non-canonical speech acts assert and complain with regard to the question responses ($\beta = 0.001$, $SE = 0.23$, $t = 0.005$, $p = .996$). In the following, we will report the results of the collapsed data of both non-canonical speech acts.

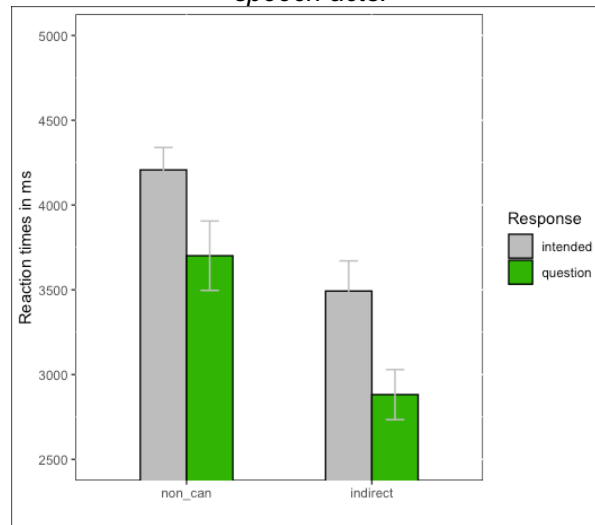
Figure 2. Question responses for non-canonical and indirect speech acts.



For the analysis of the reaction time data (only for correct responses), we excluded all data points exceeding 10000 ms and deceeding 500 ms. (2.8% data loss). The mean reaction time was 3658.7 ms (SD 1754.2 ms). The results of a linear mixed effects model with condition (CAN|NON_CAN) as fixed effect, and with by-participant intercepts as random factor show a significant main effect of condition, in that canonical speech acts are processed faster (3659.2 ms) than the non-canonical speech acts (4064.1 ms) ($\beta = 472.0$, $SE = 118.2$, $t = 3.99$, $p < .001$). For the indirect speech acts, it is the other way around. Results of a linear mixed effects model with condition (DIRECT|INDIRECT) as fixed effect, and with by-participant intercepts as random factor show a significant main effect of condition, in that the direct speech act is processed more slowly (3608.2 ms) than the indirect speech act (3161.0 ms) ($\beta = -399.8$, $SE = 133.0$, $t = -3.01$, $p < .01$, see Figure 3). Interestingly, the reaction times to canonical and direct speech acts do not differ.

A linear mixed effects model on the responses (intended vs. question) to the non-canonical and indirect speech acts shows two main effects. First, responses to non-canonical speech acts (4064.1 ms) are generally slower than to the indirect speech act (3161 ms) ($\beta = 890.0$, $SE = 141.5$, $t = 6.30$, $p < .001$). Second, question responses (3333.1 ms) are faster than intended responses (3997.7 ms) ($\beta = -346.8$, $SE = 155.4$, $t = -2.23$, $p = .026$). There is no significant interaction between both factors, indicating that question responses are faster than intended responses for non-canonical as well as for indirect speech acts (see Figure 3).

Figure 3. Reaction times for intended and question responses for non-canonical and indirect speech acts.



2.3 Discussion

Our driving hypothesis was that the two speech act classes NC-SAs and I-SAs are actually two distinct pragmatic and psychological phenomena. This hypothesis is in line with the classic claims by Searle (1975), but it contradicts more recent conceptions in the linguistics literature, according to which NC-SAs such as rhetorical questions are I-SAs as well. We tested our hypothesis in the domain of non-canonical questions versus indirect requests performed by utterances with question syntax. What we found is that participants interpreted the speech act more often as a question when it instantiated an I-SA than when it was a NC-SA. This suggests that conventionalization as a cue for interpretation is less strong than explicit force-indicating elements (e.g., modal particles) as part of the sentence. In all cases, the question interpretation was faster than the intended assertion/complaint/request-interpretations. Note that all sentences had the form of an interrogative sentence, which, apparently, was taken as the first source of information during sentence comprehension. This is compatible with the claim by Meibauer (2019) and others that the category 'sentence type' plays a major role in interpreting both indirect speech acts and non-canonical questions. However, while the I-SAs and NC-SAs have in common that the question interpretation has a processing advantage over the intended interpretation, both types of speech acts differ when they are compared to their direct/canonical counterparts. The I-SA was processed faster than its direct counterpart. The NC-SAs, however, were processed slower than their canonical counterparts. This suggests that elements indicating illocutionary force (e.g., modal particles or *why-like-what wh*-elements) help hearers to correctly identify the intention of a speaker, but that these formal markers make the comprehension process harder when they contradict other formal information such as word order and sentence type.

References (selected)

- Holtgraves, T., & Ashley, A. (2001). Comprehending illocutionary force. *Memory & Cognition*, 29, 83–90.
- Meibauer, J. (2019). What is an indirect speech act? Reconsidering the literal force hypothesis. *Pragmatics & Cognition*, 26, 61–84.
- Searle, J. R. (1975). Indirect speech acts. In Cole, P., Morgan, J. L. (eds.), *Syntax and Semantics. Speech Acts*, vol. 3., 59–82. New York: Academic Press.
- Trotzke, A., & Czipionka, A. (2022). The pragmatics of surprise-disapproval questions: An empirical study. *Linguistics Vanguard*, 8, 239–249.

Incremental negation processing with positive questions under discussion

Oksana Tsaregorodtseva*¹, Elena Albu*¹, and Barbara Kaup¹

¹University of Tübingen – Germany

Abstract

Short abstract. Previous research has shown that the processing difficulty associated with negation can be modulated by context. In the absence of a supportive context, negative sentences tend to answer a positive question under discussion (QUD) which leads to a two-step interpretation strategy. However, negation is processed incrementally when used in a context in which it addresses a negative QUD. The present study provides new evidence according to which negation can be incrementally processed when it addresses a positive polar QUD in a supportive context generated by discourse markers (*As expected, John has (not) washed the car*).

Introduction. Negative sentences are generally harder to process compared to affirmative sentences (for an overview, see Kaup & Dudschig, 2020). However, the processing difficulty can be modulated by context. When used without any contextual support, negative sentences (*John hasn't washed the car*) are processed in a two-step fashion (Kaup et al., 2006; Kaup et al., 2007): comprehenders first represent the non-factual object state (clean car) and only subsequently the factual object state (dirty car). One possible explanation is that, in the absence of any contextual information, negative sentences tend to answer a positive question under discussion (QUD) (Has John washed the car?). In contrast, negation is processed incrementally when used in a context in which it addresses a negative QUD (Tian et al., 2010, 2016; Wang et al., 2021), expressed either by means of cleft sentences (*It was John who didn't cook the spaghetti*) or wh- questions (*Which fruit isn't peeled?*).

Research question and predictions. In two behavioral experiments, we investigated whether negative sentences can be processed incrementally when addressing a positive polar QUD. To that end, in a probe recognition task (Fig. 1), affirmative and negative sentences were used in the absence (Experiment 1) or in a supportive context (Experiment 2) generated by discourse markers (*As expected, John has (not) washed the car*). These markers render the contextual expectation salient (It was expected that John would (not) wash the car), which is always congruent with the actual object state (Table 1). In both experiments, affirmative and negative sentences address a positive QUD (Has John washed the car?). For affirmative sentences, we expect faster response times (RTs) for the factual compared to the non-factual object states in both experiments. In contrast, for negative sentences we expect different patterns of responses. If negation is processed in a two-step way, we expect a crossover interaction between the factors *Polarity (aff/neg)* and *Depicted object state (factual/non-factual)*, with faster RTs for the non-factual (clean car) than for the factual object state (dirty car). If negation is processed incrementally, we expect a main effect of *Depicted object state*, with faster RTs for the factual (dirty car) compared to the non-factual object state (clean car).

*Speaker

Results. RTs on correct picture-present trials were analyzed by means of a linear mixed effects model (Table 2). As predicted, participants responded faster to the factual object states in the affirmative condition in both experiments, while different patterns of responses emerged in the negative condition. *Experiment 1* (N = 104; 20 men; *Age* = 37.26; *SD* = 12.39; online, English items) revealed a crossover interaction between the factors *Polarity* and *Depicted object state* ($\chi^2(1) = 9.50, p = .002, \beta = -21.02, t = -3.08$), a main effect of *Polarity* ($\chi^2(1) = 10.36, p = .001, \beta = -21.93, t = -3.22$) but no main effect of *Depicted object state* ($\chi^2(1) = 2.47, p = .116, \beta = -10.75, t = -1.58$). In line with the two-step procedure, these findings suggest that participants responded faster to non-factual (clean car) than to the factual object states (Fig. 2a). By contrast, *Experiment 2* (N = 88; 27 men; *Age* = 39.82; *SD* = 13.95; online, English items) showed the reversed pattern with two main effects of *Depicted object state* ($\chi^2(1) = 23.77, p < .001, \beta = -41.78, t = -4.90$) and *Polarity* ($\chi^2(1) = 16.74, p < .001, \beta = -35.07, t = -4.11$). There was a significant ordinal interaction this time ($\chi^2(1) = 11.03, p = .001, \beta = 28.39, t = -3.32$), which suggests that participants responded faster to the factual (dirty car) compared to the non-factual object state in the negative condition (Fig. 2b). To receive more information about the pattern of responses in the negative conditions, a post hoc test was performed. This showed an interaction between the *Depicted object state* and *Context* (no context Exp. 1/context Exp. 2) ($\chi^2(1) = 4.51, p = .034, \beta = 16.84, t = 2.12$), which replicates previous findings according to which the processing of negative sentences is strongly modulated by context (Fig. 2c).





Discussion. All in all, the present paper corroborates previous results which indicate that context strongly influences the processing of negative sentences. Furthermore, it provides new evidence showing that, in a supportive context, negation can be incrementally processed when it addresses a positive polar QUD.

References

- Kaup, B., & Dudschig, C. (2020). Understanding negation. In *The Oxford handbook of negation*.
- Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033-1050.
- Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly journal of experimental psychology*, *60*(7), 976-990.
- Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *Quarterly journal of experimental psychology*, *63*(12), 2305-2312.
- Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context—why and when we represent the positive argument. *Language, Cognition and Neuroscience*, *31*(5), 683-698.
- Wang, S., Sun, C., Tian, Y., & Breheny, R. (2021). Verifying Negative Sentences. *Journal of Psycholinguistic Research*, *50*(6), 1511-1534.

Keywords: negation processing, grounded cognition, pragmatics, language comprehension

Table 1: Example items for the probe recognition task in Experiments 1 & 2

Polarity	Depicted object state	Experiments 1 & 2	
		Sentence	Display
Affirmative	factual	(As expected) John has washed the car.	
	non-factual	(As expected) John has washed the car.	
Negative	factual	(As expected) John hasn't washed the car.	
	non-factual	(As expected) John hasn't washed the car.	

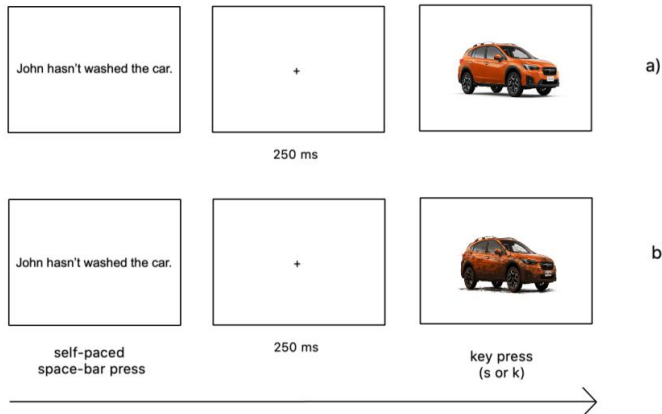


Figure 1: The time course of a typical negative trial; a) non-factual object state b) factual object state

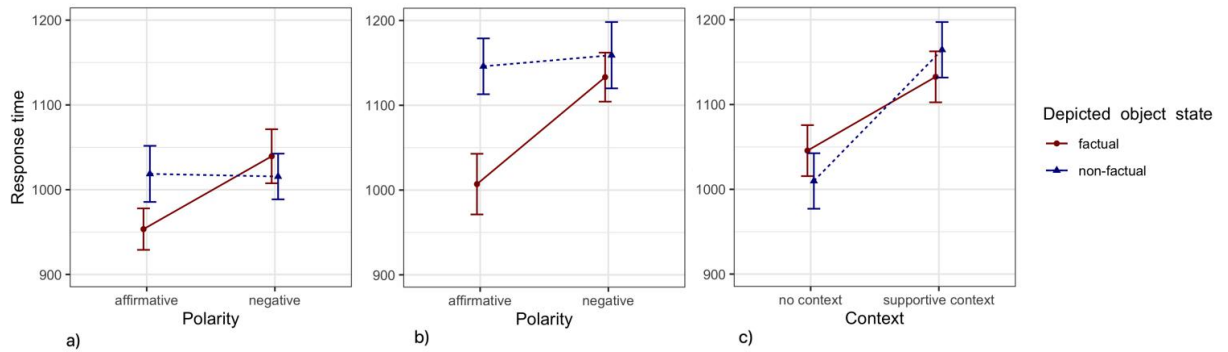


Figure 2: Response times Experiments 1 & 2, and Post hoc; error bars denote 95% confidence intervals; a) Experiment 1 b) Experiment 2 c) Post hoc

Table 2: Linear mixed effects models in Experiments 1 & 2 and Post hoc

Exp. 1 & 2	The base model: $rt \sim \text{Polarity} + \text{Depicted object state} + (1 \text{item} + 1 \text{participant})$
	The best model: $rt \sim \text{Polarity} * \text{Depicted object state} + (1 \text{item} + 1 \text{participant})$
Post hoc	The base model: $rt \sim \text{Context} + \text{Depicted object state} + (1 \text{item} + 1 \text{participant})$
	The best model: $rt \sim \text{Context} * \text{Depicted object state} + (1 \text{item} + 1 \text{participant})$

The (not-)at-issue status of character viewpoint gestures

Sebastian Walter (University of Wuppertal)

Co-speech gestures contribute *not-at-issue meaning* by default (Ebert & Ebert, 2014). However, the experimental study that provided evidence in favor of this claim (Ebert, Ebert, & Hörnig, 2020) only investigated *observer viewpoint gestures* (OVGs). It is therefore questionable whether *character viewpoint gestures* (CVGs) also contribute not-at-issue meaning by default as they differ from OVGs in certain aspects, such as informativity (Parrill, 2010). Findings from the experimental rating study at hand suggest that although CVGs and OVGs both contribute not-at-issue meaning by default, CVGs are more at-issue than OVGs.

1 Introduction

Perspective plays an integral role in the interpretation of many different lexical items, such as relational expressions (e.g., *left* and *right*). There is a general tendency to interpret expressions as speaker-oriented (Harris, 2012). There are exceptions to this general tendency, however, as for example in instances of *free indirect discourse* where some individual's thoughts/utterances are reported without any overt marking. Therefore, the perspective of this individual becomes highly prominent in these utterances (Hinterwimmer, Patil, & Ebert, 2021). Perspective is also encoded in gesture. McNeill (1992) distinguishes between CVGs and OVGs. CVGs depict an event from a character's perspective where usually the whole body is involved when producing the gesture. A character distinct from the speaker is thus impersonated. OVGs, by contrast, encode a perspective as if an event was observed from a distance. (1) illustrates this. Square brackets indicate gesture-speech alignment.

- (1) Peter overslept this morning. He then had to [run] to the bus station. + CVG illustrating running by moving the whole body/OVG illustrating running by moving two fingers

Due to the whole body being involved when performing CVGs, they have been argued to be more informative than OVGs (Parrill, 2010) as speakers can use their whole bodies to encode gestural information. Ebert and Ebert (2014) argue that speech-accompanying gestures behave similar to appositives (Potts, 2005) and are therefore analyzed as *conventional implicatures* (CIs) in their approach. CIs project through negation (2a) and cannot be directly denied in discourse (B in (2b)). Instead, a discourse interrupting element has to be used to deny their content (B' in (2b)). Therefore, they contribute meaning that is not-at-issue, i.e., their content is not on the table for discussion (Farkas & Bruce, 2010).

- (2) a. It is not the case that I brought [a bottle of water] to the talk. #A small one is enough for me. + "big" co-speech gesture
b. A: I brought [a bottle of water] to the talk. + "big" co-speech gesture
B: #No, the bottle isn't big.
B': Hey, wait a minute! Actually, the bottle isn't that big.

(cf. Ebert & Ebert, 2014)

The claim that gestures contribute not-at-issue meaning has been experimentally validated in Ebert et al. (2020). A shortcoming of their study was that they only tested for gestures which contributed information about an object's size or shape, which can therefore be classified as OVGs. It is unclear whether these findings can be transferred to CVGs without any adjustments as the two gesture types differ, for example, in size. Moreover, Hinterwimmer et al. (2021) hypothesize that the overall preference for CVGs observed in their experimental study might

be due to them differing from OVGs in their at-issue status. The study presented in this paper investigates this hypothesis. It is hypothesized that although CVGs and OVGs both contribute not-at-issue meaning by default, CVGs are more at-issue than OVGs.

2 Experimental study

2.1 Methods

2.1.1 Participants

60 native speakers of German participated in the study. All of them were recruited via Prolific.

2.1.2 Materials

Videotaped utterances were paired with either a matching or a mismatching picture (Figure 1). The utterances occurred in three conditions: accompanied by a CVG (3a), an OVG (3a), or a verbalization of the gestures (3b). Since unembedded verbal material is at-issue, the latter condition serves as a control condition. A sample item is given in (3). Crucially, the gestures only differed in the perspective they encoded. The study thus had a 3x2 design with the factors MODE (CVG, OVG, Verbal) and MATCH (matching picture, mismatching picture). There were 18 experimental items and 24 filler items.

Figure 1: The matching and the mismatching picture for the experimental item in (3)



(a) Matching picture



(b) Mismatching picture

- (3) a. Letzten Mittwoch hatte ich den ganzen Tag Termine überall in der Stadt. Nachdem einer der Termine länger dauerte als gedacht, musste ich mich [richtig beeilen].
 ‘Last Wednesday I had many appointments throughout the whole city. After one of the appointments took longer than expected, I had to [hurry a lot].’
CVG: running using the whole body
OVG: running with two fingers
- b. Letzten Mittwoch hatte ich den ganzen Tag Termine überall in der Stadt. Nachdem einer der Termine länger dauerte als gedacht, musste ich mich richtig beeilen **und rennen**.
 I myself really hurry and run
 ‘Last Wednesday I had many appointments throughout the whole city. After one of the appointments took longer than expected, I had to hurry a lot **and run**.’

The picture in Figure 1a matches the boldfaced parts of the utterances in (3), whereas the picture in Figure 1b does not match the boldfaced parts in (3). Based on the hypothesis that CVGs and OVGs differ with respect to their at-issue status, an interaction for the rating differences between matching and mismatching items in the CVG and OVG condition is predicted (Kroll & Rysling, 2019), which has been called *mismatch effect* in Ebert et al. (2020). Moreover, since CVGs are hypothesized to contribute not-at-issue meaning by default, it is predicted that the mismatch effect is stronger for Verbal items than for CVG items.

2.1.3 Procedure

Items were evenly distributed onto six lists according to a Latin Square design. Participants first saw the picture, then the video. They had to rate on a 7-point Likert-scale how well the picture

and the utterance matched (1 = entirely unmatching; 7 = entirely matching). The questionnaire was construed using SoSci Survey.

2.2 Results

Figure 2: Mean values and standard deviations (SDs) for each condition

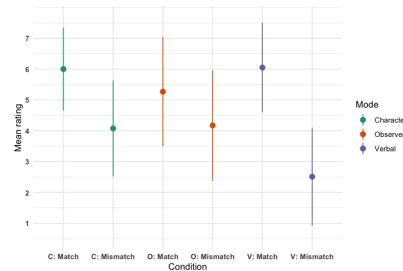


Figure 2 shows that the rating difference between match and mismatch was higher for CVG items (match: mean = 6.00, SD = 1.35; mismatch = 4.08, SD = 1.56) than for OVG items (match: mean = 5.27, SD = 1.77; mismatch: mean = 4.17, SD = 1.79). The rating difference between match and mismatch for Verbal items (match: mean = 6.05, SD = 1.45; mismatch: mean = 2.51, SD = 1.58), however, was higher than for CVG and OVG items. An ordinal mixed effects model was fitted onto the data. The output of the model is given in Table 1. It shows an interaction for the factor MATCH and the pairwise comparison between CVG and OVG for the factor MODE ($p < 0.001$). Additionally, it shows an interaction for the factor MATCH and the pairwise comparison between CVG and Verbal for the factor MODE ($p < 0.001$).

Table 1: Ordinal mixed-effects model with Mode and Match as fixed effects and participants and items as random intercepts

	Estimate	Std. Error	z value	Pr(> z)
Match	-2.857	0.139	-20.60	<0.001 ***
Mode – CVG vs. OVG	-0.423	0.136	-3.11	<0.01 **
Mode – OVG vs. V	0.394	0.143	2.76	<0.01 **
Match:Mode – CVG vs. OVG	1.176	0.273	4.30	<0.001 ***
Match:Mode – OVG vs. V	3.193	0.293	10.90	<0.001 ***

3 Discussion

The significantly lower mismatch effect for CVG items as opposed to Verbal items confirms the prediction that CVGs contribute not-at-issue meaning by default, in line with previous research (Ebert et al., 2020). Moreover, the higher mismatch effect for CVG items in comparison to OVG items confirms the hypothesis that the two gesture types differ with respect to their at-issue status. In order to bring these two findings together, a gradient approach to at-issueness (Barnes, Ebert, Hörnig, & Stender, 2022) instead of a binary approach has to be assumed. This ensures that CVGs and OVGs can both contribute not-at-issue meaning by default, but still differ with respect to their at-issue status. More specifically, CVGs are more at-issue than OVGs because the mismatch effect was significantly higher for CVG items as compared to OVG items. This interpretation is covered by the finding that not-at-issue content has a less severe impact on truth conditions than at-issue content (Kroll & Rysling, 2019).

Furthermore, the results presented can potentially account for the findings of the study reported in Hinterwimmer et al. (2021) where the authors tested for the hypothesis that gestural and linguistic perspective are preferably aligned unless there are intervening pragmatic factors overwriting this default. They conducted a forced-choice study where participants saw two identical videotaped utterances per trial, one being accompanied by a CVG and one being accompanied by an OVG. The utterance was either narrated from a character's perspective or from a narrator's, i.e., an observer's, perspective. Contrary to their hypothesis, they found an

overall preference for CVGs. This finding can be explained by the data reported here. Assume that, following Grice's (1975) cooperative principle and the conversational maxims, in particular the maxim of quantity, speakers are expected to prefer utterances which are maximally informative as long as no maxim violations take place. Furthermore, not-at-issue content does not seem to be subject to the maxim of quantity.

- (4) Q: Who is at the door?
A: Peter, a famous movie director, is at the door.

Example (4) shows that the content of the appositive does not address the QUD and is therefore not-at-issue (Koev, 2018). Surprisingly, the utterance of the appositive does not seem to violate the maxim of quantity since no conversational implicature arises from the utterance of the appositive. This observation can be related to the experimental results presented here and also to Hinterwimmer et al. (2021) by assuming that the preference for linguistic and gestural perspective to be aligned can be overwritten by the principle that speakers prefer maximally informative utterances if there are no violations of the maxim of quantity. Since CVGs are more at-issue than OVGs, CVG content is more subject to the maxim of quantity than OVG content. Hence, a preference for CVGs to be aligned with a character's perspective in speech is predicted by both principles. However, a CVG preference is also predicted for utterances where an observer's perspective is expressed as they are more at-issue than OVGs and therefore more subject to the pragmatic principle that speakers prefer maximally informative utterances. This principle thus overwrites the preference for linguistic and gestural perspective to be aligned, resulting in the observed overall preference for CVGs irrespective of the linguistically expressed perspective. A more systematic investigation of the relation between not-at-issue content and the Gricean maxims is left to future research.

References

- Barnes, K., Ebert, C., Hörnig, R., & Stender, T. (2022). The at-issue status of ideophones in German: An experimental approach. *Glossa: A Journal of General Linguistics*, 7(1), 1–39.
- Ebert, C., & Ebert, C. (2014). *Gestures, demonstratives, and the attributive/referential distinction*. Talk given at *Semantics and Philosophy in Europe 7*.
- Ebert, C., Ebert, C., & Hörnig, R. (2020). Demonstratives as dimension shifters. In *Proceedings of Sinn und Bedeutung 24* (pp. 161–178). Osnabrück: University of Osnabrück.
- Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, 27(1), 81–118.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts* (pp. 41–58). Leiden, The Netherlands: Brill.
- Harris, J. A. (2012). *Processing Perspectives* (Unpublished doctoral dissertation). University of Massachusetts Amherst, Amherst, MA.
- Hinterwimmer, S., Patil, U., & Ebert, C. (2021). On the interaction of gestural and linguistic perspective taking. *Frontiers in Communication*, 6, 1–15.
- Koev, T. (2018). Notions of at-issueness. *Language and Linguistics Compass*, 12(12), e12306.
- Kroll, M., & Rysling, A. (2019). The search for truth: Appositives weigh in. In K. Blake, F. Davis, K. Lamp, & J. Rhyne (Eds.), *Proceedings of Semantics and Linguistic Theory 29* (pp. 180–200). Los Angeles, CA: University of California, Los Angeles.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
- Parrill, F. (2010). Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668.
- Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford, UK: Oxford University Press.

The effects of information structure and sentence structure on sentence processing

Hans Wilke¹, Jet Hoek², Hannah Rohde¹

¹ School of Philosophy, Psychology & Language Sciences, The University of Edinburgh

² Department of Language and Communication, Radboud University Nijmegen

There is extensive evidence that comprehenders prefer given information to precede new information in a sentence. This principle has primarily been tested by considering information-structural features encoded in syntax, e.g., given information expressed in definite NPs. We carried out a self-paced reading experiment to revisit the given-before-new principle and disentangle new-/givenness from syntactic features. Additionally, we consider the effects of clause ordering and the mapping between a clause's information status (given/new) and its type (matrix/subordinate). We find that given-before-new sentences are processed faster, and that this effect is even stronger when the given information is hosted by a subordinate clause.

1 Introduction

The literature on sentence processing has identified various principles that govern the processing of clauses (Scholman et al., 2022). The current project investigates two of these principles: the given-before-new principle (Gundel, 1988; Halliday, 1967a, 1976b; Prince, 1981) and the clause structure principle (Diessel, 2005, 2008; Fodor et al., 1974; Gibson, 1998; Holmes, 1973; Jansen, 2008; Troost et al., 2008). The given-before-new principle posits that comprehenders prefer given information to precede new information. While there is extensive evidence that this is the case, this principle has primarily been tested by considering the information-structural features encoded in syntax (e.g., definite NPs represent given information, indefinite NPs represent new information: Bock, 1977; Clifton & Frazier, 2004). The current project revisits the given-before-new principle while focusing on the discourse-status of information and analysing new-/givenness independently of syntactic structure. The clause structure principle proposes that sentences are harder to process if the main clause follows the subordinate clause, rather than vice versa. Both the given-before-new principle and the clause structure principle have been studied individually, in isolation, but little is known about the interaction between these principles in processing. While there have been theoretical proposals and evidence from offline studies that subordinate clauses – especially sentence-early and preposed subordinate clauses – tend to be more likely hosts for given information, and main clauses for new information (Diessel, 2001), this has not been sufficiently tested in online processing. Recent work, however, shows that locating given information in a subordinate clause indeed leads to faster processing times, but only when the subordinate clause is an adverbial (Scholman et al., 2022).

We ask whether the preferred mapping between a clause's information status (given/new) and its type (matrix/subordinate) extends to other types of subordinate clauses. We measure this mapping (henceforth "clause-type mapping of information") in appositive relative clauses (ARC). Corpus studies have shown that ARCs generally contribute new information to discourse (Loock, 2007, 2010). Gibson et al. (2005) analysed the processing of ARCs in sentence-early versus sentence-final position and did not find a difference between the two positions. However, they did not explicitly manipulate the information status of the ARCs, but rather assume that ARCs are, by virtue of being ARCs, taken as presenting new information. In addition, their study focused on the reading times of the ARCs and not the entire sentence. This leaves open the question of how information structure and sentence structure influence

the processing of full sentences with subordinate clauses. Our study aims to replicate Gibson et al.'s (2005) study with a more explicit manipulation of the information status of ARCs, while simultaneously exploring the given-before-new principle, the clause structure principle, and their interaction at the level of the full sentence.

2 Method

We conducted a self-paced reading experiment in which participants saw short narratives presented in chunks (moving-window paradigm). The order of clauses in the target sentence (matrix-ARC, ARC-matrix) was crossed with information order (given-new, new-given) and clause-type mapping of information (given matrix/new ARC, new matrix/given ARC). The content of either the matrix clause or the ARC was made discourse-old/inferred given information by providing a context preceding the target sentence. The other clause in the target sentence then was the only clause containing content which was discourse-new. Consider the following example in (1) in which the target sentence represents the matrix-ARC clause order, with given-before-new, yielding a given matrix/new ARC mapping (see the appendix for an overview of the other conditions for this item):

Context:

- (1) My aunt loves to be part of the rumor mill, and just like my mom, takes any opportunity to engage in the latest stories. Because of this, I always pay close attention to what I'm saying around her. At my birthday party,

Target sentence:

my aunt was gossiping with my mom_[SENTENCE-EARLY MATRIX CLAUSE, GIVEN INFORMATION],
who was drinking rum & coke_[SENTENCE-FINAL ARC, NEW INFORMATION].

Spillover region:

As I walked by, I heard they were talking about me. My mom got startled and spilled her drink all over my aunt.

Our study consisted of 32 items in four conditions (a 2x2 design captures all three factors as there is overlap between them) and we recruited 237 self-reported native speakers of American English. After excluding participants who failed to perform above chance on the attention checks we included, we analysed the data of 195 participants. Our main interest was the reading times for the full sentences, but we also measured and analysed the reading times of the individual clauses that make up the target sentence to probe whether a preference for ARCs to contain new information is reflected in processing. It should be noted that the clause-structure principle and clause-type mapping of information hypothesis make competing predictions: if an order of matrix-before-subordinate overlaps with given-before-new, clause-type mapping of given information in a subordinate and new information in a matrix clause cannot be realised. In addition to expecting an overall preference for given-before-new, we expect that the clause-structure principle is more likely to hold as it would overlap with the observation that ARCs generally contain new information.

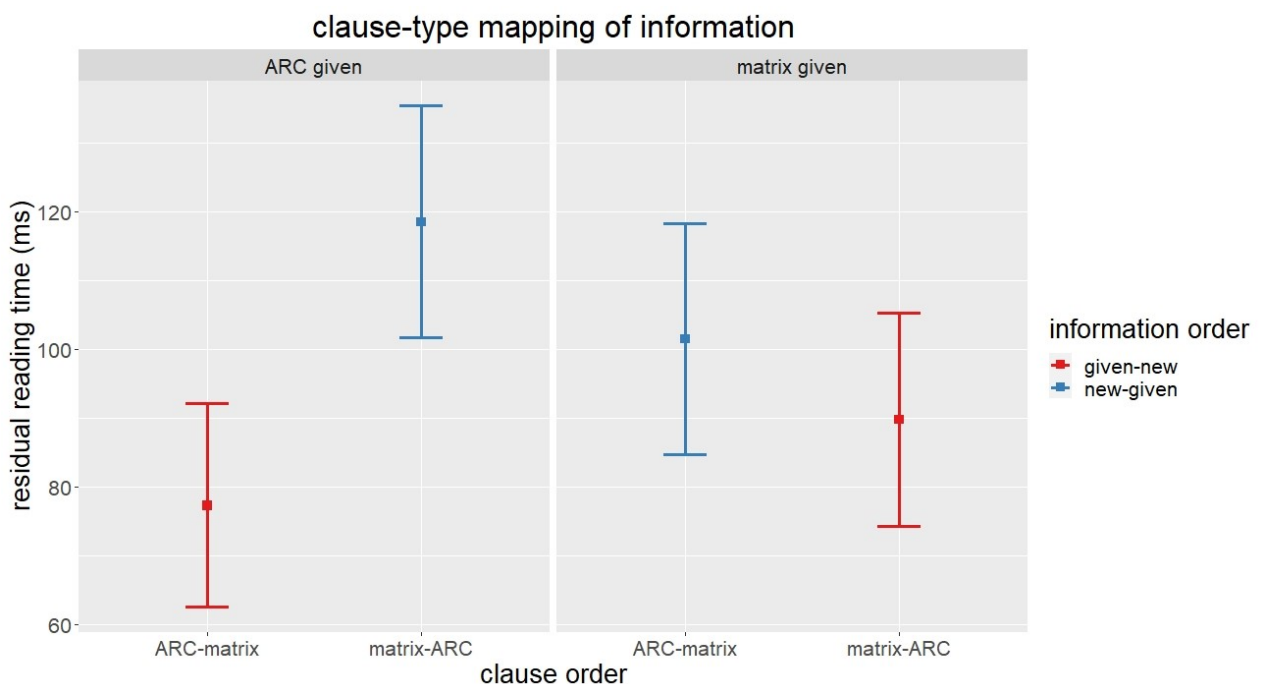
3 Results and conclusion

When reading times of the full sentence are considered, a preference for given-before-new is found (Figure 1). This extends prior work, showing that the given-before-new principle holds for sentences in which information status is manipulated at a discourse-level rather than by the syntactic nature of clauses. Furthermore, we find an interaction effect between clause-type mapping of information and clause ordering. This suggests that given information

expressed in sentence-early ARCs facilitates fastest processing, a finding that is in line with Scholman et al. (2022). We find no evidence to support the clause structure principle.

While no evidence for a direct effect of clause structure or information status was found at the level of the entire sentence, both of these had an effect at the clause level. When the reading times of individual clauses were considered, both main clauses and appositive relative clauses were read faster when their position was sentence-early, and when they contained given information. This suggests that even though ARCs might generally be more likely to contain new information, no preference for this generalisation is reflected in processing. That said, further research is needed to investigate this. Our results show that given information is always processed faster than new information, independent of position and/or clause type. How to reconcile these results with the corpus evidence that ARCs do generally contain new information remains an open question.

Figure 1: Residual reading times for the entire target sentence as a function of clause-type mapping of information, clause order and information order.



References

- Bock, J. K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior*, 16(6), 723–734.
- Clifton, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & cognition*, 32(6), 886–895.
- Diessel, H. (2001). The Ordering Distribution of Main and Adverbial Clauses: A Typological Study. *Language* 77(2), 433–455.
- Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43(3), 449–470.
- Diessel, H. (2008). Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, 19(3), 465–490.
- Fodor, J., Bever, A., & Garrett, T. (1974). *The psychology of language: an introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies.

- Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, 16(2), 313–353.
- Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in Syntactic Typology*, 17(1), 209–239.
- Halliday, M.A.K. (1967a). Notes on Transitivity and Theme in English. Part 1. *Journal of Linguistics* 3(1): 37–81.
- Halliday, M.A.K. (1967b). Notes on Transitivity and Theme in English. Part 2. *Journal of Linguistics* 3(2): 199–244.
- Holmes, V. M. (1973). Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior*, 12(3), 285–293.
- Jansen, F. (2008). Order of mention and the position of temporal clauses in Dutch complex sentences. In W. Ramm & C. Fabricius-Hansen (Eds.), *Linearisation and Segmentation in Discourse. Multidisciplinary Approaches to Discourse 2008 (MAD 08)* (pp. 77–85). Oslo: Dept. of Literature, Area Studies and Europ. Languages, Univ. of Oslo.
- Loock, R. (2007). Appositive relative clauses and their functions in discourse. *Journal of pragmatics*, 39(2), 336–362.
- Loock, R. (2010). *Appositive relative clauses in English: Discourse functions and competing structures* (Vol. 22). John Benjamins Publishing.
- Prince, E.F. (1981). Toward a Taxonomy of Given-New Information. In P. Cole (ed.), *Radical Pragmatics*. Cambridge – London: Academic Press: 223–256.
- Scholman, M. C. J., Blything, L., Cain, K., Hoek, J., & Evers-Vermeul, J. (2022). Discourse Rules: The Effects of Clause Order Principles on the Reading Process. *Language, Cognition and Neuroscience*, 1–15.
- Troost, C., Jansen, F., & Sanders, T. J. M. (2008). Bepaalt cognitieve economie de plaats van temporele en causale bijzinnen. *Tijdschrift voor Taalbeheersing*, 30(1), 75–88.

Appendix – other conditions for example (1)

- new ARC before given matrix clause:*
- (2) My aunt loves to be part of the rumor mill, and just like my mom, takes any opportunity to engage in the latest stories. Because of this, I always pay close attention to what I’m saying around her. At my birthday party, **my aunt, who was drinking rum & coke, was gossiping with my mom.** As I walked by, I heard they were talking about me. My mom got startled and spilled her drink all over my aunt.
- new matrix clause before given ARC:*
- (3) My mom, like my aunt, is a big fan of drinking rum. She thinks she is really good at hiding it by adding some coke to it. Everybody knows what is actually in her glass of course. A few weeks ago, at my birthday party, **my aunt was gossiping with my mom, who was drinking rum & coke.** As I walked by, I heard they were talking about me. My mom got startled and spilled her drink all over my aunt.
- given ARC before new matrix clause:*
- (4) My aunt, like my mom, is a big fan of drinking rum. She thinks she is really good at hiding it by adding some coke to it. Everybody knows what is actually in her glass of course. A few weeks ago, at my birthday party, **my aunt, who was drinking rum & coke, was gossiping with my mom.** As I walked by, I heard they were talking about me. My mom got startled and spilled her drink all over my aunt.

Contrastive topic marking with German *dagegen* ('in contrast') and *wiederum* ('in turn')

Regina Zieleke (University of Tübingen)

In my talk, I will discuss the German connectives *dagegen* ('in contrast') and *wiederum* ('in turn') and their function to signal contrast via contrastive topics. Corpus data reveal that, unlike other contrastive connectives, *dagegen/wiederum* are, in fact, restricted to conjuncts involving contrastive topics. I therefore argue that *dagegen/wiederum*'s contribution to discourse coherence is to signal or 'highlight' a contrastive relation already available structurally via contrastive topic marking. Data from an acceptability rating study corroborate my claims showing that (i) *dagegen/wiederum* are rated as unacceptable in conjuncts lacking contrastive topics and (ii) their absence does not reduce acceptability of the connection.

1 Introduction

Past research has shown that there are different kinds of contrastive relations that pattern with different (specified) contrastive connectives. In my talk, I will discuss the German connectives *dagegen* ('in contrast') and *wiederum* ('in turn') and their function to signal a contrastive relation defined by (information) structural alternatives.

This relation has been labelled inter alia 'semantic opposition contrast' (e.g. Lakoff 1971), 'formal contrast' (e.g. Jasinskaja 2012), or 'contrastive comparison' (e.g. Breindl et al. 2014) and information structural properties of the contrasted conjuncts in terms of parallel topic-comment- or topic-focus-structure have been widely discussed as pivotal (e.g. Sæbø 2003, Umbach 2005, Breindl et al. 2014). I will go one step further and argue that the alternatives involved in this contrastive relation stem from contrastive topic marking, as showcased by the conditions of use for *dagegen* and *wiederum*.

Example (1) shall serve as an illustration: Due to the focus alternatives 'not liking coffee' and 'drinking tea', (1a) can be interpreted as contrastive (hence the perfectly acceptable marking by *aber* ('but') in (1b)). The use of *dagegen* or *wiederum* in (1c), in turn, is heavily marked.

- | | | |
|-----|--|--|
| (1) | Peter mag keinen Kaffee. | 'Peter doesn't like coffee.' |
| | a. Er trinkt Tee. | 'He drinks tea.' |
| | b. Aber er trinkt Tee. | ' But he drinks tea.' |
| | c. Er trinkt ? dagegen / ? wiederum Tee. | 'He drinks, ? in contrast / ? in turn, tea.' |

If the connectives' use in this context is acceptable at all, the verb and its complement would have to be prosodically marked by a hat contour, cf. (1d). In a version with the object fronted as in (1e), *dagegen* and *wiederum* become equally acceptable:

- | | |
|----|---|
| d. | Er /TRINKT dagegen / wiederum \TEE. |
| e. | /TEE dagegen / wiederum \TRINKT er. |
| f. | /TEE \TRINKT er. |

Both versions, (1d) and (1e), share the property of contrastive topic marking in Büring's (2016) sense: the (prosodical and/or syntactic) splitting of the verb 'to drink' and its complement 'tea' marks them as non-exhaustive alternatives to 'to not like' and 'coffee', respectively (cf. also

Krifka 2008). As the asyndetic (1f) shows, this effect is independent of the connectives' presence.

The goals of my talk are twofold: first, I argue that a connection with German *dagegen* and *wiederum* requires a specific information structural pattern that can be analyzed in terms of information structural alternatives, viz. contrastive topics¹ in the two conjuncts. The 'semantic opposition' or 'comparison' effects ascribed to such connections, then, can be traced back to the conventional implicature triggered by contrastive topic marking. Second, since the effect of contrastive topics stands on its own, I argue that the contribution of *dagegen* and *wiederum* to discourse coherence is to simply signal or 'highlight' a contrast which is already available structurally. The two connectives can thus be regarded as markers of (Information) Structural Contrast elicited by contrastive topic marking.

2 Empirical evidence

There are two sources of empirical evidence that we will discuss. First, corpus data reveal that *dagegen* and *wiederum* – in contrast to other contrastive connectives such as German *dennoch* ('yet/nevertheless') – occur in conjuncts with a limited set of information structural properties in the conjuncts. Second, experimental data from an acceptability judgment study support the eligibility of *dagegen* and *wiederum* for contrastive topic marking, while also showing that they are, in fact, perfectly omissible.

2.1 Corpus data

In data by Zieleke (to appear), 100 sets of data for each contrastive connective² have been extracted from DeReKo sub-corpus *die Zeit* and annotated for topic development. Adopting a broad notion of topic as a 'point of departure' in Jacobs' (2001) or Chafe's (1976:50: "the topic sets a *spatial, temporal, or individual framework* within which the main predication holds"), the two contrasted conjuncts have been annotated for the topics as well as the relation between the two topics. The possible categories of topic development were (i) contrastive topics (non-exhaustive alternatives), (ii) topic promotion (Daneš's 1970 linear progression, cf. footnote 1), (iii) continuous topics, and (iv) unrelated topic shift (newly introduced, non-alternative, non-promoted topics).

The results show that, while contrastive topics only play a marginal role for contrast with other connectives such as *dennoch* (5%), they make up the entirety of connections with *dagegen* (100%) and the majority of connections with *wiederum* (53% the rest belonging to category (ii) topic promotion, cf. footnote 1).

2.2 Acceptability judgment

36 native speakers of German were presented with 36 items showing a small discourse consisting of a context sentence and two contrastive conjuncts as shown in (2). They were asked to rate the acceptability of the second conjunct among a Likert scale from 1 (very

¹ *Wiederum* is also eligible for connections involving what Daneš (1970) called linear progression:

- (i) Peter such seinen Kaffee. Der_{Kaffee} **wiederum** steht draußen.
'Peter is searching for his coffee. The coffee, in turn, is outside.'

Since the focus of our talk is on contrastive topics, we will put this use on the side. Note, however, that this connection also relies on information structural properties and is therefore compatible with our claims on the nature of contrast with *dagegen* and *wiederum*. See Zieleke, submitted, for a suggestion on a notion of Structural Contrast incorporating both information structural make ups.

² In fact, data on six German contrastive connectives (*dagegen* ('in contrast'), *wiederum* ('in turn'), *dennoch* ('yet/nevertheless'), *trotzdem* ('nevertheless'), *jedoch* ('however/yet'), *allerdings* ('however')) were analyzed in order to identify the role of information structural components of contrastive conjuncts for the type of contrast expressed.

unacceptable) to 7 (very acceptable). This conjunct was manipulated for two variables: On the one hand, the conjunct contained either *dagegen*, *wiederum* or no connective (= asyndetic version). On the other hand, the information structural makeup of the conjunct varied. In the ‘parallel’ condition illustrated in (2a), the second conjunct provided a parallel information structure with contrastive topics (aboutness or frame-setting) and contrastive foci (lexico-semantic alternatives or negation). In the ‘independent’ condition shown in (2b) the conjuncts involved a different, but coherent and contrastive continuation with diverging information structure. This diversion was induced by a shift from referential to frame topics in combination with non-contrasting verbs as in (2b), passive constructions orthetic *es gibt* (‘there are’) – sentences.

(2) Context:

Der Chef schickt zwei seiner Mitarbeiter zu einem Kundentermin.
 ‘The boss sends two of his employees to a customer meeting.’

First Conjunct:

[Herr Schmidt] fährt mit dem Auto.
 ‘Mr. Smith goes by car.’

Second Conjunct:

a) Parallel:

[Herr Müller] fährt **dagegen / wiederum / Ø** mit dem Fahrrad.
 ‘Mr. Muller goes ... by bike.’

b) Independent:

[Im Stau] bereut er **dagegen / wiederum / Ø** seine Entscheidung.
 ‘Stuck in traffic, he_{Mr. Smith} ... regrets his decision.’

In both conditions, the second conjunct can be connected by the underspecified contrastive connective *aber* (‘but’). 36 fillers were built analogously to the experimental items and served as control items.

The mean ratings are summarized in Figure 1 below. The three trellises represent the three connective-conditions, *dagegen*, *wiederum* or asyndetic, each showing the mean ratings in the parallel condition on the left (cf. (2a)) and the independent condition on the right (cf. (2b)).

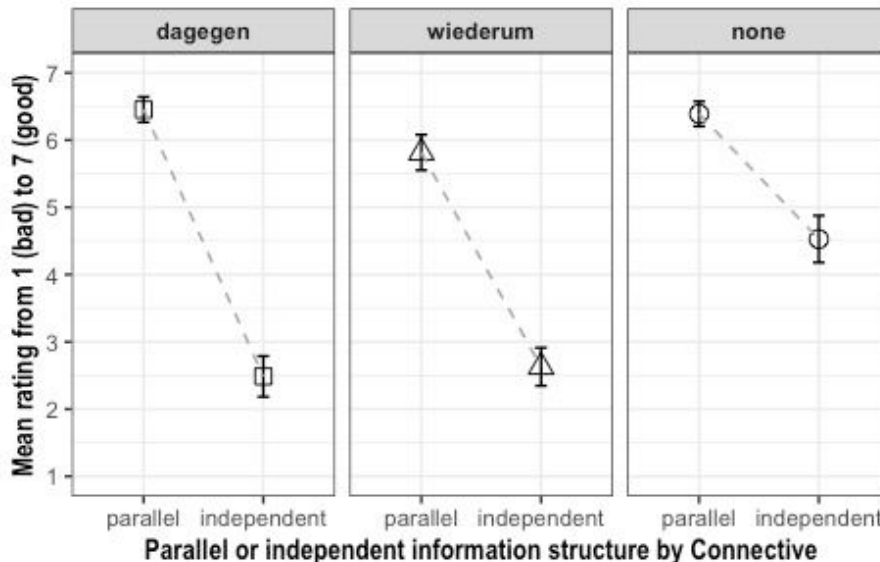


Figure 1 Mean ratings for *dagegen*, *wiederum*, and no connective

The plot allows for the following observations: first, all three connective-conditions are less acceptable in the independent condition than in the parallel one involving contrastive topics.

Second, the difference in acceptability ratings varies between the connectives *dagegen* and *wiederum* on the one hand and asyndetic connection on the other. While connections with *dagegen* or *wiederum* in the independent (yet contrastive!) condition received mean ratings of 2.5 and 2.6, respectively, the asyndetic continuation received mean ratings of 4.5. In other words, contrast without a contrastive information structure is less accessible without a contrastive marker, but still acceptable, whereas it becomes unacceptable with the markers *dagegen* or *wiederum*. Finally, the mean ratings in the parallel condition involving contrastive topics is similar across all three trellises, with 6.5 and 5.8 for *dagegen* and *wiederum*, respectively, and 6.4 for the asyndetic connection.

3 Discussion

The results of the acceptability judgment study corroborate both my claims on the nature of contrast with German *dagegen* and *wiederum*. The two connectives do require conjuncts with contrastive topics, otherwise they are rated as unacceptable. Moreover, contrastive topics are such strong markers of contrast on their own that (further) explicit marking by connectives is not required. The contribution of these connectives to discourse coherence is thus to signal a contrast already available via (information) structural means.

References

- Breindl, E., Volodina A. & Waßner, U.H. (2014). *Handbuch der deutschen Konnektoren 2*. Semantik der deutschen Satzverknüpfen. Berlin: Mouton de Gruyter.
- Büring, D. (2016). (Contrastive) topic. In Caroline Féry & Shinichiro Ishihara (eds.), *The oxford handbook of information structure*, 64–85. Oxford University Press.
- Daneš, F. (1970). Zur linguistischen Analyse der Textstruktur. *Folia Linguistica* 4(1–2). 72–78.
- Jasinskaja, K. (2012). Correction by adversative and additive markers. *Lingua* 122(15). 1899–1918.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276.
- Lakoff, R. (1971). *If's and but's about conjunction*.
- Sæbø, K. J. (2003). Presupposition and contrast: German *aber* as a topic particle. *Proceedings of Sinn und Bedeutung* (7). 25–271.
- Umbach, C. (2005). Contrast and information structure: A focus-based analysis of *but*. *Linguistics* 43(1). 207–232.
- Zieleke, R. (submitted). Contrast via information structure. On topic development with German *aber* in post-initial position. In Brysbaert, Jorina & Karen Lahousse (eds.). *On the role of contrast in Information Structure: Definition, types, encoding and annotation* (Trends in Linguistics. Studies and Monographs). Berlin: De Gruyter.
- Zieleke, R. (to appear). Three dimensions of contrast. Evidence from German contrastive discourse markers. PhD dissertation, University of Cologne.

Author Index

- Abeillé Anne, 57–64
Albu Elena, 133–135
- Bader Markus, 2–5, 101–104
Barnes Kathryn, 6–13
Beck Sara, 14–17
Benz Anton, 25–28
Bevivino Dorotea, 18–21
Bilbiie Gabriela, 22–24
Bleotu Adina Camelia, 25–28
Braun Bettina, 109–112
Brozba Gabriela, 25–28
Brunetti Lisa, 29–32
- Christensen Ken Ramshøj, 93–96
Collart Aymeric, 33–36
Copin Mireille, 97–100
Cortés Rodríguez Álvaro, 105–108
Cuneo Nicole, 37–40
- De Swart Peter, 121–124
Destruel Emilie, 41–44
Dieuleveut Anouk, 45–48
Dimroth Christine, 109–112
Drenhaus Heiner, 81–84
- Ebert Cornelia, 10–13
Ehrmantraut Luise, 49–52
- Faghiri Pegah, 53–56
Flecken Monique, 53–56
Frank Stefan, 121–124
Fricke Lea, 41–44
- Goldberg Adele, 37–40
- Hassen Amal, 57–60
Hemforth Barbara, 61–64, 97–100
Ho Deborah, 69–72
Hoek Jet, 140–143
Hsu Haley, 69–72
Hubers Ferdy, 125–128
- Istrate Marius Fabian, 61–64
- Kaiser Elsi, 65–72
- Kaup Barbara, 133–135
Kehl Andreas, 73–76
Kovač Iva, 77–80
Krifka Manfred, 89–92
- Lemke Tyll Robin, 81–84
- Mao Ruihua, 85–88
Modarresi Fereshteh, 89–92
Müller Christiane, 93–96
- Nyvad Anne Mette, 93–96
- Onea Edgar, 41–44
- Portele Yvonne, 2–5, 101–104
Post Claire, 69–72
Pozniak Céline, 97–100
- Raue Natascha, 105–108
Reich Ingo, 81–84
Reimer Laura, 109–112, 129–132
Rohde Hannah, 140–143
Rouse Madeline, 69–72
- Schoenmakers Gert-Jan, 77–80, 113–116
Schäfer Lisa, 81–84, 117–120
Storbeck Jesse, 65–68
Suijkerbuijk Michelle, 121–128
- Thiberge Gabriel, 29–32
Trotzke Andreas, 129–132
Tsaregorodtseva Oksana, 133–135
Turco Giuseppina, 18–21, 97–100
- Van Lier Eva, 53–56
- Walter Sebastian, 136–139
Weber Andrea, 14–17
Wilke Hans, 140–143
- Zieleke Regina, 144–147
Zimmermann Malte, 41–44