ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45 (3)

https://doi.org/10.18148/zs/2026-i3

Table of contents

Articles

Article 2026-2010

Quantitative perspectives on mismatches in ellipsis
Introduction to the special issue
MAX BONKE & GABRIELA BÎLBÎIE

Article 2026-2011

Gapping from a historical point of view
MARIANNA LOHMANN

Article 2026-2012

Gender mismatch and ellipsis
Comparing French adjectives and nouns
EMMA KIOUS & ANNE ABEILLÉ

Article 2026-2013

The identity condition of right node raising in German
Experimental evidence
LISA SCHÄFER & INGO REICH

Article 2026-2014

Structural priming and preposition omission compete in fragment responses to questions
JOANNA NYKIEL

Article 2026-2015

Pragmatic constraints on the complementizer *que* 'that' in Spanish embedded gapping
MAX BONKE

Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

Publisher

PubliKon
Deutsche Gesellschaft für Sprachwissenschaft – Linguistic Society of Germany

Co-Editors-in-Chief

Ingo Feldhausen
ATILF-CNRS & Université de Lorraine
UMR 7118, UFR Sciences Humaines et Sociales
54000 Nancy, France

Mathias Scharinger
Philipps-Universität Marburg
Institut für Germanistische Sprachwissenschaft
Pilgrimstein 16
35037 Marburg

Guest Editors

Gabriela Bîlbîie
University of Bucharest
Faculty of Foreign Languages and Literatures
Edgar Quinet 5-7
010017 Bucharest, Romania

Max Bonke
Universität zu Köln
Institut für deutsche Sprache und Literatur I
Albertus-Magnus-Platz
50923 Köln, Germany

Editors

Victor Edgar Onea Gaspar
Deutsche Sprache & Linguistik – Universität Graz
Mozartgasse 8/II
8010 Graz

Semra Kizilkaya
Fakultät für Linguistik und Literaturwissenschaft – Universität Bielefeld
Universitätsstraße 25
D-33615 Bielefeld

Anna Kocher
Digital Romanische Sprachwissenschaft
Katholische Universität Eichstätt-Ingolstadt
Ostenstraße 28a
85072 Eichstätt

Oliver Schallert
- Technical Editor -
Germanistische Linguistik – Department Germanistik und Komparatistik
Friedrich-Alexander-Universität Erlangen-Nürnberg
Bismarckstraße 1
D-91054 Erlangen

Sarah Schimke
Department I - Germanistik, Komparatistik, Nordistik, Deutsch als Fremdsprache
Institut für Deutsche Philologie
Ludwig-Maximilians-Universität München
Schellingstraße 3 RG
80799 München

Melanie Weirich
Sprechwissenschaft & Phonetik
Friedrich-Schiller-Universität Jena
07737 Jena

Technical support

Miriam Butt
Technical Editor (OJS)
Department of Linguistics – Universität Konstanz
Universitätsstraße 10
78464 Konstanz, Germany

Sebastian Danisch
Publikon
https://www.kim.uni-konstanz.de/publikon/
78457 Konstanz, Germany

Jana & Matthias Müller
Copy-Editing

SiliconChips Services
Copy-Editing & Produktion

Scientific advisory board

- Artemis Alexiadou (Humboldt University of Berlin, Germany)
- Sabine Arndt-Lappe (Trier, Germany)
- Amalia Arvaniti (Radboud University, Netherlands)
- Mathieu Avanzi (Neuchâtel, Switzerland)
- Ermenegildo Bidese (Trento, Italy)
- James Peter Blevins (University of Cambridge, UK)
- Ian Cunnings (Reading, UK)
- Christine Czinglar (Jena, Germany)
- Antje Dammel (Münster, Germany)
- Cristina Flores (Minho, Portugal)
- Peter Gallmann (Friedrich Schiller University of Jena, Germany)
- Hans-Martin Gärtner (Budapest, Hungary)
- Daniel Gutzmann (Bochum, Germany)
- Katharina Hartmann (Goethe University Frankfurt, Germany)
- Roland Hinterhölzl (Università Ca' Foscari Venezia, Italy)
- Anke Holler (Georg August University of Göttingen, Germany)
- Bradley Hoot (De Paul University, USA)
- Ángel Jiménez-Fernández (Sevilla, Spain)
- Baris Kabak (Julius Maximilian University of Würzburg, Germany)
- Simon Kasper (Düsseldorf, Germany)
- Felicitas Kleber (München, Germany)
- Theodoros Marinis (University of Konstanz, Germany)
- Laia Mayol (UPF Barcelona, Spain)
- Judith Meinschäfer (Free University of Berlin, Germany)
- Silvina Montrul (University of Illinois at Urbana–Champaign, USA)
- Christine Mooshammer (Humboldt University of Berlin, Germany)
- Natascha Müller (University of Wuppertal, Germany)
- Dennis Ott (Ottawa, Canada)
- Maria del Carmen Parafita Couto (Leiden, Netherlands)
- Michela Russo (Lyon, France)
- Petra Schulz (Goethe University Frankfurt, Germany)
- Antonella Sorace (Edinburgh, UK)
- Elisabeth Stark (University of Zurich, Switzerland)
- Volker Struckmeier (Wuppertal, Germany)
- Benedikt Szmrecsanyi (Katholieke Universiteit Leuven, Belgium)
- Luka Szucsich (Humboldt University of Berlin, Germany)

– Judith Tonhauser (Universität Stuttgart, Germany)
– Maria del Mar Vanrell (Mallorca UIB, Spain)
– Ruben Van de Vijver (Heinrich Heine University Düsseldorf, Germany)
– Hélène Vinckel-Roisin (Nancy, France)
– Ralf Vogel (Bielefeld University, Germany)
– Heike Wiese (Humboldt-Universität Berlin, Germany)
– Martina Wiltschko (UPF Barcelona, Spain)
– Katharina Zahner-Ritter (Trier, Germany)

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 1
https://doi.org/10.18148/zs/2026-2010

Max Bonke (corresponding author), Universität zu Köln, Germany, mbonke@uni-koeln.de
Gabriela Bîlbîie, University of Bucharest, Romania, gabriela.bilbiie@lls.unibuc.ro

Quantitative perspectives on mismatches in ellipsis: Introduction to the special issue

Abstract: We provide a brief introduction to the special issue, sketch the main topics, and provide an overview of the contributions.

Keywords: ellipsis, mismatch, quantitative methods, experimental syntax, corpus linguistics

Cite as: Bonke, Max & Gabriela Bîlbîie. 2026. Quantitative perspectives on mismatches in ellipsis: Introduction to the special issue. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2010.
Received: 30 September 2026; **accepted:** 01 October 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

Ellipsis has been of particular interest to linguistic theory because it raises fundamental questions about the relationship between form and meaning. When a speaker produces an elliptical construction, interpretation depends on material that is not overtly realized but must be recovered by the hearer from the conversational context. Generally, such recovery proceeds without great effort. The clear difference between what is pronounced and what is understood has made ellipsis a rich source for investigating the architecture of grammar. In particular, the apparent ease with which meaning is recovered from absent material has shed light on the interaction between syntax, semantics, discourse, and language processing.

Debates on ellipsis have prominently featured two key questions concerning the nature of the representation at the ellipsis site. The first question concerns syntactic representations: does ellipsis involve unpronounced structure, or can interpretation proceed without positing silent material? Competing approaches to this question have provided structural and non-structural answers, offering different views on how the grammar relates form to meaning. The second question concerns the relationship between an ellipsis and its antecedent, and in particular, the extent to which formal identity or structural parallelism is required in ellipsis. Is strict formal identity with some overtly expressed antecedent required to license ellipsis? Or, alternatively, can the content of the ellipsis site diverge in form from its antecedent, and if so, to what degree? Addressing these issues has not only advanced our understanding of ellipsis itself but has also informed broader debates about grammatical representation and the mechanisms underlying language comprehension.

While theoretical analyses have yielded important insights, the empirical basis of these debates has changed considerably as research progressed. As with many linguistic phenomena, early work largely relied on introspective judgments for constructed examples. Introspection remains an indispensable tool for generating hypotheses to this day, but as ellipsis research progresses, researchers have complemented introspective evidence with corpus data and experimental methods. These quantitative procedures have often proven advantageous compared to exclusively introspective approaches, both in ellipsis research and in the study of grammar more generally. They did so by making an evaluation of theoretical predictions possible against broader and systematically collected datasets. As a consequence, robust generalizations may be distinguished from isolated observations, enabling a more precise evaluation of competing theoretical proposals.

One venue for advancing the empirical turn in ellipsis research has been the workshop series *Experimental and Corpus-Based Approaches to Ellipsis* (*ECBAE*), which has been held (semi-) annually since 2017 in varying locations across Europe and the US. The present issue contains a selection of papers presented at the sixth edition of *ECBAE*, which was hosted at the 21st

International Congress of Linguists in Poznań, Poland, in 2024. The contributions cover a range of elliptical structures on the clausal level (gapping, stripping, fragments, Right-Node Raising (RNR)) in a number of different European languages (French, Spanish, English, Early and Modern German). All contributions in this issue address one of the most persistent and central questions in ellipsis research: how much identity between an ellipsis and its antecedent is actually required, or, put differently, to what extent mismatches between the two are possible. Methodologically, each contribution provides a rich quantitative basis for the discussion of its respective phenomenon, drawing on analyses of (historical) corpora, judgment data, reading times, or production data obtained through, for example, sentence completion tasks.

Marianna Lohmann investigates gapping in Early New High German (ENHG, 1350–1650) by focusing on those omissions of finite verbs in which ENHG differs from Modern Standard German (MSG). In comparison to MSG, ENHG places weaker restrictions on antecedent–ellipsis identity. ENHG allows not only mismatches in person/number-marking akin to MSG, but also permits mismatches in grammatical function, voice, or the exact lexical choice of the auxiliary as in (1).

(1) *da ist er eine Zeitlang still gelegen/ vnnd ~~hat~~ sich mit seinem*
there is he a time.span still lain and has REFL with his
Kriegßvolck Erquicket/.
troops refreshed
'He rested there for a while and recovered himself together with his troops.'
(Baumbank.UP, Johann Bange 1599, s488)

Drawing on the Baumbank.UP treebank (Demske 2019), all instances of these types of mismatches in coordinated clauses are identified and analyzed. The corpus study shows that these mismatches occur most frequently in the late ENHG period and in Upper German texts, a development which is similar to the development of subordinate clauses lacking a finite verb, known as afinite constructions. However, following Lohmann (2026: Chapter 7), it is argued that mismatched ellipses as in (1) are genuine instances of coordination ellipses, especially gapping, and should be kept apart from afinite constructions. This analysis is underpinned by data from Yiddish and Old English, which permit similar mismatches in coordination. Finally, the loss of mismatched gapping in Modern High German is attributed to the subsequent process of standardization, which is supported by evidence from prescriptive grammars.

The contribution by **Emma Kious** and **Anne Abeillé** focuses on gender mismatches in French stripping with predicative adjectives and human nouns. Earlier claims for other languages (Merchant 2014; Sprouse et al. 2021) posit that the two parts of speech differ in their tolerance of mismatches, raising the expectation that French adjectives tolerate mismatches like in (2) more readily than nouns (e.g., *chanteur* 'singer-M' vs. *chanteuse* 'singer-F').

(2) *Elle était vieille, et lui aussi ~~était~~ ~~vieux~~.*
she be-PST.3SG old-F.SG and he too be-PST.3SG old-M.SG
'She was old, and him too.'

Combining evidence from several contemporary French corpora with three acceptability judgment experiments, the study systematically investigates the effects of predicate type, mismatch, constituent order, noun morphology, and gender stereotypes on the acceptability and distribution of stripping ellipsis. Corpus data reveal that gender mismatches are attested for both predicative adjectives and, against expectations, nouns. The experiments further confirm that such mismatches are generally acceptable, with ratings at least as high as comparable anaphoric constructions. These findings challenge analyses that derive French nominal ellipsis through deletion under strict morphological identity, instead suggesting either that grammatical gender in French role nouns behaves like an inflectional agreement feature that can be ignored under ellipsis, or that stripping is best analyzed within a direct interpretation framework without syntactic reconstruction.

Lisa Schäfer and **Ingo Reich** investigate verb ellipsis by RNR in German complement clauses. They focus on mismatches in morphological person and number features such as in (3).

(3) *dass ich den Zeitschriftenartikel ~~lese~~ und Jan das Buch liest*
that I the journal.article read.1SG and Jan the book read.3SG
'that I read the journal article and Jan reads the book.'

They compare divergent predictions made by existing analyses that range from strict phonological identity via weaker identity requirements all the way to unrestricted acceptance of mismatches. The competing predictions are examined in five experiments that combine detailed acceptability data with results from sentence production and comprehension tasks. The production data reveal a clear mismatch penalty: speakers become progressively less likely to produce RNR as the morphological difference between antecedent and ellipsis site increases. By contrast, acceptability remains high even with substantial person and number mismatches, where RNR yields comparable patterns to the gapping and full-clause baselines. The paper argues that although speakers avoid producing stronger mismatches in RNR, such mismatches are nevertheless grammatical. The findings thus challenge theories that require strict phonological identity, instead supporting analyses in which the grammar allows for person and number mismatches in RNR. More broadly, the study demonstrates that production and acceptability data can reveal complementary aspects of the constraints governing mismatches in ellipsis.

Joanna Nykiel investigates structural priming in elliptical versus non-elliptical responses to English adjunct questions, asking whether structural priming effects are equally robust in both types of responses. The investigated paradigm is shown in (4) for the preposition *at* in the temporal adjunct *at 5 o'clock*.

(4) A: *At what time do you close?* (Prepositional question)
A': *What time do you close?* (Non-prepositional question)
B: *(At) 5 o'clock.* (Fragment answer)
B': *We close (at) 5 o'clock.* (Full clause answer)

The paper analyzes 868 question-response pairs originally collected in an elicitation experiment by Chia et al. (2020). While structural priming effects emerge across the dataset as a whole, the effects disappear for fragment responses. Instead, fragments overwhelmingly favor the omission of the preposition *at* for both prepositional and non-prepositional antecedent questions, despite the resulting structural mismatch in the case of prepositional antecedents. The paper proposes a processing-based account revolving around the specific form of the temporal adjunct in the question-response pairs. It argues that the readily accessible lexical content of the adjunct (*5 o'clock*) makes its temporal function transparent, rendering the preposition redundant as a retrieval cue. The preference of preposition omission in fragment responses is therefore analyzed as the outcome of competing processing pressures, with cue redundancy overriding the impact of structural priming that is observed both in the full set of responses and in the set of fragment responses to non-prepositional questions.

The paper by **Max Bonke** concerns a special property of gapping in Spanish: in contrast to languages like English or German, Spanish allows complementizers in embedded gapping. This is illustrated for the complementizer *que* 'that' in (5).

(5) *El cree que el mundo es su empresa y que los mexicanos ~~son~~ sus lacayos.*
He thinks that the world is his company and that the Mexicans are his lackeys
'He thinks that the world is his company and that the Mexicans are his lackeys.'

The paper asks whether *que* in embedded gapping is constrained by pragmatic properties such as factivity, assertion, or at-issueness. Previous work (Bonke & Repp 2022) showed that the acceptability of *que* depends on the embedding verb: it is fully acceptable with verbs of saying and thinking but degraded with other types. Under the hypothesis that *que* contributes a distinct component to sentence meaning, the paper investigates whether its presence affects the interpretation of embedded gapping. A by-verb correlation analysis of acceptability and independently obtained interpretation data suggests that factivity plays no role. Two rating experiments reach similar conclusions: the first applies a factivity diagnostic to gapping with and without *que*, while the second employs exploratory diagnostics to target assertion, at-issueness, and reportiveness. Across all measures, *que* shows no meaningful effect on interpretation, suggesting that *que* in embedded gapping unlikely impacts factivity or related pragmatic notions. Instead, the complementizer likely contributes no independent meaning in this construction.

Taken together, the contributions in this special issue suggest that identity constraints on ellipsis are less rigid and more nuanced than some structural approaches have assumed. By integrating quantitative evidence ranging from diachronic corpora to multi-method experimental paradigms, they highlight two main avenues of current research. First, morphological and structural mismatches (such as in agreement, grammatical function, or argument structure) may be acceptable even when production routines or historical standardization processes penalize them. This suggests that identity restrictions often arise from flexible, discourse-driven interpretation mechanisms rather than strict constraints rooted in grammar. Second, the presence or absence of overt material (such as functional complementizers or redundant prepositions) is modulated by processing efficiency, discourse transparency, and register. These collective findings point toward a gradient, multi-factorial view of ellipsis, in which mismatches may be compatible with grammatical constraints, while processing, frequency, and discourse factors shape their usage and distribution. Methodologically, these studies demonstrate that dynamic quantitative approaches drawing on multiple methods are essential for capturing how form, meaning, and processing pressures interact in ellipsis.

Acknowledgements: We thank the authors for contributing to this special issue and the presenters at *ECBAE* 6 in Poznań. MB gratefully acknowledges the CIPL travel grant to attend the *International Congress of Linguists* 2024, which hosted the *ECBAE* workshop. We furthermore thank all the anonymous reviewers for their valuable comments and evaluations of the papers in this issue, as well as the *ZS* head editors, especially Mathias Scharinger, for guiding the editorial process.

References

Bonke, Max & Sophie Repp. 2022. Complementizer deletion in embedded gapping in Spanish. *The Linguistic Review* 39(3). 525–555. https://doi.org/10.1515/tlr-2022-2096.

Chia, Katherine, Hannah Hetzel-Ebben, Maxwell Adolph, Madison Amaral, Maynor Arriaga, Holly Booth, Victoria Boudreau, Jennifer Carpenter, Carolina Cerra, Megan Clouden, Jacob Cryderman, Rida Darji, Jolie Dollison, Nicholas Franco, Lawrence Ghougasian, Laurel Hamilton, Kimiyo Karosas, Casey Kenoyer, Victoria Krenz, Samantha Lancaster, Melissa Ma, Grace Markwell, Faith Montoya, Rebecca Nadler, Samara Pinto, Melissa Rojas, Daniela Sarmiento, Chloe Stitik, Jasmine St. John, Mariana Valencia, Kayla Walker, Emma Wells, Julia Wolf, Destiny Wright & Michael P. Kaschak. 2020. Examining the factors that affect structural repetition in question answering. *Memory and Cognition* 48(6). 1046–1060. https://doi.org/10.3758/s13421-020-01036-2.

Demske, Ulrike. 2019. Referenzkorpus Frühneuhochdeutsch: Baumbank.UP. Universität Potsdam. https://doi.org/10.57754/FDAT.c7wr3-5k848.

Lohmann, Marianna. 2026. *Koordination im Frühneuhochdeutschen: Eine qualitative und quantitative Korpusuntersuchung*. Potsdam: Universitätsverlag Potsdam. https://10.25932/publishup-68092.

Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua* 151. 9–32. https://doi.org/10.1016/j.lingua.2014.01.008.

Sprouse, Jon, Troy Messick & Jonathan D. Bobaljik. 2021. Gender asymmetries in ellipsis: An experimental comparison of markedness and frequency accounts in English. *Journal of Linguistics* 58(2). 345–379.

ZS ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 2
https://doi.org/10.18148/zs/2026-2011

Marianna Lohmann, Freie Universität Berlin, Dahlem Center for Linguistics, Germany,
marianna.lohmann@fu-berlin.de

Gapping from a historical point of view

Abstract: Gapping is the optional omission of at least the finite verb in coordination. While in Modern Standard German (MSG) mismatches in person-/number-marking between the gap and its antecedent are the only ones attested, Early New High German (ENHG) also permits mismatches in 1) the verbal item (*sein* 'be' vs. *haben* 'have'), 2) verbal function (lexical vs. grammatical), and 3) voice (active vs. passive). Moreover, subordinate clauses without a finite verb are characteristic of ENHG. Such *afinite constructions* also occur in coordination structures. Previous studies have linked mismatched coordination ellipses to these afinite constructions. Based on data from the ENHG treebank, this study focuses on mismatch types in which ENHG differs from MSG. I argue that ellipses in main clause coordination should be analysed as genuine coordination ellipses, particularly gapping, rather than as phenomena resembling afinite constructions, even if they mismatch. I further claim that approaches to gapping that do not require full phonological identity between the gap and its antecedent can also analyse mismatched gapping in ENHG. Finally, there are indications that the standardisation process of High German may have contributed to the loss of these additional types of mismatched gapping, thus restricting gapping in MSG.

Keywords: ellipsis, gapping, right node raising, afinite construction, mismatches, Early New High German

Cite as: Lohmann, Marianna. 2026. Gapping from a historical point of view. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2011.
Received: 09 October 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

In Modern Standard German (MSG), finite verbs can be omitted in coordinated main (1a) and subordinate clauses (1b).[1]

(1) Gapping

a. *Maria* ***las*** *eine Zeitung und Peter* ~~***las***~~ *eine Zeitschrift.*
Maria read a newspaper and Peter read a magazine
'Maria read a newspaper and Peter (read) a magazine.'

b. *(Ich weiß,) dass Maria eine Zeitung* ***las*** *und Peter eine Zeitschrift* ~~***las***~~.
I know that Maria a newspaper read and Peter a magazine read
'I know that Maria read a newspaper and Peter (read) a magazine.'

The meaning of the ellipsis can be reconstructed based on the linguistic context formed by the coordination structure (see e.g. Aelbrecht 2015: 563 and Klein 1993: 766).[2] Thus, the overt finite verb *las* 'read' in the first conjunct licenses the gap in the second conjunct (Reich 2011: 1854–1861). A coordination ellipsis in which at least the finite verb is omitted in the second conjunct is called *gapping* (Ross 1967, 1970). This corpus study examines gapping from a historical point of view.

In Early New High German (ENHG, 1350–1650), ellipses of finite verbs seem to be less restricted than in MSG, see (2).

(2) *da* ***ist*** *er eine Zeitlang still gelegen/ vnnd* ~~***hat***~~ *sich mit seinem Kriegßvolck Erquicket/.*
there is he a time.span still lain and has REFL with his troops refreshed
'There he rested for a while and refreshed himself together with his troops.'
(Baumbank.UP, Johann Bange 1599, s488)

According to Sapp (2011: 34), reflexive verbs like *sich erquicken* 'refresh oneself' select the auxiliary *haben* 'have' in ENHG. Thus, the lexical verbs of the conjuncts select different auxiliaries to

[1] Antecedents are highlighted in bold, gaps with strikethrough. In afinite constructions, I adhere to the conventions of earlier studies and mark the position of the missing verb with an underscore. If the omission of the finite verb is possible in Modern English, I mark the gap in the translation with curly brackets.

[2] To make the discussion, in particular of the historical data, easier to understand, I refer to the initial conjunct as the first conjunct and the non-initial conjunct as the second conjunct, although, strictly speaking, the conjuncts are neither in the very first nor in the second position, as there may be linguistic material preceding them or between them.

form present perfect in (2), creating a mismatch between the gap (*hat* 'has') and the overt finite verb *ist* 'is' in main clause coordination.[3]

In ENHG, mismatched ellipses occur also in coordinated subordinate clauses, see (3).

(3) *aber weil es gar schwaerlich* ***war*** *auff zu kommen/ wegen des Windes/*
but because it at.all hard was up to get because the wind
welcher gemeinlich das gantze Jahr von Sudost kompt/ vnd der fluß
that usually the whole year from south.east comes and the river
auß Congo vns bey 6. Wochen auff gehalten ~~***hat***~~*: mußten wir auff 30. meil*
from Congo us to 6 weeks held.up has had we to 30 miles
saeglen.
sail
'We had to sail thirty nautical miles because it was hard to get up because of the wind that blows from south-east all year round, and the river from Congo held us up to six weeks.'
(Baumbank.UP, Samuel Braun 1624, 61)

In the first conjunct, the clause contains the copula *war* 'was' in perfect tense. However, there is no copular construction in the second conjunct. In addition, the second conjunct is in active voice. Here, the participle *auff gehalten* 'stopped' is used to form the perfect tense, and, as a transitive verb, it selects the auxiliary *haben* 'have' (Sapp 2011: 34).[4] In (3), the overt finite verb *war* 'was' and the gap *hat* 'has' thus differ in their function (copula vs. auxiliary).

Previous studies on ENHG link mismatched ellipsis to non-coordinated clauses that lack a finite verb (Breitbarth 2005b; Thomas 2019), as in (4).

(4) *Alß wir vns vm Mittagszeiten zuoruck nach dem Portu wider gewendet_/*
as we us at midday back to the harbour again turned
ersahen wir zur lincken von ferrne ain Schiff/
spotted we at left from afar a ship
'As we made our way back to the harbour at midday we spotted a ship in the distance on the left-hand side.'
(Baumbank.UP, Leonhart Rauwolf 1582, 99)

[3] One should be aware of the fact that existence of state verbs like *liegen* 'lie' select both *haben* 'have' and *sein* 'be' in ENHG and MSG as in (i), see also Table 1 in Section 3.1.

(i) *Ich bin/habe gesessen/gelegen/gestanden*
I was/have sitting/lying/standing
'I was sitting/lying/standing.'

[4] Sapp (2011: 34) points out that not all instances of clauses with transitive participles can be regarded as perfect tense cases. It is also possible that the auxiliary *sein* 'be' forms the passive voice. However, it is unlikely that the second conjunct in (3) is in the passive rather than the active voice. I therefore stand by this interpretation.

In (4), only the participle *gewendet* 'turned' is overt, but there is no finite auxiliary verb.[5] Note that this finite verb ellipsis is not licensed by a coordination relation, but by subordination. Dependent clauses that lack finite verbs (typically auxiliaries) in ENHG are referred to as *afinite constructions* (Biener 1925; Schröder 1985; Demske-Neumann 1990; Ebert et al. 1993; Breitbarth 2004, 2005a,b; Senyuk 2014; Blum 2018; Thomas 2019; Demske 2022).

In previous studies, both ellipses – mismatched ellipses as in (2) and (3), and afinite constructions (4) – were considered as one phenomenon. Breitbarth (2005b), who focuses mainly on afinite constructions in her study, claims that afinite constructions and coordinated clauses with mismatched ellipses exhibit a "lesser degree of spatio-temporal anchoring" (Breitbarth 2005b: 127). Thus, the function of creating a dependency relation is not only attributed to afinite constructions (Admoni 1967; Demske-Neumann 1990; Breitbarth 2004, 2005a,b; Senyuk 2014), but also to the ellipsis of coordinated clauses, so that "one of the clauses is really dependent on the other" (Breitbarth 2005b: 149). For Thomas (2019: 17), however, afinite constructions and mismatched ellipses in coordination have in common that the ellipsis is a learned stylistic feature.

While Breitbarth (2005b) and Thomas (2019) consider finite verb ellipses in coordination from the perspective of afinite constructions, this paper takes up a coordination-based approach, which is briefly mentioned by Breitbarth (2005b: 63) as an alternative analysis: "the auxiliary ellipsis in coordinated main clauses may potentially be an extension of Gapping."[6] However, with regard to the diachronic development of gapping, I follow Lohmann (2026: Chapter 7), reverse Breitbarth's statement, and argue that gapping in MSG is the result of a restriction of (historical) gapping. Thus, there are three central questions underlying this study: 1) What are the similarities and differences of mismatched ellipses in coordination and ellipses in afinite constructions? 2) In what properties do mismatched ellipses in coordination behave akin to gapping? 3) Why have (certain types of) mismatched gapping disappeared in MSG?[7]

To answer these research questions, I conduct a corpus study based on the Baumbank.UP (Demske 2019), the largest historical German treebank to date. Due to its balanced structure, the corpus (527,597 tokens) enables both diachronic (1350–1650) and diatopic (Middle West,

[5] For an alternative analysis of the lexical verbs at the end of subordinate clauses as ambiguous between participles of present perfect and finite verbforms in preterite tense see for instance Biener (1925: 296) and Blum (2018).

[6] Here, Breitbarth (2005b) quotes Behaghel (1928), probably referring to the idea mentioned at the beginning of section 2.4.3 in Breitbarth (2005b). The idea of Behaghel (1928: 491) is that afinite constructions developed "due to an extension of regular conjunction reductions" (Breitbarth 2005b: 56). However, in making this statement, he does not refer explicitly to mismatched ellipses in coordination, but to coordination ellipses in general.

[7] Whereas mismatches in person-/number-marking still occur in MSG (see Section 2.1), other types disappeared. It would therefore be promising to investigate gapping in non-standard varieties of Modern German, given the clear preference for mismatched gapping in the Upper German area in ENHG, see Table 2 in Section 3.3.

Middle East, Upper West, and Upper East) investigations on ENHG data. In contrast to Breitbarth (2005b: 113) and Lohmann (2026: Chapter 7), mismatches in person-/number-marking are not taken into account, since they appear in MSG, see Section 2.1. Instead, this study focuses on mismatches that diverge from MSG, which is mismatches in 1) the verbal item itself (*sein* 'be' vs. *haben* 'have'), 2) verbal function (lexical verb vs. grammatical verb), and 3) voice (active vs. passive). The quantitative analysis revealed 58 instances of mismatched ellipses among 8,397 coordinated structures containing the conjunction *und* 'and'. These ellipses show an increased frequency during the 16th century. This trend reflects the emergence of afinite constructions (cf. Breitbarth 2005b: 63; Blum 2018: 73–81; Lohmann 2026: 226). Moreover, in the Baumbank.UP afinite constructions predominantly appear in Upper German (69%) (cf. Blum 2018: 73–81; Lohmann 2026: 227). The same holds true for mismatched ellipses (71%). However, unlike afinite constructions, mismatched ellipses also occur in main clause coordination, exclusively in second conjuncts (cf. Lohmann 2026: 245). Here, objects can also be elided, similar to gapping.

I propose analysing mismatched ellipses in coordinated main clauses separately from ellipses in subordinate clauses (coordinated and non-coordinated). In main clause coordination, I argue that the ellipsis is genuine gapping. I will refer to this type of mismatched ellipsis as *mismatched gapping*. Further evidence supporting the claim that mismatched ellipses should be analysed as gapping, rather than as afinite constructions, comes from two West Germanic languages. Yiddish, which is closely related to German, exhibits mismatched gapping comparable to ENHG (cf. Birnbaum 1988: 67; Jacobs 2005: 262). Additionally, gapping in Old English differs from gapping in Modern English, too (Pérez Lorido 1996: 336). Crucially, neither Yiddish nor Old English have afinite constructions. The occurrence of mismatched gapping in these languages therefore suggests that the phenomenon in ENHG may not be directly related to afinite constructions, but rather to coordination.

Furthermore, there is evidence to suggest that the standardisation of High German may have influenced gapping, thereby resulting in the loss of the patterns 1)—3) in MSG. Firstly, the phenomenon of mismatched gapping has been observed in Yiddish, a language that has not undergone a standardisation process. Secondly, prescriptive grammarians have made statements about the appropriate use of ellipsis in German, in which they advise to avoid certain mismatches.

The paper is structured as follows: After outlining the general topic, current research questions and main results in Section 1, Section 2 provides the background on finite verb ellipses in High German and addresses their diachrony. Section 3 outlines the corpus study and Section 4 discusses the results in a broader context. Section 5 concludes by summarising the main points presented in this paper and raising some further research questions.

2 State of the art on ellipses of finite verbs

The existence of ellipsis, what constitutes an ellipsis, and how meaning arises when no linguistic item is overt are topics that are still being discussed in linguistics (cf. Aelbrecht 2015; van Craenenbroeck & Temmerman 2018). Most research on ellipsis has focused on modern languages. New insights have been gained by testing the grammaticality of elided vs. non-elided language items, for example. It should be noted, however, that the study of historical languages such as ENHG can only be based on corpus data. Consequently, it is not possible to conduct grammaticality tests on this data. Before presenting my approach to dealing with the corpus data from the 13th to 17th centuries in Section 3, I will discuss the current state of research on finite verb ellipses. I begin by outlining the key characteristics of finite verb ellipses in MSG (Section 2.1), followed by a discussion on ellipses in historical language periods of High German (Section 2.2).

2.1 Ellipses of finite verbs in Modern Standard German

Here, I focus on context-controlled ellipsis (Klein 1993: 766), which is also referred to as "a[ntecedent-based] ellipsis" (Reich 2011: 1854–1861). In coordination, the meaning of the elided part can be reconstructed from the linguistic context (cf. Aelbrecht 2015: 563; Klein 1993: 766).[8] The coordination structure therefore provides the linguistic context for the ellipsis.[9] Depending on the order of the antecedent and the ellipsis, the ellipsis types are terminologically distinguished. 1) The antecedent of the ellipsis is in the first conjunct and the ellipsis itself is in the second conjunct, see (5). This ellipsis of finite verbs is called *gapping* (Ross 1967, 1970).

(5) *Maria* ***las*** *eine Zeitung und Peter* ~~***las***~~ *eine Zeitschrift.*
Maria read a newspaper and Peter read a magazine
'Maria read a newspaper and Peter (read) a magazine.'

2) The antecedent is in the second conjunct and the ellipsis is in the first conjunct, as in (6). This ellipsis is known as *right node raising* (RNR)[10] (Postal 1974).[11]

[8] There is also the opinion that ellipses do not exist, which means nothing else than the phonetically realised linguistic items are interpreted (see Ginzburg & Sag 2000; Culicover & Jackendoff 2005).

[9] Lechner (2004: 97–113) shows similar properties of gapping in coordination and comparatives. A few examples here do indeed show gapping in non-standard varieties of Modern German.

[10] In Ross (1970: 253), it is referred to as *backward gapping*.

[11] In contrast to gapping, RNR does not have to affect minimally the finite verb. Items that are constituents or smaller than constituents can be elided in first conjuncts, see (i).

(i) Maria hat [DP ein Zebra und zwei Giraffen] gesehen, aber Petra hat [DP ein Kamel und fünf Giraffen] gesehen
Maria has one zebra and two giraffes seen but Petra has one camel and five giraffes seen
'Maria saw one zebra and two giraffes, but Petra saw one camel and five giraffes.'
(Hartmann 2015: 507)

(6) *(Ich weiß,) dass Maria eine Zeitung* ~~***las***~~ *und Peter eine Zeitschrift* ***las***.
I know that Maria a newspaper read and Peter a magazine read
'I know that Maria read a newspaper and Peter read a magazine.'

In coordination of main clauses, the finite verb cannot be elided in the first conjunct. This is shown in (7). Thus, RNR cannot apply here.[12]

(7) **Maria* ~~***las***~~ *eine Zeitung und Peter las eine Zeitschrift*
Maria read a newspaper and Peter read a magazine
'Maria read a newspaper and Peter read a magazine.'

In coordination of subordinate clauses, RNR (6) and gapping (1b) (repeated here again in (8)) are possible.

(8) *(Ich weiß,) dass Maria eine Zeitung* ***las*** *und Peter eine Zeitschrift* ~~***las***~~.
I know that Maria a newspaper read and Peter a magazine read
'I know that Maria read a newspaper and Peter (read) a magazine.'

When gapping occurs in coordinated subordinate clauses, not only the finite verb but also the complementizer has to be elided in the second clause (cf. Repp 2009a,b; Bonke 2024: Appendix A), see (9).

(9) **(Ich weiß,) dass Maria eine Zeitung* ***las*** *und dass Peter eine Zeitschrift* ~~***las***~~.
I know that Maria a newspaper read and that Peter a magazine read
'I know that Maria read a newspaper and (that) Peter (read) a magazine.'

If gapping is defined as the omission of at least the finite verb, then the ellipsis known as *stripping*[13], see (10), is a subtype of gapping (Repp 2009a).

(10) *Maria* ***las*** *eine Zeitung und Peter* ~~***las***~~ *(auch)* ~~*eine Zeitung*~~
Maria read a newspaper read Peter read also a newspaper
'Maria read a newspaper and Peter (read a newspaper), too.'

For a complete overview of the differences and similarities between gapping and RNR in MSG see for instance Reich (2011).

[12] In main clause coordination, RNR can apply on the non-finite verb, as in (i). Here, gapping and RNR are combined.

(i) *Maria* ***hat*** *eine Zeitung* ~~***gelesen***~~ *und Peter* ~~***hat***~~ *eine Zeitschrift* ***gelesen***
Maria has a newspaper read and Peter has a magazine read
'Maria read a newspaper and Peter read a magazine.'

[13] Some studies refer to it as *bare argument ellipsis* (Hankamer 1979).

Although gapping allows for person-/number-mismatches (Wilder 1994: 308), see (11a), tense, mode and aspect of antecedent and ellipsis have to be identical (Repp 2009b: 229–230), see (11b).

(11) Gapping

a. *Ich* ***trinke*** *das Bier und du* ~~***trinkst***~~ *den Wein.*
I drink the beer and you drink the wine
'I drink beer and you (drink) wine.'

b. **Beckham* ***spielte*** *gestern und ich* ~~***spiele***~~ *morgen.*
Beckham played yesterday and I play tomorrow
'Yesterday, Beckham played and I will play tomorrow.'
(Reich 2011: 1860)

For RNR, Christ (2011: 378) argues that there is no strict phonological identity required between ellipsis and antecedent but a phonological quasi-identity. While most of the data he discusses concern mismatches between constituents (and smaller items) regarding the grammatical categories of case and number (Christ 2011: 381–397), he also mentions the category of person, which is marked by finite verbs (Christ 2011: 397–400). Christ (2011: 397) also argues that irregular verbs, such as *trinkst* 'drink' in (12a), are less frequently elided in the first conjuncts, instead serving as overt antecedents for ellipses.

(12) Right node raising

a. *(Ich weiß), dass ich immer Bier* ~~***trinke***~~ *und du immer Wein* ***trinkst****.*
I know that I always beer drink and you always wine drink
'I know that I am always drinking beer and you are always drinking wine.'

b. **(Ich weiß), dass Beckham gestern* ~~***spielte***~~ *und ich morgen* ***spiele****.*
I know that Beckham yesterday played and I tomorrow play
'I know that yesterday Beckham played and I will play tomorrow.'

However, the data discussed in Christ (2011: 397–400) cannot be easily categorised into grammatical and ungrammatical, so it is not possible to draw valid conclusions from them. Tense and mood mismatches between ellipsis and antecedent are without doubt ungrammatical in RNR, see (12b) for a tense mismatch (Christ 2011: 407). Compared with gapping (13a), RNR is not limited to coordination structures (13b).

(13) Subordination

a. **Maria* ***hat*** *eine Zeitung gelesen, während Peter ferngesehen* ~~***hat***~~*.*
Maria has a newspaper read while Peter watched.tv has
'Maria read a newspaper while Peter watched TV.'

b. *Maria hat eine Zeitung* ~~***gelesen***~~*, während Peter zwei Zeitschriften* ***gelesen*** *hat.*
Maria has a newspaper read while Peter two newspaper read has
'Maria read a newspaper while Peter read two magazines.'

Since ellipsis is the strictest way of de-accentuating linguistic expressions, it is used to contrast linguistic material, leading to accentuating the remnants (cf. Féry & Hartmann 2005). However, not all conjuncts can be contrasted in MSG. As illustrated by example (14), RNR is ungrammatical in *asymmetric coordination* structure. The object *die Puppe* 'the doll' cannot be elided at the right edge of the first clause (Bonitz 2013: 45).

(14) **Gestern kämmte Sophia* ***~~die~~ Puppe*** *und verkleidete* ***die Puppe***
yesterday combed Sophia the doll and dressed the doll
'Yesterday, Sophia combed the doll and she dressed her.'
(Bonitz 2013: 45)

Since a finite verb must generally be overt in the second conjunct's main clause position in asymmetric coordination, gapping is not possible here (cf. Reich 2009a: 38). Further, in asymmetric coordination, the events are interpreted as one merged event (cf. Höhle 2019b: 204), or, according to Barnickel (2017: 91), the second event is in "immediate posteriority". Thus, the second conjunct depends on the first conjunct (cf. Reich 2009b: 207). In MSG, there are two types of asymmetric coordination structures (cf. Höhle 2019a,b). (15a) is characterised by an obligatory subject gap in the second conjunct. In (15b), the second conjunct does not exhibit verb-final word order. Instead, there is V2 word order.

(15) a. *Hoffentlich sieht uns keiner und zeigt uns an.*
hopefully sees us nobody and reports us on
'Hopefully, no one sees us and reports us.'
(Höhle 2019b: 202)

b. *Wenn jemand nach Hause kommt und da steht der Gerichtsvollzieher vor der Tür.*
when someone at home comes and there stands the bailiff at the door
'When someone comes home and the bailiff is standing at the door.'
(Höhle 2019a: 296)

Asymmetrical coordination structures therefore cannot express a contrastive reading of conjuncts (cf. Reich 2009a: 115) and do not allow gapping and RNR (cf. Höhle 2019b: 204). Thus, gapping and RNR are characteristic properties of (symmetrical) coordination structures, whereas RNR can also occur in subordination.

2.2 Ellipses of finite verbs in the history of German

In Old High German (OHG, 750–1050) coordination structures occur with two coordinators *inti* 'and' and *joh* 'and' (cf. Eroms 2010; Oppermann 2025). Oppermann (2025: 124) argues that

gapping is present in Old High German (contra Eroms 2010: 301). This ellipsis occurs with *inti* 'and', see (16), but there are no data with *joh* 'and' (Oppermann 2025: 124).[14]

(16) *Got* ***ist*** *dher quhad, endi got* ***~~ist~~*** *dher deta.*
God is who spoke and God is who did
'God is the one who spoke, and God is the one who created/acted.'
(Is 4,4 as cited from Oppermann 2025: 127)

RNR is also present in OHG (17), but not very frequently (Oppermann 2025: 124–125).[15]

(17) *dér ímo zuíualt.* ***~~líutet~~*** *únde énemo fieruált* ***líutet***
which him twofold rings and that.one fourfold rings
'which will sound twofold for this one and fourfold for that one.'
(Mus IV,18 as cited from Oppermann 2025: 125)

RNR occurs in Middle High German (MHG, 1050–1350) (Oppermann 2025: 142–143), so does gapping (Oppermann 2025: 144–146), see (18).

(18) *da din herze mitte gilabit* ***wart****. un gitrostit* ***~~wart~~***
there your heart with refresh became and comforted became
'With which your heart was refreshed and comforted.'
(Muri 15va,12f. as cited from Oppermann 2025: 144)

In Early New High German (ENHG, 1350-1650), gapping (19a) and RNR (19b) appear as well (Demske 2008: 303–304; Lohmann 2026: 112–116).

(19) a. *Jedoch* ***vlevzt*** *colera von im in die orn vnd melancolia* ***~~vlevzt~~***
however flows cholera from him in the ears and melancholy flows
da tz den augen auz Vnd fleuma ***~~vlevzt~~*** *da zue der nasen*
there to the eyes out and phlegm flows there to the nose
'However, cholera flows from him into the ears and melancholy flows out to the eyes and phlegm flows to the nose.'
(Baumbank.UP, Konrad von Megenberg 1348–1350, 96)

b. *auch mit dem Geschuetz auff dem Wahle dermassen vnter sie*
also with the gun at the rampart so.much among them
Geschossen/ daß Arm vnd Schenckel in die Lufft geflogen ***seyn~~/~~*** *vnd.*
shot that arms and thighs in the air blown.up are and
vngefehr 1700. Frantzosen Tod blieben ***seyn/***
approximately 1700 Frenchmen dead left are
'There was so much shooting among them on the rampart that arms and thighs were blown up and Frenchmen were left dead.'
(Baumbank.UP, Johann Bange 1599, 1008)

[14] The highlighting with bold and strikethrough is by M. L.
[15] As RNR does not necessarily have to elide the finite verb, but can instead omit other linguistic items such as constituents (see (10) in footnote 10), Oppermann (2025) does not count RNR involving only elided finite verbs separately. Consequently, there is no quantitative information in Oppermann (2025) on this point.

Throughout the history of High German, evidence of gapping and RNR has thus been found. The ellipses mentioned so far are phonologically identical to their antecedents. However, in addition to ellipsis under identity, in ENHG finite verbs can also be elided if there is no identity between the ellipsis and the antecedent. This can be either in terms of lexical entry, for example, different auxiliaries, see (20a), or in terms of grammatical feature marking, for example, different number marking, see (20b) (cf. Breitbarth 2005b: 113; Demske 2008: 304–305).

(20) a. *vnd die drey gebrueder* ***sein*** *auf solch ersuechen khumen/ vnd* ~~***haben***~~
and the three brothers are on such request come and have
das gepiet zwischen einander auß geteilt
the land among eachother divided
'On this request, the three brothers came and they divided the land among them.'
(Baumbank.UP, Sigmund von Herberstein 1557, 83; cf. Breitbarth 2005b: 5)

b. *drumb siht man itzt/ das welschlandt/ fast wust* ***ist****/*
therefore sees one now that Italy almost depredated is
kloster vorstoret ~~***sind***~~*/ bistumb vortzeret* ~~***sind***~~*/ prelaturn vnd*
monasteries destroyed are dioceses consumed.up are prelatures and
aller kirchen tzinße gen Rome tzogen ~~***ist***~~
all churches tax towards Rome drawn is
'therefore, one can see now that Italy is almost completely depredated, that monasteries are destroyed, dioceses are consumed up, that money from all churches and prelatures is drawn to Rome.'
(Luther 1520, C3^{b}, 29—32, as cited from Breitbarth 2005b: 5)

There are two corpus studies that consider mismatched ellipses in ENHG. However, the main focus of these studies, Breitbarth (2005b) and Thomas (2019), is on afinite constructions, see (4) repeated here in (21), i.e. dependent clauses that lack a finite verb in ENHG.

(21) *Alß wir vns vm Mittagszeiten zuoruck nach dem Portu wider gewendet_/*
as we us at midday back to the harbour again turned
ersahen wir zur lincken von ferrne ain Schiff/
spotted we at left from afar a ship
'As we made our way back to the harbour at midday we spotted a ship in the distance on the left-hand side.'
(Baumbank.UP, Leonhart Rauwolf 1582, 99)

It is the absence of finite verbs, mostly auxiliaries, that both mismatched ellipses in coordination (20a,b) and afinite constructions (21) have in common.

One quantitative study of auxiliary ellipsis is that of Breitbarth (2005b). The corpus contains 25 printed texts from 1450 to 1800. This time span is divided into 50-year segments, with five texts in each segment. 286,000 words were analysed, with 45,000 words per segment (Breitbarth

2005b: 13–15). In total, there are 2,474 auxiliary ellipses (Breitbarth 2005b: 15). These include (non-coordinated) afinite constructions and, as Breitbarth (2005b: 58, fn. 26) points out, coordinated main and subordinate clauses with ellipsis of finite verbs that deviate from gapping and RNR under identity. Breitbarth (2005b: 146–148) refers to these as "non-parallel coordination ellipses". 55% of auxiliary ellipses are mismatched ellipses in coordination (Breitbarth 2005b: 59).[16] These mismatches can be in type of auxiliary selected, number-marking, or tense (Breitbarth 2005b: 59–60). In 68% of the coordination data, an overt auxiliary is entirely missing (Breitbarth 2005b: 61). In her analysis, Breitbarth (2005b) does not differentiate between ellipsis in coordination of main and subordinate clauses. The Appendix contains the total number of instances (n = 168) and relative frequencies for each text (Breitbarth 2005b: 161–162), which show the instances of main clauses counted separately. Unfortunately, there is no mention of how many cases do not match in terms of grammatical features compared to lexical entries in main clause coordination.

A second recent study on auxiliary ellipsis is by Thomas (2019). The corpus contains data of GerManC (1650–1800), parts of the Bonn Corpus of Early New High German (1350–1600) and a corpus of texts from 1490–1550, which Thomas (2019: 17) compiled herself. However, this study only investigates mismatched ellipses in subordinate clauses (coordinated and non-coordinated) (Thomas 2019: 44). Thomas (2019) therefore does not collect any data at all on main clause coordination.

In addition to the auxiliary ellipsis, mismatched ellipses in coordination and afinite constructions share a similar frequency development. Afinite constructions spread rapidly in 16th century, but disappeared quickly after 1700 (Breitbarth 2005b: 43). The same applies to mismatched ellipses in coordination. They "emerge and disappear along with" afinite constructions (Breitbarth 2005b: 6). However, Breitbarth (2005b) includes ellipses in both coordinated main clauses and coordinated subordinate clause, as well as in non-coordinated subordinate clauses. It is therefore unclear whether this development would be confirmed by analysing coordinated and non-coordinated data separately.

It is argued that, for afinite constructions, the ellipsis creates a dependency relation (Admoni 1967; Demske-Neumann 1990; Breitbarth 2004, 2005a,b; Senyuk 2014). Thus, Breitbarth (2005b: 149) assumes that in coordination with a mismatched ellipsis, such as in (20a), "one of the clauses is really dependent on the other" due to the ellipsis. Referring to research on asymmetric coordination in MSG by Höhle (2019a) and Büring & Hartmann (1998), Breitbarth (2005b: 146–147) argues that (20a) exhibits the characteristics of asymmetric coordination, see Section 2.1. One such feature is an obligatory subject gap. Secondly, the order of the conjuncts cannot be switched, a feature which may also be evident in (20a) (Breitbarth 2005b: 146–147).

[16] Please note that in Breitbarth (2005b), afinite constructions and mismatched ellipses in coordination are subsumed under the term auxiliary ellipsis.

(22) **vnd die drey gebrueder* ~~***haben***~~ *das gepiet zwischen einander auß*
and the three brothers have the land among eachother
geteilt vnd ***sein*** *auf solch ersuechen khumen*
divided and are on such request come
(Breitbarth 2005b: 147)

Although it is true that there is a subject gap in the second conjunct of (20a), (22) is not a suitable test of historical data, as these are extracted from a corpus. As Labov (1972: 100) rightly points out, corpus data are "bad data", in the sense that they can neither be manipulated, as in (22), nor can their grammaticality be tested. This also serves as an argument against Thomas (2019: 17) view that auxiliary ellipsis in ENHG is a purely learned stylistic phenomenon, "[b]ecause AE [Auxiliary Ellipsis] cannot easily be categorised as grammatical or ungrammatical". If a stylistic feature does not reflect the grammar of a language, how can the language be used by its producers and understood by its perceivers?

Furthermore, asymmetric coordination is characterised by the absence of three properties that are otherwise all found in symmetric/parallel coordination: 1) no across-the-board (ATB) movement (cf. Büring & Hartmann 1998: 178); 2) no ellipses, except of subject gaps (cf. Höhle 2019b: 204); and 3) no contrastive reading (cf. Reich 2009a: 115), see also Section 2.1. Does (20a) show one or more of these features? Regarding 1), the possibility of ATB-Movement can just be checked by manipulating the data as it is shown in (23).

(23) [*Einen Wagen*]$_i$ *kaufte Hans* t_i *und meldete ihn/**t_i *sofort an*
a car bought Hans and registered it immediately on
'Hans bought a car and he registered it immediately.'
(Büring & Hartmann 1998: 178)

This property cannot therefore be proven to be either right or wrong based on historical data. The same applies to property 3), see Section 2.1. In (20a), based on world knowledge, one would say that the events expressed by the conjuncts occur one after the other. However, we do not know whether speakers in the Early Modern period interpreted the events expressed by the conjuncts as one merged event, i.e. did not contrast them, or whether, on the contrary, they used the ellipsis precisely to emphasise the remnants, as Pérez Lorido (1996: 333) assumes with regard to the occurrence of mismatched ellipses in Old English. However, it is well known that the coordinator *und* 'and' allows just for a progressive order of events (cf. Bar-Lev & Palacas 1980: 144). Therefore, asymmetric coordination is only appropriate with the coordinator *und* 'and' (cf. Reis 1993: 214; Reich 2009a: 146–148). If one wants to compare mismatched ellipses with asymmetric coordination, one could also check whether mismatched ellipses only occur with the coordinator *und* 'and', see Section 4.

Diachronically, asymmetric coordination with subject gaps in second conjuncts might have been part of German grammar throughout the history. There is data that shows similar structures like asymmetric coordination with subject gaps in OHG, MHG (Oppermann 2021; Oppermann 2025: Chapter 5), and in ENHG (Demske 2008: 305–308; Lohmann 2026: 106–112), see (24).

(24) *Mit dieser Antwort sein die Stend noch nicht content/ vnd wollen biß*
with this answer are the parties yet not satisfied and want till
Mitwoch noch ein Schrifft vbergeben
Wednesday yet a document submit
'The parties are not yet satisfied with this response and would like to submit a document by Wednesday.'
(Aviso 1609, 47.22, as cited from Demske 2008: 306)

Unlike in asymmetric coordination (24), the finite verb is also elided in (20a). However, regarding property 2), it is unclear whether data with mismatched ellipsis such as in (20a) exhibit further ellipses, such as elided objects. This is a property of gapping, and therefore of symmetric coordination, see (10) in Section 2.1, but not of asymmetric coordination (cf. Höhle 2019b: 204), see (25).[17]

(25) **Morgen zeigt Karl dem Onkel die Briefmarken und bietet* ~~***sie***~~ *ihm*
tomorrow shows Karl the uncle the stamps and offers them him
zum Verkauf an.
for sale on
'Tomorrow, Karl will show the uncle stamps and he will offer him the stamps for sale.'
(Höhle 2019b: 204)

Breitbarth (2005b: 63) suggests that "ellipses in coordinated main clauses may potentially be an extension of gapping". This quote originally refers to Behaghel (1928: 491), who, however, mentions that afinite constructions may originate from coordinated clauses lacking a finite verb. As Breitbarth (2005b) analyses both afinite constructions and mismatched coordination as one type of auxiliary ellipsis, she reinterprets the quote, despite him not explicitly referring to mismatched ellipses in coordination. Furthermore, Breitbarth (2005b: 63) states that "one would probably subsume the auxiliary ellipses in coordinated subordinate clauses under the uncoordinated ones", i.e. afinite constructions. Thus, the question of whether mismatched ellipses and afinite constructions are one and the same phenomenon, or whether they should be treated separately, remains unanswered.

I take up Breitbarth's (2005b: 56–63) assumption on mismatched ellipses in coordination, as well as Thomas's (2019: 17) hypothesis that auxiliary ellipses constitute a learned stylistic device in ENHG. However, I reverse this. Based on Lohmann (2026: Chapter 7), I claim that mismatched ellipses are genuine instances of coordination ellipses, especially gapping. Furthermore, there

[17] The marking of the ellipsis of *sie* 'them' is by M. L.

is evidence that their disappearance may have been stylistically enforced by prescriptive grammarians. The current occurrence of gapping in MSG, see Section 2.1, might therefore be a consequence of the restriction of gapping over the historical development of German.

To demonstrate this, I distinguish between non-coordinated subordinate clauses lacking a finite verb and coordinated clauses. In contrast to Breitbarth (2005b), Demske (2008), Thomas (2019) and Lohmann (2026: Chapter 7), however, I focus on mismatches that deviate from MSG in this paper. Thus, I do not consider ellipses that only show person-/number-mismatches, see Section 2.1, but ellipses with mismatches in 1) the verbal item itself (*sein* 'be' vs. *haben* 'have'), 2) the verbal function (lexical verb vs. grammatical verb), and 3) voice (active vs. passive). Furthermore, earlier research has not specified how they evaluate their data with coordination ellipses as mismatched ellipses. In my study in Section 3, I follow Lohmann (2026: 87–91) and consult the auxiliary selection hierarchy (ASH) developed by Sapp (2011) for ENHG to decide which auxiliary verb could be selected by the lexical verb in the elided conjunct.

3 Corpus study

This study is partly based on Lohmann (2026: Chapter 7). The corpus examined is the Baumbank.UP (Demske 2019). It contains 26 ENHG prose texts, which adds up to more than 600,000 tokens.[18] It is currently the largest historical corpus of German. The time span of 1350–1650 is divided in 50-year segments. In each time span, there are prose texts covering the four big dialectal High German areas: Middle West, Middle East, Upper West and Upper East. Each cell of the corpus, for example 1350–1400 Middle West, contains approximately 20,000 tokens. Thus, the corpus is balanced in terms of both time and area. The Baumbank.UP is parts-of-speech-tagged due to the Historical TagSet (HiTS) system (Dipper et al. 2013). The syntactic annotation is based on TIGER (Brants et al. 2004), but an additional label, AS, has been implemented for afinite constructions.

3.1 Method

I focus on clauses that are conjoined by the coordinator *und* 'and'. It is the most common coordinator in ENHG. The spelling variant *vnd*[19] solely accounts for 3.2% of all tokens.[20] In addition, *und* 'and' shows all properties of a coordinating conjunction since OHG (cf. Oppermann

[18] As this study is based on Lohmann (2026), the 2019 version is referenced here. In the 2019 version, the corpus has 600,775 tokens, punctuation included. By then the corpus was not relaunched online. All texts were therefore accessed via TIGERSearch (Lezius 2002). After 2019, some of the punctuations were deleted, so now the corpus has 600,569 tokens and can be accessed via TüNDRA (Martens 2013) and ANNIS (Krause & Zeldes 2016).
[19] There are approximately 13 spellings of *und* 'and' in the corpus, see Lohmann (2026: 99, fn 1).
[20] For comparison: the TIGER corpus has 888,238 tokens (punctuation included) (Brants et al. 2004). Here, *und* 'and' accounts for 1.9% of all tokens (Lohmann 2026: 99, fn 2).

2025). In contrast, *aber* 'but' became a coordinator later in the history of German (cf. Ferraresi 2018). Thus, the quantitative analysis is based on sentences conjoined by *und* 'and' to ensure coordination structures.

In the corpus, coordination of clauses is tagged as CS. In order to distinguish between afinite constructions and ellipsis in coordinated clauses, I focus on clauses tagged as S and governed by CS. As *und* 'and' is also tagged as junction in the corpus, I search for CS that governs a coordinator tagged as KON. Thus, it ensures that syndetic coordination is in focus. In total, there are 10,361 instances in which the coordinator combines sentences. Narrowing down the query to each spelling of the coordinator *und* 'and', there are 8,397 instances.[21] In the corpus, the verbs *haben* 'have' and *sein* 'be' are tagged with the prefix VA for auxiliaries, regardless whether their function is grammatical or lexical.[22]

To determine whether the ellipsis matches the antecedent, I check the conjuncts due to the (predictable) function of the finite verbs (lexical or grammatical). If a grammatical function is probable, I follow Sapp's (2011) study. The grammatical function can be either expressing tense or creating passive voice. To form the perfect tense, transitive and reflexive verbs select *haben* 'have' (Sapp 2011: 34). For intransitives, I use the ASH (Sapp 2011: 40), see Table 1.

As Table 1 illustrates, the split auxiliary system "remained relatively stable" (Sapp 2011: 40). There is variation regarding continuation and existence of state verbs. Continuation of state verbs like *stehen* 'stand', *sitzen* 'sit', *liegen* 'lay' show "variation by semantic context and dialect" (Sapp 2011: 40). In the Upper German region, there is a clear preference for *sein* 'be' (Sapp 2011: 39). Existence of state verbs select more often *sein* 'be' than *haben* 'have' in ENHG (Sapp 2011: 40). According to Sapp (2011: 34), it should be noted that examples such as (26) express passive voice rather than perfect tense.

(26) so *ist mir auch von meinem herrn nit geboten*
so is me also by my lord not asked
'So neither is it asked of me by my lord.'
(Buch von Troja 13, as cited from Sapp 2011: 34)

In (26), the transitive verb is a participle *geboten* 'asked' building together with *ist* 'is' passive voice.[23] Thus, one has to be aware of this possible interpretation.

[21] Due to the fact that the corpus has no lemma annotation, each spelling variant of *und* 'and' was taken into account.
[22] The query syntax conducted via TIGERSearch (Lohmann 2026: 230): #1:[cat="CS"] & #2:[pos="KON"] & #3:[cat="S"] & #4:[pos="VAFIN"] & #5:[pos=/V.*P.*/|pos="ADJD"] & #1 > #2 & #1 > #3 & #3 > #4 & #3 >* #5. It is set that *haben* 'have' or *sein* 'be' is present as a finite verb in at least one sentence (VAFIN), and that either a participle (V.*P.*) or a predicative adjective (ADJD) is present in one sentence. Notice that the query does not fix that both VAFIN and participle or predicative adjective must be present in the same instance. It can be either VAFIN or participle/predicative adjective.
[23] In MSG, stative passive is built by *sein* 'be' and processual passive by *werden*. The development of passive voice is discussed, for instance, in Eroms (1990, 1992) and Kotin (1998).

Table 1: Auxiliary selection hierarchy adapted from Sapp (2011: 40)

Sorace (2000)	**ENHG** **Sapp (2011: 40)**	**MSG** **Keller & Sorace (2003)**
change of location *entkommen* 'escape'	*sein* 'be'	*sein* 'be'
change of state *wachsen* 'grow'	*sein* 'be'	*sein* 'be' (*haben* 'have')
continuation of state *verweilen* 'stay'	*haben* 'have'	*haben* 'have' (*sein* 'be')
existence of state *liegen* 'lay'	*sein* 'be' (*haben* 'have')	*haben* 'have' /*sein* 'be'
uncontrolled processes *zittern* 'shiver'	*haben* 'have'	*haben* 'have'
controlled processes motional *tanzen* 'dance'	?	*sein* 'be' (*haben* 'have')
non-motional *reden* 'talk'	*haben* 'have'	*haben* 'have'

In the following sections, I discuss matching ellipses qualitatively only. Ellipses with mismatches, however, are presented quantitatively using descriptive statistics, since the focus of the study is on these.

3.2 Matching finite verb ellipsis in coordination

Finite verb ellipsis matching its antecedent appears in coordinated main clauses (27a) and in coordinated subordinate clauses (27b).

(27) a. *Dasselb hauß* ***ist*** *zu stund an nider gerissen/ vnd ain Kirchen* ***ist*** *an dieselb stat gesetzt*
the.same house is at hour on down torn and a church is at the.same place put
'The same house has been torn down at the hour and on the same place a church has been put.'
(Baumbank.UP, Sigmund von Herberstein 1557, 268)

b. *Vnd merke doby wie wol ein edel mensch der kunig was das er*
and realise thereby how well a noble person the king was that he
das alles von gotte ***hat****/ vnd nit von jm selber* ~~***hat***~~*...*
that all from God has and not from him self has
'I realise what a noble man the king was, that he has all of that from God and not from himself.'
(Baumbank.UP, Marquard von Lindau 1450, 6049)

Following Sapp (2011: 34), both conjuncts express passive voice in (27a). Thus, the auxiliary verb *ist* 'is' is omitted in the second conjunct. In (27b), the verb *hat* 'has' as a full lexical verb is missing. Both examples are therefore genuine gapping instances, as the ellipsis and antecedent match phonologically. Further, I consider data as (28) to be genuine instances of gapping since person-/number-mismatches do occur in MSG, see section 2.1.

(28) die *welt* ***ist*** *mer gekrucziget vnd ich* ~~***bin***~~ *der welte* ~~***gekrucziget***~~***/***
the world is me crucified and I am the world crucified
'The world is crucified to me and I am crucified to the world.'
(Baumbank.UP, Marquard von Lindau 1450, 843)

In addition to gapping, RNR also appears in the corpus, see (19b) repeated in (29).

(29) *auch mit dem Geschuetz auff dem Wahle dermassen vnter sie*
also with the gun at the rampart so.much among them
Geschossen/ daß Arm vnd Schenckel in die Lufft geflogen ***seyn****/ vnd.*
shot that arms and thighs in the air blown.up are and
vngefehr 1700. Frantzosen Tod blieben ***seyn****/*
approximately 1700 Frenchmen dead left are
'There was so much shooting among them on the rampart that arms and thighs were blown up and Frenchmen were left dead.'
(Baumbank.UP, Johann Bange 1599, 1008)

Due to the ENHG ASH in Table 1, the ellipsis in the first conjunct matches the overt finite verb *seyn* 'are'. The event *in die Lufft geflogen* 'blown up in the air' expresses a change of location. This example is therefore a genuine occurrence of RNR.

3.3 Mismatched finite verb ellipses in coordination

Here, I focus on finite verb ellipsis that deviate from MSG and in this sense do not match their antecedent. In Baumbank.UP (Demske 2019), there are 58 instances of such mismatches, see Table 2:[24]

[24] There are three occurrences of person-/number-mismatches in the data. Two examples appear in coordination of subordinate clauses, and one example occurs in main clause coordination. Unlike Lohmann (2026: Section 7.3), I do not take these into account, as I am focusing on mismatched ellipses that deviate from MSG. However, person-/number-mismatches do occur in MSG, see Section 2.1.

Table 2: Mismatched ellipses per time and area

	Time						
Area	**1350–1400**	**1400–1450**	**1450–1500**	**1500–1550**	**1550–1600**	**1600–1650**	**Total**
Middle West	0	0	0	0	3	4	7 (12%)
Middle East	0	0	0	0	11	0	11 (18.9%)
Upper West	0	1	0	5	10	2	18 (31%)
Upper East	0	0	0	0	18	4	22 (37.9%)
Total	0	1 (1.7%)	0	5 (8.6%)	42 (72.4%)	10 (17.24%)	58 (100%)

Of the 58 mismatched ellipses, around 72% (n = 42) occur in the period 1550–1600. Around 68% (n = 40) occur in Upper German (n = 18 in Upper West; n = 22 in Upper East). In terms of clause type, half of the cases occur in coordination of main clauses (n = 29). Examining the order of the antecedent and the ellipsis reveals that, in main clause coordination, the ellipsis only occurs in second conjuncts. In coordinated subordinate clauses, both orders appear, see Table 3.[25]

Table 3: Direction of mismatched ellipses

Ellipses	**Main clause**	**Dependent clause**
forward	yes (n = 29)	yes (n = 3)
backward	no	yes (n = 24)
No antecedent	no	yes (n = 2)
Total	n = 29	n = 29

Besides that, there are two examples of coordinated subordinate clauses that have no antecedent at all.[26] In terms of mismatch types, there are mismatches with full lexical verbs, as in (30).

[25] One reviewer suggested that text-type might be a factor that influences the use of ellipsis. The Baumbank.UP (Demske 2019) is not balanced according to text-type. This factor cannot be proved or disproved. However, in the time slot 1550–1600 there are three different types of prose: texts for description (n = 3), texts for leisure (n = 11+18), and texts about realia (n= 10). At this stage, I leave it to further research to investigate the influence of this factor.
[26] See for instance example (43c) in Lohmann (2026: 242).

(30) *die* ***haben*** *ein eygne Scytische sprach vnd yr schrifft* ***~~ist~~*** *nit vngleich*
they have a own Scythian language and their script is not unlike
der Griechischen
the Greek
'they have their own Scythian language and their script is similar to the Greek.'
(Baumbank.UP, Sebastian Franck 1534, 491)

In (30), *haben* 'have' is overt in the first conjunct, in the second conjunct the finite verb is omitted. Due to *vngleich* 'unlike' as predicative and the nominal phrase *yr schrifft* 'their script' as subject, it is more likely a copular construction in the second conjunct with the copula *sein* 'be' missing.[27]

In addition, there are mismatches with grammatical verbs, see example (2) repeated in (31).

(31) *da* ***ist*** *er eine Zeitlang still gelegen/ vnnd* ***~~hat~~*** *sich mit seinem*
there is he a time.span still lain and has REFL with his
Kriegßvolck Erquicket/.
troops refreshed
'He rested there for a while and recovered himself together with his troops.'
(Baumbank.UP, Johann Bange 1599, s488)

In (31) the auxiliary *sein* 'be' is used to form perfect tense in the first conjunct. However, the reflexive verb *sich erquicken* 'refresh oneself' selects *haben* 'have' in ENHG (Sapp 2011: 34). As shown in (32), the auxiliary *haben* 'have' can also function as the antecedent for the ellipsis.

(32) *in der Tauf* ***hat*** *Volodimer seinen Namen veraendert/ vnd* ***~~ist~~*** *BASILIVS*
in the baptism has Volodimer his name changed and is Basilius
genant worden/
named been
'in the baptism, Volodimer has changed its name and he has been named Basilius.'
(Baumbank.UP, Sigmund von Herberstein 1557, 168)

Furthermore, (32) shows that the conjuncts can differ in voice. While the first conjunct expresses the active voice with the auxiliary *hat* 'has' overtly, the second conjunct is in the passive voice (processual), indicated by *worden*. In the next section, I discuss these results in a broader context. I will also present evidence that might explain why this kind of mismatched gapping has been lost diachronically.[28]

[27] One reviewer suggested analysing the example as an object predication in the sense of *vnd die haben yr schrifft nit vngleich der Griechischen* 'and they have a script not unlike the Greek one'. With this interpretation, there is no mismatch between the gap and its antecedent. I searched the entire corpus for examples of *gleich*/*ungleich* 'equal/unequal' in a predicative function, particularly in Sebastian Franck's text, in order to evaluate the proposed interpretation. Despite this investigation, no examples were found that correspond to this interpretation. However, I cannot categorically rule it out, as one cannot draw conclusions based on the absence of evidence in corpora.

[28] It may be that these mismatched ellipses still occur, for example, in Upper German varieties. To the best of my knowledge, there have been no studies to date on coordination and coordination ellipses in non-standard German varieties. However, some examples of ellipses in non-standard German can be found in Lechner (2004: 97–113).

4 Discussion

In the preceding section, the results of the corpus study were presented. The Baumbank.UP treebank data demonstrated that gapping and RNR are applicable when the elided finite verb and the antecedent match regarding their phonological identity. However, 58 instances of mismatched ellipses have been identified. Despite the Baumbank.UP corpus being twice the size of Breitbarth's (2005b), there are fewer mismatched ellipses. Breitbarth (2005b: 59) extracted 55% of 2,474 auxiliary ellipses. However, this study diverges from Breitbarth's (2005b) in its methodological approach: it focused exclusively on clauses conjoined by *und* 'and'; distinguished between coordinated and non-coordinated clauses; and it only considered mismatches deviating from MSG. However, mismatched gapping has also been observed with other coordinators, as shown in (33).[29]

(33) *Dieser Großfuerst* ***hat*** *auch seyn zeiten/ das CASANISCH Thatterisch*
that grand.duke has also his time the Kazan Tatars
Khuenigreich an gefallen/zu wasser vnnd Lanndt/ aber ***ist*** *vngethoner*
kingdom attacked at water and land but is undone
sachen abgezogen
things withdrawn
'In his time, the Grand Duke attacked the Kingdom of the Kazan Tatars from both sea and land, but withdrew without having achieved anything.'
(Baumbank.UP, Sigmund von Herberstein 1557, 310)

The search for other coordinators like *aber* 'but' and the consideration of asyndetic coordination, as Breitbarth (2005b: 151) may have done, might lead to more mismatched gapping instances than I presented in Section 3.3.

Additionally, (33) contradicts Breitbarth's (2005b: 149) comparison of mismatched ellipses to asymmetric coordination. The coordinator *aber* 'but' expresses adversative semantics. A contrastive relation between conjuncts is not possible in asymmetric coordination in MSG due to the lack of parallelism of the conjuncts (Reich 2009a: 115). Thus, other coordinators than *und* 'and' do not form asymmetric coordination (cf. Reis 1993: 214; Reich 2009a: 146–148). What Breitbarth (2005b: 149) presumes being formed by the mismatched ellipsis in example (20a), a dependency relation like in asymmetric coordination, is therefore most likely neither a dependency relation of the conjuncts, nor is it the dependency created by the ellipsis itself. It is rather the coordinator *und* 'and' that forces a sequential or resulting interpretation of the conjuncts. Bar-Lev & Palacas (1980: 144) attribute this to the temporal semantics of the conjunction. Txurruka (2003) presumes a blocking function of subordination relations in terms of pragmatics. Generally, the level (semantic, syntactic and/or pragmatic) on which coordination applies is

[29] According to Ferraresi (2018) *aber* 'but' is already a coordinator in ENHG.

still being debated (cf. Altshuler & Truswell 2022). However, both assumptions, semantics and pragmatics, would explain why changing the order of the conjuncts is not allowed without drawing a connection to asymmetric coordination. Thus, it is the coordinator that forces one to interpret the event of the second conjunct as occurring after the event of the first conjunct, or to interpret it as a result of the event of the first conjunct. Due to the coordinator *und* 'and', there is a progressive reading of the coordination that prohibits changing the order. Nothing depends on the ellipsis itself.

Nevertheless, is there a link between mismatched ellipses and afinite constructions, as it is assumed by Breitbarth (2005b) and Thomas (2019)? For comparison, non-coordinated afinite constructions must be considered. Here, I refer to the study of Blum (2018), who analysed afinite constructions in the Baumbank.UP, excluding coordinated instances of it (Blum 2018: 8). A secondary analysis of the 1,102 afinite construtions in Blum (2018: 73–81) shows that these clauses occur most frequently in 1550–1600 and are more common in Upper High German (69%) than in Middle High German (31%) (Lohmann 2026: 226–227). These results reflect the findings of mismatched ellipses in Table 2. I assume that these are two different types of ellipses, which, however, occur more frequently in Upper German and time slot 1550–1600 for the same reason. Both fall into the period and region of the expansion of the perfect tense, which began in the Upper High German area and spread over to the Middle High German region (cf. Lindgren 1957: 44). When perfect tense was common, afinite constructions and mismatched ellipses became more frequent. Thus, the preterite almost disappeared in Upper High German, whereas in Middle High German, both preterite and perfect tense are still used (Fischer 2020: 113).

In addition to ENHG, there is another West Germanic language that exclusively uses present perfect tense and exhibits similar cases of mismatched gapping as ENHG. In Yiddish, the auxiliary in the second conjunct of a main clause coordination can be omitted, even if it diverges from the auxiliary occurring in the first conjunct (Birnbaum 1988: 67; Jacobs 2005: 262), see (34a) and (34b).

(34) Yiddish

a. *mir zenen gekumen un gegesen*
we were come and eaten
'we came and ate'
(Jacobs 2005: 262)

b. *die feigl hobn gezungen un avekgeflojgn*
the birds have sung and flown.away
'the birds sang and flew away'
(Jacobs 2005: 262)

According to Schäfer (2023), Yiddish is a Germanic language that originated in the Middle Ages and has preserved older language structures from High German, especially from oral varieties.

It uses present, perfect, double perfect, future, as well as active and passive voice. It also has a split auxiliary system (Schäfer 2022: 377). In contrast to MSG, it has remained unaffected by standardization. Moreover, to my knowledge, Yiddish has no afinite constructions.

In addition, there is a second West Germanic language that does not exhibit afinite constructions but rather displays mismatched gapping (Pérez Lorido 1996: 334–336), see (35).

(35) *& heora gemetting* ***wæs*** *eft æt Trefia þære ie & eac*
and their encounter was afterwards at Trevia the river and also
Romane gefliemed & swþor forslagen, & Hannibal gewundod
Romans put.to.flight and greatly slaughtered and Hannibal wounded
'And their encounter was at the River Trevia, and the Romans put to flight and much slaughtered, and Hannibal wounded'
(Oros. 100.6 as cited from Pérez Lorido 1996: 335–336)

The Old English example in (35) shows a mismatch in the function of the verb *wesan*. In the first conjunct, it is used as a lexical verb "and as an auxiliary of the passive in the second one" (Pérez Lorido 1996: 336). This kind of gapping produces a "humorous effect" in Modern English (Pérez Lorido 1996: 336). In Old English, it is a "proof of relative opacity of the rule of coordination deletion in Old English" (Pérez Lorido 1996: 336). Gapping is used "to emphasize a certain part of it [= sentence]" (Pérez Lorido 1996: 333). Irrelevant parts are deleted, "as long as they had appeared previously in discourse" (Pérez Lorido 1996: 333).

Ellipses in coordination, gapping and RNR, are the strictest way of de-accentuating linguistic expressions in order to place the contrastive focus on the remaining linguistic material in the coordination (cf. Féry & Hartmann 2005). Thus, any existing theory on gapping that is not based on complete phonological identity between ellipsis and antecedent provides a possible analysis for the historical data. Hartmann (2000) assumes that gapping is the deletion of assertion. Repp (2009b: 207) proposes that "[g]apping is the coordination of two sentences where the elliptic conjunct copies the anchoring of the proposition to the factual world from its non-elliptic antecedent" (Repp 2009b: 207). Thus, tense, aspect, and mood have to be identical (Repp 2009b: 230). This definition of gapping also applies to the historical data, specifically on the mismatched ellipses in main clause coordination, since these factors are identical in the examples. I therefore assume that these are genuine cases of gapping. However, mismatched ellipses in coordinated subordinate clauses are rather conjoined afinite constructions (cf. Breitbarth 2005b: 62–63). The data in Table 3 confirm the observation of Breitbarth (2005b: 62). It is forward deletion only in main clause coordination, akin to gapping. In coordinated subordinate clauses, however, not only are both orders of ellipsis and antecedent possible, which could still be explained by coordination ellipses, gapping and RNR, but in two examples, there is no antecedent at all (cf. Breitbarth 2005b: 61). Since afinite constructions – finite subordinate clauses probably without

finite verbs – can occur independently of coordination, the ellipsis of such constructions within coordination structure can be more easily explained by assuming that afinite constructions are coordinated by themselves.

Furthermore, example (36) of mismatched gapping in coordinated main clauses might show an elided object, too:

(36) *Als das geschah/ [...]* ***ist*** *der Basilius* ***dem*** ***Khuenig*** *zu fueß* gefallen/
when that happend is the Basilius king to feet fallen
vnd ~~***hat ihn***~~ *gebetten/ jme sein notturft fur zu bringen* zu/ *uergonnen/ [...]*
and has him pleased him his livelihood for to bring to grant
'When this happened, Basilius fell at the king's feet and begged him to grant him his livelihood.'
(Baumbank.UP, Sigmund von Herberstein 1557, 201)

However, if theories of coordination ellipses based on MSG observations can analyse historical data that is no longer common in MSG, why have the mismatched gaps disappeared diachronically?

There is evidence that the standardization process of High German has influenced gapping in German and may have led to the loss of the mismatched patterns. First, mismatched gapping is attested in Yiddish, see (34a,b), a language that has not undergone a standardization process. Second, there are statements by prescriptive grammarians about the appropriate use of gapping (Adelung 1782; Grimm 1837):

> Die Hülfswörter sind in den zusammengeseßten Zeiten das wahre und eigentliche Verbum, daher man sie nicht verschweigen sollte, außer, wenn mehrere mit einander verbundene Verba einer Zeit, Person und Rection einerley Hülfswort haben sollten: [...] Die Fälle, wo die Zusammenkunft mehrerer Hülfswörter Mißklang und schleppende Dehnung verursacht, lassen sich bey ein wenig Geschmack und Uibung leicht vermeiden. (Adelung 1782: 397)

> "Auxiliary verbs are the true and actual verbs, which is why they should not be omitted, except when several verbs are linked together that express equal tense, person and government, and select the same auxiliary. [...] With a little taste and practice, it is possible that situations can be avoided in which the combination of several auxiliary verbs leads to discord and sluggish elongation."

Early modern grammarians strived not only for a unified language across regions (von Polenz 2020: 24), but also for a "correctness of language in German" (von Polenz 2020: 109). Under normative pressure, the German language moved away from the spoken language, and language variants were lost (von Polenz 2020: 113). Thus, the loss of mismatched gapping may also be a phenomenon that has been influenced by the standardisation process. However, a systematic

analysis of normative grammars is lacking. It would provide insight into the influence of prescriptivism on gapping. In addition, it is unclear whether mismatched gapping has completely disappeared in Modern German. Further studies on coordination and coordination ellipsis in German varieties could yield interesting findings. These tasks remain open for future research.

5 Conclusion

In this paper, I have investigated gapping from a historical point of view. For this purpose, data was extracted from the Baumbank.UP. Due to the balanced structure of the treebank, both diachronic and diatopic conclusions can be drawn. It has been shown that mismatched gapping occurs more frequently in late ENHG, even when considered separately from afinite constructions. Furthermore, it appears more often in Upper German prose texts than in Middle German prose texts. It has also been shown that these findings mirror those of afinite constructions. However, I have argued that this is related to the spread of the perfect tense, rather than being an indication that the two phenomena should be regarded as a single ellipsis phenomenon. This assumption is supported by two other West Germanic languages that also exhibit mismatched gapping but lack afinite constructions. Since gapping in MSG shows mismatches with the antecedent due to person-/number-marking, a theory relying on complete phonological identity is unsuitable. Theories that instead consider identity at a different linguistic level could also explain the mismatches that occur in coordinated main clauses in ENHG. Regarding mismatches in coordinated subordinate clauses, I have argued that they can be more easily explained when analysed as coordination of afinite constructions. Note that Oppermann (2025: 127) classified data as gapping in OHG if "both conjuncts feature the same lemma as finite verb and are sufficiently parallel insofar as they contain two contrast pairs". It is therefore unclear whether mismatched ellipses of finite verbs, whether due to either grammatical features or lexical entries, were already present in OHG. As Oppermann (2025) relies on this property throughout her study, there is also no evidence of mismatched ellipses in MHG. Further research on this topic is therefore required. Additionally, a systematic analysis of gapping in later periods of High German would shed light on the influence of prescriptivism on gapping, as would the investigation of coordination ellipses in non-standard varieties in German. These remain tasks for future research.

Acknowledgements: I thank the audience at the "Experimental and corpus-based approaches to ellipsis" section organised by Gabriela Bîlbîie and Max Bonke at the 21st International Congress of Linguistics in September 2024 in Poznań for the discussion, and for helpful and inspiring comments I thank three anonymous reviewers.

Abbreviations

ASH Auxiliary selection hierarchy
ATB Across the board
ENHG Early New High German
MHG Middle High German
MSG Modern Standard German
OHG Old High German
RNR Right node raising

References

Adelung, Johann Christoph. 1782. *Deutsche Sprachlehre: Zum Gebrauch der Schulen in den Kgl. Preussischen Landen*. Wien: v. Trattner.

Admoni, Vladimir G. 1967. Der Umfang und die Gestaltungsmittel des Satzes in der deutschen Literatursprache bis zum Ende des 18. Jhs. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 89. 144–199.

Aelbrecht, Lobke. 2015. Ellipsis. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and analysis: An international handbook*, vol. 1. 562–594. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110377408.562.

Altshuler, Daniel & Robert Truswell. 2022. *Coordination and the syntax-discourse interface*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780119880239.001.0001.

Bar-Lev, Zev & Arthur Palacas. 1980. Semantic command over pragmatic priority. *Lingua* 51(2). 137–146. https://doi.org/10.1016/0024-3841(80)90004-2.

Barnickel, Katja. 2017. *Deriving asymmetric coordination in German: A nonmonotonic approach*. Leipzig: Universität Leipzig dissertation.

Behaghel, Otto. 1928. Deutsche Syntax: Eine geschichtliche Darstellung, vol. 3. Satzgebilde. Heidelberg: Winter.

Biener, Clemens. 1925. Von der sogenannten Auslassung der Kopula in eingeleiteten Nebensätzen. *Die neueren Sprachen* 33. 291–297.

Birnbaum, Salomo. 1988. *Grammatik der jiddischen Sprache: Mit einem Wörterbuch und Lesestücken* (5., erg. Aufl.). Hamburg: Buske.

Blum, Simon. 2018. *Zur Entstehung der afiniten Konstruktion im Frühneuhochdeutschen*. Berlin: Humboldt-Universität zu Berlin master's thesis.

Bonitz, Petra-Kristin. 2013. *Subjektellipsen in Koordinationsstrukturen. Theoretische Fundierung und empirische Erkenntnis*. Göttingen: Georg August-Universität Göttingen dissertation.

Bonke, Max. 2024. *Complementizers in ellipsis. A cross-linguistic study with special focus on Spanish*. Köln: Universität zu Köln dissertation.

Brants, Sabine, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith & Hans Uszkoreit. 2004. TIGER: Linguistic interpretation of a German corpus. *Journal of Language and Computation* 2. 597–620. https://doi.org/10.1007/s11168-004-7431-3.

Breitbarth, Anne. 2004. Auxiliary drop in Early Modern German. *Linguistics in the Netherlands* 21(1). 36–46. https://doi.org/10.1075/avt.21.07bre.

Breitbarth, Anne. 2005a. Auxiliary drop as subordination marking. *Linguistics in the Netherlands* 22(1). 37–47. https://doi.org/10.1075/avt.22.06bre.

Breitbarth, Anne. 2005b. *Live fast, die young – the short life of Early Modern German auxiliary ellipsis.* Tilburg: Universiteit van Tilburg dissertation. https://www.lotpublications.nl/Documents/115_fulltext.pdf, 8 September 2026.

Büring, Daniel & Katharina Hartmann. 1998. Asymmterische Koordination. *Linguistische Berichte* 174. 172–201.

Christ, Rüdiger. 2011. Linkstilgung und phonologische Quasi-Identität. *Linguistische Berichte* 228. 371–411. https://doi.org/10.46771/2366077500228_1.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199271092.001.0001.

Demske, Ulrike. 2008. Symmetrische und asymmetrische Koordination im Frühneuhochdeutschen. In Waldemar Czachur & Marta Czyzewska (eds.), *Vom Wort zum Text. Studien zur deutschen Sprache und Kultur. Festschrift für Professor Józef Wiktorowicz zum 65. Geburtstag*. 299–309. Warszawa: Uniwersytetu Warszawskiego.

Demske, Ulrike. 2019. *Referenzkorpus Frühneuhochdeutsch: Baumbank.UP*. Universität Potsdam. https://www.uni-potsdam.de/de/guvdds/baumbankup, 8 October 2026, https://doi.org/10.57754/FDAT.c7wr3-5k848.

Demske, Ulrike. 2022. Silent heads in Early New High German. In Ulrike Freywald, Horst J. Simon & Stefan Müller (eds.), *Headedness and/or grammatical anarchy?*. 123–151. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.6973523.

Demske-Neumann, Ulrike. 1990. Charakteristische Strukturen von Satzgefügen in den Zeitungen des 17. Jahrhunderts. In Anne Betten & Claudia M. Riehl (eds.), *Neuere Forschungen zur historischen Syntax des Deutschen*. 239–252. Tübingen: Niemeyer. https://doi.org/10.1515/9783111708751.239.

Dipper, Stefanie, Karin Donhauser, Thomas Klein, Sonja Linde, Stefan Müller & Klaus-Peter Wegera. 2013. HiTS: ein Tagset für historische Sprachstufen des Deutschen. *Journal for Language Technology and Computational Linguistics Special Issue* 28(1). 85–137. https://doi.org/10.21248/jlcl.28.2013.170.

Ebert, Robert Peter, Oskar Reichmann, Hans-Joachim Solms & Klaus-Peter Wegera. 1993. *Frühneuhochdeutsche Grammatik*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110920130.

Eroms, Hans-Werner. 1990. Zur Entwicklung der Passivperiphrasen im Deutschen. In Anne Betten & Claudia M. Riehl (eds.), *Neuere Forschungen zur historischen Syntax des Deutschen. Referate der internationalen Fachkonferenz Eichstätt 1989*. 82–97. Tübingen: Niemeyer. https://doi.org/10.1515/9783111708751.82.

Eroms, Hans-Werner. 1992. Das deutsche Passiv in historischer Sicht. In Ludger Hoffmann (ed.), *Deutsche Syntax: Ansichten und Aussichten*. 225–249. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110622447-010.

Eroms, Hans-Werner. 2010. Additive und adversative Konnektoren im Althochdeutschen. In Arne Ziegler (ed.), *Historische Textgrammatik und Historische Syntax des Deutschen: Traditionen, Innovationen, Perspektiven*, vol. 1. 279–303. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110219944.277.

Ferraresi, Gisella. 2018. Adverbial Connectives. In Agnes Jäger, Gisella Ferraresi & Helmut Weiß (eds.), *Clause structure and word order in the history of German*. 82–122. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198813545.003.0006.

Féry, Caroline & Katharina Hartmann. 2005. The focus and prosodic structure of German right node raising and gapping. *The Linguistic Review* 22(1). 69–116. https://doi.org/10.1515/tlir.2005.22.1.69.

Fischer, Hanna. 2020. The development of the perfect in selected Middle and New Germanic languages. In Robert Crellin & Thomas Jügel (eds.), *Perfects in Indo-European Languages and Beyond*. 95–122. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.352.04fis.

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative investigations: The form, meaning and use of English interrogatives*. Stanford: CSLI Publications.

Grimm, Jacob. 1837. *Deutsche Grammatik*, vol. 4. Göttingen: Dieterich.

Hankamer, Jorge. 1979. *Deletion in coordinate structures*. New York: Garland.

Hartmann, Katharina. 2000. *Right node raising and gapping: Interface conditions on prosodic deletion*. Philadelphia, Amsterdam: John Benjamins.

Hartmann, Katharina. 2015. Coordination. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and analysis: An international handbook*, vol. 1. 478–514. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110377408.478.

Höhle, Tilman N. 2019a. Assumptions about asymmetric coordination in German. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle* (2., durchg. Aufl.). 295–310. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.2588359.

Höhle, Tilman N. 2019b. Subjektlücken in Koordinationen. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle* (2., durchg. Aufl.). 193–218. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.2588353.

Jacobs, Neil G. 2005. *Yiddish: A linguistic introduction*. Cambridge: Cambridge University Press.

Keller, Frank & Antonella Sorace. 2003. Gradient auxiliary selection and impersonal passivization in German: An experimental investigation. *Journal of Linguistics* 39(1). 57–108. https://doi.org/10.1017/S0022226702001676.

Klein, Wolfgang. 1993. Ellipse. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung*, vol. 1. 763–799. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110095869.1.12.763.

Kotin, Michail. 1998. Die Herausbildung der grammatischen Kategorie des Genus verbi im Deutschen: Eine historische Studie zu den Vorstufen und zur Entstehung des deutschen Passiv-Paradigmas. Hamburg: Helmut Buske.

Krause, Thomas & Amir Zeldes. 2016. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31 (1). 118–139. https://doi.org/10.1093/llc/fqu057.

Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1(1). 97–120. https://doi.org/10.1017/S0047404500006576.

Lechner, Winfried. 2004. *Ellipsis in Comparatives*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110197402.

Lezius, Wolfgang. 2002. *Ein Suchwerkzeug für syntaktisch annotierte Textkorpora*. Stuttgart: Universität Stuttgart dissertation.

Lindgren, Kaj Brynolf. 1957. *Über den oberdeutschen Präteritumschwund*. Helsinki: Suomalaisen Tiedakademia.

Lohmann, Marianna. 2026. *Koordination im Frühneuhochdeutschen: Eine qualitative und quantitative Korpusuntersuchung*. Potsdam: Universitätsverlag Potsdam. https://doi.org/10.25932/publishup-68092.

Martens, Scott. 2013. TüNDRA: A web application for treebank search and visualization. In Proceedings of The Twelfth Workshop on Treebanks and Linguistic Theories. Sofia. 133–144. https://bultreebank.org/wp-content/uploads/2017/06/TLT12Proceedings-compressed.pdf, 8 October 2026.

Oppermann, Sophia Jana. 2021. Subject gap coordination: A diachronic view. *Journal of Historical Syntax* 5(21). 1–38. https://doi.org/10.18148/hs/2021.v5i16-25.58.

Oppermann, Sophia Jana. 2025. *Coordination Structures in Old and Middle High German*. Berlin: De Gruyter. https://doi.org/10.1515/9783111544618.

Pérez Lorido, Rodrigo. 1996. On the grammatical domain of gapping in Old English: Syntax and pragmatics. *Diachronica* 13(2). 319–346.

Postal, Paul M. 1974. *On raising: One rule of English grammar and its theoretical implications*. Cambridge, MA: The MIT Press.

Reich, Ingo. 2009a. *"Asymmetrische Koordination" im Deutschen. Studien zur deutschen Grammatik*. Tübingen: Stauffenburg.

Reich, Ingo. 2009b. Ist Asymmetrische Koordination wirklich ein Fall von Asymmetrischer Koordination? In Veronika Ehrich, Christian Fortmann, Ingo Reich & Marga Reis (eds.), *Koordination und Subordination im Deutschen*. 203–222. Hamburg: Buske. https://doi.org/10.46771/978-3-87548-956-9.

Reich, Ingo. 2011. Ellipsis. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*. 1849–1874. Berlin: De Gruyter. https://doi.org/10.1515/9783110255072.

Reis, Marga. 1993. Satzfügung und kommunikative Gewichtung: Zur Grammatik und Pragmatik von Neben- vs. Unterordnung am Beispiel 'implikativer' und-Konstruktionen im Deutschen. In Marga Reis (ed.), *Wortstellung und Informationsstruktur*. 203–250. Tübingen: Niemeyer. https://doi.org/10.1515/9783111658469.203.

Repp, Sophie. 2009a. Koordination, Subordination und Ellipse. In Veronika Ehrich, Christian Fortmann, Ingo Reich & Marga Reis (eds.), *Koordination und Subordination im Deutschen*. 245–265. Hamburg: Buske. https://doi.org/10.46771/978-3-87548-956-9.

Repp, Sophie. 2009b. *Negation in gapping*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199543601.001.0001.

Ross, John Robert. 1967. Constraints on variables in syntax. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Ross, John Robert. 1970. Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics: A collection of papers*. 249–259. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783111350219.249.

Sapp, Christopher D. 2011. Auxiliary selection in the Early New High German perfect tenses. *Groninger Arbeiten zur Germanistischen Linguistik* 53(2). 29–43.

Schäfer, Lea. 2022. Auxiliary selection in Yiddish dialects. *Journal of Germanic Linguistics* 34(4). 341–384. https://doi.org/10.1017/S1470542722000010.

Schäfer, Lea. 2023. Yiddish. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.946.

Schröder, Werner. 1985. *Auxiliar-Ellipsen bei Geiler von Kaysersberg und bei Luther*. Stuttgart: Steiner.

Senyuk, Ulyana. 2014. *Zum Status relativähnlicher Sätze im Früneuhochdeutschen: Ein Beitrag zum Wesen der Subordination im älteren Deutsch*. Potsdam: Universität Potsdam dissertation.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76(4). 859–890. https://doi.org/10.2307/417202.

Thomas, Victoria J. 2019. *Auxiliary ellipsis in Early Modern German 1350–1800*. Manchester: University of Manchester dissertation.

Txurruka, Isabel Gómez. 2003. The natural language conjunction and. *Linguistics and Philosophy* 26(3). 255–285. https://doi.org/10.1023/A:1024117423963.

van Craenenbroeck, Jeroen & Tanja Temmerman. 2018. Ellipsis in natural language: Theoretical and empirical perspectives. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*. 1–16. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198712398.013.1.

von Polenz, Peter. 2020. Geschichte der deutschen Sprache (11., überarb. Aufl.). Berlin: De Gruyter. https://doi.org/10.1515/9783110485660.

Wilder, Chris. 1994. Coordination, ATB and ellipsis. *Groninger Arbeiten zur Germanistischen Linguistik* 37. 291–329.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 3

https://doi.org/10.18148/zs/2026-2012

Emma Kious (corresponding author), Laboratoire de Linguistique Formelle, Université Paris Cité, France, emma.kious@u-paris.fr

Anne Abeillé, Laboratoire de Linguistique Formelle, Université Paris Cité, France, anne.abeille@u-paris.fr

Gender mismatch and ellipsis: comparing French adjectives and nouns

Abstract: According to Merchant (2014), Greek adjectives and human nouns behave differently under ellipsis: while adjectives allow for gender mismatch, human nouns may fall into three classes: kinship and nobility nouns (*adherfos/aderfi* 'brother/sister') which allow no mismatch, role nouns (*dhaskalos/dhaskala* 'teacher.M/F') which allow asymmetric mismatch, and syncretic nouns (*jatros* 'doctor.M/F') which allow symmetric mismatch. We compare adjectives and nouns in French. We conducted two quantitative studies on several French corpora (FrTenTen, ORFEO: spoken-part, Frantext: after-1980, DUEL), we found more cases of adjective ellipsis (N=562) than of noun ellipsis (81 cases). For predicative adjectives (with human subjects), we also found a higher rate of gender mismatch: 40.14% (224/558 in the written corpora) than for predicative nouns (17.3%, N=14 in the written corpora). The same factors (with human pronominal subjects favoring mismatch) can be observed for adjectival and nominal predicates. We also show that French human nouns allow for more gender mismatches than previously thought, even for kinship nouns (*fils/fille* 'son/daughter') and unrelated pairs (*homme/femme* 'man/woman'). We conducted three experiments to test the acceptability of gender mismatches under ellipsis, compared with anaphora, for adjectives and role nouns. We found no gender mismatch penalty under ellipsis, no robust order effect, and a role of gender biases. This suggests that adjectives and (gendered) role nouns behave similarly under ellipsis, allowing for symmetric gender mismatch under ellipsis in French.

Keywords: ellipsis, mismatch, French, corpus data, experiment

Cite as: Kious, Emma & Anne Abeillé. 2026. Gender mismatch and ellipsis: comparing French adjectives and nouns. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2012.
Received: 29 September 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1. Introduction

Significatio ex nihilo (Merchant 1999), ellipsis illustrates a discrepancy between form and meaning, where the interpretation of the elliptical clause (also called "target" clause) depends on the content of a previous clause (the antecedent) or the situation (Poppels 2022). It omits forms while preserving the clausal meaning of the utterance and makes it possible to construct an autonomous and acceptable utterance.

Stripping ellipsis, or Bare Argument ellipsis (Culicover & Jackendoff 2005), is a head ellipsis construction with a unique nominal remnant (*Paul*) and a polarity adverb (*too*), which follows a complete antecedent clause with a correlate (*Mary* in 1a).

(1) a. *Mary is happy and Paul too.*
b. *Mary is happy and* [[*Paul too*]$_{\text{FocP}}$ - ~~is happy~~]$_{\text{CP}}$.
c. *Mary is happy and* [[*Paul*]$_{\text{NP}}$ *too*]$_{\text{S}}$.

While syntactic derivation theories proposed by Ross et al. (1969) or Merchant (2001) posit movement and identity under deletion of the second clause (1b), Abeillé & Kim (2022) adopt a referential, or direct interpretation approach (Ginzburg & Sag 2000) (1c). The direct interpretation approach posits no syntactic reconstruction: the elliptical predicate (*be happy* in 1a) is inherited from the antecedent clause (or the context) with the remnant's content as argument (Nykiel & Kim 2021).

1.1. Mismatch and ellipsis

Mismatches in grammatical form between antecedent and elliptical clauses are usually possible and found cross-linguistically. In English as in French gapping, inflectional feature mismatches are attested (e.g., number) (2a), but interpreted features, such as tense, are not (2b).

(2) a. *Paul va à Paris et ses enfants* [*vont*] *à Rome.*
Paul go.PRS.3SG to Paris and his.PL children [go.PRS.3PL] to Rome
'Paul is going to Paris and his children to Rome.'
(Abeillé et al. 2014: 240)

b. #*Paul est arrivé hier et Marie* [*arrivera*] *demain.*
Paul be.PRS.3SG arrive.PTCP.M.SG yesterday and Marie [arrive.FUT.3SG] tomorrow
'Paul arrived yesterday and Marie tomorrow.'
(Abeillé et al. 2014: 240)

A corpus study by Abeillé & Kim (2022) found similar mismatches in English and French stripping (3). They found that for French stripping, for instance, 70% of *aussi* ('too') and 63% of *non plus* ('neither') fragments feature mismatches in person, number, or gender.

(3) a. *I'm beginning to like the guy. You too* [*are beginning to like the guy*] ? [COCA, 1993]
b. *Elle était ravie et moi aussi* [*j'-étais ravi*].
she be.PST.3SG thrilled.F.SG and me too I-be.PST.1SG thrilled.M.SG]
'She was thrilled and me too.'
[Frantext, Dorin, 1084]

Mismatches between antecedent and fragment challenge syntactic reconstruction approaches (with identity-under-deletion): structuralist theories of ellipsis posit an identity relation between the ellipsis and its antecedent. This notion of identity is built on connectivity effects which, according to Culicover & Jackendoff (2005), give empirical support in favour of some invisible structure at the ellipsis site. Yet, the exact nature of this identity relation has been, and remains, the topic of numerous debates: from Fiengo & May's (1994) syntactic identity theory of 'vehicle change' (the ellipsis site and its antecedent allow for a mismatch if both elements are referentially equivalent), to Merchant's (2001) semantic identity condition with the notion of e-givenness (based on the givenness of elliptical clauses), to hybrid identity theories such as Chung et al.'s (2006) no-new-words constraint or Thoms' (2015) inferential antecedent accommodation proposal or Griffith's (2019) non-isomorphic reconstruction.

Unlike identity-based theories, referential theories rely on contextually accessible information. Building upon Ginzburg & Sag (2000), referential theories consider ellipsis to involve a discourse antecedent with a uniquely identifiable thematic referent, integrated into the Common Ground. Ellipsis is then a discursive reference, akin to anaphora (Hankamer & Sag 1976). Such an approach has received empirical support from exophoric VP ellipsis (Miller & Pullum 2014), exophoric stripping ellipsis (Abeillé & Kim 2022), and case differences between elliptical and non-elliptical clauses: *Who is coming? – Me/*I. – I/*Me am coming.*

1.2. Previous work on gender mismatch under stripping

There has been some experimental work on gender mismatch and stripping ellipsis, but no dedicated corpus work, to the best of our knowledge. In a written corpus study of English and French stripping (with pronominal remnants), Abeillé & Kim (2022) found different cases of mismatch, including one gender mismatch in French, for an adjectival predicate (4).

(4) *Cette fois je suis vraiment amoureuse, lui aussi* [*est vraiment amoureux*].
this.F time I be.PRS.1SG really in.love.F.SG him too [be.PRS.3S really in.love.M.SG]
'This time, I am really in love, him too.'
[Frantext, Prin 2005]

In Spanish, Aparicio et al. (2015) compared mismatch penalties for predicative adjectives under ellipsis ('surface anaphora', per Hankamer & Sag 1976) (5a) and under ('deep') anaphora, with the Spanish clitic *lo* (close to the French expletive pronoun *le*, neither varying in gender) (5b). Deep anaphora, which involves pragmatic control of a salient discursive antecedent, should be more permissive than ellipsis, or surface anaphora, which involves syntactic control.

(5) a. *El alumno es alto y la profesora también*
the.M.SG student.M.SG be.PRS.3SG tall.M.SG and the.F.SG teacher.F.SG too
[*es alta*].
[be.PRS.3SG tall.F.SG]
'The student is tall and the teacher is too.'
(Aparicio et al. 2015: 2)

b. *El alumno es alto y la profesora también lo*
the.M.SG student.M.SG be.PRS.3SG tall.M.SG and the.F.SG teacher.F.SG too it
es.
be.PRS.3SG
'The student is tall and the teacher is that too.'
(Aparicio et al. 2015: 5)

In two eye-tracking experiments on Spanish, Aparicio et al. (2015) found that mismatching features on the predicative adjectives carried a higher processing cost under ellipsis, suggesting that these features are not ignored for ellipsis resolution. Their results also pointed towards a stronger mismatch penalty for feminine antecedents (Feminine-Masculine ordered conjuncts). But they did not distinguish number and gender features.

Looking at modern Greek, Merchant (2014) suggested that gender mismatches under stripping ellipsis are allowed for adjectives (6a), but not necessarily for nouns (6b,c). He distinguishes three classes of human nouns: nobility and kinship nouns which prohibit gender mismatches (6b), non-alternating role nouns which allow for them (6c), and alternating role nouns which only allow asymmetric mismatches (6d) in Masculine-Feminine order.

(6) a. *I Maria ine ikani, ala O Alexandros dhen ine*
the.F.SG Maria be.PRS.3SG capable.F but the.M.SG Alexandros not be.PRS.3SG
[*ikanos*].
[capable.F]
'Maria is capable, but Alexandros is not.'
(Merchant 2014: 2)

b. **O Petros ine adherfos, ke I Maria episis* [*ine adherfi*].
the.M.SG Petros be.PRS.3SG brother and the.F.SG Maria too [be.PRS.3SG sister]
* 'Petros is a brother and Maria too.'

*I Maria ine adherfi, ke I Petros episis [ine adherfos].
the.F.SG Maria be.PRS.3SG sister and the.M.SG Petros too [be.PRS.3SG brother]
* 'Maria is a sister and Petros too.'
(Merchant 2014: 4)

c. O Petros ine kalos jatros, ala I Maria ine
the.M.SG Petros be.PRS.3SG good-M-SG doctor-SG but the.F.SG Maria be.PRS.3SG
mia kakia [jatros].
a.F.SG bad.F.SG [doctor.SG]
'Petros is a good doctor, but Maria is a bad one.'

I Maria ine kali jatros, ala O Petros ine
the.F.SG Maria be.PRS.3SG good.F.SG doctor.SG but the.M.SG Petros be.PRS.3SG
enas kakos [jatros].
a.M.SG bad.M.SG [doctor.SG]
'Maria is a good doctor, but Petros is a bad one.'
(Merchant 2014: 7)

d. O Petros ine kalos dhaskalos, ala I Maria ine
the.M.SG Petros be.PRS.3SG good.M.SG teacher.M.SG but the.F.SG Maria be.PRS.3SG
mia kakia [dhaskala].
a.F.SG bad.F.SG [teacher.F.SG]
'Petros is a good teacher, but Maria is a bad one.'

*I Maria ine kali dhaskala, ala O Petros ine
the.F.SG Maria be.PRS.3SG good.F.SG teacher.F.SG but the.M.SG Petros be.PRS.3SG
enas kakos [dhaskalos].
a.M.SG bad.M.SG [teacher.M.SG]
* 'Maria is a good teacher, but Petros is a bad one.'
(Merchant 2014: 8)

Sudo & Spathas (2020) refine Merchant's categories for Greek nouns, taking a gender competition approach where the masculine gender does not trigger any social gender inferences. The asymmetry in (6d) is then due to the feminine form implying the masculine form, while the reverse does not hold.

Looking at nominal ellipsis in Brazilian Portuguese, Nunes & Zocca (2009) also proposed a similar three-way distinction between human nouns for Brazilian Portuguese: those which license gender mismatches under ellipsis in both directions (7c, d), those (7e, f) whose masculine form sometimes allows a feminine subject (with the reverse being impossible), and those (7a, b) which allow no gender mismatches.

(7) a. **O Drácula* é *conde* *e* *a* *Mina* *também* é
the-m-sg Dracula be.PRS.3SG count.M.SG and the.F.SG Mina also be.PRS.3SG
[*condessa*].
[count.F.SG]
* 'Dracula is a count and Mina is too.'

b. **A Mina é condessa e o Drácula também* é
the.F.SG Mina be.PRS.3SG count.F.SG and the.M.SG Dracula also be.PRS.3SG
[*conde*].
[count.M.SG]
* 'Mina is a countess and Dracula is too.'

c. *O Paulo é médico e a Fernanda também é*
the.M.SG Paulo be.PRS.3SG doctor.M.SG and the.F.SG Fernanda also be.PRS.3SG
[*médica*].
[doctor.F.SG]
'Paulo is a doctor and Fernanda is too.'

d. *A Fernanda é médica e o Paulo também é*
the.F.SG Fernanda be.PRS.3SG doctor.F.SG and the.M.SG Paulo also be.PRS.3SG
[*médico*].
[doctor.M.SG]
'Fernanda is a doctor and Paulo is too.'

e. ?*O Paulo é ator e a Fernanda também é*
the.M.SG Paulo be.PRS.3SG actor.M.SG and the.F.SG Fernanda also be.PRS.3SG
[*atriz*].
[actor.F.SG]
? 'Paulo is an actor and Fernanda is too.

f. ??*A Fernanda é atriz e o Paulo também é*
the.F.SG Fernanda be.PRS.3SG actor.F.SG and the.M.SG Paulo also be.PRS.3SG
[*ator*].
[actor.M.SG]
?? 'Fernanda is an actress and Paulo is too.'
(Nunes & Zocca 2009: 216)

For Bobaljik & Zocca (2011), nobility (e.g. **conde*/**condesa*) and kinship nouns (e.g. *#tio*/*#tia*) always carry gender presupposition and allow for no mismatches. Alternating role nouns (e.g. ?*ator*/*#atriz*) show an asymmetric mismatch. For the latter, this asymmetry suggests a morpho-syntactic distinction between a marked [feminine] and an unmarked [masculine], where only [feminine] triggers a female referent presupposition. Other alternating nouns, in Brazilian Por-

tuguese, may also fall into the third class of nouns (which allows for symmetric mismatch), as it includes not only common gender nouns (as in Greek) but also role nouns (*médico/a*) that inflect for gender like adjectives, and do not carry gender presupposition.

Building upon Bobaljik & Zocca's (2011) and Merchant's (2014) proposals, Sprouse et al. (2022) empirically tested English nobility nouns (*prince/princess*) (8a, b), and alternating role nouns (*actor/actress*) (8c, d) under ellipsis.

(8) a. **John is a prince and Mary is* [*a prince*] *too.*
b. **Mary is a princess and John is* [*a princess*] *too.*
c. *John is an actor and Mary is* [*an actor*] *too.*
d. **Mary is an actress and John is* [*an actress*] *too.*
(Sprouse et al. 2022: 6)

A first acceptability study tested 16 pairs of nouns, with 8 belonging to a possible kinship and nobility class as in (8a, b), and 8 to a possible actor/actress class as in (8c, d), following a 2×2×2 design (with ±ORDER ±MISMATCH and ±ELLIPSIS as factors), and found an interaction between +MISMATCH and +ELLIPSIS for 7 of the 16 pairs. A second study tested 58 gendered pairs as in (8c, d) (with a 2×2 design: ±ORDER, ±MISMATCH), and found mismatch asymmetries for 35 of them, with a high inter-participant variation. They did not make a comparison with non-elliptical or anaphora counterparts and did not take gender biases (associated with profession names) into account. They did not confirm Haspelmath's relative frequency hypothesis, according to which the higher frequency of masculine forms makes Masculine-First ordering more acceptable and more efficient for communication.[1]

The present study seeks to further explore the question of gender mismatches under stripping ellipsis for adjectives and nouns in French, using both corpora (spoken and written) and experiments.

[1] Gandón-Chapela & Gallardo-del-Puerto (2025) tested gender mismatch under VP ellipsis: participants had to choose between strict and sloppy readings. While the sloppy reading was always preferred, it was more so with gender match (i) than with gender mismatch, and slightly more so in the Feminine-Masculine order (ii) than in (iii). Their results show participant sensitivity to gender mismatches under ellipsis.

(i) *Peter has voted for himself and Robert has too.*
a. *Robert has voted for Peter.* (strict reading)
b. *Robert has voted for Robert.* (sloppy reading)

(ii) *Elaine has voted for herself and Peter has too.*
c. *Peter has voted for Elaine.* (strict reading)
d. *Peter has voted for Peter.* (sloppy reading)

(iii) *Peter has voted for himself, and Elaine has too.*
e. *Elaine has voted for Peter.* (strict reading)
f. *Elaine has voted for Elaine.* (sloppy reading)

1.3. Testing French gender

Comparing French adjectives and nouns, the former have inflectional gender while the second are usually considered as having inherent gender (Haspelmath 2002; Zwanenburg 1988). We expect more gender mismatch with adjectives than with nouns.

But adjectival and nominal categories in French are somewhat permeable (Noailly 2021) and human nouns usually come in gendered pairs, with the same paradigmatic structure as adjectives (Bonami & Boyé 2019). Although adjectival gender is inflectional, the status of nominal gender is disputed for human nouns: derivational (Zwanenburg 1988) or inflectional (Spencer 2002). As inflectional (agreement) features should be ignored for ellipsis resolution (Merchant 2010), whether nouns pattern with adjectives or not for gender mismatches under ellipsis will shed light on this debate.

In the following, we distinguish grammatical gender (which triggers agreement) as a formal feature classifying nouns, from semantic gender as a conceptual classification according to a noun's lexical meaning, and social gender as a way one presents oneself in society, hinging upon world knowledge. For human nouns, grammatical gender encodes different meanings, implications, and connotations (Richy & Burnett 2019). For languages with grammatical gender, Corbett (1991)'s Semantic Core Generalisation states that for human nouns, social, and grammatical genders tend to correspond. Alternating noun pairs fall under this generalisation (9a): masculine grammatical gender corresponds to male social gender, and feminine grammatical gender to female social gender. Epicene nouns (9b) and mixed groups (9c), on the other hand, may not.

(9) a. *C'* *est* *une* *danseuse.*
that.SG be.PRS.3SG a.F.SG dancer.F.SG
'She is a dancer.'
C' *est* *un* *danseur.*
that.SG be.PRS.3SG a.M.SG dancer.M.SG
'He is a dancer.'

b. *Ces* *personnes* *sont* *excellentes.*
these.PL people.F.PL be.PRS.3PL excellent.M.PL
'Those people are excellent.'

c. *Ces* étudiants *sont* *impressionnants.*
these.PL student.M.PL be.PRS.3PL impressive.M.PL
'Those students are impressive.'

Role nouns also come with social biases, and stereotypical gender information may become part of the word's meaning (Mickus et al., 2019). A woman dancer (*danseuse*) is more stereotypical

than a man one (*danseur*), while a man banker (*banquier* 'banker.M) is more stereotypical than a woman one (*banquière* 'banker.F') (Misersky et al. 2014).

Based on corpus studies and experiments, we investigate which factors are at play for the production and comprehension of gender mismatches under stripping ellipsis. We will directly compare predicative adjectives and nouns in French, with the hypothesis that gender mismatch under ellipsis is possible for adjectival predicates but more constrained for nominal predicates. If French resembles (Brazilian) Portuguese more than (modern) Greek, another hypothesis is that symmetric mismatch may be possible beyond syncretic nouns. We will compare production data (from both written and spoken corpora) and comprehension data (from experiments).

2. Corpus study of French ellipsis

2.1. Methodology

We used different French corpora, of varying registers and modality: Frantext Contemporain (French literary texts after 1980, ATILF), FrTenTen (a French web-based corpus from the TenTen family, Kilgarriff et al. 2004), ORFEO (the spoken part, Benzitoun et al. 2016) and DUEL (a multimodal corpus of recorded interactions, Hough et al. 2016). As they are not annotated for ellipsis, extraction from the corpora was semi-automatic.[2]

We found 643 adjectival and 81 nominal ellipsis occurrences across the corpora, most of them in the written corpora (only five stripping utterances are found across spoken corpora).

We annotated occurrences for a variety of semantic, phonological and syntactic factors and, when such information was available, for the gender of the author or speaker (see Table 1). Beyond annotating the presence of mismatches (in gender and number), as well as the conjuncts order (Masc-Masc, Fem-Fem, Masc-Fem, Fem-Masc), the targets were annotated for morpho-phonology. For adjectival predicates, we distinguished homophones (e.g. *ému/e* 'moved.M/F') and syncretic pairs (e.g. *timide* 'shy.M/F') from alternating pairs. For nominal predicates, we contrasted common gender (such as *enfant* 'child.M/F') and epicene nouns from alternating pairs. We did this with the hypothesis that syncretism would favour mismatches. We also considered the relative frequency of the masculine form and the feminine form, (*Freq_Masc-Freq_Fem*), as well as the difference between the two forms and the lemma frequency. We used log-transformed frequencies from the corpus of French subtitles Lexique (New et al. 2001). We also annotated the subject of the full

[2] For Frantext, ORFEO and DUEL, we looked for all occurrences of *aussi/non plus* ('too/neither'), manually sorting through some 80k occurrences and keeping only stripping utterances (manual selection allowed us to account for disfluencies and large contexts). For FrTenTen, we used a CQL (Corpus Query Language) search for occurrences of *aussi/non plus* ('too/neither') in the vicinity of verbs such as être/sembler/para*itre* ('be/seem/appear') using 2–5-word jokers, and kept stripping constructions.

(antecedent) clause and the subject of the elided (target) clause for their category, nominal or pronominal, and pronominal subjects were annotated for their gender marking (in French, 1st and 2nd person do not inflect in gender), and humanness. For predicative nouns, we checked whether the target was a bare noun (10a) or had a determiner (which agrees in gender, 10b).

(10) a. *Mon grand-père était pasteur. Le vôtre aussi* [était
my.SG grandfather be.PST.3SG pastor.M.SG the.M.SG yours.SG too [be.PST.3SG
pasteur].
pastor.M.SG]
'My grandfather was a pastor. Yours too.'
[Frantext, de Monferrand 1990]

b. *Je suis le chef et le dernier de mon clan, toi aussi*
I be.PRS.1SG the.SG chief.M.SG and the.SG last.M.SG of my.SG clan you too
[*tu es le chef et le dernier de ton clan*].
[you be.PRS.2SG the.SG chief.M.SG and the.SG last.M.SG of your.SG clan]
'I am the chief and last of my clan, you too.'
[Frantext, Bauchau 1990]

Table 1: Summary of the main annotation factors

Annotation target	Annotation factor	Factor levels
Correlate, Remnant	Category	Nominal/Pronominal
	Gender markedness	Marked (3rd person/Unmarked (1st, 2nd person)
	Humanness	Human/Non-human
Constituent under ellipsis	Lexical frequency	Masc. /Fem./(Masc_freq-Fetn freq)/Lemma
	± Gender mismatch	Match/Mismatch
	± Number mismatch	Match/Mistmatch
	± Bare noun	Determiner/Bare
	Order	Masc-Masc/Fem-Fem/Masc-Fem/Fem-Masc
Occurrence	Author gender	Female/Male/Unknown

2.2. Hypotheses

We expect a lower mismatch rate for nominal predicates than for adjectives: gender is indisputably an inflectional feature for adjectives, but this is more controversial for French noun pairs. Overall, with human subjects, we expect more Masculine-First order, regardless of mismatch, due to a general ordering tendency for humans (Abeillé et al. 2018; Da Cunha 2024).

In mismatch cases, there are several hypotheses. For adjectives, we expect no interaction with ordering (as for other inflectional features). For human nouns, Merchant (2014) predicts more

Masculine-First, if the masculine noun is compatible with a female subject, but not vice versa. But referential theories would predict more Feminine-First if the (elided) masculine noun is more accessible than the feminine one at the ellipsis site. We also expect an interaction with lexical frequency, if masculine forms are more frequent than their feminine counterparts (Haspelmath 2020).

Given the specificities of French, we hypothesise that syncretic adjectives and common gender nouns favour gender mismatches under ellipsis, followed by homophone pairs: in such cases, mismatches would be less salient and therefore easier to produce. According to Noailly (2021), 18% of French adjectives are syncretic for gender (e.g. *timide* 'shy.M/F') and 66% are homophones (e.g. *joli/e* 'pretty.M/F') in standard French, leading us to posit that unmarked inflection—orthographically or phonologically—would favour mismatches, in line with identity theories of ellipsis. According to Bonami & Boyé (2019), half of human nouns are syncretic or common gender in (spoken) French.

We also expect that pronominal subjects unmarked for gender (1st and 2nd person) will favour mismatches, given that gender would only be marked by agreement. In contrast and in line with Corbett's (2023) Generalised Agreement Hierarchy, we expect gendered human subjects to disfavour mismatches, as they carry inferrable social gender.

2.3. Corpus results

2.3.1. Adjectival predicates under ellipsis

For predicative adjectives, we observe an overall high rate of gender mismatches[3]: 40.03% (N=225), and a higher count of Masculine-First occurrences (N=349, 62.1%) (13). Although the Feminine-First order is less frequent (N=213, 37.9%), the mismatch rate for Feminine-First is higher (55.4%, N=118/213) (14a), than with Masculine antecedents (30.7%, N=107/349). This order effect is statistically robust: Feminine-First is a robust predictor of gender mismatches. We report these results in Figure 1 and detail Bayesian inference models in Appendix A.[4] These results may be explained by the accessibility hypothesis if masculine referents are more frequent overall, they are more accessible, and thus more easily retrieved in the case of mismatches.

We found no robust effect of adjectival form: non-syncretic adjectives, regardless of the morphophonological distance (11, 13a–c) do not seem to disfavour gender mismatches. The same holds

[3] Number mismatch rates are much lower, with 19.2% (N=108) and, for occurrences with a number mismatch, the gender mismatch rate is very similar to that of the entire sample (41.6%, N=45), which would indicate no superadditive mismatch penalty for occurrences such as (11) below.

[4] We consider there to be robust and compelling evidence for an effect of a predictor if the probability of a posterior distribution estimate being different from 0 is superior to 95%.

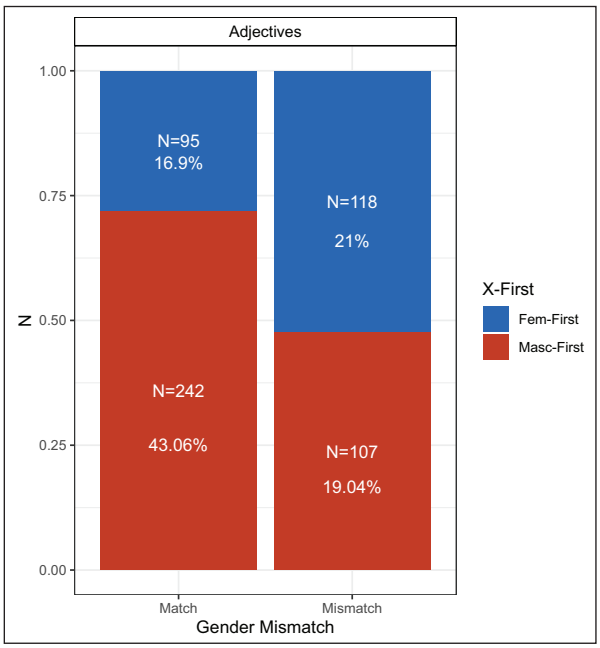


Figure 1: Order preferences for adjectival corpus data showing counts and proportions

for syncretic adjectives (12a) and homophones (12b), which do not affect gender mismatch rates: although theory predicts an advantage for such cases, the lack of syncretism effects may simply be due to the fact that adjectives in general, where gender is inflexional, do not penalise gender mismatches, regardless of more fine-grained morphophonological properties.

(11) *Astruc est fou de bonheur, elles aussi* [*sont folles de bonheur*].
Astruc be.PRS.3SG crazy.M.SG of joy, they.F.PL too [be.PRS.3PL crazy.F.PL of joy]
'Astruc is overjoyed, them too.'
(Frantext, Gréco 1982)

(12) a. *Mon père était athée, moi aussi* [*j'étais athée*].
my.SG father be.PST.3SG atheist.M.SG me too [be.PST.1SG atheist.F.SG]
'My father was an atheist, me too.'
(Frantext, Montferrand 1990)

b. *Il était* ému, *moi aussi* [*j' étais* émue].
he be.PST.3SG moved.M.SG me too [I be.PST.1SG moved.F.SG]
'He was moved, me too.'
(Frantext, Chaix 1985)

(13) a. *Elle était vieille, et lui aussi [était vieux].*
she be.PST-3SG old.F.SG and him too [be.PST.3SG old.M.SG]
'She was old, and him too.'
[Frantext, Bianciotti, 1985]

b. *Je suis devenue folle et Nick aussi [est devenu fou].*
I be.PRS.1SG become.F.SG crazy.F.SG and Nick too [be.PRS.3SG become.M.SG crazy.M.SG]
'I became crazy and Nick too.'
[Frantext, Delperdange, 1992]

c. *La photo est belle, le garçon aussi [est beau].*
the.F.SG picture.F.SG be.PRS.3SG beautiful.F.SG the.M.SG boy too [be.PRS.3SG beautiful.M.SG]
'The picture is beautiful, the boy too.'
[Frantext, Barthes, 1980]

Looking at subject category and gender markings, we found that pronominal correlates robustly favour gender mismatches, and observe a mismatch rate of 51.25% (N=103). In our corpus, while only a minority of pronominal subjects are marked for gender (13a), we observe a slightly higher gender mismatch rate than the baseline for predicative adjectives. Namely, we found a 60.5% mismatch rate (N=49) for pronominal correlates, and a 57.1% mismatch rate (N=20) for pronominal remnants (11), contra our hypothesis. Although this factor is not robust, we discuss this result in section 2.4.

We looked at subject humanness, with the hypothesis that the uninterpretability of grammatical gender for inanimates would favour mismatches. We found no main effect of subject humanness (with rates similar to baseline for correlates (44.9%, N=169) and remnants (42.4%, N=173) alike), but inferential modelling shows a robust interaction between order and subject humanness: in the Feminine-First order, human remnants (13a, b) favour gender mismatches for predicative adjectives. Despite the strong correlation between grammatical and social gender for humans (Corbett 1991), humanness does not penalise mismatches.

To further investigate interactions with social gender, we looked at a possible role of author gender in their production of mismatches, when such information was available. Most of the occurrences were produced by male authors (N=231, or 57%), but we observe a tendentially higher mismatch rate for female authors (48.3%, N=84) than for male (36.8%, N=85), although this is not a robust effect. In cases of mismatch, we found no effect of author gender on ordering preferences: Feminine-First is preferred for male (54.11%, N=46) and female (53.57%, N=45) authors alike.

2.3.2. Nominal predicates under ellipsis

In our corpora, we found fewer nominal predicates under ellipsis (N=81) than predicative adjectives (N=562), with 15 occurrences of gender mismatches. With a rate of 18.5%, gender mismatches are rarer with predicate nouns than with adjectives, but are still produced (14a, b). Given the size of our dataset, we were not able to reach robustness thresholds for reliable predictors, and instead present tendencies and observations.

(14) a. *Je me sens un habitant du quartier […] Moi aussi* [*je me sens une habitante du quartier*].
I me feel.PRS.1SG a.M.SG inhabitant.M.SG from neighbourhood me too [I me feel.PRS.1SG a.F.SG inhabitant.F.SG from neighbourhood]
'I feel like an inhabitant of the neighbourhood. Me too.'
[CFPP]

b. *Ses parents* étaient *musiciens […] Vous aussi* [*étiez musicienne*]*?*
her.PL parents be.PST.3PL musician.M.PL you.SG too [be.PST.2PL musician.F.SG]
'Her parents were musicians. You too?'
[Frantext, Garat 2006]

c. *Elle avait l' air d' une enfant prise en faute. Moi aussi* [*j' avais l' air d' un enfant pris en faute*]
she have.PST.3SG the.SG look.M.SG of a.F.SG child.F.SG taken.F.SG in fault me too I have.PST.1SG the.SG look.M.SG of a.F.SG child.F.SG taken.F.SG in fault
'She had the look of a child caught red-handed. Me too.'
[Frantext, Férey 2012]

d. *Babet n' était pas cuisinière… Marie non plus* [*n' était pas cuisinière*].
Babet NEG be.PST.3SG NEG cook.F.SG Marie neither NEG be.PST.3SG NEG cook.F.SG
'Babet was not a cook… Marie neither.'
[Frantext, Chandernagor 1995]

Previous literature on gender mismatches under ellipsis highlighted the specificity of role nouns with regards to ordering constraints: in Romance languages more specifically, for alternating role nouns, mismatches should incur a stronger penalty in the Feminine-Masculine order.

Our results, in line with Sprouse et al. (2022), seem to indicate that the Masculine-Feminine ordering may indeed be preferred, found in 62.96% of cases (N=51/81), but does not account for all cases, indicating that ordering is a preference rather than a constraint. The overall preference for Masculine-First (75.31%, N=61/81) (14a), regardless of mismatch, in line with previous observations on the ordering of conjuncts, may also play a role. When Feminine is in first position

(N=20/81) (14c), we observe a mismatch rate of 25% (N=5/20), while the mismatch rate for Masculine-First is slightly lower (16.39%, N=10/61): this seems to indicate that Feminine-First may favour gender mismatch for nominal predicates, although, again, these contrasts are not robust.

Since some nominal lemma frequencies were not available in Lexique (New et al. 2001), we analysed the lexical frequency of 57 occurrences (45 lemmas, out of the 67 total lemmas in the corpus). Overall, only one role noun (*cuisinier/ère* 'cook.M/F') was more frequent in its feminine form than its masculine form and occurred in a matching occurrence (14d). When the masculine form is the most frequent of the pair (N=19), we find four cases of mismatch, and no Feminine-First ordering.

For analysis purposes, we distinguished cases where gender iconic pairs of nouns are possible (e.g. *habitant/e* 'inhabitant.M/F') which make up 40% of our sample (N=33), from those where such pairs are not available (N=48). In the latter case, gender may be arbitrary (as with inanimates), pairs may not be possible, as for *gentilhomme* ('gentleman') or common gender nouns (e.g. *sociologue* 'sociologist.M/F').

As with predicative adjectives, we do not observe a preference for homophones gender mismatch production (mismatch rate of 16.6%, N=8, close to the overall sample rate). We note here some atypical cases, such as *gentilhomme* ('gentleman') (15c), for which there exists no feminine form: the noun is lexically specified for [masculine] and may only refer to males. We also note (15a), the *homme/femme* 'man/woman' suppletive pair, as well as the kinship nouns *fils/fille* 'son/daughter' (15b). Both of these examples feature surprising gender mismatches, notably despite the impossibility of mismatch for kinship nouns predicted by Merchant (2014).

(15) a. *Vous êtes un homme très occupé*[5]. *En dépit* des *apparences: moi*
you be.PRS.2PL a.M.SG man very busy.M.SG in spite of appearances me
aussi [*je suis une femme très occupée*].
too [I be.PRS.1SG a.F.SG woman very busy.F.SG]
'You are a very busy man. Despite appearances: me too.'
[Frantext, Garat 2008]

b. *Maria Campallo serait une fille de disparus. Votre copain*
Maria Campallo be.COND.3SG a.F.SG daughter of missing.M.PL your friend.M.SG
Luz/Orlando aussi [*serait un fils de disparus*].
Luz/Orlando too [be.COND.3SG a.M.SG son of missing.M.PL]
'Maria Campallo would be a daughter of missing persons. Your friend Luz/Orlando too.'
[Frantext, Férey 2012]

[5] A reviewer notes that another reconstruction would be possible in (15a) *je suis très occupée* ('I am very busy.F'), but this would not work for (15b) where a noun is needed. A common gender noun might be possible (*serait un enfant de disparus* 'would be a child of missing persons'). Both these examples show that the gender of the antecedent noun can be ignored under ellipsis.

c. *Je vois que vous êtes, monsieur, maintenant, gentilhomme de*
I see.PRS.1SG that you.PL be.PRS.2PL Sir now gentelman by
naissance […] *et moi aussi* [*je suis maintenant gentilhomme de*
birth and me too I be.PRS.1SG now gentleman by
naissance].
birth
'I see that you are, Sir, now, a gentleman by birth, and me too.'
[Frantext, Koltès 1988]

Determiners, in French, agree with the noun they specify: the presence of a determiner marks gender when the noun does not alternate, and makes it possible to disambiguate nominal and adjectival uses (16). In that sense, we expected bare nouns, closer to adjectives, to favour gender mismatches, but we found identical gender mismatch rates of 18.5% with and without determiner.

Given the lack of semantic motivation for grammatical gender of non-human referents, we expected non-human subjects to favour gender mismatches. Our corpus had very few non-human subjects (6 correlates and 7 remnants) (16b), and no gender mismatches with non-human subjects.

Overall, our corpus data show a predominance of (human) pronominal subjects. We observe a non-robust but tendentially higher-than-baseline mismatch rate for pronominal correlates (26.7%, N=8) and remnants (30.4%, N=7) alike. For unmarked pronominal subjects (16c)[6], this rate appears to (24%, N=6 for correlates, 28.6%, N=6 for remnants), which, if verified through inferential modelling on larger samples, would indicate that gender-marked pronominal (16a) subjects favour mismatches.

(16) a. *Il ne faut pas me prendre pour un con, ni elle non plus*
IMPRS not must NEG me take.INF for a.M.SG idiot.M.SG nor her either
[*il ne faut pas la prendre pour une conne*].
[IMPRS not must NEG her.REFL take.INF for a.F.SG idiot.F.SG]
'Don't take me for a fool, nor her either.'
[Frantext, Bayon 1987]

b. *Tous les livres de Queneau ne sont pas oulipiens, tous les*
all the.PL books of Queneau not be.PRS.3PL NEG oulipian.M.PL all the.PL
miens non plus [*ne sont pas oulipiens*].
mine either [not be.PRS.3PL NEG oulipian.M.PL]
'All of Queneau's books are not oulipian, all of mine neither.'
[Frantext, Perec, 2004]

[6] Note that in cases where the remnant is unmarked for gender, the referent's gender was recovered from the context in the annotation phase.

c. *T' es le fils d' un voleur et moi aussi [je suis*
you be.PRS.3PL the.M.SG son of a.M.SG thief.M.SG and me too I be.PRS.3PL
le fils d' un voleur].
the.M.SG son of a.M.SG thief.M.SG
'You are the son of a thief. Me too.'
[Frantext, Portante 1993]

Where such information was available, we looked at author gender as a possible factor influencing mismatches and ordering preferences. While most authors are men (77%, N=30), mismatch rates appear to be similar for male (23.3%, N=7) and female authors (22.3%, N=2), although given the size of our sample, this was not verified by our model. From Figure 2, we can observe that the proportion of masculine and feminine subjects may differ depending on author gender. Masculine-Everywhere is the predominant case for male authors, who never produce the Feminine-Everywhere. Female authors, on the other hand, appear to equally use Masculine-Feminine and Masculine-Masculine, with no Feminine-First mismatches.

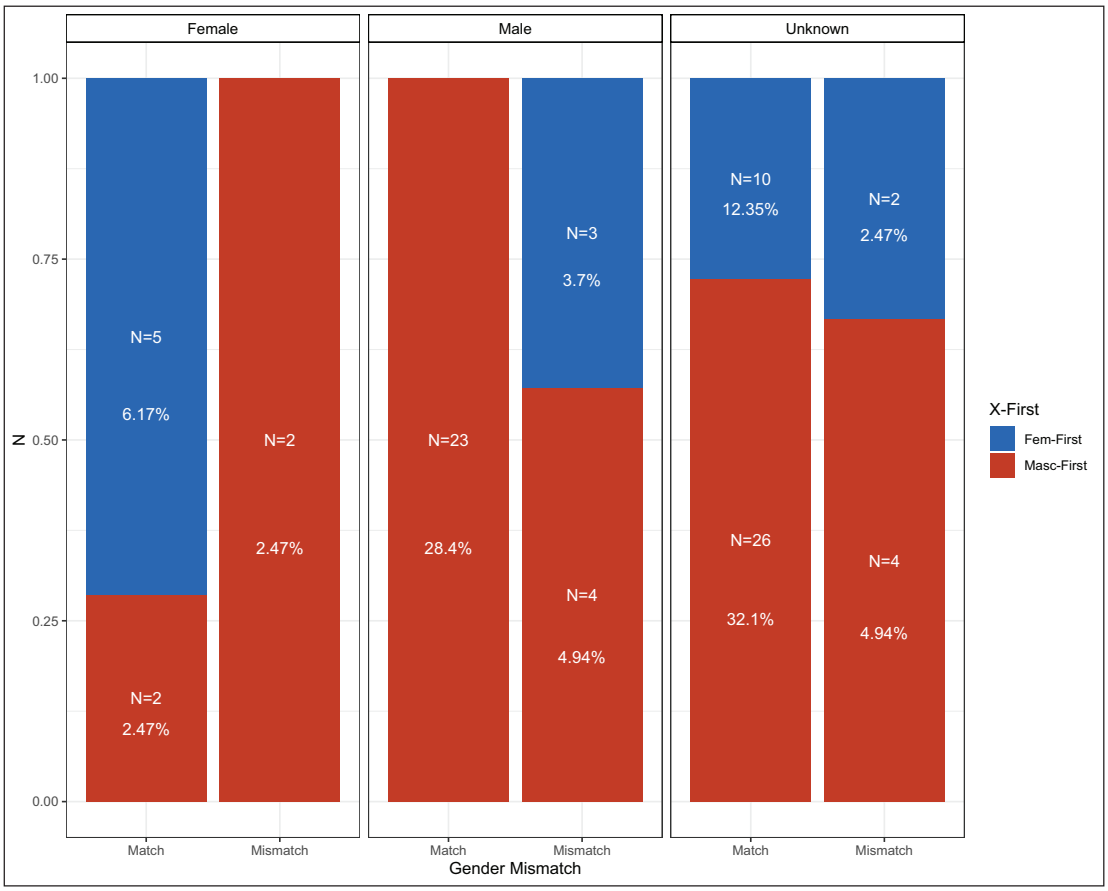


Figure 2: Order preferences per author gender for nominal corpus data

2.4. Interim discussion

Our corpus findings reveal that gender mismatch is well attested in French adjectival and nominal predicate ellipsis, in both orderings. We also found several tendencies for the production of gender mismatches under ellipsis of predicative adjectives (mismatch rate of 40.03%, N=225) and nouns (mismatch rate of 18.5%, N=15), which we present below.

Overall, animate and pronominal subjects favour mismatches, and author gender has an effect on the ordering of the conjuncts. Formal factors of the target predicate such as syncretism, homophony, as well as lexical frequency, had no robust effect, but some trends emerge, for which we evoke several hypotheses, ahead of experimental data.

First, contrary to initial hypotheses, pure indexicals (1st and 2nd person pronouns) which are unmarked for gender do not seem to favour mismatches, while marked pronouns do. While these pure indexical assume some associated demonstration for referent identification, 3rd person pronouns are directly referential, with no extra demonstration needed (Kaplan 1989). For these pronouns, which are not bound variables, gender features may be syntactically optional, but they are semantically interpreted, with the contrast between [feminine] and [masculine] being highly accessible.

The lexical frequency hypothesis posits frequency-based coding efficiency; and the ellipted material must be accessible in addressees' active memory. Haspelmath's (2020) theory would predict differential marking between the Masculine-First ordering and its marked counterpart, Feminine-First. Corpus data reveal a preference for the Feminine-First pattern in cases of mismatch, despite a strong mention bias in favour of masculine referents. The frequent ellipsis of masculine predicates (Feminine-First) is unsurprising given that they are more easily predicted and more accessible, per lexical frequency hypotheses.

Another possible explanation for ordering preferences may be syntactic: biases have been observed in French coordinate structures by Abeillé et al. (2018) and Da Cunha (2024), where male speakers clearly preferred Masculine-First orders, hinting at a syntactic gender bias correlated to the masculine social gender. Within stripping constructions, with two conjunct clauses, it is not surprising to observe the same preference for Masculine-First (in line with Aparicio et al. 2015, Sprouse et al. 2022, and Gandón-Chapela & Gallardo-del-Puerto 2025): this may be an effect of coordinate structures rather than a product of ellipsis or gender mismatches.

Our data show that gender mismatches are possible for French predicate adjectives and nouns, in both orders, even for alternating and suppletive pairs, suggesting grammatical gender may be an inflectional feature on both adjectives and human nouns. Given the small sample of predicative nouns, definitive conclusions are difficult based on corpus data alone. This is why we also ran acceptability studies.

3. Experiments on French ellipsis

We ran experiments with predicative adjectives and role nouns, to evaluate relative penalties (or meliorations) of gender mismatches per clause type (stripping ellipsis, *le* anaphora, and full repetitions), ordering, and participant gender, among other factors.

3.1. Experiment on predicative adjectives

Having collected corpus-based data on elliptical utterances, we collected acceptability data comparing 'deep' and 'surface' anaphora (Hankamer & Sag 1976), to see whether gender mismatched predicative adjectives are more acceptable in anaphoric or elliptical conditions.

3.1.1. Methodology

We ran an online acceptability judgment task (1–10 scale), coded and hosted on IbexFarm (Drummond 2007). After ethics approval by University Paris Cité Ethics Committee, we recruited 94 self-declared native French speakers on the RISC website (https://expesciences.risc.cnrs.fr/). After filtering out participants who did not meet the necessary criteria[7], we analysed the data from 89 participants, 57 of whom were women (64%), 30 were men and 2 were non-binary.

The experiment followed a 2×2 Latin square design with ±ELLIPSIS and ±MISMATCH as factors, for a total of 16 target items. The experiment also included 12 control items from An & Abeillé (2021) which tested gender agreement of adjectives, and 16 filler items on extraction out of *wh-* questions.[8]

For the target items, the +ELLIPSIS condition featured a simple stripping construction (Table 2), and the anaphoric condition featured the clitic pronoun *le*. Order was used as an intra-item variable, with half of the items following a Masculine-First order. We used human, non-pronominal subjects (using names with unambiguous social and grammatical gender). Only adjectives with an audible gender alternation were used (non-syncretic, non-homophones).

3.1.2. Hypotheses

We expect an overall preference for ellipsis: in an acceptability task, Bîlbîie (2023) found an interaction between gender mismatches and sentence types in Romanian, where non-elliptical

[7] Each item was followed by a simple yes/no comprehension question, participants who scored below 85% accuracy were not kept for analysis. At the end of each experiment, participants were presented with a short meta-data questionnaire (age, gender, native language, education level and country of residence): we thus also excluded non-native speakers.

[8] Experimental items (including filler and control items), data and scripts can be found on the project's OSF repository: https://osf.io/5azyd/.

Table 2: Example items for each condition (Experiment 1)

	UTTERANCE	±MISMATCH	ORDER	Item
1a	Ellipsis	Mismatch	Masc-Fem	Paul est grand, et Marie aussi.
1b	Ellipsis	Match	Masc-Masc	Paul est grand, et Michel aussi. *Paul is tall.M, and Marie/Michel too.*
1c	Anaphora	Mismatch	Masc-Fem	Paul est grand, et Marie l'est aussi.
1d	Anaphora	Match	Masc-Masc	Paul est grand, et Michel l'est aussi. *Paul is tall.M, and Marie/Michel (it) is too.*
2a	Ellipsis	Mismatch	Fem-Masc	Alice est belle, et Victor aussi.
2b	Ellipsis	Match	Fem-Fem	Alice est belle, et Léa aussi. *Alice is beautiful.F, and Victor/Léa too.*
2c	Anaphora	Mismatch	Fem-Masc	Alice est belle, et Victor l'est aussi.
2d	Anaphora	Match	Fem-Fem	Alice est belle, et Léa l'est aussi. *Alice is beautiful.F, and Victor/Léa (it) is too.*

utterances carried a significant repetition penalty. These results were likely due to a redundancy penalty, which Amsili et al. (2016) argued to be heightened by the presence of the additive adverb *aussi* 'too'. Similarly, results from Aparicio et al. (2015) showed ellipsis sensitivity to an antecedent's syntactic representation, with results highlighting different processes for anaphora and ellipsis.

In line with syntactic, semantic, and discursive parallelism constraints in coordinate structures (Frazier et al. 1984; Bîlbîie 2021), matching cases should be more acceptable overall, given that they satisfy all relevant constraints. Building on Hankamer & Sag's (1976) deep and surface anaphora distinction, we expect a stronger parallelism constraint for surface anaphora (stripping), where mismatches should incur a higher cost. If ellipsis patterns with anaphora, we should not observe any interaction between the utterance type and the mismatch acceptability, with an overall preference for +MATCH.

Finally, the preference for Masculine-First in nominal coordination (Abeillé et al. 2018) as well as syntactic preferences for masculine subjects (da Cunha & Abeillé 2022, da Cunha 2024) lead us to hypothesise that Masculine-First will be rated higher in all conditions. In line with corpus results, this may also interact with participant gender.

3.1.3. Results

All conditions received high acceptability ratings, far above ungrammatical controls (shown in Figure 3). We found a robust main effect of mismatch (a mismatch penalty), as well as a robust main effect of the utterance type (a preference for ellipsis), compatible with a redundancy penalty caused by the pronominal *le*, which retrieves the material in the antecedent clause, as well as the repeated copula. Interaction coefficients show no robust interaction between ellipsis and mismatch, no robust effect of ordering, and no robust interaction between order and ellipsis. We report these results in Figure 3 and detail all model results in the Appendix (Appendix B). Regarding order, we found a non-robust tendency for Masculine-First to be preferred, across ±MISMATCH and ±ELLIPSIS.

Mismatches under ellipsis were rated higher than ungrammatical controls (Figure 3), and tendentially higher than under anaphora (non-robust interaction), in both orders, unlike in Spanish, where Aparicio et al. (2015) found that the Feminine-Masculine order was more penalised than other mismatch conditions (Masculine-First, and number mismatches). Our results are compatible with both the direct interpretation approach (without syntactic reconstruction), and the deletion-under-identity approach (assuming identity ignores adjective agreement inflectional features).

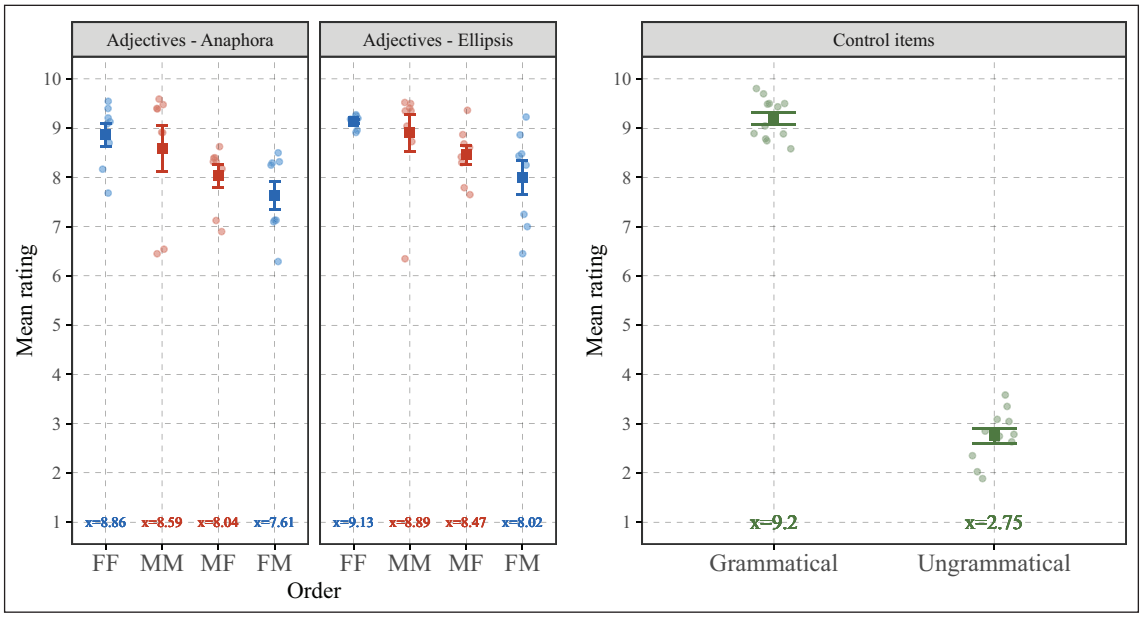


Figure 3: Mean judgments per sentence, order for adjectives (Experiment 1) and control items (right). Dots represent means per item, and boxes represent means per condition

We found no robust effect of participant gender and only observe a tendency for female participants to give slightly lower ratings to mismatches, especially under anaphora. Male participants showed a slight preference for +MATCH under anaphora, and -MATCH under ellipsis. We found no robust effect of gender on order: all participants seem to prefer Masculine-First across all conditions.

3.2. Experiments on predicative nouns

We sought to test the acceptability of gender mismatches for predicative role nouns, while accounting for gender stereotypes.

3.2.1. Methodology

To control the effect of gender stereotypes on role noun interpretation, we used normed scores from Misersky et al. (2014) to distinguish non-biased role nouns (with stereotypicality scores indicating that they do not carry stereotype information, e.g. *étudiant/e* 'student.M/F') from male-biased (e.g. *chirurgien/ne* 'surgeon.M/F') and female-biased (e.g. *infirmier/ère* 'nurse.M/F') role nouns. Experiment 2 tested 24 non-biased role nouns, and Experiment 3 tested 24 biased ones (12 female-biased and 12 male-biased role nouns).

Participants were randomly assigned to one of the two experiments, with post-hoc counterbalancing. Both experiments were online acceptability judgment tasks (1–10 scales), coded and hosted on IbexFarm (Drummond 2007). Following ethics approval, we recruited 85 self-declared native French speakers, and kept 83 for analysis, following the exclusion of participants who had obtained low comprehension scores, using the same methodology as for Experiment 1. Among these participants, 40 were women (48.2%), 41 were men (49.4%), and 2 were non-binary (2.4%), for a median age of 30.

Each experiment followed a 2×3 Latin Square design with UTTERANCE (Ellipsis/Anaphora/Full repetition) and ±MISMATCH as factors, totalling 6 conditions for 24 target items (Table 3). For both experiments, the 24 target items were supplemented with 12 control items (used for Experiment 1) and 16 filler items on the French active/passive alternation.[9]

In the Full repetition condition, both the determiner and the noun are repeated: for -MATCH, we expect this condition to serve as an ungrammatical control. For each item, we used nominal human subjects, with gender-unambiguous names. For Experiment 2, order was encoded as an intra-item variable. For Experiment 3, order was congruent with stereotypicality: items featuring a male-biased role noun followed a Masculine-First order, female-biased ones were Feminine-First.

[9] Experiment items (including filler and control items, Miserksy scores), data and scripts can be found on the project's OSF repository: https://osf.io/5azyd/.

Role nouns were selected for each experiment according to their Misersky score (Misersky et al. 2014): non-biased nouns were those with a score between 0.45 and 0.55; biased nouns had scores above 0.6 (female bias) or below 0.4 (male bias). We also controlled for morpho-phonological alternations, excluding syncretic and homophone forms (epicenes and common gender nouns): half of the noun pairs had a final consonant alternation (e.g. *assistant/assistante* 'assistant.M/F'). For the remaining nouns, all of which ended in *-eur* in the masculine, half of these took *-euse* (e.g. *chanteur/chanteuse* 'singer.M/F'), and half took *-rice* (e.g. *instituteur/institutrice* 'school-teacher.M/F') endings in the feminine. These correspond to the most frequent alternations for gender iconic pairs in the Morphalou database (Bonami & Boyé 2019). Twelve grammatical controls were also included, identical to those described in Section 3.1.1.

Table 3: Example items for each experimental condition

	UTTERANCE	±MISMATCH	Item
1a	Ellipsis	Match	Jeanne est une lectrice rapide, et Clara aussi.
1b	Ellipsis	Mismatch	Jeanne est une lectrice rapide, et Corentin aussi. *Jeanne is a fast reader.F, and Clara/Corentin too.*
1c	Anaphora	Match	Jeanne est une lectrice rapide, et Clara l'est aussi.
1d	Anaphora	Mismatch	Jeanne est une lectrice rapide, et Corentin l'est aussi. *Jeanne is a fast reader.F, and Clara/Corentin (it) is too.*
1e	Full repetition	Match	Jeanne est une lectrice rapide, et Clara est une lectrice rapide aussi.
1f	Full repetition	Mismatch	Jeanne est une lectrice rapide, et Corentin est une lectrice rapide aussi. *Jeanne is a fast reader.F, and Clara/Corentin is a fast reader.F too.*

3.2.2. Hypotheses

We only took (non-syncretic) alternating role nouns and tested several hypotheses. If French follows Greek (Merchant 2014) and English (Sprouse et al. 2022), these nouns should license mismatches only in the Masculine-Feminine order: ellipsis should be penalised, compared to anaphora, in the Feminine-Masculine order. If one assumes syntactic reconstruction and gender as a derivational feature on nouns, ellipsis with mismatch should be rated like full repetition. If on the other hand, gender is an agreement inflectional feature for these nouns like for adjectives and/or ellipsis is interpreted directly (without syntactic reconstruction), we expect no ellipsis mismatch penalty, in either order.

We do not test common gender nouns, but we expect the mismatch penalty to grow with the formal distance between the two forms. Moreover, given that gender stereotypes are inherent to the interpretation of grammatical gender for human nouns (Richy & Burnett 2019; Mickus et al. 2019), we expect an effect of gender stereotypes on the acceptability of all utterances, with the hypothesis that the stereotype violation adds a superadditive penalty in cases of mismatching gender.

Finally, as was the case for adjectival predicates in French (Experiment 1) and Spanish (Aparicio et al. 2015), we expect an overall preference for the Masculine-First condition, interacting with participant gender.

3.2.3. Results

For both experiments, we found a robust main effect of utterance types, with a repetition penalty. We observe an acceptability cline, such that ellipsis is preferred to anaphora (regardless of ±MISMATCH), and both are preferred to full repetition. For both experiments, we also find a robust main effect of +MISMATCH corresponding to a mismatch penalty across all conditions. For both experiments, we found no robust effect of noun morphology, unlike what was proposed for English (Sprouse et al. (2022) and Greek (Merchant 2014), namely that symmetrical mismatch is only allowed for common gender nouns. It seems more in line with Bobaljik & Zocca's (2011) proposal for Brazilian Portuguese, that some alternating role nouns have inflectional gender like adjectives. But while half of the nouns we tested have a consonant gender alternation (*assistant/e* 'assistant.M/F') like most French adjectives (*grand/e* 'tall.M/F'), the other half has alternating types (*chanteur/euse* 'singer.M/F', *lecteur/rice* 'reader.M/F') with no equivalent for adjectives. No difference was found between these two morphological subtypes in both experiments[10].

Observations on mean judgments show MISMATCH interacting with gender biases: biased role nouns (Experiment 3) received lower ratings under mismatch than non-biased role nouns (Experiment 2). For both experiments, mismatched full repetitions are close to ungrammatical controls (Figure 4).[11]

For non-biased nouns (Experiment 2, Figure 4), as with predicative adjectives (Experiment 1), mismatches under ellipsis received higher ratings than mismatches under anaphora, in both orders. This casts doubt on the syntactic reconstruction hypothesis for ellipsis and/or the analysis of nominal gender as derivational. We also found a robust interaction between MISMATCH and anaphora/ellipsis, both corresponding to an acceptability increase compared to the MISMATCH:full repetition condition, which acted as an acceptability penalty ceiling.

[10] Item-per-item results for all three experiments are available on the project's OSF repository.
[11] Inferential modelling results are detailed in Appendix B.

While order was randomly encoded intra-item for Experiment 2, we found a robust main interaction between order and utterance type: Feminine-Masculine elliptical and anaphoric items were preferred to Feminine-Masculine full repetition sentences. In the full repetition condition, ratings were at ceiling for a mismatch penalty: masculine nouns are not suitable with a feminine subject, and feminine nouns are not suitable with a masculine subject (Figure 4). The suspension of this gender competition effect under ellipsis is in line with Sudo & Spathas's (2020) account.

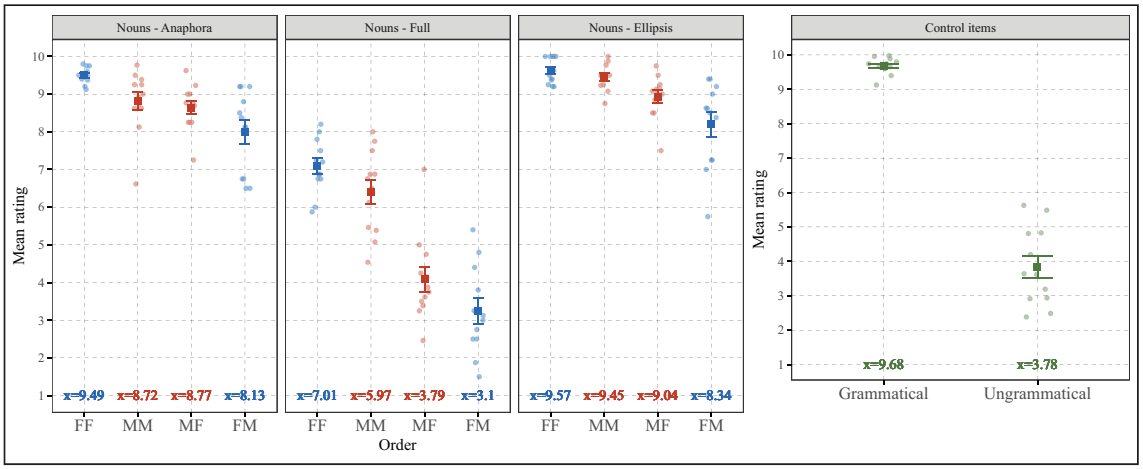


Figure 4: Mean judgments per utterance, order for non-biased nouns (Experiment 2) and control items. Dots represent means per item, and boxes represent means per condition

For biased nouns (Experiment 3, Figure 5), we found a robust MISMATCH penalty (stronger than for non-biased nouns) across all conditions, with a tendential melioration for anaphora compared to ellipsis, although both are robustly preferred to full repetition sentences. For this experiment, order was congruent with stereotype: male-biased nouns only appeared in the Masculine-First order, and female-biased nouns in the Feminine-First order. We found a robust order effect across all conditions (both directions showed equal MISMATCH penalties, with no preference between Masculine or Feminine antecedents). We found no robust interaction of order and utterance type, but observe tendencies in Figure 5.

We found no robust effect of participant gender for either experiment, but plot acceptability tendencies in Figure 6, where we observe an overall stronger penalty of MISMATCH from male participants.

Overall, experimental results are in line with corpus results and confirm the acceptability of gender mismatches under nominal ellipsis, with a tendential preference for Masculine-First for predicate nouns, for all corpus authors and experiment participants. Corpus data follow judgments for MISMATCH, with a clear preference for MATCH in experimental conditions. As with corpus data, we observe no robust effect of predicate noun morphology.

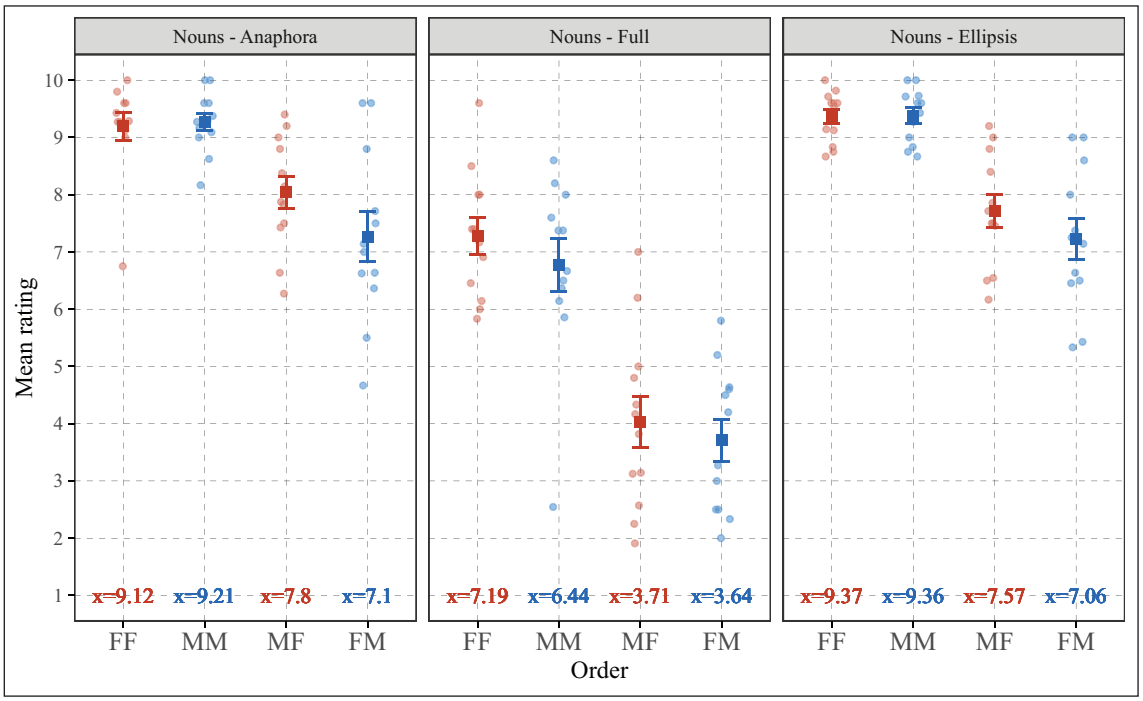


Figure 5: Mean judgments per utterance and order for biased nouns (Experiment 3). Dots represent means per item, and boxes represent means per condition

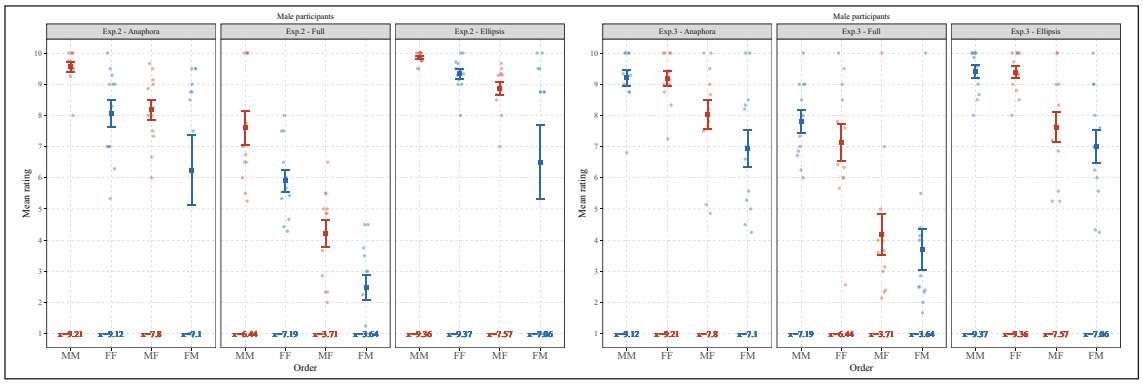


Figure 6: Mean judgments per utterance, order, and for male participants (Experiments 2 and 3)

4. Discussion

The interaction between utterance type and mismatch acceptability, which emerges for predicative uses of both adjectives and nouns, seems to confirm the hypothesis that adjectives

and human nouns pattern together in French. This is at odds with Merchant's (2014) proposal for Greek and Sprouse et al.'s (2022) for English. Our results are compatible with two possible theories. Either gender is inflectional (an agreement feature) for French predicative nouns as it is for adjectives, which would allow an analysis in terms of syntactic reconstruction under identity; or, one might assume that nominal ellipsis falls under a direct interpretation approach, with antecedent accommodation similar to pronominal resolution.

Moreover, it appears that ellipsis and anaphora pattern alike in French, which challenges previous predictions based on syntactic reconstruction at the ellipsis site. For French adjectives and nouns, the mismatch penalty under ellipsis is even smaller than under anaphora. Parker (2019) offers a possible explanation to this penalty asymmetry. Anaphora resolution based on a cue-based retrieval mechanism rests upon a cognitive principle which encodes elements in active memory. Cues allow us to extract and activate an antecedent when its correspondence with the anaphoric form is highest: according to Parker (2019), mismatches and partial matches disturb sentence processing of anaphora, which may incur acceptability penalties.

Regarding order effects, several explanations are possible at this stage, compatible with markedness and accessibility theories. It is possible to consider them a consequence of syntactic markedness, where the formally marked construction (Feminine-First) corresponds to the informationally marked construction (Comrie 1989).

For French corpus and experimental data, the Masculine-Everywhere preference is maintained, and while Feminine-everywhere is rare in corpora, it is rated above mismatched conditions by participants in experimental conditions.

In cases of mismatch, ordering preferences are reversed between corpus data (Feminine-First) and experiments (Masculine-First). This effect may be attributed to speaker gender: male authors account for an overwhelming majority of our sample, while female participants are more numerous, but this likely interacts with other factors, as Feminine-Masculine order is more frequent than Masculine-Feminine in corpora (despite a majority of male authors), while the reverse is true in experiments (despite a majority of female participants). It is possible that the grammatical masculine form is unmarked for men (if only in production data) which, consequently, renders the masculine form more accessible for ellipsis; for women, both forms would be more symmetrical in their markedness. The preference asymmetries we found are also compatible with a difference between production and comprehension: if our active memory stores some ungendered lemma, and not the grammatical form, the lemma (which is identical to the masculine form in French) is then more accessible during production.

5. Conclusion

We found that gender mismatches are attested in French corpora for predicative adjectives and nouns, and are judged acceptable in experimental studies, in both orders (Masc-Fem, Fem-Masc), across morphological types and gender stereotypes. Experimental results also reveal an interaction between mismatch acceptability and utterance type, with a general ellipsis preference for nouns and adjectives alike:

+MISMATCH:

ADJECTIVES: Ellipsis > Anaphora ≫ Ungrammatical controls

NOUNS: Ellipsis ≥ Anaphora ≫ Ungrammatical controls ≥ Full repetition

Gender mismatch under predicate ellipsis (stripping) was rated higher than with predicate anaphora, unlike what was found for Spanish adjectives (Aparicio et al. 2015). This is compatible both with deletion-based theories and direct interpretation of ellipsis for adjectives, since adjective gender is an agreement feature (contextual, inflectional) to be ignored under ellipsis. But for nouns, this challenges previous theories (Merchant 2014; Sprouse et al. 2022), which derive ellipsis by deletion-under-identity and assume nominal gender as inherent or derivational. This seems more in line with Boballjik & Zocca's (2011) proposal for Brazilian Portuguese that gender is inflectional for some alternating role nouns. The *médico/médica* class would have to be extended for French to nouns which do not morphologically patterns like adjectives, and we do not find evidence for an asymmetric *ator/atriz* class.

Overall, for corpus and experiments alike, we observe no robust effect of the morphophonology of nouns, nor do we observe a robust effect of lexical frequency. For adjectives, corpus data show that syncretic and homophone adjectives are not a factor favouring mismatches under ellipsis, and experimental data confirm that mismatched alternating pairs are fully acceptable under ellipsis, unlike with full repetition.

Contrary to Greek and English, Masculine-Feminine ordering was not the only possible order for gender mismatch for (alternating) nouns under ellipsis in French. In our corpus data, Fem-Masc ordering was among the factors favouring gender mismatch under ellipsis for both adjectives and nouns. In our experimental data, both orderings were acceptable, with a preference for Masc-Fem for ellipsis and anaphora alike, for both adjectives and nouns. The main factor disfavouring gender mismatch in our experiments was nominal gender stereotypes, which was not taken into account in previous studies.

The present study accounts for several factors which are at play for the production and comprehension of gender mismatches under stripping ellipsis, and brings empirical support for these conclusions, in a quantitative syntax (corpus studies) as well as an experimental syntax

(acceptability studies) approach. Two conclusions are suggested, which are not incompatible: the gender of French alternating role nouns is contextual and inflectional (Spencer 2002; Bonami & Boyé 2019) and thus ignored under ellipsis, and/or predicate ellipsis (stripping) is directly interpreted with no syntactic reconstruction (Ginzburg & Sag 2000; Culicover & Jackendoff 2005; Nykiel & Kim 2021). The second conclusion would apply to other languages as well, which remain to be tested empirically.

Acknowledgements: We thank the anonymous reviewers, as well as the audience at Experimental and Corpus-Based Approaches to Ellipsis, for their helpful comments and discussions.

Funding: This work has received support from the investment programme "France 2030" as part of the IdEx programme (ANR-18-IDEX-0001) implemented by Université Paris Cité, under which the inIdEx project EFL is conducted.

Appendix

Appendix A

We adopt a Bayesian framework for our inferential statistical modelling, to translate data towards a precise estimate of the proportional distribution of the data. We used the 'brms' package (Bürkner 2017) in R, and the Markov chain Monte-Carlo sampling method (MCMC) to get a representative sample of the posterior probability distribution.

To model corpus data, which are count data, we use the Bernoulli likelihood, corresponding to a 0 (-MISMATCH) or 1 response with a probability of θ (in the log-odds space). We use weakly informative priors (normal distribution centred on 0) and default priors for thresholds and standard deviations of random effects (occurrence and author ID).

For each model, we ran four sampling chains of 2000 iterations, with 1000 warm-up iterations, for a total of 4000 chains per parameter tuple.

Posterior probabilities are obtained using the hypothesis testing function of brms, based on evidence ratios. Posterior predictive checks can be found in the project's OSF.

Model 1: Adjectival Gender mismatch

> GEN_MISMATCH~XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER-)+(XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER)|AUTHOR_

ID)+(XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER)|LEMMA)

Reference levels of predictors are set to 'Match' (GENDER_MISMATCH & NUMBER_MISMATCH), 'alternating' adjectives (non HOMOPHONE, non SYNCRETIC), 'nominal' 'non-human' and 'gendered' subjects (SUBJ1, SUBJ2), and 'unknown' AUTHOR_GENDER.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
$\beta_{\text{Fem-First}}$	2.08	0.78	[0.55;3.62]	1*
$\beta_{\text{Nb-Mismatch}}$	0.32	0.77	[−1.2;1.82]	0.67
$\beta_{\text{Homophone}}$	0.15	0.72	[−1.29;1.49	0.59
$\beta_{\text{Syncretic}}$	−0.94	0.78	[−2.43;0.6]	0.88
$\beta_{\text{Pro-subj1}}$	1.58	0.78	[0;3.12]	0.98*
$\beta_{\text{Pro-subj2}}$	0.60	0.80	[−0.99;2.16]	0.78
$\beta_{\text{Human-subj1}}$	1.03	0.75	[−0.45:2.5]	0.91
$\beta_{\text{Human-subj2}}$	0.13	0.79	[−1.4;1.72]	0.56
$\beta_{\text{Ungen-subj1}}$	0.14	0.82	[−1.47;1.76]	0.57
$\beta_{\text{Ungen-subj2}}$	0.46	0.82	[−1.15;2.06]	0.71
$\beta_{\text{Female-author}}$	0.43	0.75	[−1.05;1.86]	0.71
$\beta_{\text{Male-author}}$	−0.88	0.71	[−2.29;0.49]	0.89
$\beta_{\text{Fem-First:Nb-mismatch}}$	−0.22	0.89	[−1.94;1.51]	0.59
$\beta_{\text{Fem-First:Homo phone}}$	0.54	0.87	[−1.1;2.28]	0.73
$\beta_{\text{Fem-First:Syncretic}}$	−0.40	0.93	[−2.22;1.36]	0.66
$\beta_{\text{Fem-First:Pro-subj1}}$	0.02	0.92	[−1.77;1.88]	0.5
$\beta_{\text{Fem-First:Pro-subj2}}$	−0.39	0.90	[−2.14;1.39]	0.66
$\beta_{\text{Fem-First:Human-subj1}}$	0.84	0.90	[−0.94;2.6]	0.82
$\beta_{\text{Fem-First:Human-subj2}}$	1.55	0.90	|−0.21;3.34]	0.96*
$\beta_{\text{Fem-First:Ungen-subj1}}$	−0.06	0.92	[−1.86;1.71]	0.53
$\beta_{\text{Fem-First:Ungen-subj2}}$	−0.59	0.92	[−2.35;1.21]	0.74
$\beta_{\text{Fem-First:Female-author}}$	−0.68	0.91	[−2.44;1.08]	0.78
$\beta_{\text{Fem-First:Male-author}}$	0.57	0.95	[−1.3;2.44]	0.72

Model 2: Nominal Gender mismatch

GEN_MISMATCH~XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER+DETERMINER)+(XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER+DETERMINER)|AUTHOR_ID)+(XFIRST*(NUM_MISMATCH+HOMOPHONE+SYNCRETIC+SUBJ1_CAT+SUBJ2_CAT+SUBJ1_HUM+SUBJ2_HUM+SUBJ1_GENDERED+SUBJ2_GENDERED+AUTHOR_GENDER+DETERMINER)|LEMMA)

Reference levels of predictors are identical to Model 1, with an additional factor DETERMINER ('no').

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
$\beta_{\text{Fern-First}}$	0.15	1.00	[−1.79;2.04]	0.56
$\beta_{\text{Nb-mismatch}}$	0.12	1.02	[−1.85;2.14]	0.54
$\beta_{\text{Homophone}}$	−0.08	0.98	[−1.98;1.87]	0.53
$\beta_{\text{Syncretic}}$	0.10	0.04	[−1.76;2.01]	0.53
$\beta_{\text{Pro.subj1}}$	0.24	0.97	[−1.65;2.15]	0.6
$\beta_{\text{Pro-subj2}}$	0.28	0.97	[−1.59.2.17]	0.61
$\beta_{\text{Human-subj1}}$	0.18	0.98	[−1.72;2.05]	0.57
$\beta_{\text{Human-subj2}}$	0.03	0.98	[−1.91;2]	0.52
$\beta_{\text{Ungen-subj1}}$	0.22	0.97	[−1.67;2.12]	0.58
$\beta_{\text{Ungen-subj2}}$	0.20	0.94	[−1.66;2.08]	0.58
$\beta_{\text{Female-author}}$	0.01	0.99	[−1.93;1.95]	0.51
$\beta_{\text{Male-author}}$	0.08	0.96	[−1.75;1.99]	0.54
$\beta_{\text{Determiner}}$	0.00	0.95	[−1.85;1.87]	0.5
$\beta_{\text{Fem-First:Nb-mismatch}}$	0.01	1.00	[−1.94;1.98]	0.5
$\beta_{\text{Fem-First:Homophone}}$	0.18	0.98	[−1.8;1.98]	0.57
$\beta_{\text{Fem-First:Syncretic}}$	0.13	1.01	[−1.82;2.07]	0.55
$\beta_{\text{Fem-First:Pro-subj1}}$	−0.21	0.99	[−2.13;1.71]	0.59
$\beta_{\text{Fem-First:Pro-subj2}}$	−0.13	0.97	[−2.06;1.8]	0.55
$\beta_{\text{Fem-First.Human-subj1}}$	0.04	1.02	[−1.95;2.06]	0.51
$\beta_{\text{Fern-First:Human-subj2}}$	0.15	1.00	[−1.81;2.12]	0.55
$\beta_{\text{Fem-First:Ungen-subj1}}$	−0.26	0.98	[−2.15;1.68]	0.61
$\beta_{\text{Fem-First:Ungen-subj2}}$	−0.07	0.97	[−1.97;1.83]	0.53
$\beta_{\text{Fem-First:Female-author}}$	−0.19	1.00	[−2.09;1.74]	0.57
$\beta_{\text{Fem-First:Male-author}}$	0.17	1.01	[−1.85;2.15]	0.57
$\beta_{\text{Fem-First.Determiner}}$	−0.08	1.01	[−2.09;1.86]	0.53

Appendix B

Inferential modelling of experimental data was done within the Bayesian framework: we ran cumulative mixed effects ordinal logistic regression models, adjusted using the Stan language and the 'brms' package (Bürkner 2017) in R. Acceptability ratings call for such models as the categorical predictors are ordered, and we seek to estimate the maximal cumulative probability of observing an ordinal variable Y (participants' judgments), given the fixed effects matrices (predictors) and random effects (participant and item). Each model considers population effects, which are those random effects, as well as the parameters of interest. Models were fit using factored predictors (using centred treatment coding), with intercepts by-participant and by-item, as well as random slopes for random effects.

Posterior predictive checks, priors, and model specifications can be found in this project's OSF.

We present summaries of fixed effect posterior distributions for the models.

Model 3: (Experiment 1)

RATING~ORDER*UTTERANCE+(ORDER*UTTERANCE|PARTICIPANT)+(ORDER*UTTERANCE|ITEM)

Reference levels of predictors are set to 'Masculine-Masculine' ORDER, and 'Full repetition' UTTERANCE.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
$\beta_{\text{Fem-Fem}}$	−0.18	0.61	[−1.38;1.02]	0.62
$\beta_{\text{Fem-Masc}}$	−1.52	0.84	[−3.1;0.15]	0.96*
$\beta_{\text{Masc-Fem}}$	−1.51	0.60	[−2.68;−0.29]	0.99*
β_{Anaphora}	4.65	0.49	[3.63;5.58]	1*
β_{Ellipsis}	4.99	0.57	[3.83;6.06]	1*
$\beta_{\text{Fem-Fem.Anaphora}}$	−0.54	0.69	[−1.89;0.82]	0.78
$\beta_{\text{Fem-Masc:Anaphora}}$	1.68	0.64	[0.31;2.89]	0.99*
$\beta_{\text{Masc-Fem:Anaphora}}$	0.88	0.73	[−0.56;2.26]	0.89
$\beta_{\text{Fem-Fem.Ellipsis}}$	−0.25	0.71	[−1.64:1.17]	0.63
$\beta_{\text{Fem-Masc:Ellipsis}}$	1.18	0.69	[−0.25:2.48]	0.95*
$\beta_{\text{Masc-Fem:Ellipsis}}$	0.74	0.71	[−0.68;2.1]	0.85

Model 4: (Experiment 1)

RATING~MISMATCH*UTTERANCE*MORPHOLOGY+(MISMATCH*UTTERANCE*MORPHOLOGY|PARTICIPANT)+(ORDER*UTTERANCE*MORPHOLOGY|ITEM)

Reference levels of predictors are set to +MISMATCH, 'consonant' MORPHOLOGY, and 'Full repetition' UTTERANCE.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
β_{Mismatch}	−2.03	0.51	[−3.02;−1.01]	1*
β_{Anaphora}	4.54	0.56	|3.39;5.61]	1*
β_{Ellipsis}	4.92	0.62	[3.65;6.1]	1*
$\beta_{\text{-eur/-euse}}$	0.10	0.34	[−0.57;0.78]	0.63
$\beta_{\text{-eur/-rice}}$	−0.07	0.34	[−0.77;0.59]	0.58
$\beta_{\text{Mismatch:Anaphora}}$	1.60	0.54	[0.53;2.67]	1*
$\beta_{\text{Mismatch:Ellipsis}}$	1.21	0.57	[0.08;2.35]	0.98*
$\beta_{\text{Mismatch:-eur./-euse}}$	0.16	0.56	[−0.92;1.26]	0.61
$\beta_{\text{Mismatch:-eur/-rice}}$	0.37	0.56	[−0.76;1.48]	0.75
$\beta_{\text{Anaphora:-eur/-euse}}$	−0.15	0.61	[−1.39;1]	0.61
$\beta_{\text{Ellipsis:-eur/-euse}}$	−0.20	0.64	[−1.51;0.98]	0.61
$\beta_{\text{Anaphora:-eur/-rice}}$	−0.29	0.61	[−1.5;0.91]	0.68
$\beta_{\text{Ellipsis:-eur/-rice}}$	−0.50	0.62	[−1.72;0.72]	0.79
$\beta_{\text{Mismatch:Anaphora:-eur/-euse}}$	−0.28	0.77	[−1.78;1.21]	0.64
$\beta_{\text{Mismatch:Ellipsis:-eur/-euse}}$	−0.08	0.80	[−1.63;1.48]	0.54
$\beta_{\text{Mismatch:Anaphora:-eur/-rice}}$	0.36	0.78	[−1.17;1.88]	0.68
$\beta_{\text{Mismatch:Anaphora:-eur/-rice}}$	0.15	0.80	[1.41;1.74]	0.57

Model 5: (Experiment 2)

RATING~ORDER*UTTERANCE+(ORDER*UTTERANCE|PARTICIPANT)+(ORDER*UTTERANCE|ITEM)

Reference levels of predictors are identical to Model 3.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
$\beta_{\text{Fem-Fem}}$	0.39	0.34	[−0.28;1.06]	0.88
$\beta_{\text{Fem-Masc}}$	−2.85	0.55	[−3.87;−1.73]	1*
$\beta_{\text{Masc-Fem}}$	−2.71	0.44	[−3.57;−1.82]	1*
β_{Anaphora}	3.87	0.35	[3.17;4.56]	1*
β_{Ellipsis}	3.98	0.38	[3.24;4.72]	1*
$\beta_{\text{Fem-Fem:Anaphora}}$	−0.60	0.49	[−1.57;0.37]	0.89
$\beta_{\text{Fem-Masc:Anaphora}}$	0.09	0.52	[−0.91;1.12]	0.57
$\beta_{\text{Masc-Fem:Anaphora}}$	0.40	0.50	[−0.59;1.36]	0.79
$\beta_{\text{Fem-Fem:Ellipsis}}$	−0.40	0.55	[−1.5;0.63]	0.77
$\beta_{\text{Fem-Masc:Ellipsis}}$	−0.19	0.50	[−1.17;0.79]	0.65
$\beta_{\text{Masc-Fem:Ellipsis}}$	−0.13	0.16	[−1.04;0.78]	0.61

Model 6: (Experiment 2)

RATING~MISMATCH*UTTERANCE*MORPHOLOGY+(MISMATCH*UTTERANCE*MORPHOLOGY|PARTICIPANT)+(ORDER*UTTERANCE*MORPHOLOGY|ITEM)

Reference levels of predictors are identical to Model 4.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
β_{Mismatch}	−3.13	0.52	[−4.12;−2.05]	1*
β_{Anaphora}	4.00	0.40	[3.18;4.75	1*
β_{Ellipsis}	4.05	0.44	[3.14;4.89]	1*
$\beta_{\text{-eur/-euse}}$	−0.10	0.41	[−0.88;0.73]	0.6
$\beta_{\text{-eur/-rice}}$	0.30	0.34	[−0.36;0.96]	0.82
$\beta_{\text{Mismatch:Anaphora}}$	0.32	0.53	[−0.71;1.35]	0.73
$\beta_{\text{Mismatch:Ellipsis}}$	−0.25	0.48	[−1.19;0.7]	0.7
$\beta_{\text{Mismatch:-eur/-euse}}$	−0.43	0.58	[−1.52;0.73]	0.78
$\beta_{\text{Mismatch:-eur/-rice}}$	−0.55	0.52	[−1.55;0.48]	0.86
$\beta_{\text{Anaphora:-eur/-euse}}$	−0.17	0.56	[−1.3;0.89]	0.6
$\beta_{\text{Ellipsis:-eur/-euse}}$	0.03	0.63	[−1.28;1.23]	0.53
$\beta_{\text{Anaphora:-eur/-rice}}$	0.13	0.51	[−0.89;1.11]	0.62
$\beta_{\text{Ellipsis:-eur/-rice}}$	0.16	0.57	[−0.98;1.29]	0.62
$\beta_{\text{Mismatch:Anaphora:-eur/-euse}}$	0.23	0.76	[−1.29;1.74]	0.62
$\beta_{\text{Mismatch:Anaphora:-eur/-euse}}$	0.20	0.75	[−1.25;1.72]	0.6
$\beta_{\text{Mismatch:Anaphora:-eur/-rice}}$	−0.27	0.69	[−1.58;1.12]	0.65
$\beta_{\text{Mismatch:Ellipsis:-eur/-rice}}$	0.28	0.71	[−1.11;1.71]	0.66

Model 7: (Experiment 3)
RATING~ORDER*UTTERANCE+(ORDER*UTTERANCE|PARTICIPANT)+(ORDER*UTTERANCE|ITEM)

Reference levels of predictors are identical to Model 3.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
$\beta_{\text{Fem-Fem}}$	0.39	0.34	[−0.28;1.06]	0.88
$\beta_{\text{Fem-Masc}}$	−2.85	0.55	[−3.87;−1.73]	1*
$\beta_{\text{Masc-Fem}}$	−2.71	0.44	[−3.57;−1.82]	1*
β_{Anaphora}	3.87	0.35	[3.17;4.56]	1*
$\beta_{\text{Ellipisis}}$	3.98	0.38	[3.24;4.72]	1*
$\beta_{\text{Fem-Fem:Anaphora}}$	−0.60	0.49	[−1.57;0.37]	0.89
$\beta_{\text{Fem-Masc:Anaphora}}$	0.09	0.52	[−0.91;1.12]	0.57
$\beta_{\text{Masc-Fem:Anaphora}}$	0.40	0.50	[−0.59;1.36]	0.79
$\beta_{\text{Fem-Fem:Ellipsis}}$	−0.40	0.55	[−1.5;0.63]	0.77
$\beta_{\text{Fem-Masc:Ellipsis}}$	−0.19	0.50	[−1.17;0.79]	0.65
$\beta_{\text{Masc-Fem:Ellipsis}}$	−0.13	0.46	[−1.04;0.78]	0.61

Model 8: (Experiment 3)

RATING~MISMATCH*UTTERANCE*MORPHOLOGY+(MISMATCH*UTTERANCE*MORPHOLOGY|PARTICIPANT)+(ORDER*UTTERANCE*MORPHOLOGY|ITEM)

Reference levels of predictors are identical to Model 4.

Parameter thresholds	Estimate	Std. dev.	95% Credible interval	Posterior probability P(β<0|β>0)
β_{Mismatch}	−3.13	0.52	[−4.12;−2.05]	1*
β_{Anaphora}	4.00	0.40	[3.18;4.75]	1*
β_{Ellipsis}	4.05	0.44	[3.14;4.89]	1*
$\beta_{\text{-eur/-euse}}$	−0.10	0.41	[−0.88;0.73]	0.6
$\beta_{\text{-eur/-rice}}$	0.30	0.34	[−0.36;0.96]	0.82
$\beta_{\text{Mismatch:Anaphora}}$	0.32	0.53	[−0.71;1.35]	0.73
$\beta_{\text{Mismatch:Ellipsis}}$	−0.25	0.48	[−1.19;0.7]	0.7
$\beta_{\text{Mismatch:-eur/-euse}}$	−0.43	0.58	[−1.52;0.73]	0.78
$\beta_{\text{Mismatch:-eur/-rice}}$	−0.55	0.52	[−1.55;0.48]	0.86
$\beta_{\text{Anaphora:-eur/-euse}}$	−0.17	0.56	[−1.3;0.89]	0.6
$\beta_{\text{Ellipsis:-eur/-euse}}$	0.03	0.63	[−1.28;1.23]	0.53
$\beta_{\text{Anaphora:-eur/-rice}}$	0.13	0.51	[−0.89;1.11]	0.62
$\beta_{\text{Ellipsis:-eur/-rice}}$	0.16	0.57	[−0.98;1.29]	0.62
$\beta_{\text{Mismatch:Anaphora:-eur/-euse}}$	0.23	0.76	[−1.29;1.74]	0.62
$\beta_{\text{Mismatch:Anaphora:-eur/-euse}}$	0.20	0.75	[−1.25;1.72]	0.6
$\beta_{\text{Mismatch:Anaphora:-eur/-rice}}$	−0.27	0.69	[−1.58;1.12]	0.65
$\beta_{\text{Mismatch:Ellipsis:-eur/-rice}}$	0.28	0.71	[−1.11;1.71]	0.66

References

Abeillé, Anne, Aixiu An & Aoi Shiraïshi. 2018. L'accord de proximité du déterminant en français. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics* 22. 1–25.

Abeillé, Anne, Gabriela Bîlbîie & François Mouret. 2014. A Romance perspective on gapping constructions. In Boas, Hans C & Gonzálvez-García, Francisco, *Constructional approaches to language* 15, 227–265. John Benjamins Publishing Company.

Abeillé, Anne & Jong-Bok Kim. 2022. 'Me too' fragments in English and French: a direct interpretation approach. *The Linguistic Review* 39(3). 495–524.

Amsili, Pascal, Ellsiepen, Emilia & Grégoire Winterstein. 2016. Optionality in the use of 'too': The role of reduction and similarity. *Revista da ABRALIN* 15(1).

Aparicio, Helena, Franich, Kathryn & Ming Xiang. 2015. Morphological feature mismatches under ellipsis: An eye-tracking study. *Proceedings of NELS* 45.

Benzitoun, Christophe, Jeanne-Marie Debaisieux & Henri-José Delofeu. 2016. The ORFEO project: A study corpus for contemporary French. *Corpus* 15. 1–18.

Bîlbîie, Gabriela. 2021. L'ellipse dans les comparatives: le cas du gapping en roumain. *Quand le syntagme nominal prend ses marques: du prédicat à l'argument*, 475–497. Éditions et presses universitaires de Reims.

Bîlbîie, Gabriela. 2023. Gender mismatches on adjectival predicates: An experimental perspective on stripping-like constructions in Romanian. *ECBAE 2023*. Amherst.

Bobaljik, Jonathan & Cynthia Zocca (2011). Gender markedness: The anatomy of a counterexample. *Morphology* 21(2). 141–166.

Bonami, Olivier & Gilles Boyé. 2019. Paradigm uniformity and the French gender system. In Baerman, Matthew et al. (eds.) *Morphological Perspectives*, 169–192. Edinburgh University Press.

Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. *Proceedings of the annual meeting of the Berkeley Linguistics Society* 31. 73–91.

Corbett, Greville. 1991. *Gender*. Cambridge University Press.

Corbett, Greville. 2023. The Agreement Hierarchy and (generalized) semantic agreement. *Glossa: A Journal of General Linguistics* 8(1).

Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. Oxford University Press.

Da Cunha, Yanis. 2024. *Le genre comme trait de proéminence en syntaxe*. Université Paris Cité PhD thesis.

Fiengo, Robert & Robert May. 1994. *Indices and identity*. MIT Press.

Frazier, Lyn, Lori Taft, Tom Roeper, Charles Clifton & Kate Ehrlich. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* 12.

Gandón-Chapela, Evelyn & Francisco Gallardo-del-Puerto. 2025. The effect of gender (mis)match on the interpretation of English reflexive pronouns in cases of VP-ellipsis by native and non-native speakers of English. *Glossa: A Journal of General Linguistics* 10(1). 1–33.

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative investigations*. Stanford: CSLI publications.

Hankamer, Jorge & Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3). 391–428.

Haspelmath, Martin. 2002. On understanding word order asymmetries. (Comments on John A. Hawkins, "Symmetries and asymmetries: their grammar, typology and parsing"). *Theoretical Linguistics* 28. 159–170.

Haspelmath, Martin. 2020. Role-reference associations and the explanation of argument coding splits. *Linguistics* 59(1). 123–174.

Hough, Julian, Ye Tian, Laura de Ruiter, Simon Betz, Spyros Kousidis, David Schlangen & Jonathan Ginzburg. 2016. Duel: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. *Proceedings of LREC'16.* 1784–1788.

Kaplan, David. 1989. Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals.

Merchant, Jason. 1999. *The syntax of silence: Sluicing, islands, and identity in ellipsis.* UCSC.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis.* Oxford University Press.

Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua. Structural approaches to ellipsis* 151. 9–32.

Mickus, Timothee, Olivier Bonami & Denis Paperno. 2019. Distributional effects of gender contrasts across categories. *Proceedings of SCL* 2. 174–184.

Miller, Phillip & Geoffrey K. Pullum. 2014. Exophoric VP ellipsis. *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag 5.* 32.

Misersky, Julia, Pascal M. Gygax, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdrova, Lisa Von Stockhausen & Sabine Sczesny. 2014. Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior research methods* 46. 841–871.

Noailly, Michèle. 2021. L'adjectif et le syntagme adjectival. In Abeillé, Anne & Godard, Danièle (eds.), *La Grande Grammaire du français.* Arles: Actes Sud.

Nunes, Jairo. & Zocca, C. 2009. Lack of morphological identity and ellipsis resolution in Brazilian Portuguese. In Nunes, Jairo (Ed.), *Minimalist essays on Brazilian Portuguese syntax*, 142, 215–236. John Benjamins Publishing Company.

Nykiel, Joanna & Jong-Bok Kim. 2021. Ellipsis. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *HPSG: The handbook*, 847–888. Berlin: Language Science Press.

Parker, D. 2019. Cue combinatorics in memory retrieval for anaphora. *Cognitive science* 43(3). e12715.

Poppels, Till. 2022. Explaining ellipsis without identity. *The Linguistic Review* 39(3). 341–400.

Richy, Célia & Heather Burnett. 2019. Jean does the dishes while Marie fixes the car: A qualitative and quantitative study of social gender in French syntax articles. *Journal of French Language Studies* 30(1). 47–72.

Ross, John R. et al. 1969. Guess who? *Proceedings from the annual meeting Chicago Linguistic Society* 5(1). 14–39.

Spencer, A. 2002. Gender as an inflectional category. *Journal of Linguistics* 38(2). 279–312.

Sprouse, Jon, Troy Messick & Jonathan D. Bobaljik. 2022. Gender asymmetries in ellipsis: An experimental comparison of markedness and frequency accounts in English. *Journal of Linguistics* 58(2). 345–379.

Sudo, Yasutada & Giorgos Spathas. 2020. Gender and interpretation in Greek: Comments on Merchant (2014). *Glossa: A Journal of General Linguistics* 5(1).

Zwanenburg, Wiecher. 1988. 'Flexion et derivation: le féminin en français.' Ronald Landheer (ed.), *Aspects de linguistique française, Hommage à Q.I.*M Mok, 191–208. Amsterdam: Rodopi.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 4
https://doi.org/10.18148/zs/2026-2013

Lisa Schäfer (corresponding author), Saarland University, Germany, lisa.schaefer@uni-saarland.de
Ingo Reich, Saarland University, Germany, ingo.reich@uni-saarland.de

The identity condition of right node raising in German

Experimental evidence

Abstract: This article examines the identity condition of right node raising (RNR) in German, focusing on how morpho-phonological mismatches in the verbal domain between the target and its correlate affect the acceptability and processing of RNR. We relate experimental evidence from a cloze study, three rating studies, and a reading time study to the predictions of the central theoretical approaches to RNR mismatches that postulate (gradual) phonological, lexeme, or stem identity. Comparing RNR to gapping and full forms as baselines shows that while complete identity between target and correlate yields the highest ellipsis rates, mismatches do not result in acceptability or processing penalties. However, speakers are less likely to use RNR in the cloze task as the mismatch becomes increasingly morphologically marked, whereas for gapping, we do not observe such a staircase-like pattern. While the rating and processing data of RNR align best with lexeme or stem identity, the cloze data are more in line with modified gradual phonological identity. Our findings suggest that RNR in German is more flexible than previously assumed, with acceptability and processing being tolerant of mismatches, but the cloze data showing greater sensitivity.

Keywords: ellipsis, right node raising, gapping, mismatches

Cite as: Schäfer, Lisa & Ingo Reich. 2026. The identity condition of right node raising in German: Experimental evidence. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2013.
Received: 16 October 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

While discussing a schema to derive constituent coordination from the coordination of two sentences, later famously known as "conjunction reduction," Chomsky (1957: 35–36) mentions in a footnote that structures like (1) cannot be derived by this rule, simply because the seemingly coordinated expressions *John enjoyed* and *my friend liked* are not syntactic constituents. In fact, he even considers the possibility that structures like (1) should be regarded as semi-grammatical, since they show properties like long pauses (between *liked* and *the*) and contrastive stress that "normally mark the reading of non-grammatical strings" (Chomsky 1957: 36).

(1) *John enjoyed and my friend liked the play.* (Chomsky 1957: 36)

However, it did not take long for a broad consensus to emerge that structures like (1) need to be accounted for in grammar in one way or another. One of the first accounts to this effect came from Ross (1986 [1967]: 108) who proposes to derive (1) from the coordination in (2) in terms of across-the-board (ATB)-movement: since both sentences contain identical copies of the object *the play*, the object can be moved out of both conjuncts at the same time, and adjoins as a single copy to the right of the coordination; see (2a). Since in this case ATB-movement only affects the rightmost (major) constituent and targets a right-peripheral position, it is called right node raising (RNR) in Postal (1974) and has been known by this name ever since (see also Sabbagh 2007 for a more recent proposal in terms of ATB-movement.).

(2) *John enjoyed the play, and my friend liked the play.*
   a. [[*John enjoyed* t_1] *and* [*my friend liked* t_1]] *the play*$_1$ (ATB-movement)
   b. [*John enjoyed* ~~*the play*~~] *and* [*my friend liked the play*] (ellipsis)
   c. [[*John enjoyed* 1] *and* [*my friend liked* 1]], 1 = *the play* (multi-dominance)
   d. [[*John enjoyed*] *and* [*my friend liked*]] *the play* (coordination)

Subsequent research has produced convincing arguments that RNR is not the result of movement: RNR is not subject to any relevant island constraints, nor does it respect constituency; see, e.g., Hartmann (2000), for relevant discussion. This is why many researchers have argued that structures like (1) are not the result of movement, but ellipsis. And to better reflect the analysis, other terms have been introduced like "leftward deletion" or "backward conjunction reduction," emphasizing the directionality of RNR in coordination, or "right-peripheral ellipsis," highlighting that RNR only targets strings at the right periphery of non-final conjuncts; see, e.g., Klein (1981), Wilder (1997), Hartmann (2000), for early proposals, and (2b) for illustration (strikeouts indicate ellipsis). More recently, observations that relate to the interpretation of adverbials and agreement properties have motivated multi-dominance (MD) analyses, which start from the assumption that, in principle, one and the same node may be dominated by two different nodes in syntax. With respect to structures like (1), it is argued that the object *the play* is in fact jointly dominated by the verb phrase (VP) of the first and the VP of the second

sentence; see, e.g., McCawley (1982), Wilder (1999), and de Vries (2009), among many others. In other words, the two conjuncts in (2) do not contain identical, albeit distinct copies of *the play* but the very same object. This is indicated in (2c) with the marker [1] that refers to the same object in all occurrences, namely *the play*. This marker is set in gray in the first conjunct to clarify that the object is only spelled out in the final conjunct. Last but not least, psycholinguistic considerations (see Philips 2003), more flexible frameworks (see Steedman 2000), and pragmatic considerations (see Ott 2025) open up analyses in terms of syntactic or pragmatic coordination see (2d), thus closing the circle that began with Chomsky.[1]

The aim of this contribution is not to add another proposal to the already existing accounts, nor to give a thorough discussion of their merits. The primary objective is empirical in nature: to provide crucial empirical data to underpin and support the discussion of identity constraints in German RNR. Existing accounts of RNR make sometimes more, sometimes less clear predictions concerning the nature of the gap in the first conjunct: ATB-movement accounts require morpho-phonological identity of the expression to be moved. Coordination-based approaches make, in principle, similar predictions, since it is one and the same expression that relates to the two conjuncts in the coordinate structure. In MD approaches, the picture is somewhat less clear: on the one hand, the node related to the gap is the same node as the node spelled out in the second conjunct. On the other hand, the phenomenon of "summative agreement" reported in Yatabe (2003), Kluck (2009), Grosz (2015), and others shows that the form of the finite verb may change from singular to plural; see (3) from Shen (2019).

(3) *John's glad that Mary, and Bill's happy that Sue, has/have been to Cameroon.* (Shen 2019: 2)

Still, all three approaches predict—without further assumptions—strict identity conditions for RNR. The deletion approach is more flexible in this respect: the preconditions for deletion could be stated, in principle, at any of the levels of phonology, morphology, syntax, semantics, or pragmatics, or any combination of these levels, depending on the specific empirical behavior of the ellipsis at hand.

In this paper, we will present experimental evidence from a cloze study, three rating studies, and a reading time study on verbal RNR in German that aim to test the identity conditions imposed on the target (gap) and its overt correlate (pivot). German is a great testing ground in this respect, since in German-dependent clauses, the finite verb typically ends up in sentence-final position and is thus always part of the characteristic gap in RNR structures; see (4).

[1] It should be noted that there are also mixed approaches to RNR that argue that RNR has two different potential sources: ellipsis and MD; see, e.g., Barros & Vicente (2011) and Belk et al. (2023). Where RNR is derived by ellipsis, the predictions are those of the ellipsis approach. And where RNR is derived by MD, the predictions are those of the MD approach. Since it is not fully clear (independently of the observed properties) when ellipsis and when MD applies, we do not discuss these approaches in any further detail here.

(4) *Er glaubt, dass Bernd in Berlin* (*arbeitet*) *und Anne in London arbeitet.*
he thinks that Bernd in Berlin (works) and Anne in London works
"He thinks that Bernd works in Berlin and Anne in London."

This allows us to systematically vary person and number features of the target and the correlate, thereby generating stimuli with different degrees of mismatches. But before we go into more detail about the logic behind the experiments, we will give a (probably too) brief overview of positions and discussions in the literature on identity constraints in RNR.

2 Identity constraints in RNR: a brief overview

In an early squib, Eisenberg (1973) argues that RNR tolerates mismatches in syntactic features like person and number if the target and the correlate are identical in phonological form, i.e., if they are syncretic (5a). If, however, a mismatch in the relevant features results in different forms, RNR is judged as ungrammatical (5b).

(5) a. *weil wir das Haus (kaufen.*1PL*) und die Müllers den Garten kaufen.*3PL
since we the house (buy.1PL) and the Millers the garden buy.3PL
(Eisenberg 1973: 418)

b. **weil Franz das Haus (kauft.*3SG*) und ich den Garten kaufe.*1SG
since Franz the house (buy.3SG) and I the garden buy.1SG
(Eisenberg 1973: 418)

From this contrast, Eisenberg concludes that phonological identity is the relevant constraint in RNR structures, as opposed to identity in morphosyntactic features. It should be noted, however, that he explicitly excludes case from this generalization, since "most people can also say" (Eisenberg 1973: 419) *mit (Kindern.*DAT*) und ohne Kinder.*ACC "with (children.DAT) and without children.ACC." In the following, we will refer to this approach as the *phonological identity (PI) approach.*

Klein (1993: 774) essentially follows Eisenberg's generalization but adds that the target and the correlate need to mean essentially the same (in a sense still to be specified). A similar point is made in Chaves (2014: 869), who argues, based on data from the literature, that RNR is blocked if phonologically identical expressions differ in meaning (6).

(6) a. **Mary fed* (*the lamb*) *and Tom enjoyed the lamb.* (Chaves 2014: 869)
b. **Robin swung* (*an unusual bat*) *and Leslie tamed an unusual bat.* (Chaves 2014: 869)

Like Eisenberg (1973), Chaves (2014) accepts mismatches in features such as case, if the target and the correlate are identical in form (i.e., syncretic). However, in contrast to Eisenberg (1973), he also accepts mismatches in number that result in different verb forms (7a), though his judgments seem to depend on the ordering of target and correlate (7b).

(7) a. *Tom thinks that the shoes* (*are …*)*, but Mary thinks that the coat is too expensive.*
Chaves (2014: 871)

b. ??*Tom thinks that the coat* (*is …*)*, but Mary thinks that the shoes are too expensive.*
Chaves (2014: 871)

Chaves (2014: 870) concludes from this data that "the identity conditions imposed by backward periphery deletion pertain to morph stems." While in some cases identity in phonological form is still taken to be a prerequisite for RNR (e.g., in case mismatches, cf. **er findet (Männer.*ACC*) und hilft Männern.*DAT "he finds men.ACC and helps men.DAT;" our judgment), this is not generally assumed, as the number mismatch in (7a) illustrates. Chaves (2014) speculates that the acceptability of form mismatches might depend on the strength of the semantic contribution of the feature (weaker semantic contributions resulting in increased acceptability). Thus, in a way, Chaves (2014) still assumes some kind of phonological identity but also considers a semantically based mechanism that allows for some degree of flexibility in phonological form. We will refer to this as the *morph stem form identity (MSI) approach.*

Like Chaves (2014), Christ (2011) also aims at weakening the phonological identity requirement, but he does so in a very different way. Christ (2011: 383) argues that the acceptability of RNR decreases the more the phonological form of the correlate deviates from that of the target. He illustrates this with RNR that targets nouns (8); the judgments are his:

(8) a. *Hans hat einen* (*Hund.*SG) *und Peter hat mehrere Hunde.*PL (Christ 2011: 383)
Hans has a (dog.SG) and Peter has several dogs.PL

b. ??*Hans hat eine* (*Maus.*SG) *und Peter hat mehrere Mäuse.*PL (Christ 2011: 383)
Hans has a (mouse.SG) and Peter has several mice.PL

c. **Hans hat einen* (*Atlas.*SG) *und Peter hat mehrere Atlanten.*PL (Christ 2011: 383)
Hans has a (atlas.SG) and Peter has several atlases.PL

In (8a), the plural correlate only differs in one respect, namely the plural suffix *-e*. According to Christ, RNR is still fine, i.e., RNR seems to tolerate a single deviation; this is also taken to be true of Umlaut as in *Vögel* "birds" vs. *Vogel* "bird." But if the plural shows Umlaut plus a plural suffix like *-e* as in *Mäuse* "mice" vs. *Maus* "mouse," see (8b), the two deviations from the singular seem to add up, and the resulting RNR structure is judged as significantly less acceptable. And finally, if the plural form even alters the stem of the target, as is the case in *Atlanten* "atlases" vs. *Atlas* "atlas," see (8c), RNR is judged as fully ungrammatical. This approach therefore assumes a gradual, even staircase-like decline of acceptability as a function of phonological deviation. We will refer to this as the *gradual phonological identity (GPI) approach.*

It is surprising that the discussion of identity constraints in RNR is almost exclusively based on introspective data, even though judgments are not always very clear. A notable exception is Shiraïshi et al. (2019). They report on two acceptability studies, one on French and the other on English, that test the acceptability of full matches and of syncretic and non-syncretic forms with a temporal mismatch (which is compared to *do so*-anaphora as a baseline). Given the above discussion on identity constraints, their results are unexpected: in English, both syncretic and non-syncretic mismatches are judged significantly better than the corresponding matches. This is particularly surprising given that Chaves (2014: 870) judges tense mismatches as ungrammatical. In French, mismatches are only marginally worse than matches. A comparison with the grammatical and non-grammatical controls shows that matches as well as mismatches need to be considered as grammatical. Shiraïshi et al. (2019) conclude that variation in both morphological and phonological form is generally tolerated in RNR and that identity is only required on the level of lexemes.[2] In the following, we will call this the *lexeme identity (LI) approach*.

Like Shiraïshi et al. (2019), we will also test full matches against morphological mismatches with syncretic and non-syncretic forms. Unlike Shiraïshi et al. (2019), however, we will use mismatches in number and person since they easily allow for a stepwise increase in morphological deviation required to test the GPI approach: in the first mismatch condition, target and correlate are syncretic forms, in the second condition, they differ in their ending, and in the third they differ in addition in the realization of the vowel within the stem. These four conditions are schematically illustrated in Table 1 using *lesen* "to read" as an example.

Table 1: Design of the four conditions: *identical*, *syncretic*, *ending*, and *ending+vowel*

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(er) liest*	*(sie) liest*	–	–	–	–	identical
2	*(wir) lesen*	*(sie) lesen*	+	–	–	–	syncretic
3	*(ich) lese*	*(sie) lesen*	+	+	+	–	ending
4	*(ich) lese*	*(er) liest*	+	–	+	+	ending+vowel

[2] It is worth mentioning here that this seems to be also true of the proposal presented in Hartmann (2000), which essentially assumes that RNR needs to meet a contrastive focus condition. This focus condition is implemented within alternative semantics, and thus—quite like the lexeme/stem approaches—semantic in nature. Hartmann's (2000) proposal seems to be more general, though, since it concerns not only single lexemes but any sequence of lexemes that is supposed to complete the first conjunct and is located at its right edge. We leave it to future research whether or where exactly the contrastive focus approach and the lexeme/stem approaches diverge in predictions.

Without further assumptions, ATB-movement, MD, and coordination (ATB-MD-C) approaches all predict that RNR should only be grammatical, if target and correlate are identical, simply because one and the same object relates to two different positions. This is, of course, a somewhat simplified representation. As already mentioned, there is the phenomenon of summative agreement that has been dealt with within MD analyses; and there is the phenomenon of conjunct agreement, where agreement is confined to the closest conjunct; see, e.g., Aoun et al. (1994), and more recently Himmelreich et al. (2024). Himmelreich et al. (2024: 77) explicitly state that syncretic forms improve person mismatches in last conjunct agreement in German. This is why we would like to leave open what exactly the predictions of these approaches are regarding condition 2. Conditions 3 and 4 are, for sure, predicted to be ungrammatical, since mismatches in person are usually taken to be ungrammatical in German; see Fuß (2018). The PI approach predicts that only conditions 1 and 2 are acceptable in German. While the PI approach predicts conditions 3 and 4 to be equally unacceptable, the GPI approach predicts a staircase-like decline in acceptability. With respect to the MSI approach, predictions are, again, not fully clear. The judgments concerning (7) and the remarks concerning number mismatches suggest that in the MSI approach the possibility and acceptability of conditions 3 and 4 is at least not ruled out. The LI approach, on the other hand, clearly predicts conditions 3 and 4 to be as acceptable as conditions 1 and 2 are. The predictions of the above-discussed approaches are summarized in Table 2.

Table 2: Predictions of the above-discussed approaches with respect to conditions 1–4

	C1: IDENTICAL	C2: SYNCRETIC	C3: ENDING	C4: ENDING+VOWEL
ATB–MD–C	✓	(*)	*	*
PI approach	✓	✓	*	*
GPI approach	✓	✓	✓	??
MSI approach	✓	✓	✓	✓
LI approach	✓	✓	✓	✓

To determine whether differences related to morphological deviations are specific to RNR, we need a reference point to compare RNR to. Gapping is ideal for this role. In verb-final structures like those we used to test the RNR mismatches, see (4), we can straightforwardly create analogous gapping examples like (9) where the gap occurs at the end of the second conjunct. In this case, gapping is superior to a full form with two overt verb forms because the full forms might be degraded for being redundant (an effect that we found in previous work, e.g., Schäfer et al. 2021). More importantly, there is a broad consensus in the literature that gapping tolerates mismatches and that all conditions presented in Table 1 are fully acceptable under gapping; see Eisenberg (1973), Klein (1993), Hartmann (2000), and many others. As a welcome side-effect of this investigation, these intuitions on gapping will be corroborated by empirical data.

(9) *Er glaubt, dass Bernd in Berlin arbeitet und Anne in London (arbeitet).*
he thinks that Bernd in Berlin works and Anne in London (works)
"He thinks that Bernd works in Berlin and Anne in London."

In the following sections, we will present empirical data on the production, perception, and processing of RNR mismatches. In Experiment 1, we begin by using a cloze study to investigate whether speakers of German accept RNR mismatches and whether the degree of morphological deviation has an influence on the ellipsis rate. In Experiments 2 and 3, we explore the perception of mismatches by collecting acceptability judgments for all four conditions, once with gapping, once with the full form as baseline. In Experiment 4, we focus on the most extreme contrast (between condition 1 and condition 4 in Table 1) in a third rating study. Finally, Experiment 5 is a reading time study in which we investigate the processing of RNR mismatches.

3 Experiment 1: production study RNR and gapping

As we saw above, RNR mismatches have been widely discussed in the theoretical literature, but the existing work has focused mostly on introspective judgments. We aim at expanding the sparse experimental work with a cloze study, which offers a controlled way of assessing how acceptable RNR mismatches are to speakers. By giving participants the option to either fill in the gap resulting from RNR or leave it blank, we can assess whether mismatches are treated as natural and retained or avoided in favor of full forms. In this way, Experiment 1 provides a production-oriented empirical perspective that is currently missing from the literature.

Our cloze study has the form of a 2 × 4 design, crossing EllipsisType and Identity. For EllipsisType, the baseline gapping is compared to RNR. Identity has the four levels sketched in Table 1: full identity of the two verb forms, syncretism, deviation in the inflectional ending, and deviation in the ending plus a change of the stem vowel. We measured the ellipsis rate, i.e., the proportion where participants left the gap empty relative to all (reducible) answers provided.

As sketched in Table 2, only the MSI and the LI approaches predict that all four levels of verbal mismatches under RNR are, in principle, fine. This means that under these approaches we expect no difference in the ellipsis rate across the Identity conditions. This is different for the PI approach, under which only phonologically matching conditions are predicted to be acceptable. The ending and ending+vowel conditions consequently should yield a lower ellipsis rate than the identical and syncretic conditions. The GPI approach predicts that the ellipsis rate varies as a function of the degree of morphological deviation: the more the forms differ, the lower the ellipsis rate. Concretely, based on Christ's (2011: 383) judgments for nominal RNR mismatches, the identical and syncretic conditions should essentially pattern with the ending condition, whereas a simultaneous deviation in ending and stem vowel should be degraded and

result in a lower ellipsis rate. However, in Christ's nominal example, adding an ending only represents a number mismatch, whereas in our condition, it is a mismatch in number *and* person. For this reason, the change in ending could represent a greater morphological deviation here, to be reflected in a reduced ellipsis rate compared to phonological identity. The ATB-MD-C approaches predict, without further assumptions, all mismatches to be degraded with the potential exception of syncretic forms. So at least the last two IDENTITY conditions, ending and ending+vowel, should yield lower ellipsis rates according to these approaches.

3.1 Materials

We constructed 24 items in the form of complex sentences, such as (10), that allow for RNR or gapping. They contain an embedded verb-final clause with two verb phrases coordinated by *und* "and." In the potential RNR conditions, we replaced the finite lexical verb at the end of the first conjunct with an empty text box (indicated with underscores in (10)), in the potential gapping conditions, it was the verb at the end of the second conjunct.[3] The subjects in the embedded clause were varied so that under lemma identity, the verb forms of the two conjuncts would either be identical (10a), syncretic (10b), differ in their ending (10c), or differ in ending and their stem vowel (10d).

(10) "For the presentation, it is agreed that…"

a. *Anna den Zeitschriftenartikel (___| liest) und Jan das Kapitel aus dem Buch (liest |___)*
Anna the journal.article read.3SG and Jan the chapter out the book read.3SG (identical)

b. *wir den Zeitschriftenartikel (___| lesen) und Anna und Jan das Kapitel aus dem Buch (lesen |___).*
we the journal.article read.1PL and Anna and Jan the chapter out the book read.3PL (syncretic)

c. *ich den Zeitschriftenartikel (___| lese) und Anna und Jan das Kapitel aus dem Buch (lesen |___).*
I the journal.article read.1SG and Anna and Jan the chapter out the book read.3PL (ending)

d. *ich den Zeitschriftenartikel (___| lese) und Jan das Kapitel aus dem Buch (liest |___).*
I the journal.article read.1SG and Jan the chapter out the book read.3SG (ending+vowel)

[3] To increase readability, we indicate the manipulation of ELLIPSISTYPE by placing the verb in parentheses in both conjuncts.

The items were combined with 72 fillers, all of which were coordination structures using *und* "and." Of these, 24 were from another study and included a context sentence and a target sentence featuring potential RNR structures with nominal antecedents and targets; inserting a word in these cases was optional, as with the critical items. The remaining 48 fillers were designed to present participants with all possible scenarios for filling the gap: optional filling, obligatory filling, and instances where no filling was possible. This should enhance the naturalness of the task and prevent participants from always filling in something or always leaving the gap blank. One third of these fillers contained gaps where insertion was optional, e.g., adding an adjective between article and noun. Another third included gaps that required completion, e.g., providing the object of a transitive verb. The final third comprised gaps where inserting an element would result in an ungrammatical structure, such as between a pronominalized subject and its corresponding finite verb in a verb-second clause. The last two categories also served as attention checks; see below. For each of the 16 fillers in a given category, half were accompanied by a context sentence like the 24 RNR fillers, while the other half did not include a context, aligning with the critical items.

3.2 Procedure

We recruited 241[4] native German speakers between the ages of 18 and 50 from the crowdsourcing platform Prolific (www.prolific.com) and paid them £3.69.[5] The experiment was implemented using the LimeSurvey survey presentation software (Limesurvey GmbH). The participants' task was to read the materials and decide whether to fill in a text box in the critical sentence. In the instructions, they were given examples for optional, obligatory, and impossible filling together with a brief explanation. Each experimental list started with three fillers, one of each category. The remaining fillers and the items were presented in pseudo-randomized order, ensuring that no two trials of the same type immediately followed each other. Each participant saw each token set only once and in one condition and each condition equally often. This way, we collected at least 30 answers per token set and condition.

3.3 Results

We used 16 fillers that required obligatory completion and 16 fillers where completion should not be possible as attention checks. Participants were categorized as inattentive if they did not complete at least 50% of the obligatory fillers, or if they completed 50% or more of the fillers

[4] Since one participant did not claim their reward on Prolific, we have data from 241 instead of the planned 240 participants.

[5] In all experiments reported in this paper, compensation corresponds to the German minimum wage at the time of the experiment, calculated based on the estimated time required to complete the experiment.

intended to remain unfilled due to ungrammaticality.[6] Based on these criteria, two participants who did not fill at least 50% of the mandatory fillers were excluded.

We manually reviewed the remaining 239 participants' answers and pre-processed the 5736 data points.[7] In a first step, we excluded any meaningless responses, e.g., English words like *first*, and responses that did not semantically or grammatically match the context, such as (11), where the singular verb does not agree with the plural subject.

(11) *Für den Kongress in Brüssel ist es praktisch, dass ich Französisch spreche*
for the congress in Brussels is it practical that I French speak
und Vanessa und Lena Niederländisch ***kann****.*
and Vanessa and Lena Dutch can.3SG

This resulted in a loss of 1.88% of the data. Grammatical correctness was required to ensure that participants accurately recognized the IDENTITY manipulation; entering a mismatching verb form could indicate a lack of attention to this aspect. Most agreement errors occurred when participants provided a verb form for a coordinated subject such as *Vanessa und Lena*. In several instances, apparently cursory reading resulted in only the second name, *Lena*, being identified as the subject, leading to the use of the 3rd singular instead of the 3rd person plural, as in (11).[8]

The remaining data were divided into three categories: ellipsis (60.18%), full form (20.43%), and other (19.39%). When the participant left the text box blank or did not enter a verb but only an adverb or a noun modifier,[9] this was formally classified as *ellipsis* (without implying that the gap should be analyzed as an elliptical reduction). A response was classified as *full form* when the participant entered the (correctly inflected) identical verb found in either the antecedent or

[6] Some participants provided meaningful responses even in these cases, which was not anticipated, such as placing lassen "to let" between a lexical and a modal verb (*den neuen Boden verlegen lassen können* "to be able to have the new floor laid"). These responses were not classified as ungrammatical or considered indicative of inattentiveness.

[7] Each data point was independently annotated by two student assistants and one of the authors. In instances where annotations differed, the final category was determined through joint discussion and agreement between both authors. We would like to thank Canan Borhofen and Sarah Galling for their help.

[8] It should be noted that some of these sentences, such as (i), can also yield a grammatical interpretation when the first conjunct's object is animate. Specifically, a part of the second conjunct's coordinated subject, e.g., *Daniel*, may be analyzed as belonging to the first conjunct's object, i.e., *ein paar Freunde und Daniel*, and then being extraposed. As a result, the second conjunct would consist solely of *Tobias einige Arbeitskollegen*, wherein *trifft* is an appropriate continuation. Given that it is uncertain whether participants interpreted these constructions in this manner or generated a mismatch analogous to (11) above, these sentences were excluded from analysis.

(i) *Heute Abend ist geplant, dass ich ein paar Freunde treffe und Daniel und Tobias einige*
today evening is planned that I a few friends meet and Daniel and Tobias some
Arbeitskollegen ***trifft****.*
work.colleagues meets

[9] For example, several participants specified the definite noun *Gewerkschaft* "trade union" by adding the proper name *Verdi*, which denotes the second-biggest German trade union.

postcedent. All other grammatically correct responses were categorized as *other*. This primarily involved alternative verbs, including synonyms such as *schauen* for *sehen* "to look." Additionally, participants occasionally employed the same verb lemma while varying tense or mood or embedding it under a modal verb (0.82% of the data). We did not consider them *full forms* because they cannot be readily omitted without losing this additional meaning aspect.[10]

Figure 1 shows the distribution of the three answer categories for each condition. Both gapping and RNR have the highest ellipsis rates in the identical condition. For gapping, the other three categories each show an ellipsis rate 8 to 10% lower, with no obvious difference between them. In RNR, the ellipsis rates seem to decrease stepwise from syncretic to ending to ending+vowel, while the full form rates increase accordingly. In the ending+vowel condition of RNR, participants produced slightly more full forms than ellipses. The rates of other answers are relatively stable across conditions from 14% to 24%. We therefore focused only on the ellipses and full forms for the following analysis and this way ensure that our reference quantity for ellipsis consists only of reducible sentences.

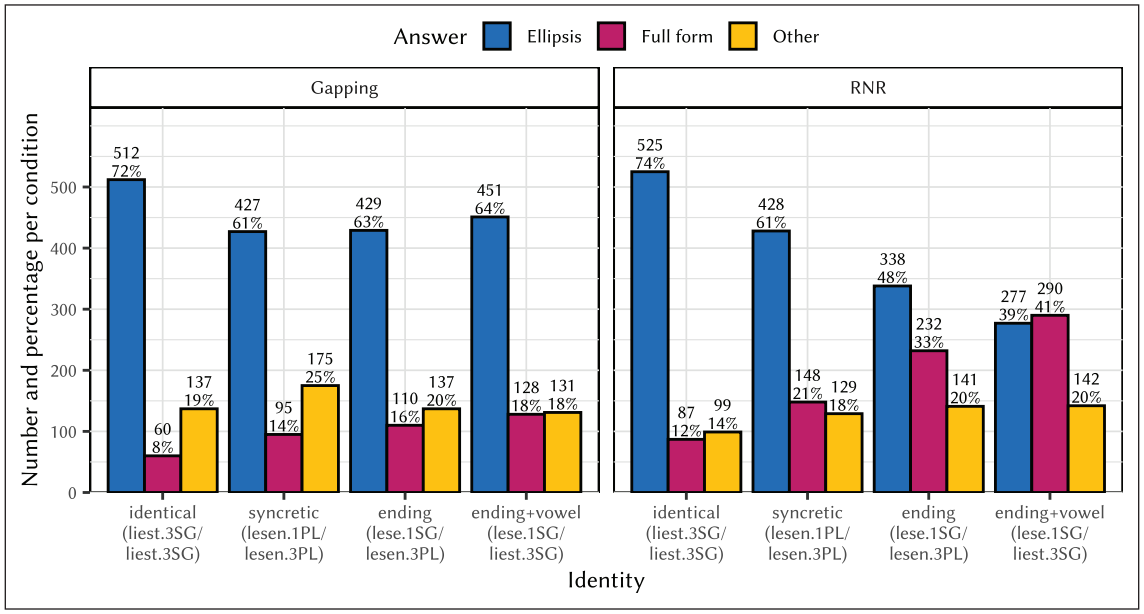


Figure 1: Absolute numbers and percentages of produced answer types per condition in Experiment 1

[10] Cases like (i) are interesting, though, since the participant replaced a potential gapping structure by an RNR construction with an elided modal verb in the first conjunct.

(i) [...] *dass wir Ski fahren und Jana und Niklas Snowboard* ***fahren wollen***.
that we ski drive and Jana and Niklas snowboard drive want

We used a logistic mixed regression model (package lme4, Bates et al. 2015) in R (R Core Team 2025). The model predicted the binary Answer (ellipsis vs. full form) with ellipsis as reference level from the dummy-coded EllipsisType (gapping as baseline with 0, RNR as 1) and from Identity. Identity was forward-coded to allow for finding a potential gradation effect: the first contrast identical.rest compared the identical condition (coded as 0.75) to the other conditions (each coded as −0.25), the contrast maphon.mmphon compared the phonologically matching conditions (identical and syncretic, coded as 0.5) to the phonologically mismatching conditions (ending and ending+vowel, coded as −0.5), and the contrast rest.vowel compared the identical, syncretic and ending conditions (each coded as 0.25) to the ending+vowel condition (coded as −0.75), i.e., to the condition with the highest deviation between antecedent and target. We also included the *z*-transformed Position of the trial in the experiment to control for potential habituation effects and considered the possible interactions between EllipsisType and Identity to account for RNR-specific effects. The model contained random intercepts for participants and items and by-participant and by-item random slopes for EllipsisType.[11] More complex models resulted in singular fits. Table 3 shows the fixed effects of the model.

Table 3: Fixed effects in the logistic mixed regression model of Experiment 1

Fixed effect	Est.	SE	*z*	*p*	
Intercept	−2.79	0.27	−10.30	< 0.001	***
EllipsisType	1.61	0.23	6.93	< 0.001	***
identical.rest	−1.17	0.25	−4.67	< 0.001	***
maphon.mmphon	−0.42	0.22	−1.92	0.055	.
rest.vowel	−0.05	0.20	−0.23	0.81	
Position	−0.30	0.05	−5.93	< 0.001	***
EllipsisType: identical.rest	−0.004	0.32	−0.01	0.99	
EllipsisType: maphon.mmphon	−0.71	0.28	−2.55	< 0.05	*
EllipsisType: rest.vowel	−0.69	0.26	−2.68	< 0.01	**

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

The simple effects of the Identity contrasts indicate how the reference level of EllipsisType, i.e., gapping, behaves with respect to the Identity manipulation: the model contains a significant simple effect of identical.rest (z = -4.67, p < 0.001). It indicates that for gapping, participants produced significantly more ellipses in the identical condition than in all three mis-

[11] The formula was Answer ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 + EllipsisType | Subject) + (1 + EllipsisType | Item).

match conditions. The effect of MAPHON.MMPHON is marginal (z = −1.92, p = 0.055), suggesting tentatively that the phonologically matching conditions lead to a higher ellipsis rate for gapping than the mismatches. Based on the visual inspection of Figure 1, this tendency seems to be driven by the identical condition. REST.VOWEL has no significant effect, i.e., there is no difference for gapping between the ellipsis rates of the ending+vowel condition compared to the rest.

The interaction terms between the IDENTITY contrasts and ELLIPSISTYPE indicate the deviation of RNR from gapping with respect to the IDENTITY manipulation. While there is no significant difference between gapping and RNR when comparing the identical condition to the rest (z = 0.01, p = 0.99), the ellipsis rate becomes significantly lower for RNR of a verb form with a different ending and a different ending plus a vowel change as compared to gapping (z = 2.55, p < 0.05) and again even lower when comparing the most extreme morphological deviation (ending+vowel) to the rest (z =2.68, p < 0.01).

Due to the forward coding of the predictor IDENTITY, the significant effect of ELLIPSISTYPE cannot be straightforwardly interpreted. While it suggests a statistically stronger preference for ellipsis across all gapping conditions compared to RNR, we refrain from concluding a general preference for ellipsis under gapping because the ellipsis rates in the identical condition are the same for gapping and RNR, and the effect seems to be mainly driven by the interaction between ELLIPSISTYPE and IDENTITY. The significant main effect of POSITION (z = -5.93, p < 0.001) shows that participants provided fewer ellipses as the experiment progressed.

As a final point of our analysis, we would like to mention the notable differences in the ratio of ellipses and other verbs between token sets and participants. Four token sets had a ratio of under 4% of other verbs, while four further reached up to 40% of other verbs. The latter are mainly token sets where participants seemed to dislike the given verb, e.g., replacing *sehen* "to see" with the synonym *schauen*. This suggests a preference to correct the verb over avoiding redundancy by producing ellipsis. There were six participants who consistently used ellipses, whereas another six always used the full form. No participant only filled in other verbs. While the strategy of only providing full forms could be a result of the participants wanting to fulfil the task "correctly", only producing ellipses might indicate that these participants are most tolerant to morphological mismatches, i.e. for them, LI seems to suffice. The statistical results remain comparable when we disregard the participants without variation.

3.4 Discussion

The purpose of our first experiment, a cloze task, was to find out if speakers accept and maintain verbal RNR mismatches or if they correct them by producing the corresponding full form. We compared the ellipsis rates of (potential) RNR structures to those of (potential) gapping structures, which served as our baseline.

The results indicate first that mismatches under gapping and RNR are frequently produced or, more precisely, not avoided by converting the elliptical utterance to a full form. Gapping exhibits an ellipsis rate of over 60% in all four IDENTITY conditions, and even RNR still shows almost 40% ellipsis in the condition with the lowest rate. For our baseline gapping, this result confirms the prevailing research opinion that gapping generally tolerates mismatches in person and number. For RNR, the result challenges the ATB-MD-C approaches, which predict that mismatches are ungrammatical and therefore not produced by speakers.

The second result is that both gapping and RNR exhibit higher ellipsis rates when there is complete identity between correlate and target compared to mismatch conditions. There is a significant mismatch penalty even in the syncretic condition and even for the baseline gapping, which, according to the literature, should tolerate any form of mismatch in person and number. The data show that while speakers produce a lot of mismatches, they display sensitivity to morphosyntactic deviations in this type of task, though they do not distinguish between different degrees of deviation in gapping.

The picture is different for RNR, where a staircase-like pattern emerges. As the morphological difference between target and correlate increases, speakers are more likely to produce the corresponding verb rather than omit it. This pattern is incompatible with the PI approach, which predicts the syncretic condition to behave like the identical condition, and with the MSI and the LI approaches, which predict all mismatches to be equally possible. At first glance, the staircase-like pattern seems to correspond to the predictions of the GPI approach: a gradual decrease in acceptability relative to the degree of morphological deviation. There are, however, two important qualifications: first, Christ (2011) considered syncretic forms and forms only differing in one aspect (like the adding of the nominal plural affix *-e*) to be acceptable as well, whereas our data already show a reduced ellipsis rate for these conditions. Second, Christ marks forms such as our ending+vowel condition with two question marks, i.e., as highly questionable. Nevertheless, the participants accepted these mismatches in almost 40% of the cases in the cloze task and did not change them to the full form. The GPI approach, if we base it literally on Christ's (2011) judgments on nominal RNR mismatches, is thus too restrictive on the one hand and too permissive on the other to explain our data in its entirety.

While the present experiment used a cloze task rather than genuine production, which may have increased participants' sensitivity to inconsistencies, and while the observed differences between items and participants suggest that factors such as verb semantics and the development of individual strategies may also have contributed to the result, we nevertheless find the clear staircase-like pattern for the ellipsis rate of RNR. In the following three experiments, we investigate whether the varying ellipsis rates are also reflected in differences in acceptability. We start by testing the complete pattern exploratorily with four IDENTITY levels, first with gapping as a baseline, then with full forms. Finally, we zoom in and focus on the strongest contrast with more data points per condition.

4 Experiment 2: gradual mismatches in RNR (and gapping)

In Experiment 2, we tested the perception of RNR mismatches to see whether it patterns with the cloze data: do speakers not only produce fewer RNR mismatches when they differ more strongly in their morphological properties, but do they also perceive them as degraded? To answer this question, we conducted an acceptability rating study in the form of a 2 × 4 design that crossed EllipsisType (gapping vs. RNR) with all four Identity conditions (identical vs. syncretic vs. ending vs. ending+vowel; see Table 1).

Experiments 2 and 3 were designed as an exploratory first step to see whether mismatches are degraded at all, and if so, to what degree, and therefore rely on a relatively small sample size given the eight experimental conditions. Although this limits the robustness and statistical power of the analyses, both experiments provide an initial overview of how speakers respond across the full set of conditions. These preliminary patterns help identify trends that warrant further investigation in the reduced Experiment 4. The items of Experiment 2 were included as fillers in another study on topic drop and therefore transformed to instant messaging dialogues (see below).[12]

All approaches discussed in Section 2 consider RNR to be acceptable if the elided form is morpho-phonologically identical to its correlate. With the possible exception of the ATB-MD-C approaches, all approaches predict syncretic forms not to be degraded either. According to the LI and the MSI approaches, this extends to all (considered) morphological mismatches in person and number. This means that we should see no difference between the Identity conditions of RNR as compared to gapping. In contrast, the PI account predicts that all phonological mismatches are degraded. Consequently, there should be a difference between the identical and the syncretic conditions of RNR on the one hand and the ending and the ending+vowel conditions on the other hand. The GPI approach derived from Christ (2011) takes it that one deviation in phonological form is still in the range of acceptable sentences; two deviations in phonological form, however, are predicted to be substantially degraded. A result in line with the GPI predictions would also best correspond to the production pattern of Experiment 1.

4.1 Materials

We used a variant of the 24 items of Experiment 1 where some token sets varied in their wording or were exchanged for reasons of greater naturalness in the different communication form, namely short instant messaging dialogues between two persons. Person A sets the discourse

[12] One reviewer suggested that the messaging format may involve greater tolerance for errors, potentially affecting judgments of RNR mismatches. We cannot fully rule out this possibility, but while certain orthographic slips are indeed more acceptable in text messages, there is no evidence that the grammatical constraints relevant to RNR are relaxed in this form of communication. Moreover, Experiments 3 and 4, which were not in the messaging format, had similar (null) results.

topic and then asks a question (12) to which person B replies (13). B's answer is always an embedded verb-final clause with a coordination using *und* "and" that either contains gapping or RNR. In (13a) the verb forms are identical, in (13b) the forms are syncretic, in (13c) only the ending differs, and in (13d) there is additionally a vowel change.

(12) A: "Again about the presentation preparation. Have you already divided the work with the others, who prepares which texts?"

(13) B: *Noch nicht ganz, aber klar ist, dass…*
B: "Not entirely yet, but it is clear that..."

a. B: … *Jan den Zeitschriftenartikel* (*liest*) *und Sara das Kapitel aus dem Buch* (*liest*). (identical)
Jan the journal.article read.3SG and Sara the chapter from the book read.3SG

b. B: … *wir den Zeitschriftenartikel* (*lesen*) *und Jan und Sara das Kapitel aus dem Buch* (*lesen*). (syncretic)
we the journal.article read.1PL and Jan and Sara the chapter from the book read.3PL

c. B: … *ich den Zeitschriftenartikel* (*lese*) *und Jan und Sara das Kapitel aus dem Buch* (*lesen*). (ending)
I the journal.article read.1SG and Jan and Sara the chapter out the book read.3PL

d. B: … *ich den Zeitschriftenartikel* (*lese*) *und Jan das Kapitel aus dem Buch* (*liest*). (ending+vowel)
I the journal.article read.1SG and Jan the chapter out the book read.3SG

The 24 items were mixed with 84 fillers. Among these, four were variants of the items with the copula verb *sein* "to be," which exhibits suppletive forms (e.g., *bin* "am" vs. *ist* "is"). We included them for exploratory purposes and will not discuss them further. In addition, 24 fillers were items from a study on topic drop. Another seven fillers were ungrammatical catch trials with two finite verbs in a single clause, serving as attention checks. The remaining 49 fillers covered a wide range of acceptability, from clearly acceptable to highly marked sentences.

4.2 Procedure

Forty-eight native German speakers aged 18 to 50 participated in the experiment. They were recruited from the crowdsourcing platform Clickworker (www.clickworker.de) and were paid €2.50. We used the LimeSurvey survey presentation software (Limesurvey GmbH) to host the experiment. The participants rated the naturalness of the target utterance on a 7-point Likert

scale (1 was *vollkommen unnatürlich* "completely unnatural" and 7 *vollkommen natürlich* "completely natural"). Following a Latin square design, the items were distributed across eight lists. Each participant rated each critical item once and in only one condition. The items and the fillers were displayed as lowercase instant messages[13] and presented in pseudo-randomized order to prevent items or fillers of the same type from appearing consecutively.

4.3 Results

We excluded six participants from the analysis because they had rated four or more of the ungrammatical control utterances with a 6 or 7 on the Likert scale, indicating potential inattentiveness. We had to discard six further data points because the ending+vowel condition of one critical item contained a typo. Figure 2 shows the mean ratings and standard errors. All conditions, including the clear mismatches, received relatively high ratings near 5 and above on the 7-point scale. The identical condition was rated higher for both gapping and RNR than the other conditions, which look very similar. For those other conditions, syncretic, ending and ending+vowel, there seems to be a slight degradation of RNR compared to gapping.

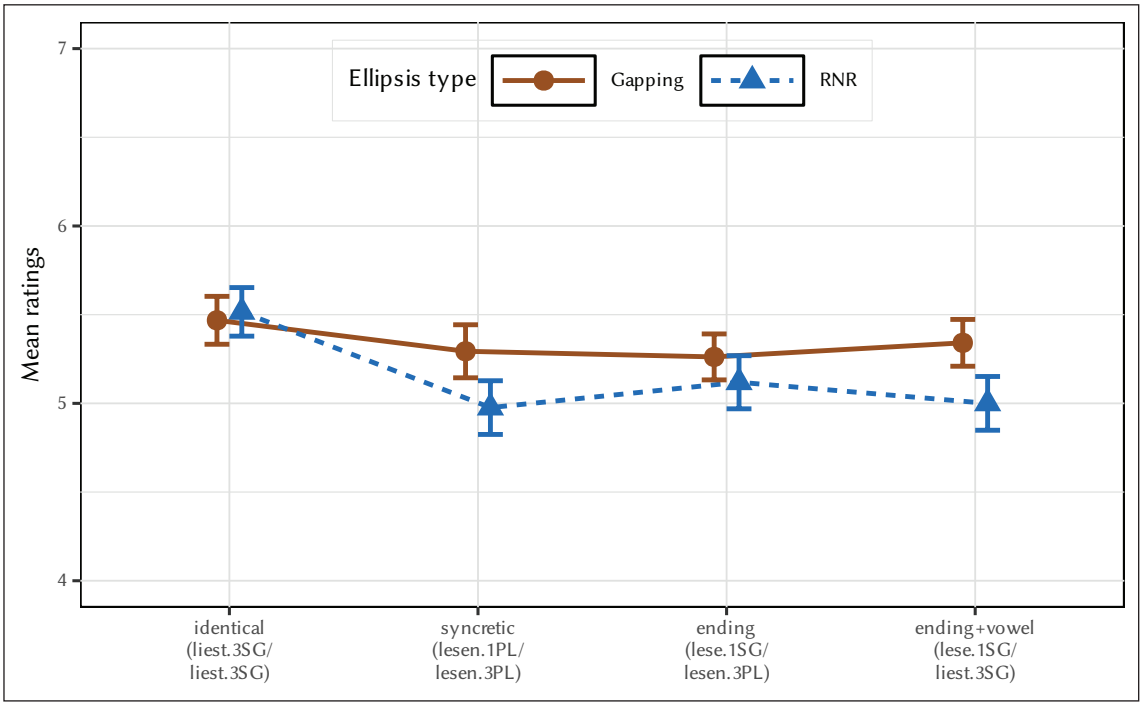


Figure 2: Mean ratings and standard errors per condition for Experiment 2

[13] This ensured that the verb was written consistently in all topic drop conditions.

We analyzed the data with cumulative link mixed models (CLMMs) from the package *ordinal* (Christensen 2023) and coded the independent variables like in Experiment 1. The acceptability scores were modeled as a function of the dummy-coded binary predictor EllipsisType (the baseline gapping coded as 0, RNR as 1) and the forward-coded predictor Identity: identical.rest compared the identical condition (coded as 0.75) to the other conditions (coded as −0.25), maphon.mmphon compared the phonologically matching conditions (identical and syncretic, coded as 0.5) to the phonologically mismatching conditions (ending and ending+vowel, coded as -0.5) and rest.vowel compared the identical, syncretic, and ending conditions (coded as 0.25) to the ending+vowel condition (coded as -0.75). We also included the *z*-transformed Position of the trial in the experiment to control for potential habituation effects and considered the possible interactions between EllipsisType and Identity to account for RNR-specific effects. The model contained random intercepts for participants and items, by-participant and by-item random slopes for the three Identity contrasts and their interactions with EllipsisType, and a by-item random slope for EllipsisType.[14] Table 4 shows the fixed effects of the model with flexible thresholds.

Table 4: Fixed effects in the CLMM of Experiment 2

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.26	0.15	−1.78	0.07	.
identical.rest	0.25	0.27	0.92	0.36	
maphon.mmphon	0.15	0.26	0.59	0.56	
rest.vowel	−0.12	0.26	−0.48	0.63	
Position	−0.06	0.07	−0.83	0.41	
EllipsisType:identical.rest	0.78	0.51	1.53	0.13	
EllipsisType:maphon.mmphon	−0.48	0.49	−0.98	0.33	
EllipsisType:rest.vowel	0.43	0.42	1.03	0.30	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

We found a marginal main effect of EllipsisType ($z = -0.83$, $p = 0.07$) suggesting a tentative preference for structures with gapping compared to those with RNR. Neither the other main effects nor the interactions were significant. There was no mismatch penalty for the baseline gapping, and the non-significant interaction suggests that RNR did not deviate from this baseline.

[14] The formula was Rating ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 EllipsisType : (identical.rest + maphon.mmphon + rest.vowel) | Subject) + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Item). When including the by-participant random slope for EllipsisType, the variance-co-variance matrix of the model was undefined. We excluded this slope because it explained the least variance of all random effects.

4.4 Discussion

In Experiment 2, we tested the acceptability of RNR (mis)matches compared to gapping (mis) matches as a baseline. As was to be expected, gapping and RNR pattern alike in the identical condition, which requires identity in morphological features and phonological form. Basically, the participants seem to find matching gapping and matching RNR equally good.

In terms of mismatches, there is a slight drop in ratings from a purely visual perspective, but this concerns both gapping and RNR, with the drop being a bit larger for RNR. This result is in line with Experiment 1, where the ellipsis rate decreased for all mismatches under gapping and RNR, but it is not predicted by any of the RNR-specific identity approaches. However, the rating differences observed are very small and do not result in statistically significant effects. With a larger sample size and more power, it is conceivable that the difference between identical and the mismatches and the differences between gapping and RNR in the identical and the ending+vowel condition might become significant, but most likely not the difference for ending, for which there is not even a descriptive tendency. Moreover, the mismatches were rated very high, with around 5 points on a 7-point scale and therefore can hardly be classified as ungrammatical.

The result of Experiment 2 is inconsistent with the ATB-MD-C approaches, which predict a general mismatch penalty. Moreover, also in RNR, just like in gapping, the mismatch conditions syncretic, ending, and ending+vowel essentially pattern alike. This is consistent with both the MSI and the LI approach, but it is evidence against the PI approach, which predicts a difference between identical and syncretic on the one hand and the two ending conditions on the other, and the GPI approach, which predicts a staircase-like pattern.

In summary, this initial exploratory rating study suggests that RNR behaves predominantly in parallel with gapping, and that RNR mismatches are acceptable and cannot be considered ungrammatical. The data pattern is compatible with the MSI and LI approaches but not with the GPI approach, which best describes the results of Experiment 1. We will return to this point and the possible differences between cloze and rating study in the general discussion. First, however, we look at all four conditions again in the equally exploratory Experiment 3, but this time we compared RNR with the full forms.

5 Experiment 3: comparing RNR structures with full forms

In Experiment 2, we used gapping as a baseline for RNR. In the likewise exploratory Experiment 3, we compared RNR to the full forms of the RNR items. This way, we wanted to ensure that the results of Experiment 2 persist in the presence of a non-elliptical baseline, i.e., we wanted to rule out that they were driven by any ellipsis-specific effects that potentially apply to both gapping and RNR.

The predictions are basically the same as for Experiment 2: the full forms should be fully acceptable, just like the RNR identical condition. Depending on the theoretical approach, the RNR mismatches should not be acceptable at all (ATB-MD-C approaches), only acceptable if the phonological form is the same (PI approach) or does not deviate too much from the correlate (GPI approach), or all be acceptable (MSI and LI approaches), which would also be in line with the result of Experiment 2.

5.1 Materials

In anticipation of the reading time experiment that was planned to test whether mismatches come with an increased cognitive effort, see Experiment 5, we made an adjustment to our materials. Measuring reading times for RNR structures requires that the area of interest, i.e., the correlate in the last conjunct, is kept constant, while the target is varied. Therefore, we consistently used a 3rd person plural verb form in the second conjunct and changed the person/number features of the overt or covert verb in the first conjunct. This modification of the IDENTITY variable is sketched in Table 5 and illustrated in (14).

Table 5: Modified IDENTITY manipulation in Experiment 3

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(sie) lesen*	*(sie) lesen*	–	–	–	–	identical
2	*(wir) lesen*	*(sie) lesen*	+	–	–	–	syncretic
3	*(ich) lese*	*(sie) lesen*	+	+	+	–	ending
4	*(er) liest*	*(sie) lesen*	+	+	+	+	ending+vowel

(14) A: "What is the status of the presentation?"
B: *Es ist abgemacht, dass...*
B: "It is agreed that..."

a. *die Mia und der Ben den Zeitschriftenartikel (lesen) und der*
the Mia and the Ben the journal.article read.3PL and the.NOM
Jonas und die Maike das Kapitel aus dem Buch (lesen) (identical)
Jonas and *the* Maike the chapter out the book read.3PL

b. *wir den Zeitschriftenartikel (lesen) und der Jonas und die Maike das*
we the journal.article read.1PL and the.NOM Jonas and the Maike the
Kapitel aus dem Buch (lesen) (syncretic)
chapter out the book read.3PL

c. *ich den Zeitschriftenartikel* (*lese*) *und der Jonas und die Maike*
I the journal.article read.1SG and the.NOM Jonas and the Maike
das Kapitel aus dem Buch (*lesen*) (ending)
the chapter out the book read.3PL

d. *die Mia den Zeitschriftenartikel* (*liest*) *und der Jonas und die*
the Mia the journal.article read.3SG and the.NOM Jonas and the
Maike das Kapitel aus dem Buch (*lesen*) (ending+vowel)
Maike the chapter out the book read.3PL

We reformulated or exchanged some token sets to improve the overall naturalness of the materials. Additionally, definite articles were inserted before proper names to clearly mark the case of the subject of the second conjunct, which was always a male name or a conjunction of a male and a female name, as nominative.

Along with the 24 items, we tested 68 fillers. Thirty-two of them were items from a study on verbal complexes with three verbs, where the focus constituent was marked by capitalization. Sixteen further fillers were pairs of *wh*-questions with prosodically (again indicated by capitalization) matching or mismatching answers. Twelve fillers contained focused direct objects in the target clause (either article or noun capitalized), and eight fillers were catch trials with severe case mismatches used as attention checks.

5.2 Procedure

Forty-eight native German speakers between the ages of 18 and 40 from Prolific (www.prolific.com) participated in the experiment and were paid £4.43. The experiment was implemented in PCIbex (Zehr & Schwarz 2023). Again, the participants rated the naturalness of the target utterances with a 7-point Likert scale (7 = "completely natural"). Because of the fillers' focus manipulation, the participants were furthermore instructed to read the dialogues out loud before rating them. Like in Experiment 2, the items were distributed across eight lists according to a Latin square design and presented in a pseudo-randomized order together with the fillers.

5.3 Results

Since no participant rated more than half of the catch trials with 5 or better, there were no exclusions. Figure 3 shows the mean ratings and standard errors for all eight conditions. Visually, the full forms and RNRs show a very homogeneous pattern, with all conditions being equally acceptable except for the full forms in the identical conditions, which seem slightly degraded.

Like in Experiment 2, we fitted a CLMM and modeled the ratings as a function of the dummy-coded predictor Form (full form as 0, RNR as 1), the predictor Identity, and the numerical *z*-transformed Position. Identity was again forward-coded: identical.rest compared identical (coded as 0.75) to the rest (−0.25), maphon.mmphon compared identical and syncretic (0.5) to ending and ending+vowel (-0.5) and rest.vowel compared identical, syncretic, and ending (−0.75) to ending+vowel (0.25). We also included all interactions of these contrasts with Form in our model, just like random intercepts for participants and items and by-subject and by-item random slopes for Form, the three Identity contrasts and their interactions.[15] Table 6 shows the fixed effects of the model with flexible thresholds. It contained only a significant main effect of Position ($z = 2.38$, $p < 0.05$), according to which participants gave better ratings as the experiment unfolded.

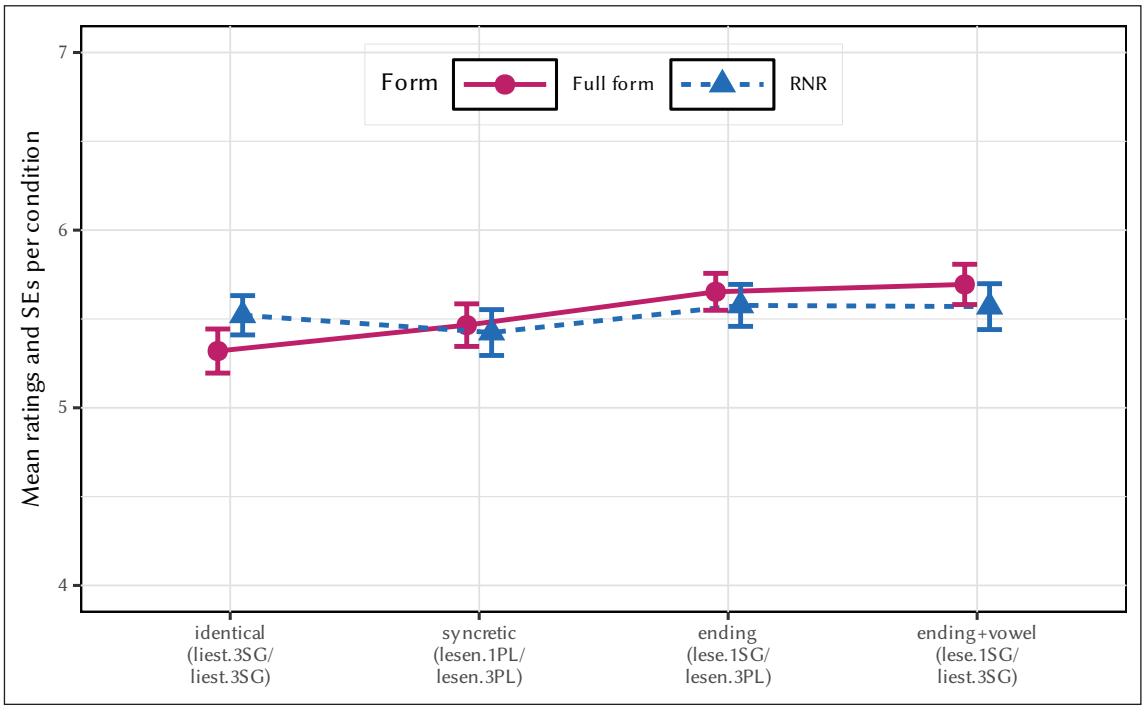


Figure 3: Mean ratings and standard errors for Experiment 3

[15] The formula was Rating ~ EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) + Position + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Subject) + (1 EllipsisType * (identical.rest + maphon.mmphon + rest.vowel) | Item).

Table 6: Fixed effects in the CLMM of Experiment 3

Fixed effect	Est.	SE	*z*	*p*	
Form	0.20	0.23	0.91	0.36	
identical.rest	−0.25	0.27	−0.93	0.35	
maphon.mmphon	−0.31	0.30	−1.01	0.31	
rest.vowel	−0.30	0.26	−1.15	0.25	
Position	0.16	0.07	2.37	0.02	*
Form:identical.rest	0.28	0.38	0.75	0.45	
Form:maphon.mmphon	0.05	0.43	0.12	0.90	
Form:rest.vowel	0.04	0.37	0.10	0.92	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

5.4 Discussion

Experiment 3 suggests that the conditions syncretic, ending, and ending+vowel pattern parallel to their full forms, and thereby provides additional support that all mismatch conditions are rated within the realm of full grammaticality. Moreover, all RNR conditions are essentially rated the same with no significant or clearly visible differences. This result is similar to the result of Experiment 2 and in line with the MSI and the LI approaches, while it runs against the PI and GPI approaches.

While Experiments 2 and 3 must be treated with caution, as they were probably underpowered with 48 subjects and 8 conditions, they consistently indicate that RNR mismatches are considered acceptable and behave similarly to the corresponding baselines gapping and full forms. In two follow-up experiments, we aimed to confirm this result with a decreased number of conditions and, thus, more data points per condition.

6 Experiments 4a and 4b: testing the ends of the scale

To address the potential power problem of Experiments 2 and 3, we reduced the number of conditions in Experiments 4a and 4b. We again used gapping as a more natural baseline and focused on the contrast that the GPI approach predicts to be most extreme, i.e., between the identical condition and the ending+vowel condition, in which the ending and stem vowel change between target and correlate. This gives us a 2 × 2 design with the two binary factors EllipsisType (gapping vs. RNR) and Identity (identical vs. ending+vowel; see Table 7).

Table 7: Design of Experiments 4a and 4b

			mismatch in		mismatch in		
	TARGET	CORRELATE	PERSON	NUMBER	ENDING	VOWEL	CONDITION
1	*(sie) lesen*	*(sie) lesen*	–	–	–	–	identical
2	*(er) liest*	*(sie) lesen*	+	+	+	+	ending+vowel

Given the results of the two previous rating studies, we expect no difference between the identical and the mismatch condition, at least no RNR-specific difference. This would be in line with the LI and the MSI approaches, which predict all (considered) morphological mismatches to be acceptable.

In contrast, the PI and the GPI accounts predict that all phonological mismatches are degraded. Consequently, there should be a difference between the identical condition of RNR on the one hand and the most extreme ending+vowel condition on the other hand. Such an interaction between Identity and EllipsisType would also be tentatively in line with the production data from Experiment 1, where we saw the staircase-like pattern of the ellipsis rate only for RNR and not for gapping. However, if the rating data patterned with the production results, we would additionally expect a (weak) main effect of Identity that shows a general mismatch penalty for both gapping and RNR in line with the reduced ellipsis rates in the cloze study for all mismatches.

We tested two variants of Experiment 4, Experiments 4a and 4b, which only differed in their fillers, which were the experimental stimuli of two studies on sluicing. Since this difference is not relevant for our research on RNR mismatches, we report both experiments in parallel below.

6.1 Materials

We tested basically the same items as in Experiment 3 in the reduced form. We simplified their structure from a dialogue to a complex sentence and removed the definite article again for greater naturalness (15).

(15) *Für das Referat ist abgemacht, dass...*
"For the presentation, it is agreed that..."

a. *Jan und Sara den Zeitschriftenartikel* (*lesen*) *und Julia und Moritz*
Jan and Sara the journal.article read.3PL and Julia and Moritz
das Kapitel aus dem Buch (*lesen*). (identical)
the chapter out the book read.3PL

b. *Jan den Zeitschriftenartikel* (*liest*) *und Julia und Moritz das Kapitel*
Jan the journal.article read.3SG and Julia and Moritz the chapter
aus dem Buch (*lesen*). (ending+vowel)
out the book read.3PL

In both experiments, 72 fillers were included: 24 instances of sluicing, in Experiment 4a with the remnant *wer* "who," in Experiment 4b with *welcher* "which," and 48 complex sentences, with half containing verb-final complement clauses and the other half causal subordinate clauses.

6.2 Procedure

We recruited further native German speakers between 18 and 40. In Experiment 4a, 47 participants[16] from Clickworker (www.clickworker.de) received €4.14 for their participation. In Experiment 4b, 72 subjects from Prolific took part and were paid £2.61. The studies were implemented in PCIbex (Zehr & Schwarz 2023). The participants' task was again to rate the naturalness of given sentences on a 7-point Likert scale (7 = "completely natural"). Additionally, they had to answer a total of eight yes–no comprehension questions asked after fillers. We used these questions as attention checks. The distribution of materials to lists and pseudorandomization was analogous to the previous experiments.

6.3 Results

In Experiment 4a, two participants were excluded because their accuracy on the comprehension questions was below 75%. In Experiment 4b, there were no exclusions. Figure 4 shows the means and standard errors for both experiments next to each other. In both experiments, gapping and RNR received almost identical ratings. From visual inspection, there seems to be a slight degradation of the mismatch condition in Experiment 4a of less than 0.25 points, while in Experiment 4b the two IDENTITY conditions are rated equally acceptable.

We analyzed both experiments again with CLMMs (Christensen 2023) that modeled the ratings as a function of ELLIPSISTYPE and IDENTITY, their interaction, and the *z*-transformed POSITION in the experiment. Both ELLIPSISTYPE and IDENTITY were dummy-coded (gapping and identical as baselines coded as 0, RNR and ending+vowel as 1). The random effect structure consisted of random intercepts for participants and items and of by-item and by-subject random slopes for ELLIPSISTYPE, IDENTITY, and their interaction.[17] The fixed effects of the model for Experiment 4a with flexible thresholds are shown in Table 8, the ones for the model of Experiment 4b with

[16] Due to a platform error, data were collected from 47 rather than the intended 48 participants.
[17] The formula was RATING ~ ELLIPSISTYPE * IDENTITY + POSITION + (1 + ELLIPSISTYPE * IDENTITY | SUBJECT) + (1 + ELLIPSISTYPE * IDENTITY | ITEM).

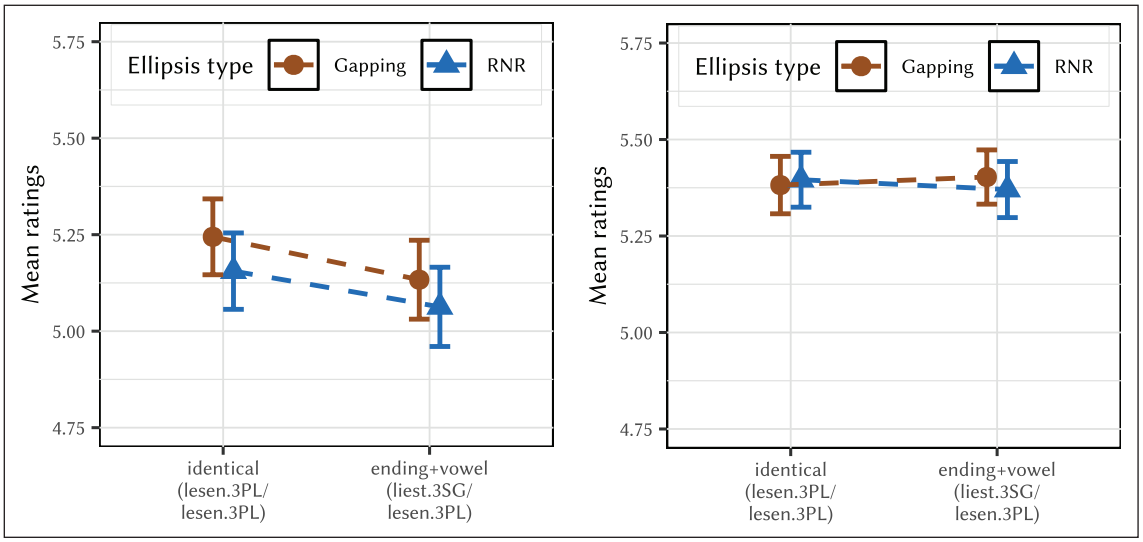


Figure 4: Mean ratings and standard errors for Experiments 4a and 4b

flexible thresholds are in Table 9. In both models, there is only a significant main effect of Position ($z_a = 4.91$, $p_a < 0.001$; $z_b = 3.83$, $p_b < 0.001$), indicating a habituation effect. There is no significant difference between the match (identical) and the mismatch (ending+vowel) condition, either in general or specifically for RNR.

Table 8: Fixed effects in the CLMM of Experiment 4a

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.12	0.20	−0.62	0.54	
Identity	−0.14	0.20	−0.69	0.49	
Position	0.30	0.06	4.91	< 0.001	***
EllipsisType:Identity	0.02	0.27	0.08	0.94	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

Table 9: Fixed effects in the CLMM of Experiment 4b

Fixed effect	Est.	SE	*z*	*p*	
EllipsisType	−0.08	0.16	−0.44	0.63	
Identity	−0.08	0.19	−0.49	0.66	
Position	0.19	0.05	3.83	< 0.001	***
EllipsisType:Identity	0.07	0.22	0.32	0.75	

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

6.4 Discussion

In both Experiments 4a and 4b, we found no difference in acceptability between an RNR match and a clear mismatch with a change in morphological ending and in stem vowel. RNR behaves completely parallel to gapping, where no differences can be seen either. The two Experiments 4a and 4b thus confirm the picture that already emerged in the exploratory Experiments 2 and 3, this time with more data points per condition: mismatches in person and number under verbal RNR are not particularly degraded in comparison to gapping or non-elliptical controls or not degraded at all. Therefore, we conclude that they are acceptable or at least not ungrammatical. Concerning RNR's identity conditions, this contradicts all approaches except for MSI and LI.

However, these approaches do not account for the pattern that we found in the cloze study in Experiment 1 and that is best compatible with (a modified) GPI account. The most extreme production effect, i.e. the bottom step of the staircase-like pattern, was not at all reflected in degraded acceptability in Experiments 4a and 4b. It might be that the task of rating, i.e. assessing acceptability, differs from that of the cloze task, in which participants have the opportunity not only to implicitly evaluate the presented sentences, but also to correct them so that they may become even more acceptable. Another possible explanation for the divergence between the cloze task and the acceptability judgment studies might be processing. The acceptability judgments occur after processing has concluded, so these ratings may not indicate issues encountered during parsing. Such issues, however, could lead speakers to spontaneously specify a full form during the cloze task. To determine whether the discrepancy between cloze and rating data is attributable to challenges in parsing, we conducted a self-paced reading study.

7 Experiment 5: self-paced reading study, RNR vs. full form

Experiment 5 is a self-paced reading study dedicated to investigating whether there is evidence for more complex processing of a mismatching verb form under RNR. Using a 2 × 2 design, the study compared the processing of RNR with that of the corresponding full form (factor Form) when the verbs were either identical or varied in their ending and vowel (factor Identity).

If the findings from the reading time study correspond to those of the cloze study, we anticipate observing processing difficulties in the mismatching ending+vowel condition. Specifically, longer reading times are expected at or after the point where participants detect the mismatch, namely, on the verb in the second conjunct or in the subsequent (spillover) material. Such processing difficulties may explain why more participants avoided ellipsis in the mismatching ending+vowel condition compared to the identical condition in the cloze task.

In our discussion of Experiment 4, we noted that the rating results, which showed no RNR-specific mismatch penalty, may not capture initial processing difficulties, as ratings are provided after parsing is complete. Alternatively, of course, these results might just indicate that participants did not experience such difficulties at all, rather than that they successfully resolved or "repaired" them. If this is in fact the case, we would expect no reading time differences for RNR between the identical and the ending+vowel conditions.

With respect to the full forms, there are in principle two possibilities. In the identical condition, the verb in the second conjunct is the same as in the first and, thus, redundant. This redundancy could either be reflected in faster reading times in the matching compared to the mismatching condition, which is less redundant, indicating easier processing or in a slow-down because the reader would expect an ellipsis but encounters a full form.

7.1 Materials

We used the same 24 items as in Experiment 4 and added a final infinite subordinate clause as spillover after the verb of the second conjunct (16). The clause always began with the conjunction *um* "in order to", a definite article, and a singular noun. These three words and the verb in the second conjunct form the critical region, on which reading times were measured.

(16) *Für das Referat ist abgemacht, dass...*
"For the presentation, it is agreed that...",

a. *Jan und Sara den Zeitschriftenartikel* (*lesen*) *und Julia und Moritz*
Jan and Sara the journal.article read.3PL and Julia and Moritz
das Kapitel aus dem Buch $\underline{lesen}_1$, $\underline{um}_2$ $\underline{den}_3$ $\underline{Aufwand}_4$ *für jeden*
the chapter out the book read.3PL to the effort for every
Einzelnen geringer zu halten. (identical)
individual lesser to keep

b. *Jan den Zeitschriftenartikel* (*liest*) *und Julia und Moritz das Kapitel aus*
Jan the journal.article read.3SG and Julia and Moritz the chapter out
dem Buch $\underline{lesen}_1$, $\underline{um}_2$ $\underline{den}_3$ $\underline{Aufwand}_4$ *für jeden Einzelnen geringer zu halten.*
the book read.3PL to the effort for every individual lesser to keep
(ending+vowel)

We also tested 84 fillers: 24 from a study comparing the present perfect and the preterite, 12 with (potential) gapping structures with parentheses, 12 coordinations where ellipsis was not licensed, and 12 complex sentences with causal subordinate clauses.

7.2 Procedure

48 German native speakers aged 18 to 40 were recruited via Prolific and paid £3.50 to participate in a PCIbex (Zehr & Schwarz 2023) study. Participants read materials word-by-word in a moving window paradigm, advancing with the spacebar, then answered balanced yes-no questions using the *j* or *f* key. Four practice trials preceded the experiment to familiarize participants with the tasks. Items were split across four lists, ensuring each token set appeared only once per participant and that all conditions were equally often represented. Again, the materials were presented in pseudorandomized order.

7.3 Results

The mean accuracy for the comprehension questions was notably high at 95.5%. As the lowest accuracy achieved was 88.1%, no participants were excluded, and data from 48 individuals were analyzed. Figure 5 shows the mean reading times and standard errors per condition for the words in the critical region. Overall, the reading times for all conditions are very similar on each of the words. On words 2 to 4, RNR seems to be read a bit slower than the full forms. On the first word, the critical verb, there seem to be slightly larger differences between conditions, however, also the error bars are larger.

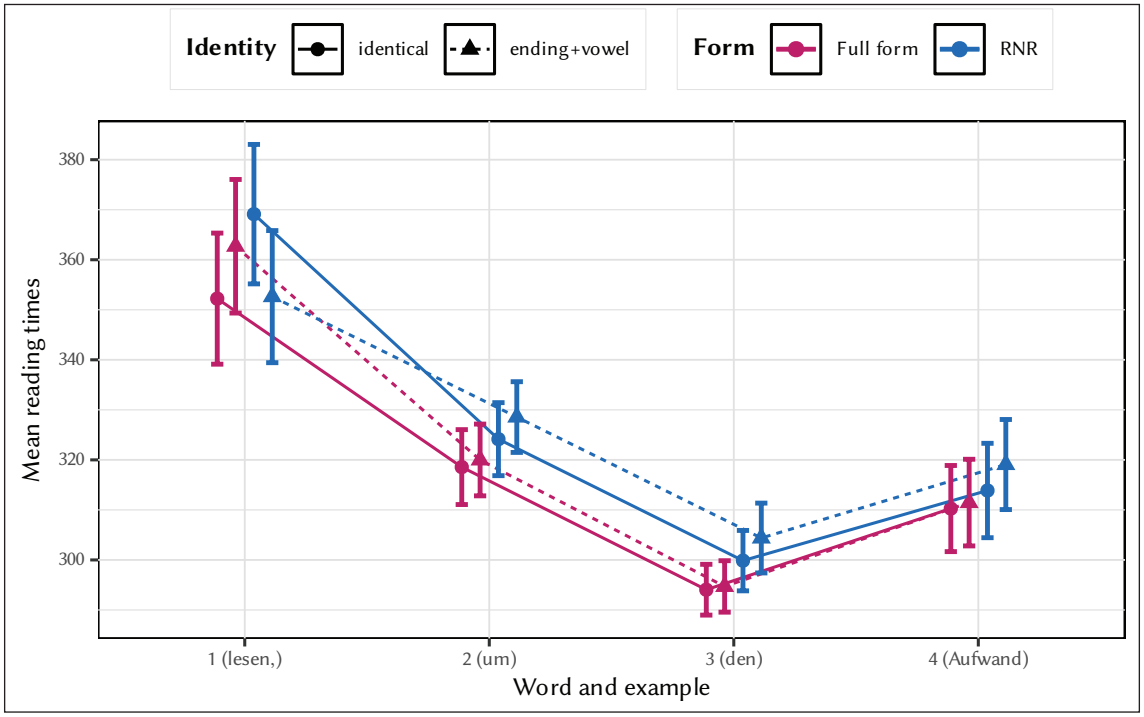


Figure 5: Mean reading times and standard errors per condition for the region of interest for Experiment 5

We conducted an analysis of the reading times for each of the four words within our designated region of interest, as indicated in (13) and in Figure 5, using separate linear mixed effects regression models (LMERs, package lme4, Bates et al. 2015) in R. All models contained the log-transformed reading times as dependent variable and as independent variables the dummy-coded Form (full form 0, RNR 1) and Identity (identical 0, ending+vowel 1), as well as their interaction and the *z*-transformed Position. The analyses of $word_1$ and $word_4$ additionally contained a fixed effect of Length, i.e., the *z*-transformed number of characters of the corresponding word.[18] This predictor was not necessary for $word_2$ and $word_3$ because they were always the subjunction *um* with two and a definite article with three letters. We reduced the random effects structure until the models converged and did not result in singular fit. The models for $word_2$ and $word_3$ contained only random intercepts for items and participants. The model for $word_1$ additionally had a by-subject random slope for Form and a by-item random slope for Identity, the model for $word_4$ had a by-subject random slope for Identity.[19]

All models contained a significant main effect of Position ($p < 0.001$), indicating that participants' reading times decreased over the course of the experiment. Only in the model for analysing $word_1$ were there additionally two marginal simple effects of Form and Identity and a significant interaction between Form and Identity (Table 10). The second verb tends to be read slower in RNR when both verbs are identical ($t = 1.74$, $p = 0.08$) and mismatches tend to be slower than matches in full forms ($t = 1.68$, $p = 0.09$). The interaction ($t = 2.24$, $p < 0.05$) qualifies these effects and suggests that the verb in the second conjunct is significantly slower in RNR structures when antecedent and target were identical compared to the full form with identical verbs. Viewed from the other side, the verb was read faster in case of a mismatch for RNR compared to the full form.

Table 10: Fixed effects in the LMER of $word_1$ (verb in second conjunct) of Experiment 5

Fixed effect	Est.	SE	*t*	*p*	
Intercept	5.71	0.06	99.85	< 0.001	***
Form	0.04	0.02	1.74	0.08	.
Identity	0.04	0.02	1.68	0.09	.
Length	0.04	0.02	1.63	0.12	
Position	−0.09	0.01	−11.86	< 0.001	***
Form:Identity	−0.07	0.03	−2.24	< 0.05	*

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < 0.1.

[18] The model calls were as follows, with the variation in random effects discussed in the following: LogRT ~ Form * Identity + Position (+ Length) + (1 | Participants) + (1 | Items).

[19] For $word_4$, we found an alternative model featuring a distinct random slope that likewise did not produce a singular fit. We compared the two models and selected the one with the lower AIC value. The parameter estimates were consistent to the second decimal point between the two models.

7.4 Discussion

In our self-paced reading study, we investigated the processing of RNR in comparison to the corresponding full form. While we did not find any relevant effects on the spillover, participants read the verb in the second conjunct faster in the mismatching ellipsis condition. This result is unexpected: the mismatch condition, which is predicted to be cognitively more demanding, is processed faster under ellipsis, while the match condition is processed more slowly. This contradicts the hypothesis that processing mismatching correlates requires greater effort.

At this point, we have no explanation for the apparent processing advantage of the mismatch.[20] What we can say, however, is that in any case the mismatch condition is not more difficult to process than corresponding full forms or than the match. Therefore, from a processing perspective, whatever the explanation for the RNR-mismatch condition is, it is not to be expected that the mismatches are rated worse compared to matches and full forms. Accordingly, our rating data from Experiments 2 to 4 and the processing data are consistent. At the same time, the results of the cloze study, i.e., the staircase-like decreasing ellipsis rates for RNR mismatches, apparently cannot be accounted for by processing. There is no evidence that RNR mismatches cause processing problems that lead speakers to more often produce a full form.

8 General discussion

In this paper, we reported on a cloze task, three rating studies, and a reading time study on RNR mismatches. Their aim was first and foremost an empirical one, namely to contribute (much-needed) empirical evidence to the discussion of identity constraints in German RNR. Another objective was to evaluate the different theoretical approaches to RNR presented at the beginning of the paper.

Even though the results differ between the more production-oriented cloze task and the other perception-oriented experiments, we can conclude from the whole picture that verbal mismatches in person and number in German RNR are not ungrammatical. In the two exploratory Experiments 2 and 3 and the two Experiments 4a and 4b, these structures received ratings that were comparable to those of gapping and the full forms, which functioned as baselines. This result is also in line with Shiraïshi et al. (2019), who did not find reduced acceptability for RNR tense mismatches in English and French. Furthermore, the cloze task in Experiment 1 showed that the participants consistently did not correct mismatches in all identity conditions, and that they did so in (almost) all conditions as frequently as they produced full forms.

[20] It is notable though that, in their rating study on English RNR, Shiraïshi et al. (2019: 8) found that "the RNR-MATCH condition is less acceptable than the RNR-MISMATCH condition ($p < 0.05$)," see their Figure 2 on the same page.

The results that the mismatch conditions pattern alike in the rating studies, that they seem not to require additional cognitive effort according to Experiment 5, and thus arguably should be regarded as grammatical are in line with semantic MSI and LI approaches. Under Chaves' (2014) MSI approach, number and person would then be features with a relatively weak semantic contribution, weak enough not to cause a reduced acceptability.

What is problematic for the MSI and the LI approaches, though, is the significantly higher frequency of ellipsis in the match condition in Experiment 1. This result is at first glance more in line with the PI approach; however, this approach would also predict the syncretic conditions to pattern alike with the identical conditions, which they did not in Experiment 1, and an acceptability penalty for the morphologically deviating conditions ending and ending+vowel, which we did not find in any of the rating studies. The GPI approach tentatively predicts the staircase-like pattern of Experiment 1, even though Christ (2011) considered the syncretic and the ending conditions to be fine as well. However, the cloze result is not reflected in any of the rating studies, where the RNR mismatches were not degraded. In other words, it seems that listeners have no problem processing this type of RNR mismatches and consider it acceptable. However, when given the opportunity to avoid such a mismatch by using a non-elliptical form, many of them take advantage of this option. No theoretical approach fully predicts this behavior. We therefore conclude that none of the current approaches can (without further assumptions) account for all the data presented in this paper.

The observed discrepancy between the results of the cloze study and the rating studies concerning RNR mismatches highlights the value of employing multiple methods to investigate apparent mismatch penalties. There are at least two options to account for this difference. The first option is that the cloze task in fact tested production rather than comprehension and that RNR mismatches are an issue for language production. This hypothesis would then have to be tested in a "real," i.e., less restricted and more spontaneous production study. The second option could be a methodological difference between the cloze task on the one hand and the rating task on the other. In the rating task, participants assess a sentence post hoc, presented as complete and attributed to another individual. In contrast, the cloze task exposes participants to a possibly incomplete sentence, granting them limited scope to amend it by inserting information at a specific position and thereby "making it their own." So, while participants may deem sentences in the rating task acceptable as utterances produced by others, the cloze task allows them to make an acceptable sentence even more acceptable (to their preferences) by changing it. The cloze task may therefore demonstrate increased sensitivity to subtle differences within high acceptability areas, at least regarding identity-related issues. These observations remain speculative and require further exploration, especially with respect to other phenomena and domains.

Acknowledgments: We thank the three anonymous reviewers for their constructive feedback on this article. We are also grateful to the two editors for their helpful guidance during the review process. Special thanks go to Robin Lemke for the in-depth discussions on the revised version of this article, which substantially improved it. We would also like to thank the student assistants of project B3, who helped in preparing the experiments and in annotating the cloze data.

Funding: Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 232722074 – SFB 1102 / Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

Abbreviations

ATB	across the board
GPI	gradual phonological identity
LI	lexeme identity
MD	multi-dominance
MSI	morph stem form identity
PI	phonological identity
RNR	right node raising
VP	verb phrase

Appendix

The materials, data, and scripts from this article can be found online at the following OSF storage: https://osf.io/f7g5k/files/osfstorage.

References

Aoun, Joseph, Elabbas Benmamoun & Dominique Sportiche. 1994. Agreement, word order, and conjunction in some varieties of Arabic. *Linguistic Inquiry* 25(2). 195–220.

Barros, Matthew & Luis Vicente. 2011. Right node raising requires both ellipsis and multidomination. *University of Pennsylvania Working Papers in Linguistics* 17(1).

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Belk, Zoe, Ad Neeleman & Joy Philip. 2023. What divides, and what unites, right-node raising. *Linguistic Inquiry* 54(4). 685–728. https://doi.org/10.1162/ling_a_00454.

Chaves, Rui P. 2014. On the disunity of Right-Node Raising phenomena: Extraposition, ellipsis, and deletion. *Language* 90(4). 834–886. https://doi.org/10.1353/lan.2014.0081.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
Christ, Rüdiger. 2011. Linkstilgung und phonologische Quasi-Identität. *Linguistische Berichte* 228. 371–411. https://doi.org/10.46771/2366077500228_1.
Christensen, Rune H. B. 2023. *ordinal – regression models for ordinal data*. R package version 2023.12-4.1. https://CRAN.R-project.org/package=ordinal, 15 September 2025.
de Vries, Mark. 2009. On multidominance and linearization. *Biolinguistics* 3(4). 344–403. https://doi.org/10.5964/bioling.8735.
Eisenberg, Peter. 1973. A note on "identity of constituents." *Linguistic Inquiry* 4(3). 417–420.
Fuß, Eric. 2018. Sprachliche variation. In Angelika Wöllstein, Peter Gallmann, Mechthild Habermann & Manfred Krifka (eds.), *Grammatiktheorie und Empirie in der germanistischen Linguistik,* 185–224. Berlin: De Gruyter. https://doi.org/10.1515/9783110490992-007.
Grosz, Patrick. 2015. Movement and agreement in Right-Node Raising constructions. *Syntax* 18(1). 1–38. https://doi.org/10.1111/synt.12024.
Hartmann, Katharina. 2000. *Right node raising and gapping: Interface conditions on prosodic deletion.* Philadelphia: John Benjamins. https://doi.org/10.1075/z.106.
Himmelreich, Anke, Melissa Jeckel & Johannes Mursell. 2024. Agreement patterns of coordination. In Anke Himmelreich, Daniel Hole & Johannes Mursell (eds.), *To the left, to the right, and much in between: A Festschrift for Katharina Hartmann,* 71–87. Frankfurt: Goethe University Frankfurt.
Klein, Wolfgang. 1981. Some rules of regular ellipsis in German. In Wolfgang Klein & Willem Levelt (eds.), *Crossing the boundaries in linguistics*, 51–78. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-8453-0_3.
Klein, Wolfgang. 1993. Ellipse. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An international handbook of contemporary research* (Vol. 1), 763–799. Berlin: De Gruyter. https://doi.org/10.1515/9783110095869.1.12.763.
Kluck, Marlies. 2009. Good neighbours or far friends: matching and proximity effects in Dutch Right Node Raising. *Groninger Arbeiten zur germanistischen Linguistik* 48. 115–158.
Limesurvey GmbH. n.d. *LimeSurvey: An Open Source survey tool.* Hamburg, Germany: LimeSurvey GmbH. http://www.limesurvey.org, 16 October 2025.
McCawley, James D. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13. 91–106.
Ott, Dennis. 2025. Rethinking right-node raising. *The Linguistic Review* 42(3). 389–418. https://doi.org/10.1515/tlr-2025-2013.
Philips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34(1). 37–90. https://doi.org/10.1162/002438903763255922.
Postal, Paul M. 1974. *On raising: One rule of English grammar and its theoretical implications*. Cambridge, MA: MIT Press.
R Core Team. 2025. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/, 15 September 2025.
Ross, John Robert. 1986 [1967]. *Infinite Syntax!* Norwood: Ablex Publishing Company. [Originally: *Constraints on variables in syntax.* PhD-Dissertation, MIT.]

Sabbagh, Joseph. 2007. Ordering and linearizing rightward movement. *Natural Language & Linguistic Theory* 25(3). 349–401. https://doi.org/10.1007/s11049-006-9011-8.

Schäfer, Lisa, Robin Lemke, Heiner Drenhaus & Ingo Reich. 2021. The role of UID for the usage of verb phrase ellipsis: Psycholinguistic evidence from length and context effects. *Frontiers in Psychology* 12. 661087. https://doi.org/10.3389/fpsyg.2021.661087.

Shen, Zheng. (2019). The multi-valuation agreement hierarchy. *Glossa* 4(1). 46. https://doi.org/10.5334/gjgl.585.

Shiraïshi, Aoi, Anne Abeillé, Barbara Hemforth & Philip Miller, (2019). Verbal mismatch in right-node raising. *Glossa: A Journal of General Linguistics* 4(1). 114. https://doi.org/10.5334/gjgl.843.

Steedman, Mark. 2000. *The syntactic process*. Boston: MIT Press. https://doi.org/10.7551/mitpress/6591.001.0001.

Wilder, Chris. 1997. Some properties of ellipsis in coordination. In Artemis Alexiadou & Tracy Alan Hall (eds.), *Studies on universal grammar and typological variation*, 59–107. Amsterdam: John Benjamins. https://doi.org/10.1075/la.13.04wil.

Wilder, Chris. 1999. Right node raising and the LCA. In Sonya Bird, Andrew Carnie, Jason D. Haugen & Peter Norquest (eds.), *Proceedings of the 18th West Coast Conference on Formal Linguistics,* 586–598. Somerville: Cascadilla Press.

Yatabe, Shuichi. 2003. A linearization-based theory of summative agreement in peripheral-node raising constructions. In Jong-Bok Kim & Stephen Wechsler (eds.), *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar,* 391–411. Stanford: CSLI Publications. https://doi.org/10.21248/hpsg.2002.20.

Zehr, Jeremy & Florian Schwarz. 2023. *PennController for Internet Based Experiments (IBEX).* https://doi.org/10.17605/OSF.IO/MD832.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 5
https://doi.org/10.18148/zs/2026-2014

Joanna Nykiel, University of California, Davis, United States, jnykiel@ucdavis.edu

Structural priming and preposition omission compete in fragment responses to questions

Abstract: This paper sets out to explore the generality of structural priming effects in elliptical responses to prepositional and nonprepositional variants of a single question: *At what time do you close?* and *What time do you close?* The responses were originally collected by Chia et al. (2020) through a language production study that had experimenters call U.S. businesses. While structural priming effects persist in the dataset including both fragment responses (*At 10 pm*) and full-sentence responses (*We close at 10 pm*), they disappear in the fragments subset, where nonprepositional responses (*10 pm*) are the dominant pattern for both nonprepositional and prepositional questions. I propose that the absence of structural priming effects here follows from the experimental items providing a distinct environment that attracts preposition omission in fragments, which here avoids repeating the preposition *at* present in prepositional questions. This environment facilitates online resolution of fragments by providing accessible antecedents in combination with reliable retrieval cues from fragments themselves, and has not yet received systematic attention in the literature on preposition omission.

Keywords: structural priming, fragments, preposition omission, cue-based retrieval, efficiency

Cite as: Nykiel, Joanna. 2026. Structural priming and preposition omission compete in fragment responses to questions. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2014.
Received: 20 October 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

Syntactic structures that speakers have previously produced themselves or processed when listening to another person are often recycled across subsequent utterances. This tendency, known as structural priming, has been confirmed through experimental (Bock 1986; Pickering & Branigan 1998; Branigan et al. 2000; Scheepers 2003, Sturt et al. 2010; Chia et al. 2020; Chia & Kaschak 2024) and corpus work (Dubey et al. 2005; Gries 2005; Szmrecsanyi 2005; Jacobs et al. 2019). For instance, when the grammar offers options, speaker B can be expected to respond to the question in (1) with a structure parallel to the one that speaker A has produced, that is, one hosting a nonprepositional fragment (NP).

(1) *A: What time do you close?*
B_1: 5 o'clock.
B_2: We close 5 o'clock.

The less expected option would be to respond with a structure hosting a prepositional fragment (PP) (*At 5 o'clock* or *We close at 5 o'clock*). These options are connected to the possibility of expressing temporal adjuncts of this kind (and corresponding interrogative phrases) as either NPs or PPs in English. But given speaker A's question, speaker B's choice here will be impacted by structural priming, as Chia et al. (2020) and Chia and Kaschak (2022, 2024) have demonstrated through language production experiments.

My goal is to examine whether structural priming effects are equally robust across two kinds of responses in Chia et al.'s (2020) data. Notice that response options for the question in (1) include full sentences (*We close (at) 5 o'clock*) and what I refer to as fragments (*(At) 5 o'clock*). While we would, in principle, not expect structural priming effects to differ between full sentences and fragments, fragments have unique properties that could be pulling in the opposite direction than structural priming effects do. To see this, consider (2), a prepositional variant of the question in (1) along with the same two responses.

(2) *A: At what time do you close?*
B_1: 5 o'clock.
B_2: We close 5 o'clock.

The B_1-response is misaligned with structural priming since a prepositional question is responded to with a non-matching structure, an NP, instead of a PP. Thus, the B_1-response should be dispreferred. However, this response could equally well have been produced by P(reposition)-omission, which optionally removes prepositions from fragments if they are present in the preceding context and is independently attested in English (e.g., Merchant 2001; Nykiel 2015, 2017, 2025; Nykiel & Hawkins 2020; Lemke 2021) and in other languages (e.g., Rodrigues et al. 2009; Nykiel 2013a; Molimpakis 2019; Hassen 2025). As this condition is satisfied (i.e., the

preposition *at* is present in the question), *at* may have been omitted from the fragment through P-omission. Thus there are two potential routes that lead to the B_1-response, with the availability of P-omission potentially competing with the pressure to keep the structure parallel across the question-and-response pair. If P-omission wins the competition, structural priming effects can be weakened. Unlike in fragments, P-omission isn't available in full sentences that correspond to fragments, as exemplified in (3).

(3) *A: Before which game did they sing the national anthem?*
B_1: The Champions League Finale.
*B_2: *They sang the national anthem the Champions League Finale.*

Therefore, the B_2-response in (2), which is a full-sentence variant of the B_1-response, can be arrived at via only one route, that is, failure to produce a parallel structure. To anticipate the results, we will see that structural priming effects play out differently in full sentences and in fragments. While they are more robust for prepositional questions than nonprepositional ones in full sentences, they are more robust for nonprepositional questions than prepositional ones in fragments, suggesting a competition between structural priming and P-omission in fragments. Drawing on existing P-omission literature, I will offer an explanation for why P-omission has the ability to override structural priming by producing structures that violate the expected parallelism in the context supplied by the A-question in (2).

The remainder of the paper is structured as follows. Section 2 provides a background on structural priming and details the design and results of the experiments conducted by Chia et al. (2020) to explore structural priming effects in dialogue exchanges depicted in (1) and (2). This sets the stage for my statistical analyses of full-sentence and fragment responses in section 4. Before turning to the analyses, I offer a theoretical background on fragments, and in particular the phenomenon of P-omission, which can look like absence of structural priming, in section 3. The results of my analyses reveal other constraints beyond structural priming that operate on fragments, depending on the questions they serve as responses to. A distinctly novel constraint is the combination of an independently processable fragment and an accessible correlate, which boosts the frequency of P-omission, trumping structural priming effects. In section 5, I discuss the ramifications of these results for work on structural priming under ellipsis, and for work on constraints operating on P-omission under ellipsis. Section 6 concludes.

2 Structural priming and Chia et al.'s (2020) study

Chia et al. (2020) explore structural priming effects in a uniquely naturalistic setting. Experimenters called U.S. businesses to inquire about their closing time by means of one of the questions depicted in (1) and (2) and immediately wrote down their responses. This experimental

paradigm is a replication of Levelt and Kelter's (1982) study of structural priming effects in Dutch, and, like that study, provides a record of entirely unscripted linguistic behavior. When engaged in conversation in this manner, speakers must first comprehend a question produced by another person and then produce a response, choosing to recycle, or not, elements of the structure of the question. Chia and Kaschak (2022), following Jacobs et al. (2019), refer to this type of priming as comprehension-to-production priming as opposed to production-to-production priming, where speakers recycle self-produced linguistic structures. Comprehension-to-production priming can be explained in terms of the existing mechanisms that are assumed to support priming: memory-based models and error-based models.

In memory-based models, an activation boost connected to prior exposure to a linguistic structure carries over to later productions. The activation boost can be short-lived and triggered by a recent encounter with a structure (Dell 1986; Pickering & Branigan 1998; Branigan et al.1999) or it can persist over longer periods of time (Reitter et al. 2011). Regardless of its duration, the structure that has received an activation boost has better odds of being reused in subsequent productions than its competitors with lower activation levels. Activation associated with a given structure can spread within a single speaker, as much as it can spread across dialogue participants. Branigan et al. (2000) and Branigan et al. (2007) provide evidence for speakers aligning their choices of linguistic structures over conversational turns, and Pickering and Garrod (2004) articulate a model of language processing in dialogue (the "interactive alignment model") to capture how dialogue participants' linguistic representations become automatically aligned over the course of a conversation at phonetic, phonological, semantic, and syntactic levels. Pickering and Garrod (2004) argue that the possibility of recycling structures with increased activation levels facilitates language production, providing a motivation for the alignment of linguistic structures.

Error-based models are based on the assumption that linguistic structures produced by speakers reflect the statistical preferences present in the discourse that surrounds them (Chang et al. 2006; Fine & Jaeger 2013; Dell & Chang 2014). While an expected structure simply confirms the current statistics, an unexpected one triggers adjustments to them due to an error signal being generated when such an unexpected structure is encountered. As a result, that structure will be more available for immediate future productions than its less surprising competitor(s). Error-based models also predict longer-term priming effects inasmuch as repeated exposure to unexpected structures leads to a gradual accumulation of statistical adjustments until no error signal is generated. Note that structures that are unexpected are typically also low-frequency structures, suggesting that priming effects are stronger for these structures.

Both memory-based models and error-based models predict priming effects in the dialogues designed by Chia et al. (2020). There are three possibilities regarding the nature of these effects. First, we might observe structural priming effects driven by the structure of the temporal adjunct

(At) what time. If the adjunct is realized as a PP in the question, then the response repeats that structure by producing a PP, either as a fragment or inside a full sentence, as in (4).

(4) *A: At what time do you close?*
B: At 5 o'clock. or *We close at 5 o'clock.*

The structure of the adjunct in the fragment response is constrained by the alternation between a PP and an NP. Given the nonprepositional question, structural priming effects arise if the response is realized either as an NP fragment or a full sentence with an NP adjunct, as in (1), repeated here for convenience as (5).

(5) *A: What time do you close?*
B: 5 o'clock. or *We close 5 o'clock.*

As the driving force behind structural priming here is the syntactic category of the temporal adjunct, the prime is no larger than the adjunct itself. It is independent of whether the syntactic structure that fragments are assigned on various theoretical accounts is sentential or not. Defining the prime this way is also aligned with the set of structural retrieval cues that are available on cue-based retrieval models of sentence processing (the structure of fragments and cue-based retrieval models are both discussed in section 3). Although priming is an entirely separate mechanism from cue-based models of sentence processing, the two converge on the processing advantage of producing structurally matching responses, as we will see in section 3. I refer to this kind of priming as adjunct-based structural priming.

The second possibility is that we might observe the lexical boost above and beyond structural priming effects. The lexical boost, which manifests as increased magnitude of structural priming, is induced by content words like verbs or nouns if they are present both in the initial experience with a structure and the subsequent reuse of that structure (Pickering & Branigan 1998; Mahowald et al. 2016). Some research points to the lexical boost being a shorter-lived effect than structural priming (Hartsuiker et al. 2008; Mahowald et al. 2016; Tooley 2020). In Chia et al.'s (2020) experiments, the lexical boost could be induced by the preposition *at*, which is present in a subset of all primes, that is, all the prepositional questions, and observed as an increase in the proportion of responses also including *at*, like those depicted in (4). Because each response immediately follows the question, the lexical boost could be detectable. However, Chia and Kaschak (2022), following the same experimental design, report no evidence for the lexical boost on the basis of a nearly identical distribution of full-sentence responses containing the preposition *at* across all questions. They report weak evidence for the lexical boost under ellipsis, with fragment responses to prepositional questions showing a 6% increase in the frequency of the preposition *at* compared to fragment responses to nonprepositional questions. The statistical analyses we will see in section 4 also cast doubt on the presence of the lexical boost.

Finally, we might observe structural priming effects at a schematic level that represents the structure of a question. Chia and Kaschak (2022) take this route when they take fragments to be sentential structures, parts of which remain silent. For them, the prime is the entire structure of a question, including the preposition placement choice.[1] If the preposition *at* is pied-piped together with its complement (*At what time do you close?*), a fragment response follows this pattern by pied-piping the preposition as well so that it escapes deletion. If the question has no preposition (*What time do you close?*), a fragment response can strand *at* within its structure, leaving it to be deleted along with the rest of the silent material. However, this view is problematic for three reasons. First, it's unclear why a structure with no preposition should prime a structure with preposition stranding rather than a structure that simply includes no preposition either. Second, the assumption that fragments are sentential structures is biased toward one kind of theory of fragments, making the priming mechanism heavily theory-dependent. Third, this view implies that overt structures can prime partially silent ones, although this is far from a given. In fact, the data we will see in section 4 reveal overall higher frequencies of full-sentence responses than fragments, suggesting that overt and partially silent structures are not primed by overt structures equally strongly.

To avoid these problems and remain theory-neutral, I assume that priming effects, if observed, are generated by the structure of a temporal adjunct and the preposition *at*. In other words, I expect to see adjunct-based structural priming, possibly in conjunction with the lexical boost produced by the preposition. Even though our understanding of the exact nature of structural priming effects differs, the patterns that Chia et al. (2020) observe are fully aligned with patterns I expect to see.

Chia et al.'s (2020) argument is that they observe predominantly structural priming effects when using the Levelt and Kelter (1982) paradigm, but they note that these effects are smaller in magnitude than could be expected. Their point of reference is the results of in-lab experiments included in Mahowald et al.'s (2016) meta-analysis of structural priming. The Levelt and Kelter (1982) paradigm yields modest structural priming effects in comparison, even when Chia et al. (2020) introduce manipulations to the experimental conditions in order to examine whether the participants' level of engagement in the conversation might make a difference. One of their experimental conditions matched Levelt and Kelter's design exactly: an experimenter called, asked the critical (prepositional or nonprepositional) question, and recorded the response. In the remaining three conditions, the critical question was preceded by either another question or a polite introduction, or followed by a clarificational question. Responses were coded separ-

[1] Some researchers have argued that priming effects may be driven by abstract syntactic structure that is fully independent of individual lexical items, as priming effects persist even when that structure includes nonce words or words embedded in inappropriate structural configurations (e.g., an intransitive verb embedded in a transitive frame) (Ivanova et al. 2012).

ately following a close scheme and an inclusive scheme. The close scheme allowed responses to be nonprepositional if the preposition *at* wasn't included in them, but other prepositions (e.g., *around 5 o'clock*) were fine. Prepositional responses required the presence of the preposition *at* on this scheme. On the inclusive scheme, prepositions like *around* were permitted in addition to *at* in prepositional responses, while no prepositions were permitted in nonprepositional responses. Chia et al.'s (2020) results reveal structural priming effects that remain unaffected by either the coding scheme or experimental condition; that is, prepositional questions were answered with prepositional responses more frequently than nonprepositional ones. Their statistical models also provide evidence for an effect of fragments, such that prepositional responses are produced less often as fragments than they are as full-sentence responses regardless of the coding scheme. I follow up on this finding, which suggests that structural priming effects do play out differently in fragments compared to sentential responses. In section 4, we will look at a subset of Chia et al.'s (2020) dataset in an effort to better understand this difference.

3 Background on fragments

Fragments are reactions to the preceding linguistic or nonlinguistic context (antecedent) that appear to be nonsentential. Most characteristically, they function as responses to *wh*-question antecedents, as in (6) and (7).

(6) *A: What are we looking at?*
B: A new TV series.

(7) *A: How did they get in?*
B: Through the back door.

In this section, I discuss the structure of fragments and the way they are processed online.

From a theoretical perspective, fragments are accounted for as genuine nonsentential utterances on the direct interpretation approach (Ginzburg & Sag 2000; Culicover & Jackendoff 2005) and as sentences that are partially unpronounced on the movement-and-deletion approach (Merchant 2001, 2004). This theoretical distinction is relevant for understanding the nature of structural priming effects detected in Chia et al.'s (2020) data. While the structure of the fragments in (6) and (7) consists of no more than the phrases themselves on the direct interpretation approach, it is sentential on the movement-and-deletion approach, as shown in (8) and (9) (strikethrough indicates the unpronounced material).

(8) *A new TV series ~~we are looking at~~.*

(9) *Through the back door ~~they got in~~.*

On the movement-and-deletion approach, the structures of a fragment and of a full sentence are aligned in Chia et al.'s (2020) experimental items except that the fragment undergoes leftward movement to escape deletion, as in (10) and (11).

(10) *(At) 5 o'clock ~~we close~~.*

(11) *We close (at) 5 o'clock.*

This is why Chia and Kaschak (2022) expect structural priming to operate at the schematic level of the full sentential structure, with a sentential antecedent priming full-sentence responses as much as it does fragments. This expectation is consistent with the movement-and-deletion approach, but not the direct interpretation approach, which predicts fragments to not be primed by sentential antecedents because their structures are nonidentical.

Adjunct-based structural priming as described in the previous section is consistent with both approaches. The only structure that plays a priming role is the structure of an adjunct, which in turn is determined by one of the syntactic categories (NP or PP) chosen for it. This is straightforward on the direct interpretation approach, as fragments are generated "as is." On the movement-and-deletion approach, the phrase moved to the sentence-initial position is whatever phrase originally occupies the adjunct position in situ.

Another issue surrounding fragments is the manner in which we process them online. When addressing this issue, I adopt the comprehender's perspective more than the speaker's because resolution of fragments requires additional resources than those normally employed for the processing of non-anaphoric material. An influential model of sentence processing, the cue-based retrieval model (McElree 2000; McElree et al. 2003; Lewis & Vasishth 2005; Lewis et al. 2006; Van Dyke 2011; Van Dyke & Johns 2012; Caplan & Waters 2013), has been shown to be compatible with the way the parser arrives at an interpretation for fragments. The steps involved in online processing of fragments are at least: (1) interpreting fragments, (2) locating their correlates (i.e., corresponding phrases) in the antecedents, and (3) retrieving fragments' sentential semantics (Culicover & Jackendoff 2005; Goldberg & Perek 2019; Harris & Carlson 2019; Nykiel et al. 2023). The second step (i.e., locating the correlate) engages a sort of "search" for the relevant representations. The search could in principle be operationalized either as a (sequential or parallel) search through all existing memory representations or a mechanism that provides direct access to the relevant representations without going through non-target representations. Martin and McElree (2011) have argued that what applies to fragments is the latter type of search, a direct-access mechanism guided by fragments' semantic and morphosyntactic features. Features that are encoded by linguistic forms on cue-based retrieval models are item-based and include lexical content, structural information like case and syntactic category, and local parenthood and sisterhood relations (Parker 2018). It remains an open question, especially in the case of

fragments, which of these features are engaged as cues during retrieval, with some researchers suggesting that structural cues are prioritized (e.g., Dillon et al. 2014; Lyu & Kaiser 2021) and others suggesting that all cues are engaged in parallel (e.g., Lewis & Vasishth 2005; Van Dyke & McElree 2006). Representations that match many retrieval cues become more highly activated, but retrieval success depends on how much overlap there is between the linguistic form initiating retrieval, the target memory representation, and distractors. If several features are shared across all of them, the result could be a slowdown in retrieving the target or downright failure to do so (Watkins & Watkins 1975; Nairne 2002; Öztekin & McElree 2007; Martin 2018).

Research on fragments shows that matching features are preferred even in simple environments. Given no distractors, speakers prefer for fragments' structural properties, like case specifications, to be aligned with the properties of their correlates (Nykiel et al. 2023) and for fragments' lexical content to be similarly aligned with the correlates' lexical content (Nykiel 2013a), suggesting that matching properties play a facilitatory role at retrieval. This research makes it plausible that syntactic category could be another candidate for serving as a retrieval cue. If it is, then adjunct-based structural priming effects could be reinforced for fragments in Chia et al.'s (2020) experimental data, such that matching questions and responses would have reliably higher frequencies for fragments than for full-sentence responses. Capturing adjunct-based structural priming from the comprehender's perspective as it does, this hypothesis aligns with the speaker's perspective, for whom a facilitatory role of structural priming is also predicted by, especially, memory-based models of priming (see section 2). This is the main hypothesis I test in section 4, but I'll add some refinements to it later in this section.

On the other hand, adjunct-based structural priming effects could be countered by a phenomenon known as P-omission, which occurs in fragments and is illustrated in (6) above (see also (3)). The response in B is an NP but could be a PP (*At a new TV series*) in its place. It is uncontroversial that P-omission is available for English fragments that have overt correlates (Merchant 2001; Chung 2006; Nykiel 2015, 2017; Nykiel & Hawkins 2020; Lemke 2021). This is so in (2), repeated here as (12), making the fragment *5 o'clock* ambiguous between two sources: P-omission and change in syntactic category, both of which constitute a violation of structural parallelism.

(12) *A: At what time do you close?*
B: 5 o'clock.

If P-omission has a high probability of occurrence following the antecedent in (12), it will pull in the opposite direction than adjunct-based structural priming (which pulls in favor of the matching PP). This brings us to the question of whether (12) does constitute an environment that attracts P-omission.

Looking at P-omission from the viewpoint of efficient communication, Nykiel and Hawkins (2020) propose that it's costly unless it applies in appropriate circumstances. Their approach is aligned with Hawkins' (2004: 25) notion of efficiency defined as the expression of maximally informative content by means of minimally complex forms ("Express the most with the least"), and with the Efficiency Hypothesis, which takes language to be a system for transmitting messages with both precision and minimal effort, where ease of production and comprehension have a role to play (Gibson et al. 2019; Futrell et al. 2020). Nykiel and Hawkins (2020) support their proposal with a robust pattern observed in crosslinguistic data: P-omission is less preferred than retention across introspective judgments reported in the literature (Stjepanovic 2008, 2012; Szczegielniak 2008; Rodrigues et al. 2009; Philippova, 2014) and predominantly less acceptable in experiments (Merchant et al. 2013; Nykiel 2013b; Molimpakis 2019; Lemke 2021; Nykiel & Kim 2022, Hassen & Abeillé 2025; Hassen 2025). Corpus data show more mixed results so far, with present-day English and French preferring P-omission (Nykiel 2017; Nykiel & Hawkins 2020; Hassen & Abeillé 2025; Hassen 2025), but earlier English dispreferring it (Nykiel 2015). This isn't surprising, given empirical evidence that repetition of redundant material can be beneficial rather than disruptive in language comprehension (Levelt 1989; Arts 2004; Maes et al. 2004; Davies & Katsos 2009; Arts et al. 2011), and is clearly beneficial both under ellipsis (Harris 2015) and when ellipsis is compared with its nonelliptical counterparts (Dickey & Bunger 2011; Kertz 2013). However, Nykiel and Hawkins (2020) argue that certain environments favor P-omission despite the benefits of not applying it;[2] of these, we will review two that are relevant: accessibility of a correlate, and independent processability of a fragment.

Even though the linguistic features of a fragment are key to successful resolution, as predicted by cue-based retrieval models, some of the responsibility lies on the accessibility of the frag-

[2] An anonymous referee suggests that P-omission, like other kinds of reductions, is consistent with information-theoretic models (Hale 2001, 2006; Levy & Jaeger 2007; Levy 2008; Demberg & Keller 2008). These models propose that more predictable material carries less information, has lower surprisal, incurs a lower processing cost, and tends to be omitted more easily than less predictable material. It seems to me that examining whether P-omission is subject to information-theoretical constraints would require precise estimates of how predictable various prepositions are in contexts more diverse than Chia et al.'s (2020) materials. However, a small experiment can offer insight into the predictability of the responses collected by Chia et al. (2020), including the responses shortened through P-omission. Toward this end, I extracted average log probability per word given the context for each response to either question from GPT2LMHeadModel. The patterns that emerged are that all prepositional responses are overall more predictable (i.e., less surprising and easier to process) than nonprepositional responses regardless of the question, and that prepositional full sentence responses are more predictable than prepositional fragments. These patterns are partially aligned with the responses actually produced in Chia et al.'s (2020) experiments: prepositional full-sentence responses are by far the most frequent at 60.9%, but prepositional fragments account only for 6% of the responses. Speakers thus end up producing responses whose predictability falls at the lower end of the scale approximately 30% of the time, with nonprepositional fragments accounting for over 90% of these responses. Overall, this predictability pattern suggests that P-omission is used despite incurring a processing cost, which is in line with Nykiel and Hawkins' (2020) proposal. The surprisal per word values for each response can be found in Table 1 in Appendix 2.

ment's correlate. The role of a correlate that Nykiel and Hawkins (2020) have in mind doesn't follow from cue-based retrieval as such but has a basis in independent research examining how memory encoding impacts the accessibility of phrases of various size and complexity for future retrieval from memory. A phrase that is semantically contentful and syntactically complex (e.g., a lexical NP) is processed more deeply and becomes more highly activated than a phrase that is less contentful and less complex (e.g., a pronoun). This creates an advantage in the form of a future accessibility boost (Craik & Lockhart 1972; Gallo et al. 2008; Hofmeister 2007, 2008, 2011; Hofmeister et al. 2007; Hofmeister et al. 2013; Karimi et al. 2014). Simply put, the degree of activation is directly proportional to accessibility for future retrieval from memory. This future accessibility boost has the additional effect that more accessible phrases are typically picked up by anaphors with low content, while less accessible phrases are picked up by more contentful anaphors (Ariel 1990; Gundel et al. 1993; Karimi et al. 2014; Troyer et al. 2016; Nykiel et al. 2023). Nykiel and Hawkins (2020) argue that P-omission represents one of the ways of reducing the content of an anaphor whose function is to pick up an accessible correlate, that is, a correlate containing a lexical NP rather than an indefinite or interrogative pronoun. Their argument finds robust support in crosslinguistic data (Rodrigues at al. 2009; Nykiel 2013b; Nykiel & Kim 2022; Hassen 2025; Hassen & Abeillé 2025). But how exactly does accessibility impact example (12)? The correlate contains an interrogative *what*-NP phrase rather than an interrogative pronoun like *when* with corresponding temporal semantics. These phrases differ in their activation, with *what time* being more highly activated due to its higher semantic content and syntactic complexity compared to *when*. This is why the entire question supplies a context in which the fragment can be easily resolved to a temporal adjunct. Nykiel and Hawkins (2020) predict that, when referring back to the accessible correlate *at what time* by means of a fragment, speakers will tend to drop the preposition *at* in order to minimize the fragment's content. Their prediction follows from one of the efficiency principles proposed in Hawkins (2004: 38), Minimize Forms, which defines a preference for linguistic forms to encode minimal information when the missing information is inferable from the context.

The second environment to favor P-omission is linked to the independent processability of a fragment. Nykiel and Hawkins (2020) define this concept as a fragment's ability to be assigned all, or as many as possible, of its intended semantic and morphosyntactic features on the basis of its own linguistic make-up. That is, the parser should ideally be able to interpret as accurately as possible the fragment's semantics and grammatical role in relation to the antecedent, features which subsequently guide the cue-based search for the correlate. All else being equal, the presence of a preposition aids in achieving this goal, as it helps clarify the semantics and grammatical role of the fragment. A key term here, drawn from Hawkins (2004: 32), is a "domain," or the smallest sequence of linguistic forms that makes it possible to process all the syntactic and semantic relations among these forms. Applying this term to fragments, we can construe them

as domains in which their properties can be processed more or less successfully, and in which prepositions play a facilitatory role, if retained. However, one difficulty that can arise is the presence of a semantic dependency between a preposition that has been retained in a fragment and a higher verbal head that is only present in the antecedent. This case is illustrated in (13) from Nykiel and Hawkins (2020: 424).

(13) *A: And you're depending on what?*
B: On people going into the booth and changing their minds?

The preposition *on* and the higher verb *depend* form a semantic dependency extending beyond the fragment itself and into the antecedent, which disrupts successful interpretation of the fragment as a domain. Nykiel and Hawkins (2020) offer empirical support for the idea that prepositions which can't be processed within the bounds of a fragment tend to be dropped instead, as their presence would unnecessarily extend the domain in which a fragment can be processed. This idea is based on another efficiency principle defined in Hawkins (2004: 31), Minimize Domains[3], which favors, among other things, the kind of domain minimization that P-omission achieves for cases of semantically dependent prepositions.

The fragment in (12) doesn't pose a processing problem of this kind, as the preposition *at* is not semantically dependent on any higher verbal head. However, there is an entirely novel pattern here: the fragment's temporal semantics and the grammatical role of an adjunct are transparent with or without the preposition. This in turn means that the domain that the fragment constitutes can be minimized from a PP to an NP with no loss of semantic or morphosyntactic information. This is thus a situation where the presence or absence of a preposition quite plausibly neither disrupts nor improves the independent processability of a fragment. It's a distinct possibility that, when faced with a combination of an accessible correlate and an easy-to-process fragment, speakers choose to drop the preposition as a redundant cue, in effect minimizing the domain in which the fragment can be processed.

Nykiel and Hawkins (2020) report how each of these environments impacts the odds of P-omission. They offer evidence that the independent processability of fragments underlies high frequencies of P-omission in English (even though it's driven by semantically dependent prepositions, which are absent from Chia et al.'s (2020) data), with an effect size much larger than that of correlate accessibility. These findings allow us to reformulate the main hypothesis to test in this study: adjunct-based structural priming effects could be overridden by a combination of fragments' independent processability and correlate accessibility effects in Chia et al.'s (2020) fragments. The specific hypothesis is that if adjunct-based structural priming effects are weak

[3] Although Minimize Domains has had much success in explaining P-omission in English, it usually applies to word order phenomena (see, e.g., Hawkins 2004, 2014; Lohse et al. 2004; Wiechmann & Lohmann 2013).

or non-existent, then questions don't predict responses; that is, we will observe similar frequencies of nonprepositional fragments for both nonprepositional and prepositional questions. However, a note of caution is due. The direction in which the independent processability of fragments might impact P-omission remains an unknown in the data. Its effects could either boost the frequencies of nonprepositional fragments in order to remove redundant retrieval cues through P-omission or have no effects at all if all fragments' independent processability is at ceiling regardless of the presence of prepositions. We will have evidence for the former scenario if no adjunct-based structural priming effects emerge. But if adjunct-based structural priming effects do emerge, this can be interpreted as evidence that fragments' independent processability plays less of a role in the retrieval process than Nykiel and Hawkins (2020) argue for, at least whenever no semantically dependent prepositions are involved. We can draw this conclusion because Chia et al.'s (2020) data has the distinct advantage that the set of factors that might be driving P-omission can be limited to exactly two: independently processable fragments and accessible correlates. This brings us to statistical analyses of a subset of Chia et al.'s (2020) data.

4 Current data and method

The data selected for the current statistical analyses is a subset of questions and responses where we can examine the competition, if any, between structural priming and P-omission. This subset consists of full-sentence and fragment responses using the preposition *at* or no preposition at all, along with the questions they are intended as responses to. That is, only the questions and responses shown in (1)–(2) and (4)–(5) are included. Following this procedure, I retrieved 868 responses, of which 317 (36.5%) are fragments and 551 (63.5%) are full sentences. Before proceeding, let me summarize the two-fold purpose behind this study. First, we are probing the data for adjunct-based structural priming, potentially reinforced by the lexical boost. Second, we hypothesized that adjunct-based structural priming and P-omission compete for fragment responses, opening up the possibility that we fail to detect adjunct-based priming effects in the fragment subset. I begin by addressing the lexical boost, and do so on the basis of raw frequency data. The presence of adjunct-based structural priming and the competition between adjunct-based priming and P-omission are then tested by means of regression modeling.

Consistent with what Chia and Kaschak (2022) report, there is little evidence for the lexical boost based on the raw frequencies of full sentences and fragment responses shown in Figure 1 and Figure 2. Among full sentence responses, 97% are prepositional when produced in response to prepositional questions and 96% are prepositional when produced in response to nonprepositional questions. Among fragment responses, the corresponding values are 20% and 12.6%.

I don't have an explanation for the absence of the lexical boost in full-sentence responses, but its near-absence in fragments is clearly tied to adjunct-based structural priming effects themselves, such that the priming work that prepositional questions are doing is likely undone by P-omission. With little evidence for the lexical boost in fragments, there can be little hope for adjunct-based structural priming effects occurring there either.

When examining adjunct-based structural priming effects, I analyze full sentences and fragments first together and then separately. The results of these analyses are discussed below. Because we're concerned with fragments here, all separate analyses of full sentences can be found in Appendix 1.

4.1 Results and discussion

I begin by testing whether adjunct-based structural priming effects are still reliable across all responses, as the current sample is smaller than the original Chia et al. (2020) dataset. The model is a Bayesian multilevel probit regression analysis predicting the response (prepositional, nonprepositional) from Question Type (prepositional, nonprepositional) in the 2x2 design, with group-level intercepts and slopes for Experimenter (i.e., the person who asked questions). The population-level factor, Question Type, was treatment-coded. This analysis, and all subsequent ones, were conducted in R with the brms package with a logit link function (R core team 2021; Buerkner 2017; 2018). All models were fitted using weakly informative priors suggested by the brms package: improper flat priors for the population-level coefficients, a Student-t prior (3, 0, 2.5) for the population-level intercept, and a half-Student-t prior for the standard deviations of group-level effects, all on the logit scale. All analyses ignored the additional conditions Chia et al. (2020) introduced into the manner their questions were asked, given that no reliable differences were found among them. As expected, the first model revealed evidence for a main effect of Question Type such that prepositional questions triggered prepositional responses more often than nonprepositional questions ($\Delta = 0.46$, $CI(\Delta) = [0.17, 0.75]$, $P(\Delta > 0) = 1$). The remaining analyses examine adjunct-based structural priming in fragments more closely.

I first test whether adjunct-based structural priming effects emerge in fragments, and if they do so more than in full-sentence responses. These effects could be more robust for fragments due to increased processing demands placed on fragments as context-dependent utterances, as discussed in section 3. Figure 1 shows the raw frequencies of fragment responses, realized as either NPs or PPs, depending on whether the question is prepositional or not. Note the robust pattern: NPs are by far the more frequent responses than PPs across the board, even though there also is a slight increase in PPs for prepositional questions, keeping their structures aligned. The model I fitted to these data confirmed that the type of question doesn't

predict the response. The model included one population-level factor Question Type (prepositional, nonprepositional), which was treatment-coded, and group-level intercepts and slopes for Experimenter. The dependent variable was Response Type (NP, PP). I found no evidence for adjunct-based structural priming effects ($\Delta = -0.54$, $CI(\Delta) = [-1.19, 0.12]$, $P(\Delta < 0) = 0.94$); there was only a trend for prepositional questions to lower the frequency of NP responses compared to nonprepositional questions. Given the lack of evidence for adjunct-based structural priming in fragments, we can't expect that structurally matching adjuncts play more of a role in fragments than they do in full sentences. This conclusion is also suggested by the raw frequencies of full-sentence responses shown in Figure 2, where primarily prepositional responses were produced across the board. This dominance of one type of response in this subset as well suggests that full responses and fragments are equally unaffected by structurally matching adjuncts.[4] I confirmed this conclusion upon fitting a Bayesian multilevel probit regression model to the dataset consisting of full-sentence responses and fragments. This model's dependent variable was a binary factor: structurally matching or mismatching adjuncts for each question-and-response pair. The population-level factor was Response Type (fragment, full sentence), which was treatment-coded. Group-level intercepts and slopes for Experimenter were included as before. This analysis yielded no reliable difference between fragments and full sentences with respect to structural match ($\Delta = 0.14$, $CI(\Delta) = [-0.14, 0.43]$, $P(\Delta > 0) = 0.84$).

To get a better sense of whether the lack of adjunct-based structural priming effects in fragments can be attributed to high frequencies of P-omission, I fitted another Bayesian multilevel probit regression model to the fragments dataset. This model was designed to test whether structurally matching pairs of questions and responses were associated with nonprepositional questions more than prepositional ones, as Figure 1 suggests. In this model, the population-level factor was Question Type (prepositional, nonprepositional), treatment-coded as before, and group-level intercepts and slopes for Experimenter were also included. The results revealed that prepositional questions favored matching responses to a much lesser extent than nonprepositional ones ($\Delta = -3.43$, $CI(\Delta) = [-4.11, -2.77]$, $P(\Delta < 0) = 1$). This effect indicates that it's specifically responses to prepositional questions that fail to align with the pattern motivated by adjunct-based structural priming. In other words, P-omission competes with adjunct-based structural priming and wins the competition much of the time.

[4] It is potentially interesting to note that, even though neither the fragment responses nor the full-sentence responses show adjunct-based structural priming effects in separate analyses, the patterns are reversed. The set of full-sentence responses reveals stronger priming effects for prepositional questions than nonprepositional questions (see Appendix 1). This difference can be partially explained away by the unavailability of P-omission in full sentences.

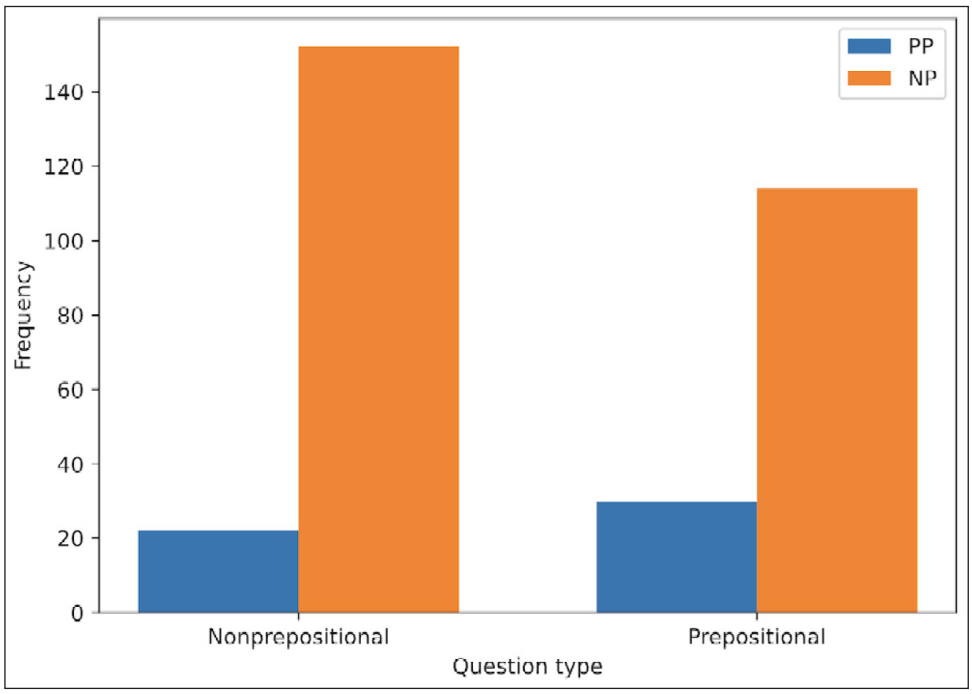


Figure 1: Raw frequencies of NPs and PPs by Question Type

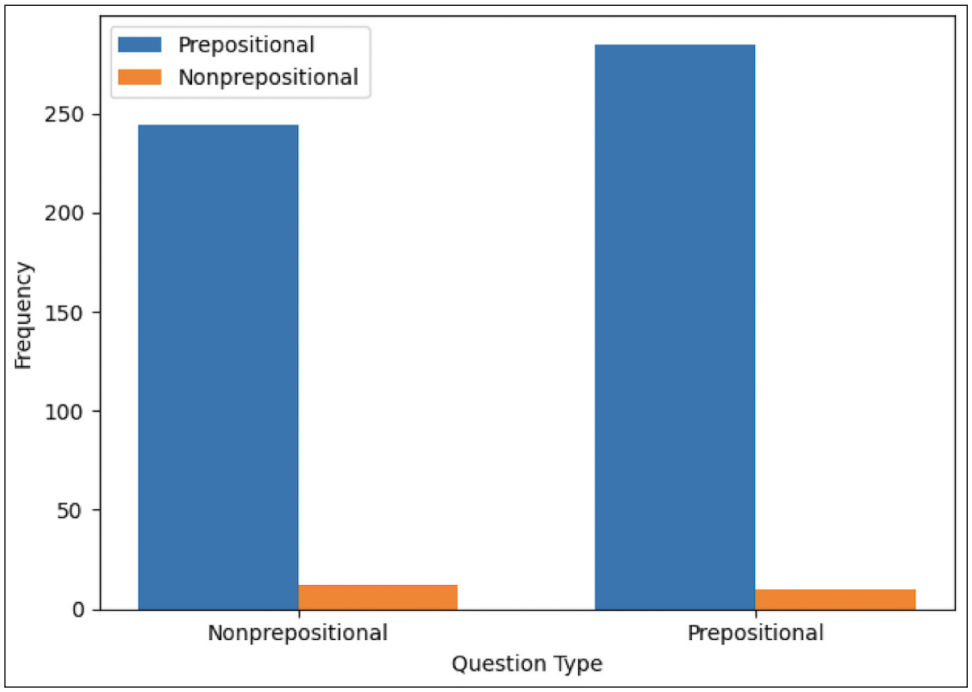


Figure 2: Raw frequencies of full-sentence responses by Question Type

5 General discussion

The current results reveal that while reliable adjunct-based structural priming effects can be seen across all responses, post-hoc analyses of fragment responses offer insight into how these effects interact with forces driving P-omission, a phenomenon that affects fragments in specific environments. I have tested and confirmed the hypothesis that adjunct-based structural priming and P-omission compete when P-omission is available, that is, when fragments serve as responses to prepositional questions. The novel finding is that P-omission wins the competition with the result that we no longer observe adjunct-based structural priming effects in the fragments dataset. This finding is both surprising and not. The unique nature of fragments as utterances relying on the surrounding context for their interpretation makes it a plausible idea that structural parallelism between them and the surrounding context has a processing advantage, as argued by Nykiel et al. (2023). On the other hand, P-omission is known as a process that readily engages when fragments are easy to interpret (Nykiel & Hawkins 2020). As discussed in section 3, there is good reason to think that fragments are indeed easy to interpret in the environment designed by Chia et al. (2020).

However, there is a dissociation between nonprepositional and prepositional questions. Robust adjunct-based structural priming effects arise in fragment responses to nonprepositional questions, where P-omission is unavailable and where there is no reason to think that something else interferes with the structure of a response. In these cases, the corresponding grammatical roles of fragments and correlates are realized by the same syntactic categories, NPs, increasing the ease with which they can be paired together during the search for the correlates, if this search engages the direct-access mechanism that cue-based retrieval models assume. Because prepositional questions pattern very differently, we are left with a situation that has ramifications for both structural parallelism on the level of individual elliptical constructions and for constraints on P-omission under ellipsis.

Although a subset of the current data reveals an actual preference for nonparallel syntactic categories, fragments aren't the only elliptical construction to tolerate such mismatch. Another example is Poppels and Miller (2023), who examine nonparallel syntactic categories under English pseudogapping. In pseudogapping, a stranded auxiliary is followed by a phrase similar to a fragment, as in (14) from Poppels and Miller (2023: 303).

(14) *They will give Dana books just as they will to Tim.*

The nonparallelism in (14) is due to the dative alternation, an alternation between the double object construction (*give Dana books*) and the prepositional object construction (*give books to Dana*). The post-auxiliary PP *To Tim* located in the right clause thus corresponds to the NP *Dana* in the antecedent clause. Compare this with the parallel PP structures in (15) (Poppels & Miller 2023: 303).

(15) *They will give books to Dana just as they will to Tim.*

What Poppels and Miller (2023) find is that although nonparallel structures do receive a penalty, the magnitude of this effect is reduced by perceived semantic similarity between the parallel and nonparallel structures. That is, when speakers perceive *give Dana books* and *give books to Dana* as semantically similar, the penalty for the nonparallel structure is reduced, pointing in the same direction as the current results: the preference for structural parallelism competes with and can be trumped by nonsyntactic constraints.

At the same time, there is general agreement that both fragments and other elliptical constructions obey structural parallelism constraints by favoring matching structures over mismatching ones (Frazier & Clifton 1998; Carlson 2002; Dickey & Bunger 2011; Carlson & Harris 2018; Kim & Runner 2018; Nykiel et al. 2023; Poppels & Miller 2023) or disallowing mismatching structures altogether (Orth & Yoshida 2025). This is so despite finer-grained construction-specific variation, where, for instance, Verb Phrase ellipsis is quite generous with respect to the range of possible structural mismatches (Arregui et al. 2006; Kertz 2013; Kim & Runner 2018; Parker 2018), but gapping appears not to be (Kim et al. 2020; Overfelt 2024; Orth & Yoshida 2025). While we're still far from understanding both the full extent of this variation and how much of it is specific to ellipsis, the current results highlight one reason why the mechanisms supporting fragments can be oblivious to structural parallelism.

The choice to produce an NP fragment as a response to a prepositional question is a principled departure from structural parallelism. I propose this departure happens not only because P-omission is available, but also because Chia et al.'s (2020) data provide one of the easy-to-process environments that have been shown to attract P-omission. Growing crosslinguistic work offers both experimental and corpus evidence to support the idea that accessible correlates have facilitatory effects on P-omission. Typical accessible correlates look like *in which department* in (16) and *to a wedding* in (17). Typical responses lack prepositions (note that the response in (17B) is a sluice), which Nykiel & Hawkins (2020) offer evidence for in English.

(16) *A: And you're in which department?*
B: Linguistics (department).

(17) *A: I'm going to a wedding.*
B: Whose (wedding)?

Nykiel (2013b) and Nykiel and Kim (2022) demonstrate experimentally that there is no penalty for P-omission in Polish exactly when correlates contain lexical NPs rather than pronouns. Hassen and Abeillé (2025) and Hassen (2025) offer French experimental and corpus data pointing in the same direction. In their corpus data, P-omission is observed almost exclusively with accessible correlates. As discussed in section 3, each question in Chia et al.'s (2020) dataset contains an accessible correlate.

Furthermore, fragment responses to Chia et al.'s (2020) questions can all be assigned their temporal adjunct semantics immediately upon being processed and independently of the antecedents. I speculated in section 3 that this could either lead to a spike in the frequency of P-omission or leave it unaltered. As we have observed the former scenario, explaining the high frequency of P-omission is clearly within the reach of Nykiel and Hawkins' (2020) processing-based proposal. Speakers perceive prepositions as redundant retrieval cues when fragments are already easy to process due to their correlates' accessibility and their own independent processability and, therefore, tend to remove them. This tendency could be seen as reflecting additive effects of correlate accessibility and the independent processability of fragments on P-omission, and represents an entirely novel finding in work focused on constraints on P-omission. It hasn't yet been explored in detail to what extent the parser does prioritize the ease with which fragments can be processed independently. As discussed in section 3, Nykiel and Hawkins (2020) propose that the independent processability of fragments is a main source of crosslinguistic differences in the frequency of P-omission and, as such, deserves scholarly attention in future research. They themselves focus on one dimension of this factor, namely, the presence of a semantic dependency between prepositions and verbal heads, and single it out as the best predictor of P-omission in English. To take this line of inquiry further, we could systematically examine how other features of fragments (e.g., repetition of the head nouns *department* and *wedding* in (13)–(14)) may impact their independent processability and how they may guide speakers' preferences for P-omission.

6 Conclusion

My goal in this paper has been a detailed examination of structural priming effects in fragment responses to two variants of a single question. While the experimental data collected by Chia et al. (2020) reveal reliable adjunct-based structural priming effects across all pairs of questions and responses (both fragment and full-sentence responses), it turns out that these effects disappear in the subset of fragment responses. I have proposed that this is due to the availability of P-omission for prepositional questions and the nature of Chia et al.'s (2020) experimental items, which provide a novel and distinctly supportive environment for P-omission.

Acknowledgements: I am grateful Michael Kaschak for sharing his entire dataset with me, which made the detailed analyses reported in this paper possible. I am also grateful to three anonymous referees for their detailed suggestions that have helped me improve this paper. Finally, I would like to thank the audience at the 7th Conference of the International Society for the Linguistics of English (ISLE 7), where an earlier version of this work was presented, for their comments.

Appendix 1

These are Bayesian multilevel probit regression analyses of the subset consisting of the full-sentence responses shown in Figure 2. The structure of the two models below corresponds to the structures of the two models of fragments described in section 4.1. The first model looked for traces of adjunct-based structural priming in full sentences, and the second model compared the preference for matching structures in prepositional questions and nonprepositional questions.

Model 1: Question Type doesn't predict the response (i.e., no adjunct-based structural priming effects)

Model structure: Question Type (prepositional, nonprepositional) as a population-level factor, with group-level intercepts and slopes for Experimenter

Results: No reliable difference between prepositional and nonprepositional questions with respect to the frequency of prepositional responses ($\Delta = 0.35$, $CI(\Delta) = [-0.51, 1.24]$, $P(\Delta > 0) = 0.79$).

Model 2: Structurally matching pairs of questions and responses are favored by prepositional questions compared to nonprepositional ones

Model structure: Question Type (prepositional, nonprepositional) as a population-level factor, with group-level intercepts and slopes for Experimenter

Results: Prepositional questions have significantly higher frequencies of structurally matching questions and responses than nonprepositional questions ($\Delta = 6.97$, $CI(\Delta) = [5.92, 8.30]$, $P(\Delta > 0) = 1$).

Appendix 2

Table 1: Surprisal per word for each response extracted from GPT2LMHeadModel

Response	Question: What time	Question: At what time
Nonprepositional fragment (NP)	6.4	6.6
Nonprepositional full sentence	6.1	6.1
Prepositional fragment (PP)	5.6	5.8
Prepositional full sentence	4.3	4.2

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Arts, Anja. 2004. *Overspecification in instructive texts*. Nijmegen: Wolf Publishers.

Arts, Anja, Alfons Maes, Leo Noordman & Carel Jansen. 2011. Overspecification facilitates object identification. *Journal of Pragmatics* 43(1). 361–374. https://doi.org/10.1016/j.pragma.2010.07.013.

Arregui, Ana, Charles Clifton Jr., Lynn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246. https://doi.org/10.1016/j.jml.2006.02.005.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387. https://doi.org/10.1016/0010-0285(86)90004-6.

Branigan, Holly P., Martin J. Pickering & Alexandra A. Cleland. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review* 6(4). 635–640. https://doi.org/10.3758/BF03212972.

Branigan, Holly P., Martin J. Pickering & Alexandra A. Cleland. 2000. Syntactic coordination in dialogue. *Cognition* 75(2). B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5.

Branigan, Holly P., Martin J. Pickering, Janet F. McLean & Alexandra A. Cleland. 2007. Syntactic alignment and participant role in dialogue. *Cognition* 104(2). 163–197. https://doi.org/10.1016/j.cognition.2006.05.006.

Buerkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. https://doi.org/10.18637/jss.v080.i01.

Buerkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411. https://doi.org/10.32614/RJ-2018-017.

Caplan, David & Gloria Waters. 2013. Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review* 20(2). 243–268. https://doi.org/10.3758/s13423-012-0369-9.

Carlson, Katy. 2002. *Parallelism and prosody in the processing of ellipsis sentences*. New York: Routledge.

Carlson, Katy & Jesse Harris. 2018. Zero-adjective contrast in much-less ellipsis: The advantage for parallel syntax. *Language, Cognition and Neuroscience* 33(1). 77–97. https://doi.org/10.1080/23273798.2017.1366530.

Chang, Franklin, Gary S. Dell & Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113(2). 234–272. https://doi.org/10.1037/0033-295X.113.2.234.

Chia, Katherine & Michael P. Kaschak. 2022. Structural priming in question-answer dialogues. *Psychonomic Bulletin & Review* 29(1). 262–267. https://doi.org/10.3758/s13423-021-01976-z.

Chia, Katherine & Michael P. Kaschak. 2024. Elliptical responses to direct and indirect requests for information. *Language and Speech* 67(1). 228–254. https://doi.org/10.1177/00238309231176526.

Chia Katherine, Hannah Hetzel-Ebben, Maxwell Adolph, Madison Amaral, Maynor Arriaga, Holly Booth, Victoria Boudreau, Jennifer Carpenter, Carolina Cerra, Megan Clouden, Jacob Cryderman, Rida Darji, Jolie Dollison, Nicholas Franco, Lawrence Ghougasian, Laurel Hamilton, Kimiyo Karosas, Casey Kenoyer, Victoria Krenz, Samantha Lancaster, Melissa Ma, Grace Markwell, Faith Montoya, Rebecca Nadler, Samara Pinto, Melissa Rojas, Daniela Sarmiento, Chloe Stitik, Jasmine St. John, Mariana Valencia, Kayla Walker, Emma Wells, Julia Wolf, Destiny Wright & Michael

P. Kaschak. 2020. Examining the factors that affect structural repetition in question answering. *Memory and Cognition* 48(6). 1046–1060. https://doi.org/10.3758/s13421-020-01036-2.

Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. In Rebecca T. Cover and Yuni Kim (eds.), *Proceedings of the annual meeting of the Berkeley Linguistics Society 31*, 73–91. Berkeley, CA: Berkeley Linguistics Society.

Craik, Fergus I. M. & Robert S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11. 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Davies, Catherine & Napoleon Katsos. 2009. Are interlocuters as sensitive to over-informativeness as they are to under-informativeness? PRE-CogSci workshop: production of referring expressions. Austin, TX: Cognitive Science Society. Online https://www.researchgate.net/publication/264873903_Are_Interlocutors_as_Sensitive_to_Over-informativeness_as_they_are_to_Underinformativeness.

Dell, Gary. S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3). 283–321. https://doi.org/10.1037/0033-295X.93.3.283.

Dell, Gary S. & Franklin Chang. 2014. The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B* 369(1634). 20120394. https://doi.org/10.1098/rstb.2012.0394.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210. https://doi.org/10.1016/j.cognition.2008.07.008.

Dickey, Michael W. & Ann Bunger. 2011. Comprehension of elided structure: evidence from sluicing. *Language and Cognitive Processes* 26(1). 63–78. https://doi.org/10.1080/01690961003691074.

Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu & Colin Phillips. 2014. The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5. 1025. https://doi.org/10.3389/fpsyg.2014.01025.

Dubey, Amit, Patrick Sturt & Frank Keller. 2005. Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 827–834. Stroudsburg, PA: Association for Computational Linguistics.

Fine, Alex B. & Florian T. Jaeger. 2013. Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37(3). 578–591. https://doi.org/10.1111/cogs.12022.

Frazier, Lynn & Charles Jr. Clifton. 1998. Comprehension of sluiced sentences. *Language and Cognitive Processes* 13(4). 499–520. https://doi.org/10.1080/016909698386474.

Futrell Richard, Roger P. Levy & Edward Gibson. 2020. Dependency Locality as an explanatory principle for word order. *Language* 96(2). 371–412. https://doi.org/10.1353/lan.2020.0019.

Gallo, David A., Nathaniel G. Meadow, Elizabeth L. Johnson & Katherine T. Foster. 2008. Deep levels of processing elicit a distinctiveness heuristic: evidence from the criterial recollection task. *Journal of Memory and Language* 58(4). 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001.

Gibson, Edward, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407. https://doi.org/10.1016/j.tics.2019.02.003.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations*. Stanford: CSLI Publications.

Goldberg, Adele & Florent Perek. 2019. Ellipsis in construction grammar. In Jeoren van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis*, 188–204. Oxford: Oxford University Press.

Gries, Stefan T. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399. https://doi.org/10.1007/s10936-005-6139-3.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of anaphoric expressions in discourse. *Language* 69(2). 274–307. https://doi.org/10.2307/416535.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of NAACL (Vol. 2)* (Pittsburgh, PA), 159–166.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30(4). 643–672. https://doi.org/10.1207/s15516709cog0000_64.

Harris, Jesse A. 2015. Structure modulates similarity-based interference: An eye tracking study. *Frontiers in Psychology* 6, Article 1839. https://doi.org/10.3389/fpsyg.2015.01839.

Harris, Jesse A. & Katy Carlson. 2019. Correlate not optional: PP sprouting and parallelism in 'much less' ellipsis. *Glossa* 4(1). 83. https://doi.org/10.5334/gjgl.707.

Hartsuiker, Robert J., Sarah Bernolet, Sofie Schoonbaert, Sara Speybroeck, Dieter Vanderelst. 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58(2). 214–238. https://doi.org/10.1016/j.jml.2007.07.003.

Hassen, Amal. 2025. *Interrogative and exclamative sluices in French and Tunisian Arabic*. University Paris Cité dissertation.

Hassen, Amal & Anne Abeillé. 2025. Preposition omission in French interrogative sluices: Empirical findings and theoretical implications. *Glossa* 10(1). https://doi.org/10.16995/glossa.15197.

Hawkins, John A. 2004. *Complexity and efficiency in grammars*. Oxford: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Hofmeister, Philip. 2007. *Facilitating memory retrieval in natural language comprehension*. Stanford University unpublished PhD dissertation.

Hofmeister, Philip. 2008. The after-effects of linguistic form choice on comprehension. Poster presented at the *21st Annual CUNY Conference on Human Sentence Processing*, Chapel Hill, 13–15 March.

Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes* 26(3). 376–405. https://doi.org/10.1080/01690965.2010.492642.

Hofmeister, Philip, Florian T. Jaeger, Inbal Arnon, Ivan A. Sag & Neal Snider. 2013. The source ambiguity problem: distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28(1/2). 48–87. https://doi.org/10.1080/01690965.2011.572401.

Hofmeister, Philip, Florian T. Jaeger, Ivan A. Sag, Inbal Arnon & Neal Snider. 2007. Locality and accessibility in wh-questions. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 185–206. Berlin: de Gruyter.

Ivanova, Iva, Martin J. Pickering, Holly P. Branigan, Janet F. Mclean & Albert Costa. 2012. The comprehension of anomalous sentences: Evidence from structural priming. *Cognition* 122(2). 193–209. https://doi.org/10.1016/j.cognition.2011.10.013.

Jacobs, Cassandra L., Sun-Joo Cho & Duane G. Watson 2019. Self-priming in production: Evidence for a hybrid model of syntactic priming. *Cognitive Science* 43. e12749. https://doi.org/10.1111/cogs.12749.

Karimi, Hossein, Kumiko Fukumura, Fernanda Ferreira & Martin J. Pickering. 2014. The effect of noun phrase length on the form of referring expressions. *Memory and Cognition* 42(6). 993–1009. https://doi.org/10.3758/s13421-014-0400-7.

Kertz, Laura. 2013. *Discourse expectations and the grammar of ellipsis*. Manuscript. Providence, RI: Brown University.

Kim, Christina S. & Jeffrey T. Runner. 2018. The division of labor in explanations of Verb Phrase ellipsis. *Linguistics and Philosophy* 41(2). 41–85. https://doi.org/10.1007/s10988-017-9220-0.

Kim, Nayoun, Katy Carlson, Michael Walsh Dickey & Masaya Yoshida. 2020. Processing gapping: Parallelism and grammatical constraints. *Quarterly Journal of Experimental Psychology* 73(5). 781–798. https://doi.org/10.1177/1747021820903461.

Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments*. Berlin: Language Science Press.

Levelt, Willem & Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14(1). 7–106. https://doi.org/10.1016/0010-0285(82)90005-6.

Levelt, Willem. 1989. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt & T. Hoffman (eds.), *Advances in neural information processing*, Vol. 19, 849–856. Cambridge, MA: MIT Press.

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. https://doi.org/10.1207/s15516709cog0000_25.

Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10(10). 447–454. https://doi.org/10.1016/j.tics.2006.08.007.

Lohse, Barbara, John A. Hawkins & Thomas Wasow. 2004. Domain minimization in English verb-particle constructions. *Language* 80(2). 238–261. https://doi.org/10.1353/lan.2004.0089.

Lyu, Jun & Elsi Kaiser. 2021. Unpacking the blocking effect: Syntactic prominence and perspective-taking in antecedent retrieval in Mandarin Chinese. *Glossa* 6(1). Article 136. https://doi.org/10.16995/glossa.5781.

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory. *Journal of Psycholinguistic Research* 29(2). 111–123. https://doi.org/10.1023/A:1005184709695.

McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. J*ournal of Memory and Language* 48(1). 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6.

Maes, Alfons, Anja Arts & Leo Noordman. 2004. Reference management in instructive discourse. *Discourse Processes* 37(2). 117–144. https://doi.org/10.1207/s15326950dp3702_3.

Mahowald, Kyle, Ariel James, Richard Futrell & Edward Gibson. 2016. A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* 91. 5–27. https://doi.org/10.1016/j.jml.2016.03.009.

Martin, Andrea. E. & Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language* 64(4). 327–343. https://doi.org/10.1016/j.jml.2010.12.006.

Martin, Andrea E. 2018. Cue integration during sentence comprehension: Electrophysiological evidence from ellipsis. *PLoS ONE* 13(11). Article e0206616. https://doi.org/10.1371/journal.pone.0206616.

Merchant, Jason. 2001. *The syntax of silence.* Oxford: Oxford University Press.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27(6). 661–738. https://doi.org/10.1007/s10988-005-7378-3.

Merchant, Jason, Lynn Frazier, Charles Clifton Jr. & Thomas Weskott. 2013. Fragment answers to questions: A case of inaudible syntax. In Louis Goldstein (ed.), *Brevity*, 21–35. Oxford: Oxford University Press.

Molimpakis, Emilia. 2019. *Accepting preposition-stranding under sluicing cross-linguistically: A noisy-channel approach.* University College London dissertation.

Nairne, James S. 2002. Remembering over the short-term: The case against the standard model. *Annual Review of Psychology* 53(1). 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131.

Nykiel, Joanna. 2013a. Wh-phrases in sluicing: an interaction of the remnant and the correlate. In P. Hofmeister & E. Norcliffe (eds.), *The core and the periphery. Data-driven perspectives on syntax inspired by Ivan A. Sag*, 253–274. Stanford, CA: CSLI Publications.

Nykiel, Joanna. 2013b. Clefts and preposition omission in sluicing. *Lingua* 123(1). 74–117. https://doi.org/10.1016/j.lingua.2012.10.011.

Nykiel, Joanna. 2015. Constraints on ellipsis alternation: A view from the history of English. *Language Variation and Change* 27(2). 227–254. https://doi.org/10.1017/S0954394515000071.

Nykiel, Joanna. 2017. Preposition stranding and ellipsis alternation. *English Language and Linguistics* 21(1). 27–45. https://doi.org/10.1017/S1360674315000477.

Nykiel, Joanna. 2025. Preposition omission under English pseudogapping. *Glossa* 10(1). 1–51. doi: https://doi.org/10.16995/glossa.15203.

Nykiel, Joanna & John A. Hawkins. 2020. English fragments, Minimize Domains, and Minimize Forms. *Language and Cognition* 12(3). 411–443. https://doi.org/10.1017/langcog.2020.6.

Nykiel, Joanna & Jong-Bok Kim. 2022. On the grammaticality of morphosyntactically reduced remnants in Polish sluicing. *Linguistics* 60(1) .177–213. https://doi.org/10.1515/ling-2021-0071.

Nykiel Joanna, Jong-Bok Kim & Rok Sim. 2023. Case-matching effects under clausal ellipsis and the cue-based theory of sentence processing. *Journal of Linguistics* 59(2). 327–360. https://doi.org/10.1017/S0022226722000068

Orth, Wesley & Masaya Yoshida. 2025. Gaps that do not sprout. *Linguistic Inquiry.* https://doi.org/10.1162/ling_a_00551.

Overfelt, Jason. 2024. Having space to sprout: Failed sprouting in sub-clausal ellipses. In Robert Autry, Gabriela de la Cruz, Luis A. Irizarry Figueroa, Kristina Mihajlovic, Tianyi Ni, Ryan Smith & Heidi Harley (eds.), *Proceedings of the 39th West Coast Conference on Formal Linguistics, Somerville, MA.* Cascadilla Proceedings Project.

Öztekin, I. & Brian McElree. 2007. Retrieval dynamics of proactive interference: PI slows retrieval by eliminating fast assessments of familiarity. *Journal of Memory and Language* 57(1). 126–149. https://doi.org/10.1016/j.jml.2006.08.011.

Parker, Dan. 2018. A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa* 3(1). 129. https://doi.org/10.5334/gjgl.621.

Philippova, Tatiana. 2014. P-omission under sluicing, [P clitic] and the nature of P-stranding. In Martin Kohlberger, Kate Bellamy & Eleanor Dutton (eds.), *Proceedings of ConSOLE XXII*, 133–155.

Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4). 633–651. https://doi.org/10.1006/jmla.1998.2592.

Pickering, Martin J., & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2). 169–190. https://doi.org/10.1017/S0140525X04000056.

Poppels, Till & Philip Miller. 2023. Remnant connectivity in pseudogapping: Experimental evidence for a direct generation approach. *Journal of Linguistics* 59(2). 293–325. https://doi.org/10.1017/S0022226722000354.

Reitter, David, Frank Keller & Johanna D. Moore. 2011. A computational cognitive model of syntactic priming. *Cognitive science* 35(4). 587–637. https://doi.org/10.1111/j.1551-6709.2010.01165.x.

Rodrigues, Cillene, Andrew Nevins & Luis Vicente. 2009. Cleaving the interactions between sluicing and preposition stranding. In D. Torck and L. Wetzels (eds.), *Romance languages and linguistic theory 2006*, 175–198. Amsterdam: John Benjamins Publishing Company

Scheepers, Christoph. 2003. Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition* 89(3). 179–205. https://doi.org/10.1016/S0010-0277(03)00119-7.

Stjepanovic, Sandra. 2008. P-stranding under sluicing in a non-P-stranding language? *Linguistic Inquiry* 39(1). 179–190. https://doi.org/10.1162/ling.2008.39.1.179.

Stjepanovic, Sandra. 2012. Two cases of violation repair under sluicing. In Jason Merchant & Andrew Simpson (eds.), *Sluicing. Cross-linguistic perspectives*, 68–82. Oxford: Oxford University Press.

Sturt, Patrick, Frank Keller & Amit Dubey. 2010. Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language* 62(4). 333–351. https://doi.org/10.1016/j.jml.2010.01.001.

Szczegielniak, Adam. 2008. Islands in sluicing in Polish. In Natasha Abner & Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project*, 404–412. Somerville, MA, USA.

Szmrecsanyi, Benedikt M. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1). 113–149. https://doi.org/10.1515/cllt.2005.1.1.113.

Tooley, Kristen M. 2020. Contrasting mechanistic accounts of the lexical boost. *Memory and Cognition* 48(5). 815–838. https://doi.org/10.3758/s13421-020-01019-3.

Troyer, Melissa, Philip Hofmeister & Marta Kutas. 2016. Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in Psychology* 7, art. 374. https://doi.org/10.3389/fpsyg.2016.00374.

Van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55(2). 157–166. https://doi.org/10.1016/j.jml.2006.03.007.

Van Dyke, Julie A. 2011. Cue-dependent interference in comprehension. *Journal of Memory and Language* 65(3). 247–263. https://doi.org/10.1016/j.jml.2011.05.002.

Van Dyke, Julie A. & Clinton L. Johns. 2012. Memory interference as a determinant of language comprehension. *Language and Linguistics Compass* 6(4). 193–211. https://doi.org/10.1002/lnc3.330.

Watkins, Olga C. & Michael Watkins. 1975. Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory* 1(4). 442–452. https://doi.org/10.1037/0278-7393.1.4.442.

Wiechmann, Daniel & Arne Lohmann. 2013. Domain minimization and beyond: Modeling prepositional phrase ordering. *Language Variation and Change* 25(1). 65–88. https://doi.org/10.1017/S0954394512000233.

ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 3, Article 6
https://doi.org/10.18148/zs/2026-2015

Max Bonke, Universität zu Köln, Germany, mbonke@uni-koeln.de

Pragmatic constraints on the complementizer *que* 'that' in Spanish embedded gapping

Abstract: Spanish, unlike English, allows the declarative complementizer *que* 'that' in contexts where the finite verb has been elided via gapping. However, the acceptability of *que* in Spanish gapping varies systematically with the type of embedding verb: some verbs permit it readily, while others show a marked degradation in acceptability. Previous work hypothesized that factivity might account for these verb-type effects, but the evidence has been based on inferences from acceptability patterns and lacks support from reliable quantitative data. This paper provides a quantitative and experimental investigation of verb-type constraints on *que* in embedded gapping. A correlation analysis reveals that the degradation of *que* aligns with independently collected measures of verb factivity. However, this correlation is driven by a subset of verbs, i.e., saying and thinking verbs, rather than representing a general pattern. Two experiments further examine the interpretive impact of *que*, testing whether its presence modulates factive inferences and exploring possible connections to additional pragmatic properties (assertion, at-issueness, reportiveness). In both experiments, the presence or absence of *que* has no systematic effect on interpretation. These results reinforce previous rejections of factivity-based accounts; however, the precise source of verb-type effects on *que* in Spanish gapping remains elusive.

Keywords: gapping, clause embedding, complementizer, factivity, quantitative pragmatics

Cite as: Bonke, Max. 2026. Pragmatic constraints on the complementizer *que* 'that' in Spanish embedded gapping. *Zeitschrift für Sprachwissenschaft* 45(3). https://doi.org/10.18148/zs/2026-2015.
Received: 16 September 2025; **accepted:** 27 August 2026; **published:** 16 November 2026
Published by: publikon / dgfs (Linguistic Society of Germany) under a Creative Commons 4.0 BY license

1 Introduction

Gapping[1] is a type of ellipsis that deletes the finite verb in the second of two coordinate clauses (plus, optionally, further identical constituents) (Ross 1968, 1970; Sag 1977; Hankamer 1979; Johnson 2009, 2018, among many others). An example of gapping is in (1), where the underscore indicates the position of the elided verb *lives*.

(1) *Mary lives in England and John _ in Wales.*

As has long been observed, gapping is subject to structural and discourse restrictions. One syntactic domain in which such restrictions are at play is in clausal embedding. When coordinated clauses with gapping serve as the complement of an embedding predicate, only the first conjunct may host a complementizer (e.g., *that*), but not the second, elliptical conjunct. (Fiengo 1974; Wilder 1994, 1997; Hartmann 2000; Taraldsen 2000; Repp 2009; Broekhuis & Corver 2019, among others). This property is illustrated in (2), where the gapping structure in (1) above constitutes the clausal complement of the embedding verb *think*. This structure is only grammatical without the second *that*.

(2) *Anne thinks that Mary lives in England and* (**that*) *John _ in Wales.*

The fact that the complementizer is ungrammatical in the second conjunct is clearly a property of gapping, since in non-elliptical coordinate embedded clauses the complementizer is grammatical; see (3).

(3) *Anne thinks that Mary lives in England and that John lives in Wales.*

In English, there is thus a categorical, ellipsis-specific ban on complementizers from clauses with gapping. Cross-linguistically, however, the possibility of complementizers in gapping is a matter of variation. Other languages with comparable restrictions to English include German (Wilder 1994, 1997; Hartmann 2000; Repp 2009) and Dutch (Broekhuis & Corver 2019), but there are also languages that allow complementizers in gapping (for an overview, see Bonke 2024: Ch. 2).

Among the languages that allow gapping with complementizers is Spanish, as shown in the actual use example in (4). This example contains a coordinate structure with an elliptical second conjunct (*los mexicanos sus lacayos*, lit. 'the Mexicans his lackeys', with elided copula verb *son* 'are'). This coordinate structure functions as the complement of the clause-embedding verb *cree* 'thinks'. Importantly, the elliptical conjunct features the complementizer *que* 'that'.

[1] This paper consists of revised sections of my dissertation (Bonke 2024: Chapters 4/5). The data and analyses for the two experiments presented in this paper are available on OSF: https://osf.io/r4shg.

(4) Spanish
El cree que el mundo es su empresa y ***que*** *los mexicanos _ sus lacayos.*
He thinks that the world is his company and that the Mexicans his lackeys
'He thinks that the world is his company and that the Mexicans are his lackeys.'
(corpus example from Bonke & Repp 2022: 528)

While Spanish allows complementizers in gapping as shown, they are not obligatory. Embedded gapping without the complementizer in the elliptical conjunct (like in English) is also possible; the sentence in (4) above would be equally grammatical without *que* in the second conjunct (*… y los mexicanos sus lacayos*, lit. 'and the Mexicans his lackeys').

This alternation of gapping with vs. without the complementizer in Spanish raises the question as to whether the presence/absence of *que* might be constrained, for example, by grammatical properties or properties of the discourse context. In an acceptability rating study, Bonke & Repp (2022) find constraints on the type of the embedding verb (e.g., *cree* 'thinks' in (4) above). Their results indicate that embedded gapping without *que* is equally acceptable across (types of) embedding verbs, but the variant with *que* is degraded in acceptability with some verb types (e.g., emotives like *lamentar* 'regret', or knowledge verbs like *saber* 'know'). Bonke & Repp (2022) attempt to establish a connection between these acceptability patterns and pragmatic properties of the verbs, focusing on factivity. The hypothesis motivating their experiments is that those verbs that easily allow for gapping with *que* are non-factive, while those that penalize *que* in gapping are factive.

Due to some unexpected patterns in the acceptability data (which are described in detail in Section 2.2), Bonke & Repp (2022) ultimately reject the hypothesis that factivity is crucial in the acceptability of gapping with *que*. However, as I argue in this paper, the empirical case for their conclusion is insufficiently supported. Acceptability data only allows for indirect inferences on the (presumed) interaction of *que* in gapping and clause meaning, which forms a cornerstone of Bonke & Repp's (2022) structural analysis of gapping. Furthermore, Bonke & Repp's (2022) categorization of verbs as factive vs. non-factive relies largely on introspection rather than robust quantitative evidence, which introduces a potential source of misclassification. These empirical issues motivate a re-examination of Bonke & Repp's (2022) gapping data under a more sophisticated, empirically validated treatment of factivity.

In this paper, I return to the factivity hypothesis, putting an empirical focus on the meaning of *que* in gapping. I present a new analysis of existing data on Spanish, which correlates Bonke & Repp's (2022) acceptability data with independently collected quantitative data on the interpretation of factivity for the relevant clause-embedding verbs (Bonke 2024: Ch. 5). I also present the results of two experiments that target the pragmatic contribution of *que* in gapping. The first experiment (Experiment 1) tests whether the presence of *que* in the elliptical conjunct affects

the interpretation of factivity. The results of the correlation analysis and Experiment 1 indicate no meaningful connection between factivity and *que* in gapping, supporting Bonke & Repp's (2022) conclusion with clear and direct empirical evidence. Due to these null results, I conducted the second experiment (Experiment 2), which applied novel diagnostics to gapping clauses with vs. without *que* to explore the hypothesized pragmatic differences more generally. Due to the novel diagnostics and exploratory nature of Experiment 2, the results are to be treated with caution; however, the results parallel the results of Experiment 1 in suggesting no effect of the complementizer on pragmatics.

This paper is structured as follows. Section 2 provides background information on factivity and outlines Bonke & Repp's (2022) study design, findings, and analysis in more detail. Section 3 contains the correlation analysis of the acceptability data and quantitative data on the factivity of the relevant clause-embedding verbs from Bonke (2024: Ch. 5). Section 4 contains Experiment 1 on factivity, Section 5 contains the exploratory Experiment 2, and Section 6 concludes.

2 Gapping, complementizers, and verb-type effects

This section goes into detail on the acceptability patterns of the Spanish complementizer *que* 'that' in embedded gapping. For an example of this structure, see (4) in the introductory section above; further examples follow in Section 2.2 below. Specifically, this section outlines how the acceptability of *que* varies depending on the type of the embedding verb. The basis of this discussion is the acceptability study by Bonke & Repp (2022), who hypothesize that a particular property of embedding verbs, factivity, is crucial for the acceptability of *que* in gapping (but end up rejecting this hypothesis). I start with a brief overview of classical approaches on factivity (Section 2.1), outlining the facts on English, which I assume generalize to Spanish without issue. After that, I turn to Spanish and the verb-type effects on *que* in embedded gapping in Section 2.2.

2.1 Factivity and factive sub-types

Kiparsky & Kiparsky (1970) first described the concept of factivity. They divided clause-embedding predicates into two groups, depending on the discourse status of their complement clause. Factive predicates presuppose the propositional content of their complement clause, while non-factive predicates do not presuppose the propositional content of their complement clause. To illustrate the difference, consider the difference between examples (5) and (6).

(5) *I regret that it is raining.* (Kiparsky & Kiparsky 1970: 147)

(6) *I believe that it is raining.*

When a speaker utters the sentence in (5), they take the proposition *it is raining* for granted and as uncontroversial in discourse. In more technical terms, they assume the propositional content of the complement clause (i.e., that it is raining) to be part of the common ground (Stalnaker 1978, 2002). This presupposed discourse status of the content of the complement clause means that the predicate that embeds it, i.e., the verb *regret*, is factive.

In contrast, someone that utters the sentence in (6) does not presuppose the content of the complement clause, i.e., they do not take the proposition *it is raining* for granted and as uncontroversial in discourse. The content of the complement clause is thus not treated as part of the common ground. Consequently, the embedding predicate *believe* is non-factive.

A core property of factive presuppositions (as of any presupposition) is that they "survive" in entailment-canceling environments. A sentence provides such an environment if it contains an element or structure that manipulates the truth conditions of an equivalent simple, affirmative sentence such that the entailments within that simple sentence disappear. The sentences in (7) to (9) show what this means based on two such elements/structures, negation and question formation. The truth conditions of the simple assertive sentence in (7) are fundamentally different from its negated version in (8) or the polar question in (9). Entailments of the simple sentence (7), e.g., *Someone regrets something*, are not present in (8) and (9). However, like the simple sentence in (7), the sentences in (8) and (9) still contain the presupposition that it is raining, triggered by the factive verb *regret*. This "survival" of the factive presupposition is referred to as "projection": the presupposition projects out of the entailment-canceling environment.

(7) *I regret that it is raining.* [Simple sentence]
Presupposes: *It is raining.*

(8) *I don't regret that it is raining.* [Negation]
Presupposes: *It is raining.*

(9) *Does anybody regret that it is raining?* [Polar question]
Presupposes: *It is raining.*

While the criterion for factivity is relatively clear (presupposition plus projection), it is not always trivial to say whether this criterion applies to an individual clause-embedding predicate. I will demonstrate this by means of an example below. As a result, the classification of a predicate as factive or non-factive may be tricky. Different contributions in the literature on factivity may put the same predicates into different categories. For a general overview of these inconsistencies in classification, I refer to Degen & Tonhauser (2022) and references therein. Here, I will focus on one proposed solution to these difficulties in classification, which splits the class of factive predicates into two sub-classes, semi-factives and true factives (Karttunen 1971),

depending on how consistently they display presupposition projection. This split may be one of several approaches to inconsistent projection patterns; however, it is especially relevant in this paper because it informs Bonke & Repp's (2022) classification of clause-embedding verbs.

The split of factive predicates into true factive and semi-factive sub-types was proposed by Karttunen (1971). Within this split, true factives are assumed to be consistent and invariant in presupposition projection, while semi-factives show variability (see Karttunen 1971; Hooper & Thompson 1973; Hooper 1975 and much subsequent literature). As an example for a semi-factive predicate, Karttunen (1971) gives the verb *discover*; see (10) and (11). Karttunen (1971) argues that *discover* projects a factive presupposition from under negation as in (10), but not from under a polar question as in (11). Thus, *discover* shows variable projection patterns, which differentiates it from other predicates like *regret*, which are assumed to show consistent projection in all entailment-canceling environments. For examples, see (8) and (9) above; cp. also Karttunen (1971) on *regret* specifically. The term "true factive" to refer to consistently presupposing predicates like *regret* is from Hooper & Thompson (1973) and Hooper (1975).

(10) *John didn't discover that he hadn't told the truth.*
Presupposes: *John hadn't told the truth.* (Karttunen 1971: 63)

(11) *Did you discover that you had not told the truth?*
Does not presuppose: *You had not told the truth.* (Karttunen 1971: 63)

This classification of factive predicates raises the question as to whether the split by presupposition behavior correlates with a difference in lexical semantics. Do semi-factives share any common features apart from their variable projectivity? And conversely, do true factives share any common features apart from their (assumed) consistent projectivity? Bonke & Repp (2022) assume that semi-factives are verbs of knowledge and perception (such as *know*, *find out*, *see*), and true factives are emotives (such as *regret*, *hate*, *disapprove*). This assumption has ample precedence in the literature (Hooper & Thompson 1973; Hooper 1975; Beaver 2010; Bîlbîie & de la Fuente 2019, among others).

2.2 Factivity and *que* 'that' in embedded gapping (Bonke & Repp 2022)

Bonke & Repp (2022) derive their hypothesis that factivity and the factive sub-types may be crucial for *que* in gapping from the existing literature on Spanish clausal complementation, where factivity and related properties are sometimes adduced to account for syntactic features in complement clauses (e.g., Demonte & Fernández-Soriano 2009; de Cuba & MacDonald 2013; Bîlbîie & de la Fuente 2019, among others; see also Haegeman 2006; Kastner 2015 on English). As we will see in this section, their experimental results suggest that the type of embedding

verb indeed plays a major role in the acceptability of *que* in gapping. At the same time, the results do not confirm that factivity is the crucial property, leading Bonke & Repp (2022) to reject the factivity hypothesis.

Bonke & Repp's (2022) acceptability study consists of four acceptability judgment experiments. They tested sentences with embedded gapping such as (12) and (13), which varied in the presence vs. absence of *que*, and in the type of embedding verb (e.g., *cree* 'thinks' in (12) vs. *desaprueba* 'disapproves' in (13)). Participants were asked to judge how natural they thought the sentences were within a provided linguistic context (1–2 sentences).

(12) *Santiago* ***cree*** *que el rey fum-a cada día y* { *que* | Ø } *la reina, a veces.*
Santiago thinks that the king smokes-IND every day and that Ø the queen at times
'Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

(13) Santiago ***desaprueba*** *que el rey fum-e cada día y* { *que* | Ø } *la reina, a veces.*
Santiago disapproves that the king smokes-SUBJ every day and that Ø the queen at times
'Santiago disapproves of the fact that the king smokes every day and that the queen smokes sometimes.'

The verbs varied in factivity (non-factive like *cree* vs. factive like *desaprueba*) and mood selection (indicative vs. subjunctive). The difference in mood selection shows in the examples above, where in (12) the embedding verb *cree* 'thinks' selects a complement clause in which the predicate appears in its indicative form (i.e., *fum-a* 'smokes-IND'). In (13), the embedding verb *desaprueba* 'disapproves' selects an embedded predicate in the subjunctive form (i.e., *fum-e* 'smokes-SUBJ').

The study contained the four verb types in Table 1. As discussed in Section 2.1, the factives split into semi-factives and true factives, a distinction that per assumption aligns with different lexical semantic classes: knowledge and perception verbs vs. emotives, respectively. These classes also differ in mood selection preferences, with knowledge and perception verbs favoring indicative complements and emotives favoring subjunctive complements. Non-factives are likewise subdivided by lexical class and mood preference. One set in the sample consists of indicative-preferring non-factives, represented by saying and thinking verbs, which are prototypical non-factives. The other set consists of subjunctive-preferring non-factives, encompassing two lexical semantic classes: desideratives (e.g., *want*, *wish*, *hope*) and verbs with implicit, lexicalized negation (e.g., *deny*, *doubt*, *prohibit*).[2]

[2] Some negation verbs, e.g., *deny* or *doubt*, belong to the class of response stance verbs, which presuppose that the propositional content of their complement is under debate in the discourse (Cattell 1978; Kastner 2015). This presupposition differs from factive presupposition, whereby the content is taken to be true and uncontroversial in discourse.

Table 1: The four verb types in Bonke & Repp (2022)

Factivity	Mood selection	Lexical semantic classes	Examples
Non-factive	Indicative	Verbs of saying and thinking	*decir* 'say' *creer* 'think'
Non-factive	Subjunctive	Desideratives, negation verbs	*querer* 'want' *negar* 'deny'
Factive	Indicative	Verbs of knowledge and perception (= semi-factives)	*saber* 'know' *ver* 'see'
Factive	Subjunctive	Emotives (= true factives)	*lamentar* 'regret' *alabar* 'praise'

Bonke & Repp (2022) hypothesize that non-factive predicates easily allow *que* in gapping, while factives do not allow *que* as easily.[3] This means that with a non-factive embedding verb like *cree* 'thinks' in (12) above, acceptability ratings for the variant with *que* in the gapping conjunct should be comparable to the ratings without *que* (and both should be within the acceptable range). With a factive embedding verb like the emotive *desaprueba* 'disapproves' in (13) above, ratings should be lower with *que* than without *que*. They furthermore hypothesize a difference between the factive sub-types, with semi-factives (knowledge/perception verbs) showing a smaller effect than the true factives (emotives) and thus occupying an intermediate position between non-factives and true factives. The two types of non-factives, i.e., the indicative-taking verbs of saying and thinking on the one hand and the subjunctive-taking desideratives and negation verbs on the other, are not expected to differ.

The experimental results are summarized in Table 2. Gapping without the complementizer received high acceptability ratings across the board and thus serves as the baseline for a by-verb-type comparison to gapping with *que*.[4] The results partially align with the hypotheses: as expected, the indicative-taking non-factives (saying/thinking verbs) overall show no significant difference with vs. without *que* (and ratings indicate high acceptability in general). Also as expected, subjunctive-taking factives (emotives, i.e., true factives) show lower ratings with *que* than without *que*; however, the ratings with *que* are not so low as to indicate outright

[3] The reasons to hypothesize that these verb types show different patterns w.r.t. the acceptability of *que* in gapping stem from observations from the literature on Spanish clause syntax. Generally, complement clauses of factives seem to be more restricted as to the syntactic structures they may contain than complement clauses of non-factives (see e.g., Demonte & Fernández-Soriano 2009: 46 f.; de Cuba & MacDonald 2013: 312, and esp. quantitative evidence from Bîlbîie & de la Fuente 2019). For the sake of conciseness, I will not go into detail on these structural properties of factive complements.

[4] Judgments were given using a slider on a continuous scale (output coding: 0–100). I will keep my discussion of Bonke & Repp's (2022) results on a coarse-grained level and therefore not provide specific numbers. However, in the correlation analysis in Section 3, I use their numeric results.

ungrammaticality. The indicative-taking factives (knowledge/perception verbs, i.e., semi-factives) occupy the expected intermediate position between the subjunctive-taking factives and the (indicative-taking) non-factives. Like the true factives, they show a degradation with *que* compared to without *que*, which, however, is smaller on average than the degradation observed for the true factives and does not show as consistently across experiments. The only verb type that does not pattern as expected is the subjunctive-taking non-factives. Recall that these verbs are predicted to be like the indicative-taking non-factives, i.e., to show no effect of *que*. However, these verbs do show a degradation with *que*, which is comparable in size to that found for the true factives.

Table 2: Summary of Bonke & Repp's (2022) results for the acceptability of embedded gapping with vs. without the complementizer *que* 'that' by the four verb types

Verb type	**Lexical semantic classes**	**Acceptability without *que***	**Acceptability with *que***	**Results as predicted?**
Non-factive + indicative	Saying/thinking	High	High	Yes
Non-factive + subjunctive	Desideratives/ negation verbs	High	**Degraded**	**No**
Factive + indicative	Knowledge/perception (= semi-factives)	High	(Mildly) degraded	Yes
Factive + subjunctive	Emotive (= true factives)	High	Degraded	Yes

It is important to note that these patterns are specific to gapping: in corresponding non-elliptical structures (*…and (that) the queen* smokes *every day*), no degradation of the complementizer was found with subjunctive-taking factives, i.e., the verb type that displays such a degradation most consistently in gapping.

Regarding gapping, Bonke & Repp (2022) furthermore provide a fine-grained examination of the individual verbs within the four types (i.e., 12 verbs per type). This examination shows that their four-way classification of verb types does not ideally capture the variability in the acceptability of gapping with *que*. From the perspective of individual verbs, none of the four verb types exhibits a uniform pattern; instead, distinctions between the types are gradual and emerge only as tendencies. For example, while the subjunctive-taking factives as a type show the largest average degradation with *que*, some of the individual subjunctive-taking factives show considerably smaller degradations, patterning more with the indicative-taking non-factives. Conversely, some of the indicative-taking non-factives show a noticeable degradation, even though taken together as a type they do not show this degradation.

In sum, Bonke & Repp's (2022) results suggest that the acceptability of *que* in gapping is constrained by properties of the embedding verb. When considering the ratings by the four verb types, it is one verb type in particular, indicative-taking non-factives (i.e., saying/thinking verbs), that allows *que* without issue. The other three verb types show different degrees of degradation with *que*.

To account for these findings, Bonke & Repp (2022) propose a syntactic analysis, which assumes a structural difference between gapping with vs. without *que*. According to this proposal, overt *que* indicates a clausal structure of the elliptical conjunct (as opposed to a smaller, sub-clausal structure without *que*). By hypothesis, the clausal structure in this proposal carries specific pragmatic meaning, which is compatible with the meaning of those embedding verbs that show little or no degradation of *que* (i.e., saying/thinking verbs), but clashes with the meaning of the other verbs. This clash in meaning comes at the cost of varying degrees of degraded acceptability, depending on how easily the required interpretation can be constructed with the individual embedding verb. Assuming that the pragmatic property is (non-)factivity, this approach derives the degradation of gapping with *que* under, for example, an emotive embedding verb from the difficulty with which a non-factive sentence meaning could be constructed with such a normally factive verb. Bonke & Repp (2022), however, reject their initial hypothesis that the crucial pragmatic meaning is factivity since the two non-factive sub-types in their experiments do not pattern alike and verb-by-verb analysis reveals heterogeneous effects within types and gradual rather than categorical differences across types.

Nevertheless, it may be premature to reject the factivity-based hypothesis altogether. Such a hypothesis may be coherent with Bonke & Repp's (2022) results if one assumes that their categorization of verbs as factive or non-factive was incorrect in a substantial number of cases. A case in point is the fact that verbs with implicit negation, which formed part of the (subjunctive-taking) non-factives in the experiments, are treated as factive in parts of the literature (e.g., Meinunger 2017). Furthermore, an altogether more sophisticated and nuanced view of factivity may be called for. Recent quantitative research on factivity (Degen & Tonhauser 2022) suggests that factivity is a gradual property rather than a categorical one as often assumed. Coupled with the assumption that some of the verbs were categorized incorrectly, the gradualness of factivity may account for the gradualness in the acceptability data. I pursue this hypothesis in the following section by means of a correlation analysis that connects Bonke & Repp's (2022) acceptability data with quantitative factivity data from Bonke (2024: Ch. 5). Additionally, I experimentally investigate the question of whether the presence of *que* in gapping influences the interpretation of factivity in Section 4.

Alternatively, we may follow Bonke & Repp (2022) in rejecting the factivity-based hypothesis in light of their results. In this case, the question arises as to what determines acceptability instead. Since the type of verb does play a crucial role, an obvious hypothesis would be that instead of

factivity, a different meaning component is the determining factor. This component should relate closely to the indicative-taking non-factives (saying/thinking verbs), since these form the only verb type that showed a negligible effect of *que* in gapping. Bonke & Repp's (2022) analysis focuses on assertion embedding in this regard; Bonke (2024: Ch. 3) mentions at-issueness and reportiveness as further potential meaning components. I explore these options in Section 5.

3 Does the acceptability of *que* 'that' in gapping correlate with the factivity of the clause-embedding verb?

In this section, I connect the acceptability data from Bonke & Repp (2022) on the complementizer *que* 'that' in embedded gapping to independently obtained, quantitative data on factivity from Bonke (2024: Ch. 5). On the basis of the individual embedding verbs, I investigate whether the degree to which *que* is degraded correlates with an experimentally obtained measure of factivity. This means that both measures must receive a numerical value for each embedding verb. For the *que*-degradation, I follow Bonke & Repp (2022) in calculating the difference between the mean acceptability rating in the experimental condition with vs. without *que*. This results in a number between −100 and 100 for every verb,[5] with large positive numbers indicating that the presence of the complementizer leads to a large degradation in acceptability, and small positive numbers indicating that there was a small degradation. Regarding the measure of factivity, I outline Bonke's (2024: Ch. 5) procedure and results in some more detail in the following sub-section, before turning to the correlation analysis of the two measures in Section 3.2.

3.1 Quantitative data on factivity for 44 Spanish verbs (Bonke 2024: Ch. 5)

To obtain a measure of factivity, Bonke (2024: Ch. 5) used (a Spanish translation of) the *certain that*-diagnostic as used by, for example, Tonhauser et al. (2018) and Degen & Tonhauser (2021, 2022). In the experiment, participants judged experimental items such as (14). The items consisted of a sentence containing clausal complementation, which varied as to the embedding verb (sketched in (14) with *afirma* 'claims' and *sabe* 'knows'). The sentences in the items were polar questions, which constituted the necessary entailment-canceling environment to investigate presupposition projection (see Section 2.1).[6] The polar question in turn was the direct speech complement of a speaker (*María*). The experimental materials contained a total of 44 embedding verbs, which also featured in Bonke & Repp's (2022) experiments on embedded gapping.

[5] Bonke & Repp's (2022) collection of acceptability data proceeded on a continuous scale, with the output coded as integers from 0 to 100.

[6] A reviewer points out that the question in (14) shows declarative subject-verb word order, which is possible in Spanish next to verb-subject word order, but may introduce special pragmatic effects. I acknowledge the possibility of such pragmatic effects; however, the crucial property of entailment cancelation is not dependent on word order in interrogatives. Non-inverted word order was used consistently in the experimental items; see Bonke (2024: Ch. 5) for details.

(14) *María: ¿Marco {afirma / sabe / …} que Martina organiza su boda ella sola?*
María: Marco claims knows that Martina organizes her wedding she only
'María: Does Marco claim / know that Martina organizes her wedding all by herself?'

Participants were asked how certain they thought the speaker was that the content of the complement clause was true; see the diagnostic question in (15) for the item in (14) above. They indicated their judgment on a continuous scale with the extremes labeled *no* 'no' and *sí* 'yes' (output coding 1–100). High ratings were thought to indicate high projectivity and therefore high factivity, since participants understood the speaker to take the content of the complement for granted and as uncontroversial in discourse (i.e., presupposed; see Section 2.1 above). Conversely, low ratings were thought to indicate low projectivity, i.e., non-presupposition.

(15) *¿Está María segura de que Martina organiza su boda ella sola?*
is María certain of that Martina organizes her wedding she only
'Is María certain that Martina organizes her wedding all by herself?'

The results show a spread of the verbs along the certainty scale. Some verbs received low certainty ratings (e.g., mean rating of around 30 for *afirma* 'claims' in (14) above) while others received high ratings (e.g., mean rating of around 90 for *sabe* 'knows' in (14) above). This spread along the scale is expected, given the inclusion of (presumed) non-factives (e.g., verbs of saying like *afirma* 'claims') and factives (e.g., knowledge verbs like *sabe* 'knows'). Crucially, however, several verbs did not appear at either end of the scale but instead received intermediate ratings (e.g., *presume* 'presumes', mean rating 57; or *niega* 'denies', mean rating 62). In fact, the results do not overall suggest the existence of well-defined distinct categories (i.e., non-factive vs. factives) at all. Instead, the verbs occupy positions along a spectrum covering the entire scale, which thus blurs the distinction between factive and non-factive.[7]

An interesting finding within these gradual results concerns the verbs with an implicit negation. Recall that Bonke & Repp (2022) classed these as non-factive, while other contributions in the literature treat at least some of them as factives (e.g., Meinunger 2017). In Bonke's (2024: Ch. 5) data, these verbs consistently exhibit intermediate certainty ratings, indicating that they pattern neither with clear non-factives nor with clear factives. The intermediate certainty ratings for this lexical class may explain their inconsistent categorization in the literature. More importantly for the case at hand, the fact that these verbs received unexpectedly high certainty ratings considering Bonke & Repp's (2022) non-factive classification may account for the observed *que*-degradation in their verb type. Recall that the negation verbs, alongside desider-

[7] Degen & Tonhauser (2022) find similar gradual results for English.

atives, formed the subjunctive-taking "non-factives". If enough of these presumed non-factives actually pattern more closely like factives than expected, factivity may still account for the degradation with *que* for their verb type after all.

3.2 Correlating certainty and acceptability ratings

Given that both the certainty ratings from Bonke (2024: Ch. 5) and the effect of gapping-*que* on acceptability as observed in Bonke & Repp (2022) are gradual across the different verbs, a hypothesis regarding their relation may be formulated in gradual terms: the more factive a verb is, the higher its degradation with *que* is expected to be. This hypothesis thus predicts a positive correlation between mean certainty ratings and the average *que*-degradation.

The mean certainty ratings and the average *que*-degradation for the 44 verbs included in both studies are plotted in Figure 1.[8] The vertical axis shows the mean certainty ratings and the horizontal axis shows the average degradation with *que* in embedded gapping (in comparison with the version without *que*). There is one verb, *imagina* 'imagines', with a slightly negative value on the horizontal axis, which means that this verb on average was rated slightly more acceptable with *que* than without.

To test whether the two measures are correlated, I fitted a linear model predicting the average certainty rating per verb from its average degradation with *que* in gapping. The right side of Figure 1 shows the regression line that resulted from this model. The line has a positive slope

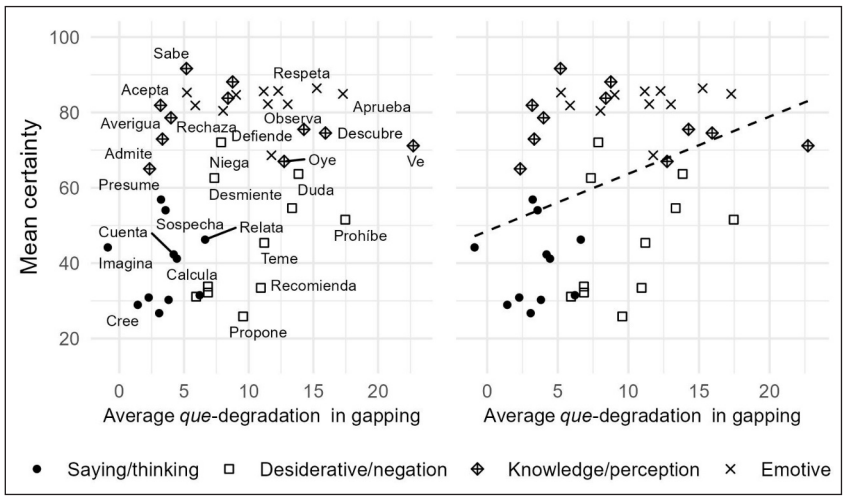


Figure 1: Correlation of the mean degradation of gapping with *que* (Bonke & Repp 2022) and mean certainty rating for the 44 verbs (Bonke 2024: Ch. 5), with labels (left) and regression line (right)

[8] All statistical analyses in this paper were done in R (R Core Team 2026). Figures were generated using ggplot2 (Wickham 2016).

(b = 1.52), which indicates that as predicted, verbs with a higher certainty rating show a higher degradation of *que*. The model indicated that this correlation is statistically significant (p = 0.02). However, the model indicated a low correlation coefficient (adjusted R^2 = 0.11), which means that despite the significant correlation of average certainty with *que*-degradation, certainty can account only for a small part (11%) of the variation within the *que*-degradation data. This lack of explanatory power also shows in the relatively large residuals in Figure 1 (i.e., the distance of the individual data points from the regression line).

The low R^2 is unexpected. If factivity really is the crucial, determining factor for the acceptability of gapping-*que*, this coefficient should be larger and the correlation should be stronger. Despite the statistical significance, the weakness of the correlation casts doubt on the role of factivity for *que*. Another problem results from the role of one particular verb type, the indicative-taking non-factives (= verbs of saying and thinking). Recall that this verb type was the only one that consistently showed only a negligible degradation of *que* (see Section 2.2 above). The low average *que*-effect for this verb type is visible in Figure 1, where the black dots representing the verbs of this type range near the left end of the horizontal scale. This verb type also seems to drive the correlation with the certainty ratings: the black dots are concentrated in an area toward the bottom left in Figure 1, which creates the impression that they are "pulling down" the lower end of the regression line.

To find out whether the indicative-taking non-factives indeed drive the correlation, let us look at the data once more without this verb type. Figure 2 shows the resulting correlation of mean certainty rating and average *que*-degradation in gapping. This figure shows the same data

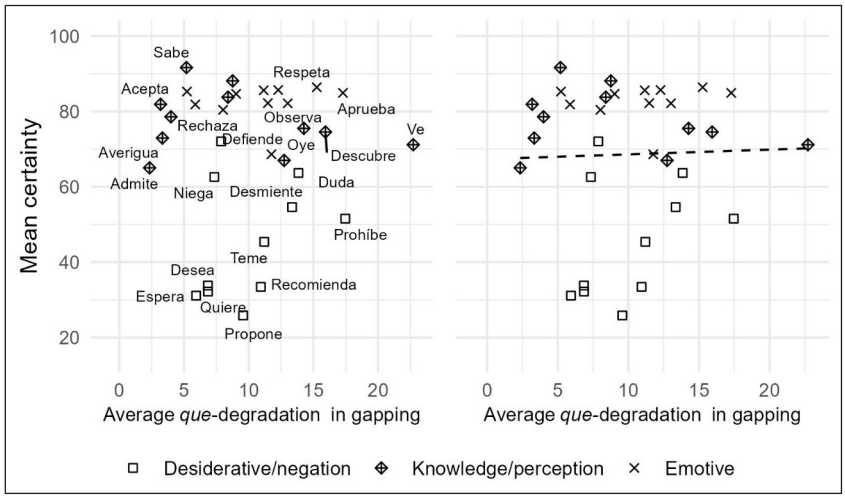


Figure 2: Correlation of the mean degradation of gapping with *que* (Bonke & Repp 2022) and mean certainty rating (Bonke 2024: Ch. 5) without the verbs of saying and thinking, with labels (left) and regression line (right)

as Figure 1 above, the difference being that the black dots representing the indicative-taking non-factives have been taken out. Without these verbs, the figure still shows a large spread of data points in both the vertical and horizontal directions. This means that the data still includes verbs with high and low certainty ratings (vertical axis) and verbs with large and little *que*-degradation (horizontal axis). If factivity determines the acceptability of *que* in gapping, the correlation of certainty ratings and *que*-degradation that we saw above should thus also surface here. However, a linear model fitted to these data indicates a much smaller effect (b = 0.13), which is visualized by the near horizontal regression line on the right side of Figure 2. The linear model furthermore indicated that the effect was not statistically significant (p = 0.87) and that the correlation coefficient was close to zero (adjusted R^2= −0.03). The degradation with *que* thus does not account for certainty ratings in this sub-set of the data, contrary to expectation.

All in all, the correlations of certainty ratings and *que*-degradation presented in this section provide evidence against the hypothesis that factivity is responsible for the *que*-effect in gapping. This conclusion holds even though there is a statistically significant positive correlation of the average certainty rating and *que*-degradation for each verb. The low correlation coefficient, paired with the driving force of the indicative-taking non-factives, shows that certainty, and thus by extension factivity, are unlikely to play a decisive role in the *que*-effect. Instead, the decisive factor is likely an inherent property of the verb type that drives the correlation, i.e., the indicative-taking non-factives (verbs of saying/thinking). This result parallels the conclusions Bonke & Repp's (2022) draw from their acceptability data, which likewise suggests that these verbs are crucial for *que* in gapping, independently of factivity. In the following sections, I try to pin down the special role of the verbs of saying and thinking by presenting two rating experiments on the meaning of *que* in gapping.

4 Experiment 1: Factivity

In the following two sections, I approach the meaning contribution of the complementizer *que* 'that' in Spanish embedded gapping. Specifically, I investigate whether there is a difference in meaning between gapping with *que* and without *que*. The motivation for this procedure stems from Bonke & Repp (2022), who suggest that the verb-type effects they observe may be accounted for under the hypothesis that *que* in gapping is indicative of an aspect of sentence meaning (see Section 2.2). The potential candidates for the relevant sentence meaning they list are (non-) factivity (which is the focus of this section), assertion, at-issueness, and reportiveness (which will be approximated in Section 5).

Bonke & Repp's (2022) reasoning leads to testable hypotheses. If *que* in gapping is associated with non-factive meaning, gapping with *que* is expected to receive a non-factive interpretation across the board. Without *que*, interpretation should be determined by the embedding verb, thus

paralleling the patterns presented in Section 3.1. This predicts that for factive embedding verbs, the effect of gapping-*que* should be strong, lowering certainty ratings to the level of non-factive verbs. Non-factive verbs should receive low ratings, irrespective of the presence of *que*.

The correlation analysis in Section 3.2 suggested that factivity is unlikely to be responsible for the *que*-effect in gapping. However, the comparison of factivity and acceptability was indirect, comparing acceptability data to independent factivity data. From a methodological viewpoint, it is thus justified to return to the issue of factivity once more by means of a more direct approach. Experiment 1 thus applies the *certain that*-diagnostic for factivity (see Section 3.1) to Bonke & Repp's (2022) gapping materials.

4.1 Design, materials, procedure, and participants

Experiment 1 had a 1×2 design, the sole factor being the presence vs. absence of *que* in the elliptical conjunct. For a sample item in both conditions, see (12). This item is identical to an item from Bonke & Repp (2022), with the only difference being that it is presented in a direct speech context. This adjustment of the items was necessary since the *certain that*-diagnostic measures the certainty of the speaker (i.e., *María* in (16), cp. Section 3.1 above).[9] Items were preceded by 1–2 sentences of conversational context.

(16) *María:* *Santiago cree que el rey fuma cada día y* {*que* | Ø} *la reina, a veces.*
María Santiago thinks that the king smokes every day and that Ø the queen at times
'María: Santiago thinks that the king smokes every day and that the queen smokes sometimes.'

The materials contained the 48 embedding verbs of Bonke & Repp's (2022) study and 24 embedded clauses. Each verb combined with two embedding clauses to form a total of 96 sentences like (16), which in the two conditions made a total of 192 test items. The items were distributed across eight lists in a Latin square design, balancing the four verb types. Each list contained 24 test items so that each participant judged each embedded clause once and half of the embedding verbs. In addition to the test items, every list featured ten declarative control items and ten question control items like (17) and (18), which provided baselines for certainty and non-certainty, respectively. Each list thus contained 44 sentences to be judged.

[9] As can be seen in (12), the experimental items did not feature entailment-canceling environments. The use of the *certain that*-diagnostic on these sentences thus raises conceptual difficulties since such environments are crucial for the investigation of (factive) presuppositions (see Section 2.1 above). The certainty ratings for the experimental items may not be straightforwardly equated with factivity, since high certainty ratings may simply reflect the speaker's certainty of their assertion rather than presupposition. However, the ratings for the individual verb types will be consistent with what would be expected for presupposition (see below).

(17) *María: Julio adora el tiempo soleado y fantástico de Gran Canaria.*
María Julio loves the weather sunny and fantastic of Gran Canaria
'María: Julio loves the sunny and fantastic weather of Gran Canaria.'

(18) *María: ¿Elena estudia toda la noche para un examen?*
María Elena studies all the night for an exam
'María: Does Elena spend all night studying for an exam?'

Participants were asked to imagine that they were attending a party hosted by their friend María. María was described as a chatty person who loved talking about other people. Upon entering the kitchen, participants were asked to imagine overhearing a statement (= the experimental items and declarative fillers) or question (= the question fillers) that María made, within the conversational context provided.[10] Participants then indicated how certain they thought María was regarding the content of the elliptical conjunct in case of the target items, or the whole sentence in case of the declarative and question controls. The diagnostic was the *certain that*-diagnostic, e.g., *¿María está segura de que la reina fuma a veces?* 'Is María certain that the queen smokes sometimes?' (see Section 3.1 above for details). To give their ratings, participants used a slider on a continuous scale with the extremes *no* 'no' and *sí* 'yes'. Responses were coded as integers from 1 to 100.

The experiment was an online questionnaire hosted on soscisurvey.de (Leiner 2024). I used prolific.com to recruit monolingual native speakers of European Spanish, who grew up speaking only their native language. In total, 80 participants completed the questionnaire. Only data from participants showing a statistically significant difference between declarative and question controls (assessed via per-participant t-tests) was included in the analysis. All 80 participants met this criterion; therefore, all data entered the analysis.

4.2 Results and discussion

The results are in Figure 3. The ratings for the filler controls show that the *certain that*-diagnostic worked as expected: declarative fillers received high certainty ratings and question fillers received low certainty ratings. The ratings for the two experimental conditions, present

[10] I acknowledge that it may be difficult to imagine being in a situation where you accidentally overhear the utterance someone has made, but where you are simultaneously aware of the conversational context of that utterance. I proceed with the assumption that the participants were not so confused by this slight incoherence in the experimental setup that it affected their ratings tremendously. A further issue with the experimental setup was the varying degrees of degradation of the items in the condition with *que*. These degraded items may have been difficult for participants to interpret regarding speaker certainty. To alleviate this difficulty, the instructions made participants aware that not all of María's utterances may sound very natural, and that they should try to give an intuitive judgment as well as they could.

vs. absent *que*, show hardly any discernible difference, let alone the predicted strong effect.[11] I conclude that *que* in gapping does not determine speaker certainty and, by extension, factivity.

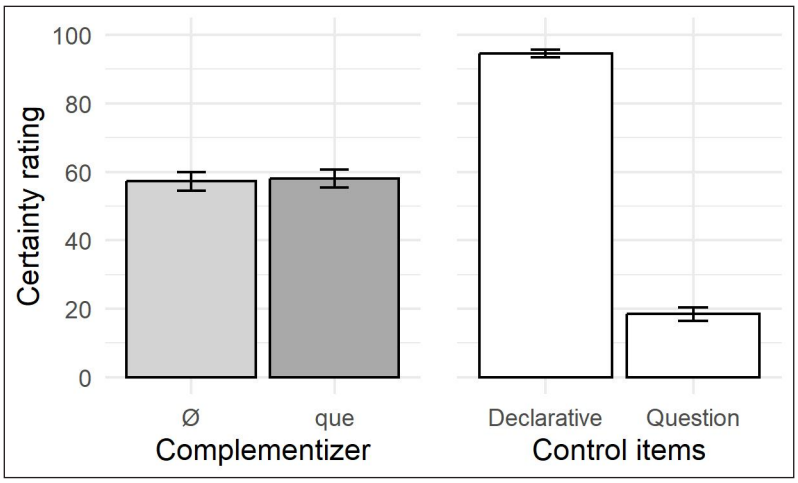


Figure 3: Mean ratings and 95% confidence intervals for experimental conditions and controls (Exp. 1)

A closer look at the four individual verb types reinforces this conclusion. As can be seen in Figure 4, the effect of gapping-*que* on certainty ratings within the verb types is just as negligible as for the overall data. The two factive verb types received higher certainty ratings than the two non-factive types, which suggests that the diagnostic likely picked up on factivity, at least in the items with embedded gapping, where the diagnostic targeted the complement clause.

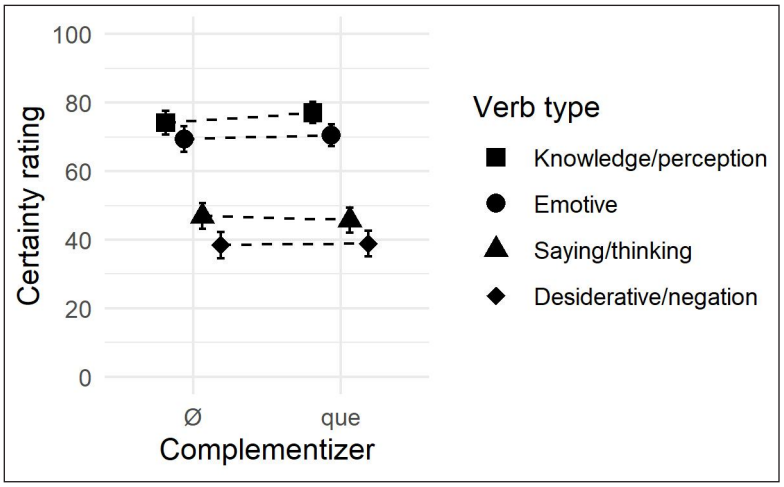


Figure 4: Mean ratings and 95% confidence intervals by verb type and experimental condition (Exp. 1)

[11] Since the size of the effect (or rather lack thereof) sufficiently illuminates the role of *que* in the interpretation of speaker certainty, I did not conduct an inferential analysis on the data. I do not believe the results of such an analysis would have yielded any insights into the questions at hand.

5 Experiment 2

Experiment 1 strongly suggested that the presence of the complementizer *que* 'that' in Spanish embedded gapping has no effect on the interpretation of factivity. This result raises the question of whether *que* affects meaning at all. To explore this issue, Experiment 2 takes a more general approach by applying further pragmatic diagnostics to gapping, which unlike the *certain that*-diagnostic for factivity, have no precedence in the literature. Thus, in Experiment 2, no robust diagnostics target any one specific pragmatic property. Instead, multiple diagnostics are applied in a coarse-grained approach to see whether any effect of *que* on interpretation can be recorded at all, which gives Experiment 2 an exploratory character.

5.1 Experimental materials, diagnostics, and predictions

The 1×2 factorial design and the crucial test items (i.e., embedded gapping with vs. without *que* and different embedding verbs) in Experiment 2 were identical to those of Experiment 1. See Section 4.1 above for a description and see example (16) in that section for a sample item. For each experimental item, there were three statements targeting the interpretation of the elliptical conjunct in the embedded clause. Participants indicated whether they agreed or disagreed with the three statements by clicking a box labeled *sí* 'yes' or another labeled *no* 'no'. For the example item in (16) above, participants gave judgments on the three statements in (19) to (21).

(19) *Possible that not*-diagnostic
Es posible que la reina no fume a veces.
is possible that the queen NEG smokes at times
'It is possible that the queen does not smoke sometimes.'

(20) Importance diagnostic
Lo más importante que dice María es que la reina fuma a veces.
the most important that says María is that the queen smokes at times
'The most important thing that María is saying is that the queen smokes sometimes.'

(21) Reportiveness diagnostic
Santiago le contó algo nuevo a María, concretamente que la reina fumaba a veces.
Santiago her told something new to María specifically that the queen smoked at times
'Santiago told María something new, specifically that the queen smokes at times.'

As mentioned above, the three diagnostics are not standard measurements of any linguistic property. However, they offer an approximation of properties that characterize verbs of saying and thinking, i.e., those verbs that consistently show no degradation of *que* in gapping (Bonke & Repp

2022, cp. Section 2.2 above) and drive the correlation of acceptability and certainty as discussed in Section 3.2. These verbs have been proposed to embed assertive (see Hooper & Thompson 1973; Hooper 1975, and much subsequent work) and at-issue content. These two interconnected properties are targeted by the *possible that not-* and importance diagnostics. The *possible that not*-diagnostic simulates the rejection of the content of the elliptical clause in dialogue, i.e., the potential for an interlocutor to reply with, e.g., *No, that's not true, the queen does not smoke sometimes.*[12] The potential to be targeted for rejection has been connected to asserted content, e.g., by Farkas & Bruce (2009: 82). The importance diagnostic probes whether the content of the elliptical clause represents the main point of the utterance, which captures at-issueness in an intuitive way, see, for example, Simons et al. (2010), who use the term "main point" in a pre-theoretical description of this concept. The reportiveness diagnostic straightforwardly relates to the lexical semantics of saying verbs and extends to many instances of thinking verbs in actual use.

The general reasoning of Experiment 2 follows that of Experiment 1: if gapping with *que* contributes its own meaning component, it should yield a consistent interpretation across verb types (as measured by the proportions of *yes* vs. *no* responses). If this meaning is compatible with verbs of saying and thinking, the presence or absence of *que* should not affect interpretation for these verbs. For the other verb types, gapping with *que* is expected to pattern like saying/thinking verbs, whereas without *que*, interpretations may differ. In other words, saying/thinking verbs should show no effect of *que*, while other verbs should match them with *que* and potentially diverge without it.

An analysis of the filler items included in the experiment revealed that the three diagnostics were likely not ideal to measure specific pragmatic properties like assertion or at-issueness. An in-depth description and critical discussion of the filler items are beyond the scope of this paper, but see Bonke (2024: Ch. 5) for details. Some of the inadequacies of the diagnostics will show in the discussion of the results. Nonetheless, the results are informative regarding the research question of this paper: if *que* in gapping indeed carries meaning, at least one of the diagnostics should suggest an effect for at least some of the verb types, even if it does not perfectly capture the specific pragmatic it targets. I will therefore proceed in describing the experiment using descriptive terms, i.e., without referring to the technical notions at-issueness or assertion.

Experimental items were distributed across eight lists (Latin square). In addition to 24 experimental items, each list contained 16 filler items, so that each participant gave three judgments for 40 sentences. The experiment was an online questionnaire hosted on soscisurvey.de, with participant recruitment via prolific.com. The data of 76 native speakers of European Spanish entered the analysis.

[12] Due to the experimental goal of collecting multiple ratings for each item, it was not feasible to present items in dialogue, i.e., in a format in which the content of the gapping clause was rejected by an interlocutor in discourse.

5.2 Results

The results for the three diagnostics are given below in Figures 5 (*possible that not*), 6 (importance), and 7 (reportiveness). As can be seen in these figures, the effect of *que* on the interpretation of these properties is very small across the board. For all three properties, these small effects obtain for the total data including all verbs (left sides of the figures), as well as for each of the four individual verb types (right sides of the figures). Even for those effects that seem to be stronger than others (e.g., the importance ratings for the saying/thinking verbs in Figure 6 below), the error bars indicating the 95% confidence intervals overlap by a considerable margin. These results do not suggest that *que* in gapping has a strong contribution to pragmatic meaning.[13]

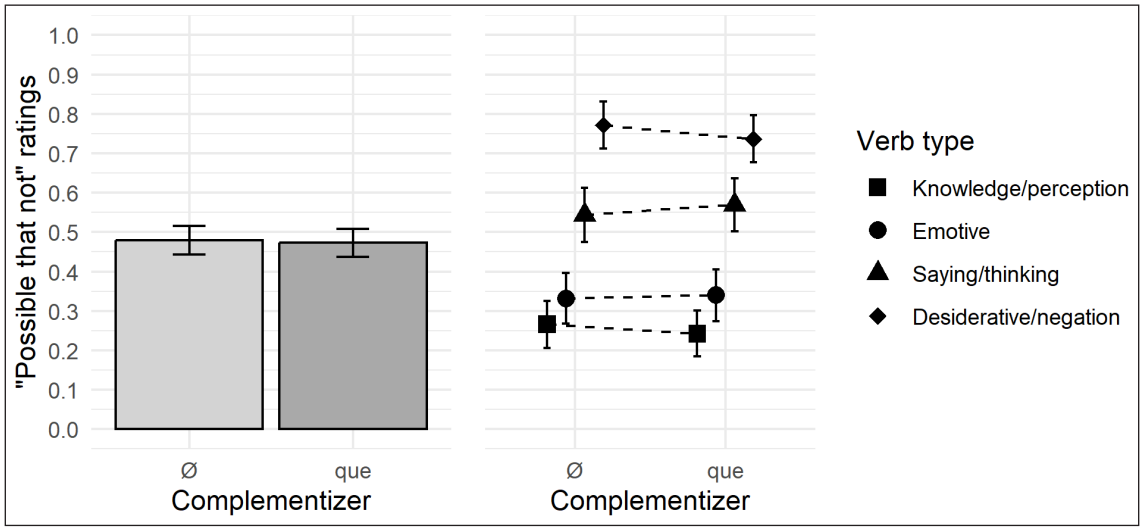


Figure 5: *Possible that not*-diagnostic, Exp. 2 (proportion of *yes*-responses and 95% confidence intervals by experimental condition); all embedding verbs (left) and the four verb types (right)

The results for the *possible that not*-diagnostic (Figure 5) broadly aligned with the initial hypotheses for three of the verb types: both factive types showed relatively low proportions of *yes*-responses with little difference between them, and the verbs of saying and thinking showed a higher proportion, as predicted. However, the desideratives and negation verbs did not match predictions, since instead of patterning with the factives, they yielded the highest proportion of *yes*-responses overall. This finding suggests that the diagnostic did not tap into a property that characterizes verbs of saying and thinking. Possibly, a negative inference associated with the desideratives and negation verbs is responsible for the high proportion of *yes*-responses.

[13] Since a visual inspection of the figures provides sufficient evidence for this conclusion, I will not provide any inferential statistics for the data of Experiment 2. I do not believe that the discussion of *p* values, had they indicated statistical non-significance or (against expectations) statistical significance, would contribute anything to our understanding of the meaning of *que* in gapping.

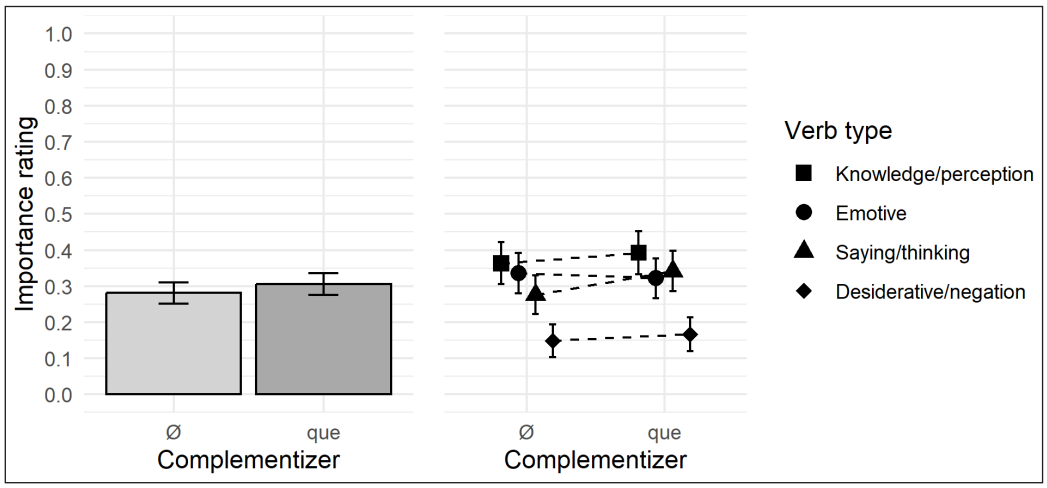


Figure 6: Importance diagnostic, Exp. 2 (proportion of *yes*-responses and 95% confidence intervals by experimental condition); all embedding verbs (left) and the four verb types (right)

The importance ratings in Figure 6 are surprisingly low (below 0.5) across the board. The indicative-taking non-factives should have received much higher ratings, given the assumptions outlined above. A likely explanation for these low ratings is the fact that the experimental items featured coordinate embedded clauses. Test items contained two coordinate clauses with contents of equal value, e.g., [*the king smokes every day*] and [*the queen smokes sometimes*] in (16) above (Section 4.1). The importance diagnostic for this item in (20) above carries a uniqueness inference triggered by the superlative *lo más importante* 'the most important thing'. This inference may have made participants look for one single "most important thing" in the sentence, which may have clashed with the presence of two equally important conjuncts. The analysis of coordinate vs. simple filler items in Bonke (2024: Ch. 5) confirms this suspicion.

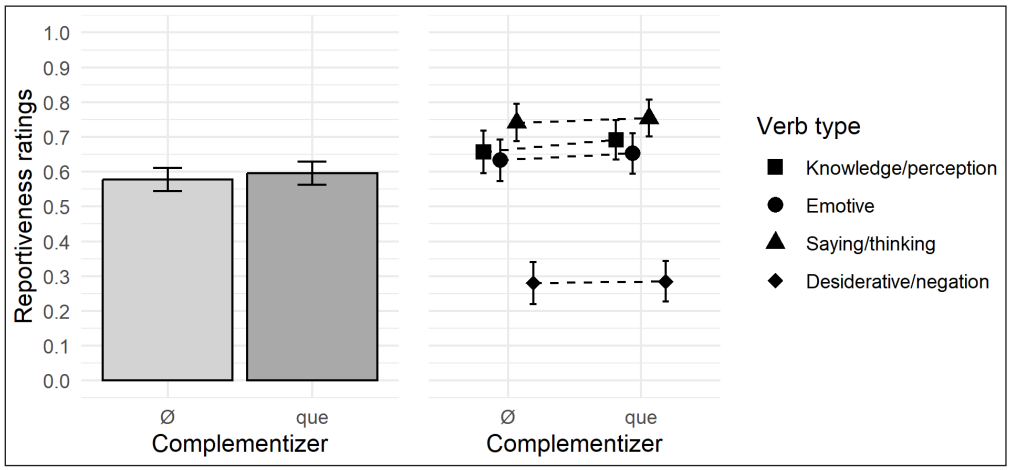


Figure 7: Reportiveness diagnostic, Exp. 2 (proportion of *yes*-responses and 95% confidence intervals by experimental condition); all embedding verbs (left) and the four verb types (right)

Finally, the reportiveness ratings are in Figure 7. Once more, the indicative-taking non-factives should receive a relatively high proportion of *yes*-responses, given that this type contains verbs of saying and thinking. While these verbs do receive the highest *yes*-proportion, the two factive types receive similarly high *yes*-proportions, which do not seem to suggest a categorical difference between these three verb types. Only the subjunctive-taking non-factives receive the expected low proportion of *yes*-responses. This result is surprising because it seems to suggest that emotive verbs (i.e., the subjunctive-taking factives) and verbs of knowledge and perception (i.e., the indicative-taking factives) may quite easily receive reportive ratings.

In sum, the results of Experiment 2 do not provide any evidence for the hypothesized effect of *que* in gapping on interpretation. All three diagnostics deliver null results, both for the pooled data for all verbs as well as the data of the four individual verb types. Additionally, a critical examination of the diagnostics suggests that they do not ideally capture specific pragmatic properties (like assertion or at-issueness). Nevertheless, the results of Experiment 2 are at least somewhat informative: whether or not *que* surfaces in the elliptical conjunct in gapping is most likely not associated with a categorical difference in sentence meaning.

6 Conclusion

In this paper, I set out to investigate the role of the complementizer *que* 'that' in Spanish embedded gapping, with a particular focus on whether its presence is constrained by pragmatic properties. Previous work (Bonke & Repp 2022) investigated the role of such pragmatic properties, especially factivity, as potential explanations for verb-type effects in the acceptability of *que*. The present study presented a new correlation analysis and two experiments to reassess this hypothesis more directly.

The correlation analysis of acceptability data with independently collected quantitative measures of factivity showed only a weak link between the two. While there was a statistically significant correlation, its explanatory power was low, and further inspection revealed that the effect was largely driven by one verb type, namely verbs of saying and thinking. Once these verbs were removed from the analysis, the correlation disappeared, indicating that factivity is unlikely to be the decisive factor in determining the acceptability of *que*.

Experiment 1 tested more directly whether the presence of *que* influences the interpretation of factivity by applying the *certain that*-diagnostic (e.g., Tonhauser et al. 2018) to gapping sentences with and without *que*. The results confirmed the conclusion of the correlation analysis by showing no systematic differences: certainty ratings were unaffected by the complementizer, suggesting that *que* does not alter the interpretation of factivity in embedded gapping.

Experiment 2 extended the investigation beyond factivity to other possible pragmatic contributions of *que* in gapping, using exploratory diagnostics. Across diagnostics and verb types, the results were again null: the presence or absence of *que* did not produce meaningful differences in interpretation. While the diagnostics themselves may not have been ideal tools for measuring specific pragmatic properties, the consistent lack of meaningful effects nevertheless suggests that *que* does not contribute to meaning in the way previously hypothesized.

The origin of the differential sensitivity to gapping-*que* exhibited by different embedding verbs as observed by Bonke & Repp (2022) thus remains unaccounted for. However, based on the present results, we may confidently conclude that the categorization of embedded verbs along pragmatic properties, especially factivity, yields no insight into that origin. Instead, it may be fruitful for future research into *que* in gapping to give up on semantic or pragmatic partitions of embedding verbs altogether and instead turn to the investigation of other potential factors. Unfortunately, the present study does not seem to provide any suggestions as to what these factors might be.

Data Availability Statement: The data and analyses for the two experiments presented in this paper are available on OSF: https://osf.io/r4shg.

Acknowledgements: I would like to thank the audience at *Experimental and Corpus-based Approaches to Ellipsis* 2024 in Poznań and three anonymous reviewers for their helpful comments and discussions. I furthermore thank Sophie Repp and Judith Tonhauser for their guidance and support in the preparation of my PhD dissertation, a part of which forms the basis for this paper.

References

Beaver, David. 2010. Have you noticed that your belly button lint color is related to the color of your clothing? In Rainer Bauerle, Uwe Reyle & Thomas Zimmermann (eds.), *Presuppositions and discourse: Essays offered to Hans Kamp*, 65–100. Leiden: Brill. https://doi.org/10.1163/9789004253162_004.

Bîlbîie, Gabriela & Israel de la Fuente. 2019. Can gapping be embedded? Experimental evidence from Spanish. *Glossa* 4(1). 110. https://doi.org/10.5334/gjgl.782.

Bonke, Max. 2024. *Complementizers in ellipsis*. Köln: Universität zu Köln dissertation.

Bonke, Max & Sophie Repp. 2022. Complementizer deletion in embedded gapping in Spanish. *The Linguistic Review* 39(3). 525–555. https://doi.org/10.1515/tlr-2022-2096.

Broekhuis, Hans & Norbert Corver. 2019. *Syntax of Dutch: Coordination and ellipsis*. Amsterdam: Amsterdam University Press.

Cattell, Ray. 1978. On the source of interrogative adverbs. *Language* 54(1). 61–77. https://doi.org/10.2307/412999.

de Cuba, Carlos & Jonathan E. MacDonald. 2013. On the referential status of embedded polarity answers in Spanish. *Hispanic Linguistics Symposium* 16. 312–323.

Degen, Judith & Judith Tonhauser. 2021. Prior beliefs modulate projection. *Open Mind* 5. 59–70. https://doi.org/10.1162/opmi_a_00042.

Degen, Judith & Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language* 98(3). 552–591. https://doi.org/10.1353/lan.0.0271.

Demonte, Violeta & Olga Fernández-Soriano. 2009. Force and finiteness in the Spanish complementizer system. *Probus* 21(1). 23–49. https://doi.org/10.1515/prbs.2009.002.

Farkas, Donka F. & Kim B. Bruce. 2009. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118. https://doi.org/10.1093/jos/ffp010.

Fiengo, Robert W. 1974. *Semantic conditions on surface structures*. Cambridge, MA: MIT dissertation. http://hdl.handle.net/1721.1/13018.

Haegeman, Liliane. 2006. Conditionals, factives and the left periphery. *Lingua* 116(10). 1651–1669. https://doi.org/10.1016/j.lingua.2005.03.014.

Hankamer, Jorge. 1979. *Deletion in coordinate structures*. New York & London: Garland.

Hartmann, Katharina. 2000. *Right node raising and gapping: Interface conditions on prosodic deletion*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.106.

Hooper, Joan B. 1975. On assertive predicates. In John P. Kimball (ed.), *Syntax and Semantics,* vol. 4, 91–124. Leiden: Brill. https://doi.org/10.1163/9789004368828_005.

Hooper, Joan B. & Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4(4). 465–497. http://www.jstor.org/stable/4177789.

Johnson, Kyle. 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry* 40(2). 289–328. https://doi.org/10.1162/ling.2009.40.2.289.

Johnson, Kyle. 2018. Gapping and stripping. In Jeroen van Cranenbroek & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis,* 562–604. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198712398.013.24.

Karttunen, Lauri. 1971. Some observations on factivity. *Papers in Linguistics* 4. 55–69. https://doi.org/10.1080/08351817109370248.

Kastner, Itamar. 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164. 156–188. https://doi.org/10.1016/j.lingua.2015.06.004.

Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics*, 143–173. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783111350219.143.

Leiner, Dominik J. 2024. *SoSci Survey (Version 3.5.02)*. https://www.soscisurvey.de, 06 March 2026.

Meinunger, André. 2017. Desires, wishes and hopes – desiderative predicates and presuppositions. *STUF - Language Typology and Universals* 70(4). 611–635. https://doi.org/10.1515/stuf-2017-0026.

R Core Team. 2026. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/, 06 March 2026.

Repp, Sophie. 2009. *Negation in gapping*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199543601.001.0001.

Ross, John R. 1968. *Constraints on variables in syntax*. Bloomington, Indiana: Indiana Univ. Linguistics Club.

Ross, John R. 1970. Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in Linguistics*, 249–259. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783111350219.

Sag, Ivan A. 1977. *Deletion and logical form*. Bloomington, Indiana: Indiana Univ. Linguistics Club.

Simons, Mandy, Judith Tonhauser, David Beaver & Craige Roberts. 2010. What projects and why. In Nan Li & David Lutz (eds.), *Proceedings of SALT 20*, 309–327. Ithaca, NY: CLC Publications.

Stalnaker, Robert. 1978. Assertion. In P. Cole (ed.), *Syntax and semantics*, vol. 9, 315–332. New York: Academic Press.

Stalnaker, Robert. 2002. Assertion. In Paul Portner & Barbara H. Partee (eds.), *Formal semantics: The essential readings*, 147–161. Oxford: Blackwell. [Reprint of Stalnaker (1978).]

Taraldsen, Knut Tarald. 2000. Rethinking gapping and the order of constituents. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og fellesskap*, 211–224. Oslo: Novus.

Tonhauser, Judith, David I. Beaver & Judith Degen. 2018. How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics* 35(3). 495–542. https://doi.org/10.1093/jos/ffy007.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer. https://ggplot2.tidyverse.org.

Wilder, Chris. 1994. Coordination, ATB and ellipsis. In Jan-Wouter Zwart (ed.), *Minimalism and Kayne's asymmetry hypothesis*, 291–329. Groningen: Rijksuniversiteit.

Wilder, Chris. 1997. Some properties of ellipsis in coordination. In T. Alan Hall & Artemis Alexiadou (eds.), *Studies on universal grammar and typological variation*, 59–107. Amsterdam: John Benjamins.