The licensing and usage of topic drop in German

Lisa Schäfer

Open Germanic Linguistics 12

language science press

Open Germanic Linguistics

Editors: Michael T. Putnam, Hanna Fischer, David Natvig, Richard Page & Laura Catharine Smith

In this series:

1. Lemke, Robin. Experimental investigations on the syntax and usage of fragments.
2. Larsson, Ida & Erik Petzell (eds.). Morphosyntactic change in Late Modern Swedish.
3. Hall, Tracy Alan. Velar fronting in German dialects: A study in synchronic and diachronic phonology.
4. Cysouw, Michael. Encyclopaedia of German diatheses.
5. Bacskai-Atkari, Julia. The syntax of functional left peripheries: Clause typing in West Germanic and beyond.
6. Nickel, Grit. Nominale Flexionsmorphologie in den ostoberdeutschen Dialekten Bayerns.
7. Schmitt, Eleonore. Frequenz. Prototyp. Schema.: Ein gebrauchsbasierter Ansatz zur Entstehung grammatischer Varianten.
8. Bech, Kristin & Alexander Pfaff (eds.). Noun phrases in early Germanic languages.
9. Somers, Katerina. How to create an early German *scriptus*: The literization approach to historical German syntax.
10. Vieregge, Annika. Bewertung, Variation und Wandel der Präpositionalkasus im Deutschen: Der Einfluss metapragmatischer Urteile auf die Rektion von Präpositionen.
11. Kinn, Kari & Michael T. Putnam (eds.). A reference guide to the syntax of North American Norwegian.
12. Schäfer, Lisa. The licensing and usage of topic drop in German.

ISSN (print): 2750-5588
ISSN (electronic): 2750-557X

The licensing and usage of topic drop in German

Lisa Schäfer

language
science
press

Lisa Schäfer. 2025. *The licensing and usage of topic drop in German* (Open Germanic Linguistics 12). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/490
© 2025, Lisa Schäfer
Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
http://creativecommons.org/licenses/by/4.0/
Zugleich Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie an der Philosophischen Fakultät P der Universität des Saarlandes im Fach Neuere deutsche Sprachwissenschaft, 2024
ISBN: 978-3-96110-516-8 (Digital)
978-3-98554-145-4 (Hardcover)

ISSN (print): 2750-5588
ISSN (electronic): 2750-557X
DOI: 10.5281/zenodo.15584708
Source code available from www.github.com/langsci/490
Errata: paperhive.org/documents/remote?type=langsci&id=490

Cover and concept of design: Ulrike Harbort
Typesetting: Lisa Schäfer
Proofreading: Amir Ghorbanpour, David Carrasco Coquillat, Elliott Pearl, George Walkden, Jean Nitzke, Lachlan Mackenzie, Maria Onoeva, Mary Ann Walter, Nicoletta Romeo, Rainer Schulze, Tabea Reiner
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
Scharnweberstraße 10
10247 Berlin, Germany
http://langsci-press.org
support@langsci-press.org

Storage and cataloguing done by FU Berlin

FREIE UNIVERSITÄT BERLIN

Es gibt gewisse Möglichkeiten die Informationstheorie in der Linguistik anzuwenden. Shannon selbst war der erste, der auf solche Anwendungen hingewiesen hat. Allen gemeinsam ist, daß sie zwar vom linguistischen Standpunkt aus nicht übermäßig vielsagend sind, aber doch ganz unterhaltsam.

Flemming Topsoe

in: Informationstheorie. Eine Einführung. (1974)
Stuttgart: B. G. Teubner: p. 62.

Contents

Acknowledgments **vii**

1 Introduction **1**
1.1 Contextualization . . . 3
1.2 Research questions . . . 5
1.3 Method . . . 6
1.4 Contributions . . . 8
1.4.1 Focus on licensing and usage . . . 8
1.4.2 Empirical approach . . . 8
1.4.3 Definition and typological perspective . . . 8
1.4.4 Prefield restriction as syntactic licensing condition . . . 9
1.4.5 Recoverability as usage condition . . . 9
1.4.6 Information-theoretic account . . . 9
1.4.7 Empirical support for the information-theoretic account 10
1.5 Structure of this book . . . 10

2 Definition and typological perspective **13**
2.1 Definition . . . 14
2.1.1 Omission . . . 14
2.1.2 Prefield restriction . . . 15
2.1.3 Omitted constituent . . . 17
2.1.4 Modality, register, and text type . . . 19
2.1.5 Topic drop and text type in the fragment corpus FraC . 21
2.2 Distinguishing topic drop from related phenomena . . . 27
2.2.1 *pro*-drop . . . 28
2.2.2 Discourse (*pro*)-drop . . . 30
2.2.3 Dialectal null subjects . . . 31
2.2.4 V1 declaratives . . . 31
2.3 Topic drop in other Germanic V2 languages . . . 34
2.3.1 Danish . . . 35
2.3.2 Dutch . . . 36
2.3.3 Icelandic . . . 37

2.3.4 Norwegian . . . 38
2.3.5 Swedish . . . 39
2.3.6 Yiddish . . . 39
2.3.7 Summary . . . 40
2.4 Register-dependent omissions in other languages . . . 40
2.4.1 Subject omission in English and French . . . 41
2.4.2 Argument omission in Russian . . . 44
2.5 Summary: a typological view on topic drop . . . 45

I Licensing of topic drop 47

3 Syntactic licensing: prefield restriction of topic drop 49
3.1 Prefield restriction and topicality . . . 53
3.1.1 Defining *topic* . . . 53
3.1.2 Relation between topic drop and topicality . . . 57
3.1.3 Contra prefield position as topic position . . . 58
3.1.4 Contra topicality as a sufficient condition . . . 59
3.1.5 Contra topicality as a necessary condition . . . 61
3.1.6 Experiment 1: topic drop of expletives . . . 66
3.1.7 Summary: topic drop and topicality . . . 75
3.2 Is topic drop restricted to the prefield? . . . 75
3.2.1 Theoretical background . . . 75
3.2.2 Experiment 2: prefield restriction . . . 78
3.3 Is topic drop possible in (potentially) embedded clauses? . . . 84
3.3.1 Theoretical background . . . 84
3.3.2 Experiment 3: topic drop in (potentially) embedded clauses 91
3.4 Is topic drop restricted to the very first element? . . . 97
3.4.1 Theoretical background . . . 98
3.4.2 Experiment 4: topic drop and preposed conjunctions . . . 104
3.5 Summary: prefield restriction . . . 116
3.6 Toward a syntactic analysis of topic drop . . . 117
3.6.1 Operator analysis . . . 118
3.6.2 *pro*-analysis . . . 122
3.6.3 PF-deletion analysis . . . 125
3.6.4 Summary: syntactic analysis . . . 127

4 Recoverability as a prerequisite for topic drop 129
4.1 Linguistic antecedents . . . 130

4.2 Extralinguistic antecedents . . . 137
4.3 Non-referential elements . . . 138
4.4 Recoverability and givenness . . . 139
4.5 Recoverability as felicity or usage condition . . . 143

II Usage of topic drop **147**

5 Previous accounts of topic drop usage **149**
5.1 Redundancy avoidance and economy . . . 149
5.2 Cohesion . . . 151
5.3 Topic avoidance . . . 152
5.4 Socio-pragmatic functions . . . 153
5.4.1 Expressiveness, condensation, and action orientation . . . 153
5.4.2 Elaborations/reformulations and responses . . . 155
5.4.3 Completion or continuation . . . 156
5.4.4 Rhetorical and social function . . . 158
5.5 Prosody . . . 159
5.6 Summary: previous accounts of topic drop usage . . . 161

6 Information-theoretic account of topic drop usage **163**
6.1 Information theory: basic concepts . . . 163
6.1.1 Information . . . 164
6.1.2 Communication and channel . . . 167
6.2 *Uniform information density* (*UID*) . . . 170
6.2.1 *Avoid troughs* principle . . . 173
6.2.2 *Avoid peaks* principle . . . 175
6.3 *Facilitate recovery* principle . . . 177
6.4 Summary: information-theoretic approach to topic drop usage . . . 179

7 Methodology of the empirical studies **181**
7.1 Production, perception, and processing . . . 182
7.2 Methodology: corpus study . . . 183
7.2.1 The fragment corpus FraC . . . 184
7.2.2 Annotation procedure . . . 185
7.2.3 FraC data sets . . . 189
7.3 Methodology: experiments . . . 192

8 Syntactic function **193**
8.1 Theoretical overview . . . 193
8.1.1 Subjects . . . 193
8.1.2 Objects . . . 193
8.1.3 Predicatives . . . 204
8.2 Previous empirical evidence . . . 205
8.2.1 Poitou (1993) . . . 205
8.2.2 Frick (2017) . . . 206
8.2.3 Ruppenhofer (2018) . . . 207
8.2.4 Trutkowski (2018) . . . 208
8.3 Information-theoretic predictions for syntactic function . . . 211
8.3.1 Subject vs. object topic drop . . . 212
8.3.2 Direct vs. indirect objects . . . 213
8.3.3 1st and 2nd person objects vs. 3rd person objects . . . 214
8.4 Corpus study of syntactic function . . . 219
8.4.1 Syntactic function in FraC-TD-Comp . . . 220
8.4.2 Syntactic function in FraC-TD-SMS . . . 221
8.5 Summary: syntactic function . . . 222

9 Topicality **225**
9.1 Information-theoretic predictions for topicality . . . 225
9.2 Experimental investigations of topicality . . . 226
9.2.1 Experiment 5: topicality (simple question method) . . . 226
9.2.2 Experiment 6: topicality (complex question method) . . . 232
9.2.3 Experiment 7: topicality (subject method, lexical verbs) . . . 237
9.2.4 Experiment 8: topicality (subject method, modal verbs) . . . 244
9.3 Summary: topicality . . . 250

10 Grammatical person, verbal inflection, and ambiguity avoidance **253**
10.1 Theoretical overview . . . 253
10.1.1 Theoretical overview of grammatical person and verbal inflection . . . 253
10.1.2 Theoretical overview of ambiguity avoidance . . . 260
10.2 Previous empirical evidence . . . 266
10.2.1 Previous empirical evidence regarding grammatical person 266
10.2.2 Previous empirical evidence regarding ambiguity avoidance . . . 270
10.3 Information-theoretic predictions for grammatical person and verbal inflection . . . 272

10.4 Corpus study of grammatical person and verbal inflection . . . 274
10.4.1 Grammatical person in FraC-TD-Comp . . . 274
10.4.2 Grammatical person in FraC-TD-SMS . . . 277
10.4.3 Inflectional ending in FraC-TD-SMS-Part . . . 279
10.4.4 Grammatical person and inflectional ending in FraC-TD-SMS-Part – logistic regression analysis . . . 283
10.5 Experimental investigations of grammatical person and verbal inflection . . . 289
10.5.1 Experiment 5: grammatical person (1SG vs. 3SG) . . . 290
10.5.2 Experiment 7: grammatical person (1SG vs. 3SG, lexical verbs) . . . 293
10.5.3 Experiment 8: grammatical person (1SG vs. 3SG, modal verbs) . . . 296
10.5.4 Experiment 9: grammatical person (1SG vs. 2SG) . . . 301
10.5.5 Experiment 10: grammatical person (1SG vs. 2SG, dialogues) . . . 305
10.6 Summary: grammatical person, verbal inflection, and ambiguity avoidance . . . 310

11 Verb type and verb surprisal 315
11.1 Theoretical overview of verb type . . . 315
11.2 Previous empirical evidence regarding verb type . . . 316
11.2.1 Poitou (1993) . . . 316
11.2.2 Androutsopoulos & Schmidt (2002) . . . 316
11.2.3 Helmer (2016) . . . 318
11.2.4 Frick (2017) . . . 320
11.3 Information-theoretic predictions for verb type and verb surprisal 322
11.4 Corpus study of verb type and verb surprisal . . . 324
11.4.1 Verb type in FraC-TD-Comp . . . 325
11.4.2 Verb type in FraC-TD-SMS . . . 328
11.4.3 Verb type and surprisal in FraC-TD-SMS-Part . . . 329
11.4.4 Verb type and surprisal in FraC-TD-SMS-Part – logistic regression analysis . . . 332
11.5 Experimental investigations of verb surprisal . . . 338
11.5.1 Experiment 11: verb surprisal (context) . . . 338
11.5.2 Experiment 12: verb surprisal (verb type) . . . 348
11.6 Summary: verb type and verb surprisal . . . 353

12 General discussion **355**
12.1 Definition and typological perspective 355
12.2 When is topic drop licensed? 356
12.2.1 Prefield restriction as licensing condition 356
12.2.2 Recoverability as felicity condition 357
12.3 When is topic drop used? 358
12.3.1 Information-theoretic account of topic drop usage . . . 358
12.3.2 Evidence for the *avoid troughs* priciple 359
12.3.3 Evidence for the *avoid peaks* priciple 361
12.3.4 Evidence for the *facilitate recovery* priciple 361
12.3.5 Summary: information-theoretic account 362
12.4 Open questions and outlook 362

References **365**

Index **401**
Name index 401
Language index 409
Subject index 411

Acknowledgments

This book is a slightly revised version of my dissertation for the degree of Doctor of Philosophy from the Faculty of Philosophy at Saarland University in Saarbrücken from 2024. A large number of people have contributed to the success of my doctorate, whom I would like to thank in the following.

First of all, I want to thank my supervisor and "Doktorvater" Ingo Reich from the bottom of my heart. Thank you for placing your trust in me, first as a student assistant, then as a research assistant substituting for parental leave, and finally as a doctoral student and project member. Thank you for always having an open door and an open ear for me, for the many constructive discussions about my project and the valuable comments on the earlier versions of this dissertation. I would also like to thank my second reviewer and co-supervisor Heiner Drenhaus. Through your persistent questioning in exactly the right places and your critical comments on my work, you helped to make this book more precise. And thank you also for the inspiring conversations we had at numerous coffee meetings. Another big thank you goes to my third reviewer Hannah Rohde. Thank you so much for the effort you went to in order to be present at my defense and thereby making that special day even more special. Thank you also for your valuable feedback on my work, from which this published version has benefited greatly. I would also like to thank the other members of my committee Stefanie Haberzettl, Elke Teich, and Robin Lemke for their time and effort. Thank you, Stefanie, for making the defense a special event with your personal words and thank you, Elke, for making my doctorate possible at all through founding and leading the collaborative research center (SFB) 1102.

For Robin, my former office neighbor and colleague in project B3 of the SFB 1102, I would like to express my special thanks. I am grateful for the enormous amount of feedback I received from you at numerous stages of this project. Thank you as well for always being available as a sparring partner for discussions about experimental stimuli, syntactic theories, and statistical analyses. In addition, you were a great travel companion on several work trips that took us to culinary delights such as the legendary Deep-Dish Pizza. I would also like to thank my amazing current and former colleagues from the Department of Neuere deutsche Sprachwissenschaft at Saarland University in Saarbrücken for proofreading my

thesis, for valuable feedback during the colloquium, and for creating a wonderful work atmosphere. Whether it's the daily visit to the Mensa, the games nights, the trips to the Christmas market, departmental hikes, or birthday and Christmas parties – I've always had and still have a great time with you. Thanks to Philipp, Nele, Jenny, Constanze, Ricarda, Fabian, Dari, Lena, Luise, Sophia, Jessica, Natascha, Julia, Joanna, Josef, Augustin, and everyone else! In addition, a big thank you also to the student assistants who supported me in this work by preparing stimuli and testing experimental setups. Thanks to Anna, Annika, Fabian, Katrin, Lea, Luise, Natascha, and Ricarda.

I would also like to thank all the members of the SFB 1102. Thanks to everyone who gave me constructive feedback on my project at several PhD Days and retreats. Thanks to Marie-Ann and the rest of the coordination team, whose patience overcame many a bureaucratic problem. I am also extremely grateful to the Deutsche Forschungsgemeinschaft (DFG) for their funding, without which this work and numerous valuable conference trips would not have been possible.[1]

Many thanks also to Michael T. Putnam and the rest of the editors of the Open Germanic Linguistics series for considering my work. Likewise, I am very grateful to the team at Language Science Press, especially to Sebastian Nordhoff and Felix Kopecky for their help with typesetting. And a big thank you to all the volunteer proofreaders who improved the language of this book considerably. All remaining errors are, of course, entirely my own.

Ein besonderes Dankeschön gilt meinen Freundinnen und Freunden, die mich über die anstrengenden Jahre immer wieder abgelenkt und aufgemuntert haben. Danke, dass ihr für mich da wart, Anne, Dennis, Hannah, Matthias, Sophie und alle anderen. Von ganzem Herzen danke ich auch meinen Eltern Ursula und Peter. Vielen Dank, dass ihr mich ermuntert habt, meinen Weg zu gehen und mich dabei bedingungslos unterstützt habt. Danke auch an meine Oma Doris, die immer an mich geglaubt hat. Und schließlich danke ich Jan – für alles. Tausend Dank, dass du mich immer auffängst und dass du mich selbst an den härtesten Tagen zum Lachen bringst.

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

1 Introduction

One fundamental property of human language is that we frequently leave certain things implicit while communicating, as exemplified in (1).

(1) a. Someone was looking for you, but I don't know who Δ.
b. I know that Robin loves Δ and Fabian hates coriander.
c. *Milliarden-Investor in Ensdorf: "*Δ *Haben uns ins Saarland*
billions-investor in Ensdorf we have us in.the Saarland
verliebt"
fell.in.love
'Billion-euro investor in Ensdorf: '(We) fell in love with the Saarland' '
(Saarbrücker Zeitung, 02/02/2023, p. 1)

These structures seem somehow "incomplete" because they lack elements that would usually be present in a "normal" sentence, such as a second verb phrase (VP) (2a), a determiner phrase (DP) in object function (2b), or the preverbal constituent (2c). The "normal", i.e., syntactically complete, sentences are referred to as *full forms* in this book.

(2) a. Someone was looking for you, but I don't know who ⟨was looking for you⟩.
b. I know that Robin loves ⟨coriander⟩ and Fabian hates coriander.
c. Milliarden-Investor in Ensdorf: "⟨Wir⟩ haben uns ins Saarland verliebt"

Despite their incompleteness, the utterances in (1) are nevertheless usually understood by their addressees – including the "omitted" part in (2). This phenomenon has been already discussed since antiquity under the term *ellipsis* (from Ancient Greek ἔλλειψις, *élleipsis*, 'omission', *Ellipsis* 2023) and continues to be an intriguing and popular research topic in modern linguistics.[1] Still, many questions sur-

[1]This is evidenced, for example, by the publication of the *The Oxford handbook of ellipsis* in 2018 (van Craenenbroeck & Temmerman 2018), by the inclusion of articles on ellipses in handbooks of different disciplines, such as Reich (2011) in *Semantics. An international handbook of natural language meaning* (von Heusinger et al. 2011) or van Craenenbroeck & Merchant (2013) in *The Cambridge handbook of generative syntax* (den Dikken 2013), and by the initiation of an annual workshop on experimental and corpus-based approaches to ellipsis in 2017 (see Bîlbîie & Nykiel 2023).

rounding ellipsis have not been fully resolved, such as which types of ellipsis should be meaningfully distinguished and why, how ellipses differ across languages, how they can be adequately described grammatically, how hearers process them, and why hearers understand elliptical utterances anyway, to name just a few.

Cross-linguistically, a number of different ellipsis types have been discussed, such as *sluicing* (e.g., Ross 1969, Merchant 2001) in (1a) or *right node raising* (e.g., Postal 1974, Hartmann 2000) in (1b). This book is concerned with the type in (1c), the omission of the preverbal constituent (to which I refer as the *prefield constituent* later in this book) in verb-second (V2) declarative clauses in most Germanic languages:[2] so-called *topic drop* (e.g., Fries 1988). A more detailed study of topic drop seems to be required because, although numerous claims about topic drop have been put forward in the theoretical literature, they have often been made in isolation, are partly contradictory, and have not or only rarely been empirically tested. Therefore, this book has the goal of investigating the licensing and the usage of topic drop thoroughly using empirical methods, especially experiments.

Example (1c) illustrates a case of topic drop where the 1st person plural subject pronoun *wir* ('we') is omitted from the clause. The normally obligatorily filled preverbal position of the declarative clause is empty. However, from the usage of the reflexive pronoun *uns* ('ourselves'), it is evident that *wir* must still be part of the sentence. I usually term the unrealized constituent *omitted* or *covert*, making it clear that I consider topic drop to be a case of ellipsis. The empty preverbal position is marked with the Greek capital letter Delta Δ and glossed and translated with an intuitive reconstruction of the omitted constituent.[3]

While the omission of a referential subject as in (1c) is indeed a particularly typical case of topic drop in German, its occurrence in the text type newspaper headline is not. This is because, as this book also evidences in line with previous research, topic drop typically occurs in certain text types. Figure 1.1, glossed and translated in example (3), therefore, shows topic drop in a more characteristic text type, namely, in a message of the instant messaging service WhatsApp, taken from the Mobile Communication Database 2 (MoCoDa2) (Beißwenger et al. 2020).[4] In the present dialogue, there is a further omission of a referential

[2] Although I only investigate topic drop in German in this book, it is to be expected that the results are at least partially transferable to the other Germanic V2 languages, provided that language-specific peculiarities are taken into account.

[3] When presenting experimental items, I indicate the variation between topic drop and the corresponding full form by putting the prefield constituent into parentheses.

[4] The still growing Mobile Communication Database 2 consisted in January 2025 of about 1 000 instant messaging (WhatsApp) chats with roughly 39 000 messages and 318 000 tokens of about 3 550 writers, which were donated for scientific usage.

subject (*ich* – 'I'), as well as the omission of a direct object (*das* – 'that') and of a non-referential subject (expletive *es* – 'it'), showing a wider range of possible topic drop constructions.

(3) a. *MB: Möchstest [sic!] du in Mittelalter deine BA*
would.like you in middle.ages your bachelor.thesis
schrieben [sic!] oder warum willst du die MP in Neuzeit
write or why want you the oral.exam in modern.era
machen
make
MB: 'Do you want to write your bachelor thesis on medieval times or why do you want to do the oral exam on modern times?'

b. *TZ: Beides Neuzeit*
both modern.times
TZ: 'Both modern times'

c. *MB:* Δ *Darf man nicht. Es sei denn das eine ist frühe*
that may one not it is.CONJ PART that one is early
Neuzeit
modern.times
MB: 'You are not allowed to do (that). Unless the one is early modern times'

d. *TZ: Wieso darf man das nicht?*🤔 Δ *Gibt doch 3 Körbe.* Δ *Nehme*
why may one that not it gives PART 3 baskets I take
einmal 19. Jh und einmal 20
once 19th century and once 20th
TZ: 'Why is that not allowed?🤔 (There) are 3 baskets. (I) take once 19th century and once 20th.'

In this book, I focus on topic drop of subjects and, to a lesser extent, objects in text messages and instant messaging chats. I investigate both its licensing and its usage.

1.1 Contextualization

To the best of my knowledge, topic drop was first discussed in modern linguistics by John Robert Ross and Marga Reis in 1982. In early works, topic drop research was mainly concerned with issues of licensing and the syntactic description of the phenomenon (see Fries 1988, for a first detailed investigation). It has also been discussed repeatedly in the generative literature, for instance, in the context of

Figure 1.1: Excerpt from a WhatsApp dialogue between two male history students, taken from MoCoDa2, chat #6WJlJ, https://db.mocoda2.de/view/6WJlJ (visited on 01/02/2025)

related null subject phenomena such as *pro*-drop but especially in relation to the question of which null category could describe it most adequately (e.g., Huang 1984, Cardinaletti 1990, Haegeman 1990, Rizzi 1994). In the German-language research, especially of the late 1980s and the 1990s, individual aspects of topic drop were repeatedly singled out (e.g., Sternefeld 1985, Oppenrieder 1987, Auer 1993, Poitou 1993, Zifonun et al. 1997). As a result, authors often put forward isolated hypotheses, concerning, e.g., the preferred occurrence of topic drop with certain grammatical functions or persons, as well as the role of topicality and the following verb. Following Auer (1993), who first discussed topic drop in corpus data, topic drop became increasingly a matter of interest in spoken language research after the turn of the millennium, where aspects of the usage and the effect of the ellipsis type came to the fore (e.g., Sandig 2000, Günthner 2000, 2006, Schwitalla 2012, Imo 2013, 2014, Helmer 2016). At the same time, topic drop was also increasingly investigated quantitatively as a typical phenomenon of the communication form text message and its successor instant message (Androutsopoulos & Schmidt 2002, Döring 2002, Frick 2017).

The short sketch of topic drop research since the 1980s illustrates the different perspectives and foci on the phenomenon of topic drop. However, what is still needed is, first, a systematic investigation of the several isolated claims and, second, an approach that simultaneously considers the licensing and the usage of topic drop, as well as their relation. Such a double perspective is the goal of this book. In other words, my aim is to determine the licensing conditions of topic drop and to answer the question of why speakers use topic drop, provided these conditions are met. At the same time, I take a more empirical and quantitative approach than previous research, which often relied on unnatural introspective examples, some of which were judged inconsistently by different authors. In addition to the corpus data already established in topic drop research, above all, I also take into account experimental results, in the form of acceptability judgments by naive participants.

1.2 Research questions

This book focuses on two central aspects of topic drop: licensing and usage. The first major research question I would like to answer is: When is topic drop syntactically licensed in German? This question can be divided into two sub-questions since, according to previous literature, there are two central conditions for topic drop, the restriction to the prefield and the recoverability of the omitted constituent. Consequently, the first sub-question concerns the restriction of topic drop to the prefield, i.e., the preverbal position of declarative V2 main clauses, and the adequate description of this restriction. While this prefield restriction is widely accepted in the literature, although it is not entirely uncontroversial, it has been formalized in different ways and captured both syntactically and in terms of information structure. The central goals of the first part of this book are to confirm the prefield restriction empirically and to determine its nature more precisely through theoretical discussions and experiments. To a lesser extent and exclusively from a theoretical perspective, I address the second sub-question: What is the role of recoverability for topic drop and how can this concept be meaningfully grasped?

The second major research question is the following: When is topic drop used, provided it is syntactically licensed? This question is motivated by the observation that, in a context where topic drop is permitted, speakers[5] generally have

[5] I use the term *speaker* to generally refer to the "producer" of language, may it be in spoken or in written form, i.e., a speaker could also be a writer. Similarly, *hearer* is used to denote the "addressee" of language.

a choice between using the ellipsis or a full form. In fact, they sometimes use one and sometimes the other in similar contexts or situations. To answer this question, I propose an information-theoretic account of topic drop usage and validate it by means of empirical studies. The idea is that speakers (do not) use topic drop to optimally distribute the hearer's processing effort across an utterance. Since the information-theoretic account is based on probabilities and general processing principles, it is transferable to different (elliptical) phenomena, as has already been done, and allows for explaining their usage in a uniform way. In my empirical studies, I investigate, first, which of the factors proposed in the theoretical literature really impact the usage of topic drop. Second, I examine whether the information-theoretic approach can provide a unified account of the previously isolated claims from the literature and whether it has additional explanatory power.

1.3 Method

To answer these two research questions, I begin with a discussion of the existing literature on the licensing and the usage of topic drop, respectively. The discussion mainly concerns the theoretical literature, which, as already indicated above, often made isolated claims about topic drop in the past. I derive testable hypotheses from these claims, which the theoretical literature itself often did not produce. In the second part of this book on the usage of topic drop, I develop an information-theoretic approach. It builds on existing approaches to information-theoretic modeling of other phenomena involving optional omissions and, thus, contributes to a unified analysis of the usage of different types of ellipsis and related phenomena. The information-theoretic account predicts that the usage of topic drop is determined by an interplay of two tendencies: (i) speakers use topic drop to omit predictable prefield constituents and (ii) speakers use the full form to reduce the processing effort on the following verb. With these principles, the approach can explain the effects of individual factors, which have been predicted by the theoretical literature, without additional assumptions, e.g., that topic drop of *ich* ('I') is so frequent because it is usually highly predictable. In addition, the approach predicts specific effects, such as the effect of the surprisal of the following verb, that cannot be explained by any other theory. I empirically tested the hypotheses from the theoretical literature and the predictions of the information-theoretic account, partly individually, and partly in combination using two different methods: a corpus study and acceptability rating experiments. This allowed me to examine both the production and perception of topic drop.

According to the information-theoretic account, they should behave in parallel since the speaker is argued to perform audience design (Bell 1984) to shape their language production for the benefit of the hearer.

With the corpus study, I focused on authentic speech data to investigate the *production* of topic drop in different text types. For this purpose, I extracted instances of topic drop from the entire text type-balanced fragment corpus FraC (Horch & Reich 2017) and compared them to corresponding full forms as reference data. In the second step, I focused on the text message subcorpus and conducted an inferential statistical analysis of topic drop in this single text type. Although there have been corpus studies of topic drop, they either lacked reference data in the form of full forms or statistical analysis or both (e.g., Auer 1993, Androutsopoulos & Schmidt 2002, Döring 2002, Helmer 2016), or the analysis was not state-of-the-art (Frick 2017). In my corpus studies, I collected reference data to be able to assess ellipsis-specific effects and used a more complex statistical method, i.e., logistic regressions, which allowed me to consider several factors simultaneously as well as to observe possible interactions.

In this book, I present the results of a total of 12 acceptability rating studies that I employed to investigate the *perception* of topic drop. I used minimal pairs to systematically test possible factors influencing topic drop and collected the ratings of linguistic non-experts for this purpose. This allowed me to also look at feature combinations that occur too rarely or not at all in the production data but for which my hypotheses derived from the literature or the information-theoretic approach nevertheless predict differences. The focus on experimental methods, in this case on rating studies, is relatively new for topic drop research. To the best of my knowledge, only Trutkowski (2018) has conducted an experimental study of topic drop so far, also a rating study. My work, thus, also contributes to a greater methodological diversity in topic drop research.

My empirical investigations of topic drop are largely focused on two text types or types of communication (see Section 2.1.4 for a terminological discussion), in which topic drop is particularly frequent. As mentioned above, I zoomed in on the text message subcorpus in my corpus study for an inferential statistical analysis, while in all but one of my experiments, I presented the stimuli in the form of instant messaging chats, the successors of text messages. This decision is motivated by the intention to investigate topic drop in text types where this ellipsis type is as natural as possible. Thus, any observed differences in the usage of topic drop should depend on the varying factors and not on the text type. The focus on text and instant messages has the consequence that the conclusions drawn in this book apply primarily to topic drop in these two text types. However, descriptively, there seem to be few differences between the results from the

whole corpus and those from the text message subcorpus. Similarly, experiment 10, which tested topic drop in spoken dialogues presented in written form, found similar effects of grammatical person as experiment 9 with an instant messaging design.

1.4 Contributions

1.4.1 Focus on licensing and usage

This book examines the two central aspects of the ellipsis type topic drop: on the one hand, its licensing, above all the question of the nature of the syntactic environment in which it is possible. On the other hand, its usage, i.e., the question of when speakers of German decide against the full form and in favor of the ellipsis, provided the latter is licensed, and why they do so. In doing so, it offers a descriptive and explanatory approach that brings together what can be described as two of the central strands of topic drop research: the generative tradition, which is dedicated to adequately describing the syntactic properties of topic drop, and the tradition that approaches usage preferences with the help of authentic speech data, be it dialogues or text messages.

1.4.2 Empirical approach

To this effect, I take a strongly empirical approach. First, I place myself in the tradition of investigating topic drop and its production through authentic language data in corpus studies. Second, I focus even more on experimental work, first, to investigate the perception of topic drop and, second, to systematically test individual factors that in some cases could not be investigated in the corpus study due to data sparsity. With these studies, I empirically test for the first time some claims that have so far only been postulated in the theoretical literature.

1.4.3 Definition and typological perspective

I present an explicit definition of topic drop in German that focuses on its elliptical nature and its prefield restriction. Based on this definition, I distinguish topic drop from similar phenomena with which it has been equated in the past, most importantly *pro*-drop and verb-first (V1) declaratives. From a typological perspective, I consider topic drop as a common phenomenon of the Germanic V2 languages and discuss similarities and differences between these languages. Furthermore, I point out the possibility of a common analysis of topic drop with register-dependent argument omissions in other languages.

1.4.4 Prefield restriction as syntactic licensing condition

Concerning the syntactic licensing of topic drop, I propose a refined prefield restriction based on the results of my experiments. These experiments are, to my knowledge, the first to systematically study the licensing of topic drop. With them, I first provide empirical evidence for the validity of the prefield restriction, which I and most of the literature assume, and then examine its properties in more detail. I argue that topicality is not a licensing condition of topic drop, and thus the label *topic drop* falls short of describing this ellipsis type in German. Instead, my experiments suggest that the prefield restriction is a positional constraint that limits this ellipsis to the preverbal position (that is, to the specifier of the complementizer phrase [Spec, CP] in generative terms) of declarative V2 clauses (contra Helmer 2016). More specifically, I argue that topic drop is not possible in any prefield position but only in a prefield that is either not c-commanded sentence-internally by a potential identifier (Rizzi 1994) or that is the highest prefield of a root clause (Freywald 2020). I suggest that, based on my data, a PF-deletion approach is best suited as an analysis of topic drop and should be preferred over the more complex operator or *pro*-approaches, which do not provide greater explanatory power.

1.4.5 Recoverability as usage condition

I argue that recoverability should not be treated as a further licensing condition of topic drop but rather as a felicity condition, which topic drop shares with other ellipsis types. I operationalize recoverability by gradual givenness (Ariel 1990, Chafe 1994). The recoverability of an expression is expected to increase with its givenness in context.

1.4.6 Information-theoretic account

Concerning the usage of topic drop, I propose an information-theoretic approach as an explanation. This approach stands in the tradition of a strand of work that explains optional omissions by the speaker's intention to distribute processing effort for the hearer as optimal as possible across utterances (e.g., Levy & Jaeger 2007, Jaeger 2010, Asr & Demberg 2015, Lemke 2021, to name just a few). The approach is based on general production and processing mechanisms. If certain properties of a phenomenon that were previously explained by specific grammatical rules can be traced back to these general mechanisms, this simplifies the grammar to be assumed.

For topic drop, I argue that its usage is determined by an interplay of the three information-theoretic principles *avoid troughs*, *avoid peaks*, and *facilitate recovery*. According to the former two, which are derived from the *uniform information density hypothesis* (*UID*) (Levy & Jaeger 2007) and its precursors (e.g., Fenk & Fenk 1980), speakers avoid regions of too low and too high processing effort for the hearer. The *facilitate recovery* principle predicts that speakers lower the hearer's processing effort by providing cues that facilitate the resolution of ellipsis.

1.4.7 Empirical support for the information-theoretic account

My empirical studies provide by and large support for these three principles. For instance, the consistently observed preference for topic drop of the 1st person singular subject pronoun *ich* ('I') can be accounted for by the *avoid troughs* principle. The high predictability of *ich* in the investigated text types text messages and instant messages results in very low processing effort and in an increased chance of omitting the predictable constituent. The corpus results for verb surprisal are in line with the *avoid peaks* principle. Topic drop is less frequent before verbs with high processing costs because the preverbal constituent seems to be needed to make the verb easier to process.[6] In the corpus study, there was some evidence of an impact of the *facilitate recovery* principle. Topic drop of the 1st person singular was shown to be more frequent when the following verb had a distinct inflectional ending, which facilitates ellipsis resolution as it unambiguously determines the reference of the omitted prefield constituent.

1.5 Structure of this book

In the following, I briefly outline the structure of this book. It consists of an introductory definition chapter followed by two parts, each addressing one of the two major research questions. Part one deals with the licensing of topic drop and part two with its usage.

In Chapter 2, I begin by proposing a working definition of topic drop. In the second step, I delimit topic drop from similar phenomena such as *pro*-drop and V1 declaratives, before I point out similarities to topic drop in other Germanic languages and to related ellipsis types in other languages.

The first part of this book is concerned with the question of topic drop licensing. In Chapter 3, I begin with the restriction of topic drop to the preverbal prefield position and work out the nature of this restriction in detail. In Section 3.1, I

[6]The experiments on verb surprisal provided no further evidence of an effect of verb surprisal.

conclude that topicality is neither a (strictly) sufficient nor a necessary condition for topic drop. In Section 3.2, I provide experimental evidence for the prefield restriction, which most of the literature and I assume, and argue against Helmer's (2017b) proposal of analyzing omissions in the middle field as topic drop as well. Then, I determine the nature of the prefield restriction more precisely with two further experiments, each devoted to a special case of topic drop: topic drop in (potentially) embedded clauses in Section 3.3 and topic drop after conjunctions in Section 3.4. I conclude that topic drop either only occurs in a prefield that is not c-commanded sentence-internally by a potential identifier (Rizzi 1994) or in the highest prefield of a root clause (Freywald 2020). In Section 3.5, I summarize my findings before I bring them together with previous generative analyses of topic drop in Section 3.6. I conclude that the PF deletion approach can account for most of the properties of topic drop in German that I have identified while it requires the fewest additional assumptions.

In Chapter 4, I turn to recoverability, often considered another condition for topic drop, and look at its properties in Sections 4.1 to 4.3. In Sections 4.4 and 4.5, I argue that the relationship between the antecedent and the omitted referential constituent that enables recoverability is best described in terms of a gradual givenness concept. Recoverability should be considered a felicity or usage condition rather than a licensing condition.

In the second part of this book, I focus on the usage of topic drop to answer the question of when topic drop is used, provided it is licensed. In Chapter 5, I first outline the previous approaches to topic drop usage, before I present my information-theoretic account in Chapter 6. After introducing basic information-theoretic concepts in Section 6.1, I turn to the *uniform information density hypothesis* (*UID*) and its precursors in Section 6.2, according to which speakers strive to distribute information as evenly as possible across an utterance. From *UID*, the *avoid troughs* principle and the *avoid peaks* principle, discussed in Sections 6.2.1 and 6.2.2, can be derived, which I argue guide the usage of topic drop. I complement them with a further principle, the *facilitate recovery* principle, discussed in Section 6.3, according to which the ellipsis resolution also affects the processing of topic drop structures and how felicitous they are.

In Chapter 7, I outline the methodology of the empirical studies presented in the second part of this book. I discuss the relation between production, perception, and processing in Section 7.1 and relate it to the information-theoretic reasoning. Sections 7.2 and 7.3 provide details about the corpus study and the experiments discussed in the subsequent chapters.

The following four chapters are then each dedicated to one or more factors that potentially influence the usage of topic drop. In each chapter, I first discuss the

state of research, as well as previous studies, outline the information-theoretic predictions, and present my own corpus and/or experimental results. Chapter 8 is concerned with syntactic function and presents corpus results, while Chapter 9, based only on experimental data, investigates whether topicality is a usage factor. In Chapter 10, I turn to grammatical person and number, verbal inflection, and ambiguity avoidance, and, finally, in Chapter 11, to verb type and verb surprisal. I investigated the factors in these two chapters with my corpus study and with experiments.

Finally, the conclusion of the book is provided by the general discussion in Chapter 12. There, I revisit the central research questions and answer each of them in light of the findings in this book. In doing so, I also discuss the strengths and limitations of the proposed information-theoretic approach and identify research desiderata.

2 Definition and typological perspective

The subject of this book is an ellipsis type in German|(, where, roughly speaking, an element that can function at least as a subject (1a) or as an object (1b) is omitted from the preverbal, clause-initial position of a verb-second (V2) declarative clause.

(1) a. Δ *Hab das schon gemacht.*
I have that already done
'(I) have done that already.'
b. Δ *Hab ich schon gemacht.*
that have I already done
'I have done (that) already.'

This phenomenon is known by a variety of terms, often implying a theoretical positioning on the part of the authors who invented or use them. Initially, Reis (1982) called it *ellipsis in telegraphese*, referring to the text type telegrams for which it is a characteristic. In the same year, Ross (1982) discussed it as *pro*(*noun*) *zap* (see also Fries 1988) thereby characterizing it as pronoun omission. Terms such as *Vorfeld-Ellipse* ('prefield ellipsis') (Reis 2000, Frick 2017) and *Vorfeld-Analepse*[1] ('prefield analepsis') (Hoffmann 1999) emphasize the positional restriction of the omission to the prefield (the preverbal position in V2 clauses, see below). The terms (*uneigentliche*) *Verbspitzenstellung* ('improper verb top positioning') (Auer 1993, Sandig 2000, Günthner 2006, Imo 2013, Ruppenhofer 2018) or *uneigentliche Verberststellung* ('improper verb first positioning') (Duden 2022: §493–494), which are mostly popular in spoken language research, describe the surface form of the phenomenon with the finite verb in clause-initial position. However, the addition of 'improper' suggests that this surface structure does not correspond to the real underlying structure.

In recent literature, the most popular term is *topic drop* (Thráinson & Hjartardóttir 1986, Sigurðsson 1989, Haegeman 1997, Jaensch 2005, Ackema & Neeleman 2007, Erteschik-Shir 2007, Haegeman 2007, Reich 2011, Trutkowski 2011,

[1] Zifonun et al. (1997) attribute the term *Analepse* ('analepsis') to Blatz (1896) and use it to refer to elliptical phenomena, which involve a linguistic antecedent as opposed to ellipsis where the omission can be reconstructed through the extralinguistic context (Zifonun et al. 1997: 569).

Volodina 2011, Schalowski 2015, Trutkowski 2016, Helmer 2016, 2017b, Ruppenhofer 2018, Freywald 2020).[2] It suggests that ellipsis targets the topic of a clause and is therefore related to similar terms such as *zero topic* (Huang 1984), *null topic* or *0-topic* (Fries 1988, Cardinaletti 1990, Eckert 1998), *Topik-Wegfall* ('topic omission') (Klein 1993), *Topik-Tilgung* ('topic canceling') (Sternefeld 1985), and *Themaanalepse* ('theme analepsis') (Zifonun et al. 1997). As the term *topic drop* is widely used, especially in international research, I use it in this book as a temporary label for the ellipsis type that is being investigated, without committing to its theoretical implications. However, I also show that *topic drop* may not be the most accurate term and propose that *prefield ellipsis* would be a more suitable designation.

2.1 Definition

For the ellipsis type that I henceforth term *topic drop*, I propose the following preliminary definition:

Definition 1 *Topic drop is the omission of a constituent from the prefield of declarative verb-second (V2) clauses.*

It is also a characteristic of topic drop in German, although not part of its definition, that it is mostly restricted to colloquial spoken language and to conceptually spoken (Koch & Oesterreicher 1985) text types. In the following, I briefly discuss the individual aspects of the provided definition, as well as the role of text type. Most of the points are examined in greater depth in the course of this book. I refer to the relevant chapters and sections as needed. At the end of Chapter 3, I present a refined definition of topic drop that captures the prefield restriction more precisely.

2.1.1 Omission

By stating that topic drop involves the omission of a constituent, I take the view that utterances with topic drop are syntactically incomplete, i.e., they are cases of ellipsis (e.g., Reich 2011, 2018). I assume that the preverbal position, which can be

[2] The term seems to have been first used by Thráinson & Hjartardóttir (1986) in the title of their paper *Pro-drop, topic-drop...: where do old and modern Icelandic fit in?*, presumably as an analogy to *pro-drop*. In the text itself, however, Thráinson and Hjartardóttir speak of *zero topics*. To the best of my knowledge, it is Sigurðsson (1989: 145) who first explicitly states that "[t]he German pronoun zap [...] is actually a 'topic-drop'."

described as the prefield in the topological field model (Drach 1937, Höhle 1986, Wöllstein 2018) or as the specifier of the complementizer phrase [Spec, CP] in generative grammar (Thiersch 1978, den Besten 1983), is still present in the clause but contains a covert or phonologically empty element instead of an overt one (see Section 3.6 for different options how to represent this). Reis (2000: 218) points out that, despite their syntactic incompleteness, utterances with topic drop are generally "functional parallel" to complete V2 structures. Helmer (2016), however, emphasizes functional differences between utterances with an overt prefield constituent and utterances with a covert prefield constituent. I address the question of when topic drop is actually used in the second part of this book.

2.1.2 Prefield restriction

Part of the above definition is the main licensing condition of topic drop, namely, its generally assumed restriction to the prefield (e.g., Fries 1988, Auer 1993, Zifonun et al. 1997, Reis 2000, Frick 2017, Freywald 2020), i.e., the position immediately to the left of the finite verb in declarative V2 clauses (2a). Omissions in the middle field are usually considered to be impossible (but cf., Helmer 2016: and see Section 3.1.3).

The prefield restriction implies that topic drop can only occur in clauses that have a prefield. This means that topic drop is possible in declarative main clauses but not in subordinate clauses with verb-final word order (2b), or interrogative and imperative clauses with V1 word order. It is also not possible to omit operator phrases from the prefield that have a specific function, such as relative or interrogative phrases, i.e., the question word in *wh*-questions with V2 word order cannot be omitted,[3] see (2c).[4] The acceptability of topic drop in potentially

[3]While it could be argued that *wh*-words asking for adjuncts (e.g., *when*, *where*, *why*) cannot be omitted because they usually provide information about what the question is asking for and cannot be recovered if omitted, this is not the case for *who* or *what*. Those *wh*-words that ask for arguments can in fact be recovered from the sentence structure. For instance, in (i) it would generally be easy to infer that the omitted *wh*-word needs to be the one asking for the subject, i.e., *wer* ('who').

(i) *Δ *Hat angerufen?*
who has called
'(Who) has called?'

Since topic drop is still not possible in this case, it seems reasonable to assume the proposed ban on topic drop of operator phrases, as suggested above. I thank Ingo Reich for this suggestion.

[4]Note that the topic drop example is identical to the polar question *Have you done that already?* but does not convey the intended meaning of asking for the time, place, or reason.

embedded V2 clauses (2d) is disputed. While Jaensch (2005) argues that it may be acceptable for some speakers and in restricted contexts, Cardinaletti (1990) rejects it altogether. In Section 3.3, I investigate this case in detail with experiment 3; the results of which motivate a specification of the prefield restriction.

From the observation that the prefield usually contains exactly one single constituent,[5] it follows that the omitted element has to be a constituent[6] and that topic drop cannot target more than one constituent per clause (2e) (Fries 1988). In Chapter 3, I examine the prefield restriction of topic drop in detail.

(2) a. **Das hab* Δ *schon gemacht.*
that have I already made
'(I) have done that already.'

b. **Tino weiß, dass* Δ *das schon gemacht hab.*
Tino knows that I that already made have
'Tino knows that (I) have done that already.'

c. *Δ *hast du das schon gemacht?*
when/where/why/… have you.SG that already made
'(When/where/why/…) have you done that already?'

d. ?*Tino weiß,* Δ *hab das schon gemacht.*
Tino knows I have that already made
'Tino knows (I) have done that already.'

e. *Δ *Hab* Δ *schon gemacht.*
I have that already made
'(I) have (that) already done.'

[5] An exception to this rule are well-known cases of so-called *mehrfache Vorfeldbesetzung* ('multiple fronting') (Müller 2005), such as (i), where two constituents occur in the prefield. Since these cases are exceedingly rare and topic drop does not seem to be able to target both constituents simultaneously, I do not discuss them here.

(i) a. *Alle Träume gleichzeitig lassen sich nur selten verwirklichen.*
all dreams concurrently let themselves only seldom realize
'Only rarely can all dreams be realized concurrently.' (Müller 2005: 299)

b. *Die Kinder nach Stuttgart sollst du bringen.*
the kids to Stuttgart shall you bring
'You are to bring the kids to Stuttgart.' (Engel 1970: 81)

[6] In German, the constituent status of a word or phrase is often tested by inserting it into the prefield (the prefield test) (Pittner & Berman 2021: 33).

2.1.3 Omitted constituent

I stated in the definition above that topic drop is the omission of a constituent. In the literature, this constituent is often equated with a pronoun (Klein 1993, Jaensch 2005, Reich 2011, Volodina & Onea 2012, Duden 2016). The Duden grammar specifies that topic drop is the omission of personal pronouns in subject function and of weakly stressed or unstressed demonstrative pronouns in subject, object, or predicative function (Duden 2016: §1378, Duden 2022: §35).[7] For instance, Volodina & Onea (2012: 214) explicitly state that verb arguments can only be omitted as definite pronouns because this ensures, in their view, that the referent is unambiguous and sufficiently activated in the discourse. I discuss this restriction in detail in Chapter 4 on recoverability. While recoverability is indeed an important condition not only for the omission of (referential) constituents but also for the resolution of pronouns, it is questionable whether topic drop can only target pronouns, as Volodina & Onea (2012) postulate.

In any case, it is difficult to determine what exactly is omitted in topic drop. In fact, we can only resolve the ellipsis by inserting a plausible prefield constituent depending on a possible antecedent. This prefield constituent is usually a pronoun. However, there are cases where it is unclear which pronoun, if any, should be inserted. For example, in (3), taken from the advertisement subcorpus of the fragment corpus FraC (Horch & Reich 2017, see Section 7.2.1 for details), the grammatical gender of the brand *lavera* is unclear, which makes it hard to pick a pronoun to substitute it.

(3) *lavera.* Δ *Wirkt*[8] *natürlich schön.*
lavera it/she/he works naturally beautiful
'lavera. (It/She/He) works naturally beautifully.' [FraC P1004–P1005]

Finally, a part of the literature argues that not only subjects and objects can be omitted but also temporal and local adverbials in the form of what Sigurðsson (2011: 289, footnote 33) terms "anaphoric light adverbials". They usually must be reconstructed as the proforms *da* ('there'/'here') or *dann* ('then') instead of a "classic" pronoun referring to a determiner phrase (DP) (Fries 1988, Sigurðsson 2011, Schalowski 2015). In sum, the claim that only pronouns can be targeted by topic drop seems to be too strong.

[7] The 2022 Duden adds that also the semantically empty expletive subject *es* ('it') can be omitted (Duden 2022: §35). See also Duden (2022: §268, §493–494) on topic drop.

[8] The slogan can be seen as a play on words, using the double meaning of *wirken* as 'to appear' and 'to work'. As this is irrelevant to the gender discussion, I only focus on the 'work'-meaning in the glossing and translation.

The case of omitted adverbials leads us directly to the second point regarding the omitted constituent, that is, its syntactic function. On the one hand, there are authors who assume that topic drop can only target verb arguments, that is, only subjects and objects (e.g., Cardinaletti 1990, Auer 1993, Jaensch 2005, Volodina & Onea 2012, Imo 2013).[9] However, there is no consensus on whether, first, only referential subjects or both referential and non-referential subjects can be omitted (see the discussion in Section 3.1.5), and, second, whether all types of objects can be omitted (equally well) (see Section 8.1). On the other hand, the view that only subjects and objects can be targeted by topic drop is challenged by Fries (1988) and Schalowski (2015), mentioned above, who argue that adjuncts, for instance in the form of adverbials, can also be omitted. A corresponding example from the dialogue subcorpus[10] of the FraC is given in (4). It also shows that resolving an omitted adverbial is often more difficult and less clear than resolving an omitted verb argument. In this particular case, it is unclear whether a temporal element (replaceable by *dann* ('then')) or a local element (replaceable by *da* ('there')) has been omitted.

(4) a. *und dann sind wir eben um zwölf Uhr spätestens in der*
and then are we PART at twelve o'clock at.the.latest in the
Filiale. [...] ja, das ist sehr schön, genau zum zum [sic!]
branch yes that is very nice exactly for.the for.the
Mittagessen. [...]
lunch
'And then we are just at twelve o'clock at the latest in the branch. [...] Yes, that's very nice, right at lunch time.'

[9] To my knowledge, there has been no discussion in the literature as to whether free datives (Hole 2014), which are not verb arguments, can also be targeted by topic drop. In the authentic example with a free dative in the prefield in (i), taken from the discussion section of a blog post, for instance, I would consider an omission of *der* to be at least marked, if not impossible. However, it is already controversial for dative objects whether they can be omitted (see Section 8.1).

(i) *Also meine Gini, ein labbi ist 16 geworden, ?*(der) backe ich jedes Jahr eine*
so my Gini a labrador.DIM is 16 become her.DAT bake I every year a
Torte!
cake
'So my Gini, a labrador turned 16, I bake (her) a cake every year!' (https://www.fashion-kitchen.com/2015/09/hackfleisch-wurstchen-torte-fur-den.html, visited on 01/02/2025)

[10] This subcorpus contains data from the Tübinger Baumbank des Deutschen/Spontansprache TüBa-D/S ('Tübingen treebank of spoken German') (Hinrichs et al. 2000), which in turn stem from the Verbmobil project (Burger et al. 2000).

b. *ja,* Δ *können wir mit den Kollegen Mittagessen gehen*
yes then/there can we with the colleagues lunch go
'Yes, (then/there) we can have lunch with the colleagues.' [FraC D348–D353]

A case similar to the omission of adverbs is the omission of *da* as part of prepositional adverbs such as *damit*, *darauf*, *davon*, *daran* etc. In colloquial speech in particular, *da* can be separated from the preposition, placed in the prefield, and omitted from there, as shown in example (5) (see Fries 1988, Reis 2000: for details).[11]

(5) *ja,* Δ *hab ich nur gerade nicht dran gedacht.*
yes there have I only now not on thought
'Yes, I just haven't thought about it right now.' [FraC S126]

While for prepositional adverbs the *da* is relatively easy to recover, for adverbs, as mentioned above, ellipsis resolution is often less straightforward. For such adjuncts, the question arises of where exactly to draw the line between a structure with omission and a possibly genuine V1 structure (see the discussion of V1 declaratives in Section 2.2.4). Therefore, I focus only on topic drop of verb arguments in this book and leave the omission of adjuncts and the placeholder *es* to future research. I mainly consider subjects, both referential and non-referential ones, as well as objects and, very marginally, predicatives.

2.1.4 Modality, register, and text type

Another property of topic drop that is also often discussed in the literature is its frequent occurrence in certain modalities, registers (Halliday 1978, Biber 1994), or text types. However, unlike null articles or null copulas, which are usually ungrammatical outside the text type headlines, topic drop can occur in several registers and text types, even if it is sometimes more, sometimes less marked, or appropriate. Therefore, this property of topic drop should not be considered as a defining criterion or a licensing condition.

It is important to note that the term *text type* is a problematic one, not least because it is often not clearly distinguished from similar concepts. For instance, in

[11] Prepositional adverbs are generally formed by combining *da* and the corresponding preposition into one lexeme (i.e., *da* + *mit* = *damit*). However, if the preposition starts with a vowel, an additional *r* needs to be inserted for phonological reasons (i.e., *da* + *r* + *an* = *daran*). If these cases are split, the remaining preposition does not only keep this *r* but also receives an additional *d* in the word-initial position, i.e., *an* becomes *ran* and then *dran*, as shown in example (5).

German, *Textsorte*, *Textklasse*, *Texttyp*, and *Textart* (all translatable as 'text type') are often used interchangeably (see, e.g., Gansel & Jürgens 2009: 65 and, more broadly, the many contributions in Habscheid 2011). In this book, I try to use the term *text type* in as general and theory-neutral a way as possible. For instance, I use it to refer to the various subcorpora in the FraC, although, for at least some of them, it may be more reasonable to assume that they contain different (more fine-grained) text types (such as the chat subcorpus, which contains both group chats in a chat room and one-on-one chats), while others would be better classified as *forms of communication* (see, for example, Dürscheid 2016).

In the literature, it is often stated that topic drop is a phenomenon of spoken German (Huang 1984, Cardinaletti 1990, Poitou 1993, Jaensch 2005, Dittmann et al. 2007, Volodina 2011), and even that it is exclusively restricted to spoken language (Volodina & Onea 2012: 219). This restriction is occasionally further specified to mean colloquial or informal spoken language (Huang 1984, Poitou 1993, Jaensch 2005). Volodina & Onea (2012: 219) argue that the more normative a text type is, the less frequent or less acceptable topic drop is. According to both authors, topic drop occurs most often in spoken informal discourse, which is characterized by "communicative proximity" (intimacy and a high degree of familiarity of the communication partners, referential proximity, etc.) (Volodina & Onea 2012: 221).

Here, Volodina & Onea (2012) refer to the influential concept of "language of immediacy and distance" developed by Koch & Oesterreicher (1985). Koch & Oesterreicher (1985: 23) propose the term *konzeptionelle Mündlichkeit* ('conceptual orality') to refer to text types that, regardless of whether they are phonically mediated, exhibit properties of the language of immediacy, such as the previously mentioned intimacy and high degree of familiarity of the communication partners, referential proximity, but also dialogue, face-to-face situation, affectivity, etc. They assume a continuum between (conceptually) spoken and written texts and classify a (written) diary entry as more conceptually spoken than a (spoken) job interview (Koch & Oesterreicher 1985: 18). Volodina (2011: 272) and Volodina & Onea (2012: 221) argue that topic drop occurs frequently in text types that they classify as conceptually spoken, for example, private letters, emails, chats, diaries, and chronicles (see Ruppenhofer 2018 and Section 2.1.5 for empirical support). The restriction to these and other text types is also mentioned in much of the remaining literature but without reference to Koch & Oesterreicher's (1985) concept of conceptual orality. For instance, Fries (1988: 27) mentions telegrams, private letters, diaries, certain types of conversations, certain literary texts such as narratives, (audio) dramas, or speech bubbles in comics. Trutkowski (2016: 188)

talks about "electronic [...] registers", Imo (2014: 153, footnote 6) about "computer-mediated communication such as chat, email or SMS communication". In particular, for the latter and their successors instant messages, topic drop is frequently described as a characteristic (Androutsopoulos & Schmidt 2002, Döring 2002, Dittmann et al. 2007, Frick 2017). According to Frick (2017: 236), speakers should be even more likely to use topic drop in instant messages than in text messages. She argues that the permanent visibility of the precontext, as well as the displayed information of whether the interlocutor is online or typing enforces the joint context orientation and, therefore, facilitates the recovery of omitted elements with linguistic or situational antecedents (see also Chapter 4).

Dittmann et al. (2007: 24–25) propose to sharply separate topic drop in text messages, which they consider to be an instance of telegraphese, from topic drop in spoken language. They refer to Auer (1993) and argue that the topic drop function of condensing adjacent utterances that he proposes (see also the corresponding discussions in my Sections 5.2 and 5.4) is not possible turn-initially, whereas topic drop in text messages can (and does) occur in this position. Since this book is mainly limited to topic drop in text and instant messages, I cannot contribute to an empirical differentiation of topic drop across text types. Still, it seems plausible to me that at least some of the effects I found for topic drop in text and instant messages also hold for other text types. This hypothesis is supported by the results of my experiments 9 and 10, which suggest that the difference between 1st and 2nd person singular in instant messages and spoken dialogues is equally irrelevant.

2.1.5 Topic drop and text type in the fragment corpus FraC

At this point, it seems reasonable to anticipate my corpus study (see Section 7.2 for an overview), which also revealed the typical occurrence of topic drop in certain text types. The FraC consists of 17 different text types. As Figure 2.1[12] and Table 2.1 show, the distribution of full forms and instances of topic drop is very different across these text types. Two parameters are relevant here: (i) the omission rate by which the bars in Figure 2.1 are arranged and (ii) the number of instances of topic drop by which Table 2.1 is ordered. Text messages are far ahead in both parameters. The text message subcorpus of the FraC contains the most instances of topic drop with 353 and has the highest omission rate of almost

[12] All bar plots in this book were created using the package ggplot2 (Wickham 2016) in R (R Core Team 2021).

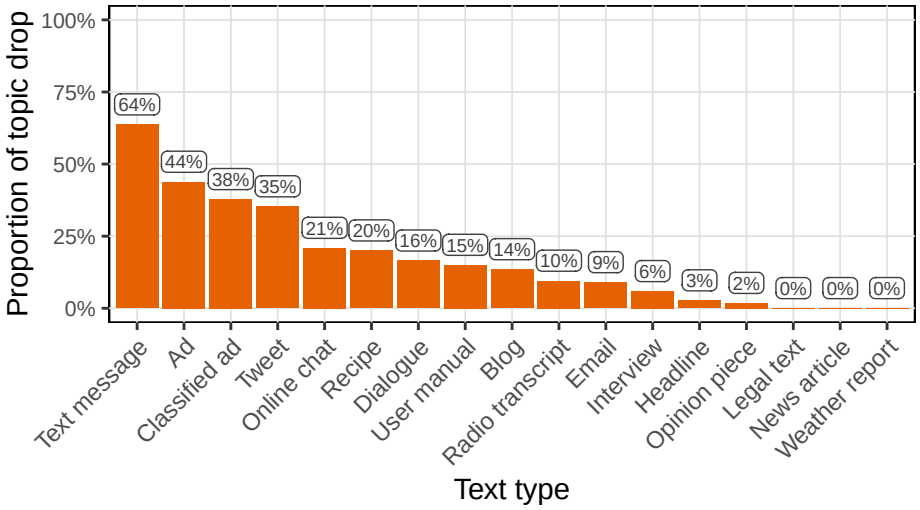


Figure 2.1: Proportion of topic drop across text types in the FraC (rounded to integers)

64%. In news articles,[13] legal texts, and weather reports, there are no instances of topic drop.

What we can see in Table 2.1 and Figure 2.1 is only partially consistent with what is assumed about the role of text type for topic drop in the literature. Two questions arise from this discrepancy: first, why is topic drop so common in text messages, and, second, why is it not more common in other text types, especially spoken ones? While an exhaustive answer to both questions must be left to future research, it is nevertheless permissible to present some initial thoughts here.

Regarding the first question, it is often discussed in the literature whether the frequent use of topic drop in text messages is a consequence of their technically

[13]The following example from a news article published online shows that topic drop is not impossible even in this text type. Possibly the more informal online setting and the lighter topic – sports, more precisely soccer – favor the occurrence of topic drop.

(i) *„Ein Gegentor lag nicht in der Luft", sagte Thomas Müller bei "Amazon Prime". „Ich*
a goal.against laid not in the air said Thomas Müller at Amazon Prime I
weiß nicht, was ich sagen soll." Δ *Hat man auch noch nicht oft erlebt.*
know not what I say shall that has one also yet not often experienced
' 'A goal against wasn't in the air,' Thomas Müller said on 'Amazon Prime'. 'I don't know what to say.' (That) hasn't been experienced very often either.' (tagesschau.de, 04/13/2022, https://www.tagesschau.de/sport/sportschau/bayern-villareal-103.html, visited on 03/13/2023)

Table 2.1: Distribution of topic drop and full forms across text types in the FraC, ordered by the number of topic drop occurrences

Text type	Full form	Topic drop	Total	Omission rate
Text message	201	353	554	63.72%
Classified ad	150	91	241	37.76%
Dialogue	431	85	516	16.47%
Online chat	280	74	354	20.90%
Tweet	121	66	187	35.29%
Ad	63	49	112	43.75%
Blog	282	44	326	13.50%
Email	386	38	424	8.96%
Radio transcript	315	33	348	9.48%
Interview	368	23	391	5.88%
User manual	51	9	60	15.00%
Opinion piece	220	4	224	1.79%
Recipe	12	3	15	20.00%
Headline	36	1	37	2.70%
News article	165	0	165	0.00%
Legal text	67	0	67	0.00%
Weather report	63	0	63	0.00%

imposed character limitation (e.g., Döring 2002: 100, Dürscheid & Brommer 2009: 7–8) and thus a mainly economic phenomenon (see Section 5.1). This assumption is questioned by the fact that omitting a personal pronoun has only little "savings potential" (Döring 2002: 108) and by the observation of Döring (2002: 103–104) that text message writers usually do not fully exploit the character limit and even fall well below it (see also Thurlow & Poff 2013: 172 for the same observation in English and Hård af Segerstad 2005: 331 for Swedish). In other words, the senders would still have enough space to make the prefield constituent explicit, and yet they do not do so. The purely brevity-based explanation falls short here (see also Dittmann et al. 2007: 45), at least synchronically. It might be possible, however, that topic drop was at least originally motivated by brevity and has become a conventionalized pattern over time. As an alternative explanation, Thurlow & Poff (2013: 172) suggest that "[t]he length and abbreviated linguistic forms of texts would seem to be more a function of the need for speed, ease of typing and, perhaps, other symbolic and pragmatic concerns, such as gender identity performance or a preference for more dialogic exchanges."

This idea is supported by a further observation from the literature according to which the linguistic and stylistic devices found in text messages, including topic drop, are similar to those of other text types. On the one hand, these are text types that have existed for a longer time, such as notes (Thurlow & Poff 2013: 173) and telegrams (Reis 1982, Barton 1998, Dittmann et al. 2007) and for which brevity and conventionalization should also play a role. Topic drop has already been discussed extensively in research in the context of telegrams under the term *telegraphese*, especially in the work by Jürgen Tesak and colleagues (e.g., Tesak & Dittmann 1991, Tesak et al. 1995, Tesak & Niemi 1997). On the other hand, Frick (2017: 230–237) and also Stark & Meier (2017) have found that topic drop continues to be used in communication via instant messaging services such as WhatsApp, which can in some ways be considered the successor of text messaging (for similarities and differences, see Dürscheid & Frick 2014). These text types are likewise often characterized by demands of speed and simplification. Future research on the role of text type for topic drop will face the task of systematically identifying similarities between text types that favor the occurrence of topic drop, such as the two previously mentioned characteristics but also the determinacy of speaker and hearer roles, as well as the general frequency of (personal) pronouns in the prefield.

The second question of why topic drop is not more frequent in some other text types in the FraC, concerns primarily the three spoken text types dialogues, radio transcripts, and interviews. While topic drop is often said to be restricted to or at least to occur mainly in spoken German (Huang 1984, Cardinaletti 1990, Poitou 1993, Jaensch 2005, Dittmann et al. 2007, Volodina 2011, Volodina & Onea 2012), these text types only have omission rates of 16%, 9%, and 6% respectively, which are clearly below the rates of several written text types. The relatively low omission rate in the radio transcripts may partially be explained by the fact that they are mainly based on pre-formulated texts that only mimic spontaneity and authenticity. It is, therefore, possible that topic drop is rather a means of spontaneous speech and that radio personalities and announcers (unconsciously?) orient themselves more strongly to standard grammar when pre-formulating their texts than do speakers of spontaneous speech, which then leads to a lower frequency of topic drop in the radio transcripts. However, this explanation does not apply to dialogues and interviews, since both corresponding subcorpora in the FraC do in fact contain transcripts of spontaneous speech. The dialogue subcorpus consists of spontaneous but elicited conversations about organizing a business trip, collected as part of the Verbmobil project (Burger et al. 2000). The interview subcorpus contains the transcripts of radio interviews conducted with politicians, scientists, prominent people, etc. conducted as part of the *SWR1 Leute* podcast.

What all three subcorpora have in common is that they were recorded either for building a corpus or for broadcasting and that the interlocutors are aware of this fact. Consequently, the conversations are mostly rather formal. While the formality in the radio transcripts varies depending on the radio station and its potential target audience, the interlocutors in the Verbmobil corpus use the formal form to address each other[14] and presumably, the interviews mostly took place between people who did not know each other personally beforehand. In these more formal settings, where the interlocutors are aware that they are being recorded, colloquial speech is likely less prevalent than in informal private conversations between friends. This hypothesis awaits an empirical investigation in future research, but it would be consistent with the second claim often made in the literature, that topic drop is limited to informal contexts or colloquial speech (Huang 1984, Poitou 1993, Jaensch 2005, Volodina & Onea 2012, Trutkowski 2016).

The explanation of formality may also apply to the text type emails in the FraC and explain why it does not exhibit a higher omission rate. The literature also claims that topic drop is particularly frequent in personal letters and their "successors", private emails (Fries 1988, Volodina 2011, Volodina & Onea 2012). However, in the email subcorpus of the FraC, there are at most 36% of private or at least informal emails, namely those between members of a band and between the same members and potential promoters. The remaining 64% are formal emails between students and their professors. While there are some cases of topic drop in these formal emails, as well (6 instances), most instances of topic drop stem from the more personal band emails (32 instances).

As discussed in Section 2.1.4, both publicity and unfamiliarity of the partners are conditions of communication that Koch and Oesterreicher associate with what they term *Sprache der Distanz* ('language of distance') (Koch & Oesterreicher 1985: 23) and with conceptually written text types. In contrast, a lack of publicity and familiarity characterizes *Sprache der Nähe* ('language of immediacy') and conceptually spoken text types in which topic drop is argued to occur frequently according to Volodina (2011) and Volodina & Onea (2012). This line of reasoning seems to be particularly fruitful for text messages, the text type with the highest omission rate and the largest number of topic drop in the FraC. Text messages, in particular, the type of private text messages contained in the FraC[15]

[14] According to Florian Schiel (p.c.), the interlocutors in the Verbmobil corpus were explicitly instructed to do so, i.e., to use formal speech and to use the formal form to address each other even though some of them knew each other from their workplace.

[15] Spycher (2004: 13) points out that text messages are also used by organizations or companies to provide clients with information, e.g., concerning train timetables, stock market news, weather reports, concerts, etc. Such text messages with a service function are not included in the FraC subcorpus.

exhibit many of the exemplary features that Koch & Oesterreicher (1985: 23) list as typical for the language of proximity: dialogue, familiarity, free development of topics, no publicity, spontaneity, involvement, expressiveness, and affectivity. However, they do not meet the criteria of face-to-face communication and interleaving of situations. Tweets, online chats, and blogs exhibit fewer of these features since they are public (except for the one-on-one chats) and less familiar, as the writer and the addressee do not necessarily know each other. The text types ads, classified ads, recipes, and user manuals meet even fewer criteria for conceptually spoken text types, but they still exhibit high rates of topic drop.

In ads and instructions, topic drop often occurs as part of an enumeration, such as in (6), or in the context of a combination between text and image where the object shown in a picture is left out in the corresponding linguistic description.

(6) a. *Die Spezialpflege mit Feuchtigkeit bindendem Urea*
the special.care with moisture binding urea
'The special care with moisture-binding urea.'
b. - Δ *glättet spürbar*
it smoothes noticeably
'(it) smoothes noticeably'
c. - Δ *lindert Juckreiz*
it relieves itching
'(it) relieves the itching'
d. - Δ *stärkt die Hautbarriere*
it strengthens the skin.barrier
'(it) strengthens the skin barrier' [FraC P986–P989]

The four instances of topic drop in the recipe subcorpus stem from comments by the recipe author, i.e., add-ons to the actual instructions, as illustrated in (7).

(7) *Ich streue immer etwas Petersilie darüber,* Δ *sieht gut aus und*
I sprinkle always some parsley thereon that looks good VPART and
schmeckt.
tastes
'I always sprinkle some parsley on top, (that) looks good and tastes good.'
[FraC R82]

In classified ads, topic drop occurs especially often in personal advertisements. At least originally, it seems to have served an economic purpose (see Section 5.1) there, as shown in example (8). Given the fact that the costs for personal adver-

tisements often depend on their length (e.g., Bachmann-Stein 2011: 98), topic drop can be considered a means to express more information in less space, similar to abbreviations such as the *NR* for *Nichtraucher* ('nonsmoker'). Further research is needed to decide whether topic drop is generally an expression of linguistic economy in this text type or whether the economic consideration led to a grammaticalization and/or conventionalization of topic drop in personal advertisements.

(8) a. *Kurz zu mir,* Δ *bin geschieden, 45, 1,78m, braune kurze Haare mit*
briefly to me I am divorced 45 1.78m brown short hairs with
grauen Ansätzen, NT, NR, männlich.
grey roots nondrinker nonsmoker male
'Briefly about me, (I) am divorced, 45, 1.78 m, brown short hair with gray roots, nondrinker, nonsmoker, male.'

b. Δ *Wohne und arbeite in Potsdam,* Δ *wäre cool wenn Du*
I live and work in Potsdam it would.be cool if you.2SG
auch in Potsdam oder Umgebung Dein zu Hause hast.
also in Potsdam or surrounding your.2SG at home have
'(I) live and work in Potsdam, (it) would be cool if you also have your home in Potsdam or the surrounding area.' [FraC N1491–N1942]

These examples show that there are potentially several reasons why topic drop does or does not occur frequently in certain text types. An explanation that only relies on (conceptually) spoken text types or economy falls short of accounting for the whole data. Here, further research is needed (but see Ruppenhofer 2018 for a first corpus study of topic drop in several text types).

2.2 Distinguishing topic drop from related phenomena

Having characterized topic drop as an ellipsis in predominantly (conceptually) spoken German, in this section, I distinguish it from (seemingly) similar phenomena to also determine what topic drop is *not*. This is of importance because it shows even more clearly what the defining or characteristic properties of topic drop are and how they may be appropriately modeled syntactically.

Following the seminal paper by Huang (1984), topic drop is considered to be one of three types of referential null subjects that are often distinguished, each characteristic of different languages. Besides the "Germanic *topic drop* type", Sigurðsson (2011: 268) lists the "Romance *pro drop* type", and the "Chinese *discourse drop* type", which is also called discourse *pro*-drop or radical *pro*-drop, referring

to the most intensively discussed representatives.[16] I discuss not only similarities and differences of topic drop in German to the two other null pronoun types *pro*-drop and radical *pro*-drop, but I also distinguish it from two phenomena relevant to German, null subjects in German dialects and V1 declaratives. I argue that it is the positional restriction of topic drop and its syntactic incompleteness that justify a distinction from the other phenomena.

2.2.1 *pro*-drop

At first glance, it may seem natural to treat topic drop as an instance of the similar phenomenon *pro*-drop, i.e., the non-realization of subject pronouns such as the Spanish *yo* ('I') in (9) (Perlmutter 1971).

(9) Δ *Ya lo he hecho.*
I already this have done
'(I) have done that already.'

The concept of a so-called pro-*drop parameter* stems from generative grammar. In languages where the parameter has a positive setting, "not only 'weather verbs', such as *rain* but also verbs that have subjects with a definite θ-role may appear at surface structure with no NP subject" (Chomsky 1981: 28). It is commonly assumed that *pro*-drop is licensed by a rich inflectional system (Taraldsen 1980, Rizzi 1982). Ackema & Neeleman (2007: 82) state that "[t]he rationale behind the hypothesis that rich agreement is necessary for pro drop is that the features of the empty pronoun would not be recoverable without it." In particular, Romance languages such as Italian and Spanish, but also Slavic languages, Hebrew and Basque (see Perlmutter 1971) are characterized as *pro*-drop languages and distinguished from non-*pro*-drop languages such as English and French that normally require pronominal subjects to be realized.[17]

Coming back to topic drop, we can observe that the utterance with *pro*-drop in (9) looks strikingly similar to the topic drop example (1a), repeated here as (10).

[16] It is important to note that Sigurðsson (2011) himself argues for a unified approach to null arguments, proposing the principle of *C/edge linking* as a requirement. See Section 3.6 for details.

[17] As Dryer (2013a) problematizes, the concept of *pro*-drop is potentially Anglo-centric because it assumes that non-realization is a deviation from the norm of realizing the subject. However, in most languages of the world (Dryer 2013a lists 437 out of 711 languages), pronominal subjects are expressed through affixes on the verb, while only 82 languages behave like English and German in that they have normally present subject pronouns. Aware of the possibly Anglo-centric view, I nevertheless follow most of the previous literature and continue to use the term pro-*drop* and speak of the subject being unrealized or covert in these languages.

However, although *pro*-drop and topic drop share the lack of realization of the subject in declarative main clauses, there are at the same time several differences that have led authors to explicitly argue against equating both phenomena (e.g., Fries 1988, Klein 1993, Trutkowski 2011, Volodina 2011).

(10) Δ *Hab das schon gemacht.*
I have that already done
'(I) have done that already.'

First, *pro*-drop is restricted to subject pronouns in the nominative case, whereas topic drop can also target objects and adverbs. Second, topic drop is positionally restricted to the prefield, i.e., the preverbal position of declarative V2 clauses ([Spec, CP] in generative terms),[18] and to main clauses, whereas *pro*-drop has no such restriction and can occur in several positions as well as in subordinate clauses, as in the Spanish example (11).

(11) a. *Boris Becker: Si* Δ *fuiste un campeón,* Δ *siempre serás un campeón*
if you were a champion you always will.be a champion
Boris Becker: 'If (you) were a champion, (you) will always be a champion' (elpais.com, 08/27/2018)[19]

b. *Wenn Δ ein Champion warst, wirst Δ immer ein Champion sein

A further difference is that the use of overt subject pronouns in *pro*-drop languages such as Spanish is often argued to fulfill special pragmatic functions such as emphasis or ambiguity avoidance (Davidson 1996, Pešková 2013). This means that example (11a), in which *tú* is unrealized, is not functionally equivalent to a variant with overt subject pronouns. This is different for topic drop, where according to Reis (2000: 218), utterances with and without the omission are functionally parallel (but cf., Eckert 1998 and Helmer 2016 for functional differences). These major differences lead me to separate the two phenomena and to discuss topic drop independently of *pro*-drop.

[18] Note that Nygård (2014) and Helmer (2016) consider topic drop to also be possible in the middle field. See Section 3.2 for details.

[19] Source: https://elpais.com/deportes/2018/08/27/actualidad/1535371649_689952.html (visited on 12/29/2024).

2.2.2 Discourse (*pro*)-drop

Related to *pro*-drop, and sometimes even equated with it, is a phenomenon called discourse *pro*-drop (Tomioka 2003, Barbosa 2011), discourse drop (Sigurðsson 2011), or radical *pro*-drop (Neeleman & Szendrői 2005, 2007). In languages such as Chinese, Japanese, and Korean (see Neeleman & Szendrői 2007 for the discussion of further languages) pronouns can be omitted rather freely as long as they are recoverable from the discourse context (Huang 1984, Neeleman & Szendrői 2007), as illustrated in the Chinese example (12).

(12) a. *A: Zhangsan kanjian Lisi le ma?*
Zhangsan see Lisi LE[20] Q
A: 'Did Zhangsan see Lisi?'
b. *B:* Δ *kanjian ta le.*
he see he LE
B: '(He) saw him.'
c. *B:* Δ *kanjian* Δ *le.*
he see he LE
B: '(He) saw (him).'
d. *B: Zhangsan shuo [*Δ *kanjian* Δ *le].*
Zhangsan say he see he LE
B: 'Zhangsan said that (he) saw (him).' (Huang 1984: 533, shortened)

Unlike the "classic" *pro*-drop languages discussed in the literature that have a rich inflectional system, the discourse (*pro*) drop languages Chinese and Japanese completely lack agreement (Neeleman & Szendrői 2007: 672). This has led many authors to distinguish between both phenomena (e.g., Huang 1984, D'Alessandro 2015), but there are also unitary approaches (e.g., Sigurðsson 2011, Duguine 2014). Regardless of whether discourse (*pro*) drop is considered to be a subtype of *pro*-drop or a phenomenon on its own, it is different from topic drop in German. While example (12b) again looks similar to the topic drop examples above – a preverbal subject is omitted from a declarative main clause –, examples (12c) and (12d) differ remarkably from topic drop in German. In both (12c) and (12d) not only the subject but also the clause-internal object is omitted, in (12d) even from a subordinate clause. The omission of two constituents per clause and the omission from a subordinate clause violate two main restrictions of topic drop in German. Therefore, I distinguish topic drop also from discourse (*pro*) drop.

[20] LE is the perfective or inchoative aspect marker (Huang 1984: 533, footnote 1).

2.2.3 Dialectal null subjects

Example (11b) showed that Standard German|(does not allow for *pro*-drop. However, in modern German dialects, there are instances of null subjects, such as (13) from Bavarian, that are regularly analyzed as *pro*-drop (Rosenkvist 2009, Axel & Weiß 2011, see also Frick 2017: 159–167 for Swiss German).

(13) *wennsd* Δ *moang wieda gsund bisd*
if.2SG you.2SG tomorrow again healthy are.2SG
'if (you) are healthy again tomorrow' (Axel & Weiß 2011: 36)

Such null subjects do not only occur in German dialects but "in all the major dialect groups within the Continental West-Germania" (Axel & Weiß 2011: 21) and were also possible in earlier Old High German (Axel & Weiß 2011) and in other older Germanic languages (Rosenkvist 2009).[21]

The properties that distinguish dialectal null subjects from topic drop are similar to those that were outlined for *pro*-drop in Section 2.2.1. Dialectal null subjects are not restricted to the preverbal position (Fries 1988: 23), and not only can they occur in main clauses with V2 word order but also in embedded clauses (Fries 1988: 23, Rosenkvist 2009: 171). Moreover, they seem to be generally restricted to pronouns in the nominative case (Fries 1988: 23), often require distinct inflectional marking on the verb (Rosenkvist 2009: 163), and often demand or are accompanied by morphological changes of, e.g., complementizers, i.e., so-called *inflected complementizers* such as *wennsd* ('if-2SG') in (13) (Fries 1988: 23, Rosenkvist 2009: 162). These characteristics clearly distinguish dialectal null subjects from topic drop.

2.2.4 V1 declaratives

In spoken German, as well as in certain written text types, we find utterances such as (14) and (15). Although they look identical to instances of topic drop because both start with a finite verb, they are in fact declarative utterances.

(14) *Kommt ein Mann in die Kneipe...*
comes a man into the bar
'A man comes into the bar...' (Önnerfors 1997: 50, shortened)

[21] There seems to be no consensus in the literature on whether null subjects in old Germanic languages follow the same principles as null subjects in modern dialects (see Axel & Weiß 2011 for a unifying view; but cf. Rosenkvist 2009).

(15) *Hab ich ihr ganz frech noch en Kuß gegeben.*
have I her quite cheekily yet a kiss given
'I gave her another kiss quite cheekily.' (Önnerfors 1997: 99, shortened)

In the literature, such cases are referred to as *V1 declaratives* (Önnerfors 1997, Reis 2000, Schwitalla 2012, Schalowski 2015) or *eigentliche Verbspitzenstellung* ('proper verb top positioning') (Auer 1993, Imo 2013). Auer (1993: 195) defines them as "clauses in which all obligatory complements are present but occur after the finite verb."[22] Zifonun et al. (1997: 637) term them "communicative minimal units in declarative mode with verb-first word order"[23] and state that they are complete because neither complements nor supplements, i.e., adjuncts, are omitted. This completeness is the main difference to topic drop and leads me (i) to use the term *V1 declaratives* for utterances such as (14) and (15), and (ii) to distinguish them from topic drop. In utterances with topic drop, arguments required by the verb or adjuncts are missing from the clause. In contrast, V1 declaratives are syntactically complete and, thus, not an elliptical phenomenon (Önnerfors 1997).

It is worth noting that the completeness of V1 declaratives has been, and still is, disputed in the literature (see Önnerfors 1997: 13–18, for an overview since the 19th century). Önnerfors (1997: 49–50) notes that Altmann (1987: 33) discusses an example similar to (14), which Önnerfors considers to be a V1 declarative, jointly with clear cases of topic drop such as (16). Altmann (1987: 33) treats both as "confounding factors when determining the verb order"[24] and as cases of "Vorfeldellipse" ('prefield ellipsis'), thereby indirectly equating topic drop and V1 declaratives.

(16) Δ *Komme morgen.* Δ *Bleibe drei Tage.*
I come tomorrow I stay three days
'(I) come tomorrow. (I) will stay three days.' (Altmann 1987: 33)

Sandig (2000: 300) also treats V1 declaratives as ellipsis but as a separate phenomenon. She argues that in V1 declaratives so-called "cohesive devices" such as (*und*) *da* or (*und*) *dann* are "saved", i.e., omitted. Several authors, such as Oppenrieder (1987) and Imo (2013), at least suggest that an element can be 'inserted'

[22] My translation, the original: "Sätze[], in denen sämtliche obligatorischen Ergänzungen vorhanden sind, aber nach dem finiten Verb stehen" (Auer 1993: 195).

[23] My translation, the original: "kommunikative Minimaleinheiten im Aussagemodus mit Verberststellung" (Zifonun et al. 1997: 637).

[24] My translation, the original: "Störfaktoren bei der Verbstellungsbestimmung" (Altmann 1987: 33).

into the presumably empty prefield, or that V1 declaratives can be 'translated' into corresponding V2 declaratives by inserting elements into the prefield (e.g., Auer 1993). As candidates for these elements, the authors most often propose the semantically relatively empty elements *da* ('there'), *dann* ('then'), and *es* ('it') (Oppenrieder 1987, Auer 1993, Poitou 1993, Sandig 2000, Imo 2013; see Önnerfors 1997: 18) but also *jetzt* ('now') (Oppenrieder 1987) or even *davon* ('thereof') or *darüber* ('thereover') (Imo 2013).

In Önnerfors (1997), all three proposals toward a more or less elliptical nature of V1 declaratives are rejected. He compares the distribution of V1 declaratives and their apparent V2 equivalents with *es* or *da* in the prefield. In several cases, he argues that the V2 variant is inadequate in context or in the text type or leads to a shift in meaning compared to the original V1 utterance (Önnerfors 1997: 54–59, see (17) vs. (14) and (18) vs. (15)). Önnerfors concludes that, since neither V2 sentences with *es* nor with *da* can adequately replace the corresponding V1 declaratives in any usage case, V1 declaratives cannot be an instance of ellipsis but are complete sentences without a prefield (Önnerfors 1997: 4).

(17) (?Es/?Da) kommt ein Mann in die Kneipe...

(18) (*Es/Da) hab ich ihr ganz frech noch en Kuß gegeben.

Önnerfors (1997) furthermore supports his claim with diachronic and cross-linguistic data. First, he argues that V1 declaratives are a very old word order option in Germanic languages. They were already present in Proto-Indo-European and Proto-Germanic, albeit being more marked in comparison to the dominant verb-final word order (Önnerfors 1997: 8). Even when several Germanic languages developed V2 word order, V1 remained available in all Germanic languages – synchronically and diachronically (Önnerfors 1997: 9). For instance, V1 word order was frequently used in Old High German and Early New High German but less so in Middle High German (Önnerfors 1997: 10). Second, he states that the V1 declaratives, which several authors argue are derived by ellipsis from V2 structures, are older than their alleged 'full forms'. V2 declaratives with *es* in the prefield developed in Middle High German, utterances with *da* are attested earlier, but both types still occurred later than the first V1 declaratives (Önnerfors 1997: 52). This makes an ellipsis interpretation of V1 declaratives impossible, at least in Old High German, and less likely for modern variants. Following Occam's razor, Önnerfors (1997: 52) proposes to prefer the simpler option, i.e., to treat V1 declaratives as sentences without a prefield, which have been in existence in the German language since the Old High German period, over a more complex approach that treats V1 declaratives as elliptical utterances where a semantically empty element is first inserted and then deleted from the prefield.

Without ultimately concluding whether all cases treated by Önnerfors (1997) as V1 declaratives are in fact such, I distinguish them from topic drop. In this book, I focus mostly on topic drop of verb arguments because, for them, it can be clearly decided whether they have been omitted from a clause. This decision has the practical consequence that I, unlike, e.g., Frick (2017), do not consider cases where a placeholder *es* could theoretically be placed in the prefield, as in (19).

(19) *(Es) liest ja nicht jeder die Frankfurter Rundschau.*
it reads PART not everyone the Frankfurter Rundschau
'Not everyone reads the Frankfurter Rundschau.' [FraC B222]

2.3 Topic drop in other Germanic V2 languages

As Rohrbacher (1999: 11) points out, all Germanic languages except English have V2 word order, i.e., in declarative main clauses the main verb is always in the second position following what Rohrbacher terms a "clause-initial phrase unit such as the subject [...] or a topicalized XP" (Rohrbacher 1999: 11).[25] In almost all Germanic V2 languages, this preverbal unit can be omitted from the clause, i.e., they allow for topic drop (Ackema & Neeleman 2007: 83) – apparent exceptions of Germanic V2 languages without topic drop are the Flemish dialects of Dutch (Haegeman 1996: 141) and some varieties of Danish (Rizzi 2000: 283).

In this section, I briefly go through some other Germanic V2 languages in alphabetical order and sketch the properties of topic drop in these languages. Such an outline could in principle serve to further sharpen the picture of topic drop in German by revealing possible language-specific peculiarities. However, in the following, an overall consistent picture of topic drop as a common ellipsis type of the Germanic languages under discussion emerges, despite several minor deviations. Here, too, the prefield restriction and the typical occurrence in certain registers and/or text types can be identified as characteristic features. Although

[25] Rohrbacher (1999: 14) points out that in sentences with a subject in the initial position followed by an adverbial, English has verb-third order (i), but that a finite auxiliary in the second position is required in direct complement questions (ii). He states that English is therefore often called a "residual V2" language, following Rizzi (1990b, 1996).

(i) a. Mary never liked trashy movies.
b. *Mary liked never trashy movies. (Rohrbacher 1999: 14)

(ii) a. *What kind of movies Mary liked?
b. *What kind of movies liked Mary?
c. What kind of movies did Mary like? (Rohrbacher 1999: 14)

this book is not typologically oriented but focuses on topic drop in German, this section is intended to open up a typological perspective. It aims at stimulating reflection on the extent to which the results obtained in this work are transferable to topic drop in other languages, without actually being able to accomplish such a transfer. I refer readers who are only interested in a summary of the typological results to Section 2.5.

2.3.1 Danish

The research on topic drop in Danish, at least the one available in English, seems to be less extensive than in most other Germanic languages. It even seems to be questionable whether topic drop is even possible in Danish. On the one hand, Haegeman (1996: 141) claims that Danish is more liberal than Dutch or German because it allows not only for the omission of referential constituents but also of expletives such as *det* in (20). In Section 3.1.5, I argue that this is not a real difference to German, since German also allows for the omission of expletives and similar elements.

(20) Δ *Regnede meget igår*
it rained much yesterday
'(It) rained a lot yesterday.' (Haegeman 1996: 141)

On the other hand, Rizzi (2000: 283) claims that in some varieties of Danish, topic drop is not possible at all. Similarly, Hamann & Plunkett (1998) state that the existence of topic drop in Danish is unclear but that it seems to be "not totally excluded on some stylistic levels and omission from the first position of the sentence is better than omission from other positions" (Hamann & Plunkett 1998: 49). They support their claim with example (21a) with an omitted 3rd person singular referential subject and example (21b) with an omitted direct 3rd person singular object.

(21) a. ???Δ *Har ikke købot boge*
he has not bought book.the
'(He) has not bought the book.'
b. ?Δ *Kender jeg ikke*
that know I not
'I don't know (that).' (Hamann & Plunkett 1998: 49, their judgments)

2.3.2 Dutch

Topic drop in Dutch is similar to topic drop in German. It only targets elements in the preverbal position or, as Thrift (2001) puts it, in the topic position (Thrift 2001: 71), it is limited to (Thrift 2001: 57) or, at least, common in spoken language (Weerman 1989: 54), and the omitted element must be recoverable from the context (Corver & Broekhuis 2016: 1329; see also Thrift 2001: 71). Topic drop in Dutch can target "subjects, (in)direct objects, complement prepositional phrases and the objects of prepositions" (Thrift 2001: 49). See (22) for an example of an omitted object.

(22) a. *Wat heb jij met dat boek gedaan?*
what have you.2SG with that book done
'What have you done with that book?'
b. Δ *Heb ik aan Marie gegeven.*
that have I to Marie given
'(That), I have given to Marie.' (Thrift 2001: 63)

Jansen (1981), cited in Thrift (2001), collected a corpus of spontaneous speech data and found that topic drop in Dutch is more common with objects, such as (22), than with subjects (Thrift 2001: 58). A similar tendency seems to be present in German as well (see Section 8.4). However, a systematic investigation with sufficient data is still pending for both German and Dutch.

In a survey[26] of 19 native speakers of Dutch, mostly students, Thrift (2001: 62) found that Dutch also patterns with German in that unstressed 1st and 2nd person objects cannot occur in the preverbal position and cannot be omitted (see Section 8.1). Interestingly, she found an effect of animacy for 3rd person direct objects that, as far as I know, has not yet been attested for German.[27] The omission of inanimate objects, such as (22), was rated as acceptable while animate direct

[26]The informants first read a context question before the target utterance was read aloud to them in at least three conditions (full forms with different word orders and a version with topic drop), apparently one after the other. Their task was to judge each utterance as "good", "ungrammatical", or "unsure" (Thrift 2001: 59). In addition to the problematic presentation of multiple conditions of the same token set in succession, which makes the manipulation evident, it remains unclear how many utterances each informant saw and how many utterances there were in total. Therefore, the results reported below should be treated with some caution given the methodology.

[27]While Reis (1982: 190) mentions animacy as a potential factor for topic drop in German, she does so in a different context. She speculates whether animate subjects can be better omitted in telegraphese. This contrasts with Poitou's (1993: 116) corpus study, who found that most of the omitted subjects in his data set were inanimate.

objects received mixed judgments from Thrift's informants (Thrift 2001: 63–64). The results of my corpus study do not generally argue against the existence of such a restriction in German, since most of the omitted direct objects refer back to abstract and inanimate propositions and verb phrases (see Section 10.4.1). However, a systematic empirical investigation of animacy is still pending for both languages.

An interesting difference between topic drop in Dutch and topic drop in German concerns 1st and 2nd person subjects. In her survey, Thrift (2001: 60) found that the omission of unstressed 1st and 2nd person subject pronouns is generally impossible except in diary contexts, in which 1st person singular pronouns can be omitted.[28] This result is surprising given the German data, where especially the 1st singular pronoun is frequently omitted in several text types (see Section 10.4.1). Thrift (2001: 72) attributes the Dutch pattern to the fact that the overt unstressed 1st and 2nd person pronouns cannot occur as topics, i.e., in the topic position, because their reference constantly shifts during a conversation, just as the roles of speaker and hearer change. At least cross-linguistically this explanation cannot suffice because turn-taking and reference shift also occur in German and other languages, where 1st and 2nd person subject pronouns can easily be omitted. Furthermore, when a speaker produces an utterance with an unrealized 1st or 2nd person subject pronoun, it is evident to whom they refer – themselves in the case of the 1st person and their interlocutor in the case of the 2nd person. Thus, it remains unclear how to explain this difference between German and Dutch. Despite this divergence, both languages show many similarities concerning the properties of topic drop, such as the positional restriction and the syntactic functions that can be targeted.

2.3.3 Icelandic

As Sigurðsson (1989) points out, in Icelandic topic drop occurs in declarative main clauses. Similar to the Germanic languages discussed so far, a subject can only be omitted from [Spec, CP] (23). Although object drop seems to be less frequent than in German (Sigurðsson 1989: 142), it is possible in Icelandic, in particular with the 3rd person singular neuter 'that', 'it' or 'this' (Sigurðsson 1989: 156). Moreover, Icelandic also allows for a special type of null objects independent of the [Spec, CP] restriction that is heavily constrained and does not seem to be equally acceptable to all speakers in all contexts (Sigurðsson 1989: 152–153).

Similar to topic drop in the other Germanic languages, topic drop in Icelandic seems to be most common in colloquial speech and in the so-called *telegraphic*

[28] Thrift (2001: 60) states that this result is in agreement with Jansen (1981).

style, which according to Sigurðsson (1989: 139) is distinctive of letters, diaries, postcards, and telegrams. As a parallel not only to German but also Swedish (see below) and in contrast to Dutch, Icelandic topic drop is particularly frequent with the 1st person singular. Ambiguous cases such as (23) are most likely interpreted as 1st person singular, which Sigurðsson attributes to the easy identifiability of the speaker (Sigurðsson 1989: 140).

(23) Δ *Veit* *það.*
I/she/he know.(1SG/3SG) it
'(I) know it.' / '??(She/He) knows it.' (Sigurðsson 1989: 140, his judgment)

2.3.4 Norwegian

For Norwegian, Eide & Sollid (2011: 333) state that topic drop is restricted to root clauses and topics. They implicitly assume a restriction to the prefield or [Spec, CP] by stating that the resulting structures have V1 word order (Eide & Sollid 2011: 346). In contrast, Nygård (2018: 10) states that topic drop occurs also "occasionally sentence-medially" (similar to Helmer 2016 for German, see Section 3.2). In their corpus of "2015 main clause declaratives from eight interviews with respondents from two dialect areas" (Eide & Sollid 2011: 342), Eide & Sollid (2011: 346) found that 8.3% of all clauses contained topic drop and that not only subjects, as in (24a), but also non-subjects, as in (24b), were omitted. Nygård (2014: 172–173) states that topic drop in Norwegian can also target objects and expletives.

(24) a. Δ *Traff HAM igjen i dag.*
I met HIM again today
'(I) met him again today.' (Eide & Sollid 2011: 333, original emphasis)
b. Δ *va ho å kjøpte hus der.*
then was she and bought house there
'(Then) she went and bought a house there.' (Eide & Sollid 2011: 346)

Both Eide & Sollid (2011) and Nygård (2014) emphasize the fact that topic drop typically occurs in certain registers and text types. Eide & Sollid (2011: 333) list spoken informal language, headlines, and diary syntax. Nygård (2014) extends the enumeration to include spontaneous speech and "'hybrid' registers such as sms, e-mails, online chats, Facebook updates, interviews and headlines" (Nygård 2014: 172). This is strikingly similar to the text types in which topic drop occurs in German (see Sections 2.1.4 and 2.1.5).

2.3.5 Swedish

According to Håkansson (1994: 50), topic drop is a common phenomenon in colloquial Swedish that, like in German, is restricted to [Spec, CP] (Mörnsjö 2002: 55, Platzack 2013: 27). Like topic drop in German, it targets not only subjects (25) and direct objects but also indirect objects, objects of prepositions, expletives, and "quasi-argument[s] or [...] frame topic[s] like här 'here', där 'there', då 'then', nu 'now' and the adjunctive så 'so, then'" (Platzack 2013: 29, see also Mörnsjö 2002). Mörnsjö (2002: 31) analyzed a collection of V1 declaratives and instances of topic drop collected from spoken language corpora and radio and television broadcasts, among others. She states that subjects and direct objects with propositional reference are the most common types of topic drop in her data, although objects with nominal reference were also omitted (Mörnsjö 2002: 57–58). However, she did not find any omitted indirect objects (Mörnsjö 2002: 62). This pattern is another similarity to topic drop in German, as Mörnsjö (2002) points out as well (see also Section 8.1). In addition, similarly to German and different from Dutch, Sigurðsson (2011: 279) claims that ambiguous utterances with topic drop and syncretic verb forms, such as (25), are often interpreted as 1st person (singular) by default, whereas 2nd and 3rd person readings are more constrained, the latter requiring a context with speaker shift.

(25) Δ *Kommer* *tillbaks imorgon.*
come.∅-AGR back tomorrow
'(I/We/She, etc.) will be back tomorrow.' (Sigurðsson 2011: 268)

One possible difference from German that Mörnsjö (2002) addresses is the possibility of 1st and 2nd person object topic drop. While topic drop of 1st and 2nd person objects is often considered to be ungrammatical in German (see the discussion in Section 8.1), Mörnsjö (2002: 70–73), admitting that she has not found any authentic examples, argues that in Swedish there is no actual ban on omitting them, but rather that pragmatic factors make it unlikely for them to be put into the prefield and omitted from there. Such a pragmatic explanation may also be beneficial for German (see the discussions in Sections 8.1 and 8.3).

2.3.6 Yiddish

Yiddish, unlike German, is claimed to have no null objects (van der Wurff 1996: 342). Rohrbacher (1999) states that it allows for referential null subjects (26) that have to be analyzed as topic drop because "they are restricted to the utterance-initial position, thus being permitted in matrix clauses without topicalization [...]

but being barred from questions [...], matrix clauses with topicalization [...], and embedded clauses" (Rohrbacher 1999: 253). Without providing numbers, Rosenkvist (2012: 437) emphasizes that topic drop is much more frequent in Yiddish than in other modern Germanic languages. He states that all subjects can be omitted in the preverbal position if an antecedent is available in discourse.

(26) Δ *Horevet iber di koykhes.*
she work over the strength
'(She) is working too hard.' (Rohrbacher 1999: 254, following Prince 1999: 83)

Interestingly, Rohrbacher (1999), following Prince (1999), and Rosenkvist (2012) argue that Yiddish does not only allow for topic drop but also for a restricted version of referential pro drop but only in the 2nd person singular, which distinguishes it from German.

2.3.7 Summary

The Germanic V2 languages discussed share with German the possibility of omitting a subject from the [Spec, CP] or prefield position of declarative main clauses. This similarity, which is most likely rooted in the general V2 property, suggests that it may be beneficial to also look at topic drop cross-linguistically in a systematic manner. This is not least due to the fact that topic drop seems to be restricted not only positionally in all these languages but also to spoken language and informal written text types. Besides these similarities, the differences between the languages, such as the apparent impossibility of object omission in Yiddish or the apparent impossibility of omitting the 1st and 2nd person in Dutch, Dutch may further motivate the joint consideration of topic drop in the Germanic languages to explain which language-specific differences lead to deviations from the cross-linguistic pattern.

2.4 Register-dependent omissions in other languages

Having distinguished topic drop from similar phenomena previously in this chapter, and having emphasized that it can be regarded as a common ellipsis type of numerous Germanic languages with the same basic properties, I would like to undertake another typological digression in the following, namely, to register-dependent argument omissions in English, French, and Russian, Russian i.e., languages that are not regarded as classic V2 languages. For these phenomena, I

would also like to present the differences and similarities to topic drop, but it is more difficult to decide at this point whether a distinction from topic drop or a joint consideration with topic drop is the more useful strategy. I do not commit myself to either option here, but I would like to suggest, as in the previous sections, a typological perspective on topic drop. There may be cross-linguistic tendencies favoring, for instance, left-peripheral positioning and/or the occurrence in certain registers or text types as characteristics of different ellipsis types.

2.4.1 Subject omission in English and French

There are also register-dependent null arguments in the non-V2 languages English (27) and French (28).

(27) A very sensible day yesterday. Δ Saw no one. Δ Took the bus to Southwark Bridge... (*The diary of Virginia Woolf*, vol. 5, 1936–41: 203; cited in Haegeman 1990: 167, shortened)

(28) Δ *M'accompagne au Mercure, puis à la gare...* Δ *s'est*
she me.accompanies to.the Mercure then to the station... she herself.is
donné souvent l'illusion de l'amour à P...
given often the.illusion of the.love to P...
'(She) accompanies me to the Mercure, then to the station... (She) often gave herself the illusion of love to P...' (Paul Léautaud, *Le Fléau. Journal particulier* 1917–1930: 69–70; cited in Haegeman 1990: 167, shortened)

The subject pronoun of a finite main clause can be omitted in spoken colloquial English (Haegeman 1997, Weir 2012) and in certain "abbreviated written registers" (Haegeman 1997: 233) of English and French (see also Haegeman 1990, 2000, Weir 2012), such as diaries, postcards, informal letters, notes, instructional writing (e.g., recipes, stage directions), text messages, and emails (Haegeman 1990, 1997, 2017, Weir 2012).

These null subjects – null objects are not possible (Haegeman 1990: 172) – are mostly restricted to main clauses, i.e., they cannot occur in subordinate clauses and only marginally in embedded clauses (Haegeman 1990: 168, Haegeman 2007: 98, Weir 2012: 108). Subjects of all grammatical persons can be omitted (Haegeman 1997: 238, Haegeman 2000: 133, Haegeman 2007: 95, Haegeman 2017: 233, Weir 2012: 106), as well as both referential and non-referential subjects (Haegeman 1997: 236, Haegeman 2017: 233), provided they are not focused (Weir 2012: 108). Both Weir (2012) and Haegeman (2000) describe the English null subjects by

and large as a left-edge phenomenon[29] (Weir 2012: 109, Haegeman 2000: 139), i.e., the subject is usually deleted in sentence-initial position. However, written, but not spoken English, as well as written French, allow for preposed adjuncts (29a) but not arguments (29b), so that the null subject is not necessarily the left-most element of the clause (Haegeman 2000: 148).

(29) a. After Dr. Krook, Δ had good lunch at Eagle with Gary [Hamp]. (Sylvia Plath, 03/06/1959: 126; cited in Haegeman 2000: 149)
b. *Dr. Krook, Δ met for lunch at Eagle. (Haegeman 2000: 149)

This difference led Haegeman, who in previous work had discussed null subjects in spoken and written English jointly (Haegeman 1997, 2013), to separate them in Haegeman (2017). With this distinction, she follows Weir (2012) who lists another distributional difference. In spoken English, null subjects cannot occur before unclitized variants of verbs that also have clitic forms such as *am*, *are*, and *is*, see (30).

(30) *Δ am thinking of leaving tomorrow. (Weir 2012: 107, his judgment)

The preceding discussion has shown that null subjects in English and French share many similarities with topic drop. In both cases, the omission is restricted to spoken language or special text types, mostly occurs at the left periphery (except for preposed adjuncts in written language) and in main clauses, and can target any grammatical person. These similarities led Haegeman (1990) to analyze null subjects in French and English as instances of topic drop. However, she corrects this view in Haegeman (1997) and Haegeman (2007) rejecting a topic drop analysis for three reasons: (i) In English and French only subjects but not objects can be null (Haegeman 1997, 2007), unlike in most topic drop languages (see Section 2.3). (ii) According to her argument, topic drop cannot target expletives and quasi-arguments as they cannot be topicalized, but these elements can be null in English and French (Haegeman 1997, 2007). (iii) A topic drop analysis would incorrectly predict that embedded null subjects should only be possible in English if no overt complementizer is present because subject extraction is possible in cases with null complementizers (Haegeman 2007: 103).

[29] Weir (2012: 109) notes that in spoken English not only subjects but also additional elements such as auxiliaries, determiners, or syllables can be omitted at the left edge of a sentence. A similar pattern has been observed in Norwegian, where topic drop sometimes seems to 'expand' so that the finite verb in the left bracket is also omitted, especially if it is an auxiliary or a modal verb (Eide & Sollid 2011: 333, Nygård 2014: 173). For German, Wilder (1996: 161–164) discusses similar cases as *Vorwärtstilgung* ('forward canceling'). I remain agnostic as to whether these cases are in fact some kind of extension of left-peripheral null subjects or topic drop, or whether they should not be more uniformly described as fragments (e.g., Morgan 1973, Merchant 2004, Reich 2007, Lemke 2021).

Concerning (i), the impossibility of omitting objects indeed seems to be the most notable difference between Germanic topic drop (but note the apparent ban on object omission in Yiddish) and the English and French left peripheral null arguments. For example, Wilder (1996: 178–179) states that, under his analysis of topic drop as *Vorwärtstilgung* ('forward deletion'), it is impossible to explain why a given, fronted object cannot be omitted in English. However, it seems reasonable to look for the cause in the word order rules of English and French, which have SVO as their dominant word order (Dryer 2013b). Therefore, the constituent that immediately precedes the verb is usually the subject and not the object in these languages. Thus, if the register-dependent omission in English and French were restricted to the immediate preverbal position, as is the case with topic drop in German, this would explain the restriction to subjects and the impossibility of omitting objects. The case of fronted objects addressed by Wilder (1996: 178–179) could then be treated analogously to cases of left dislocation in German, which cannot be targeted by topic drop either (see Section 3.4.1). In both cases, the left dislocated element is topicalized and, as, for example, Ward & Birner (1994: 161) discuss for English, made salient or emphasized, a function that would just no longer be fulfilled if it were omitted.

Regarding (ii), the literature review in Section 2.3 made it clear that topic drop can also target expletives and quasi-arguments at least in Swedish and Danish. In Section 3.1.5, I show that omitting expletives is also possible in German and present corresponding corpus data, which I complement with experimental evidence in Section 3.1.6. This poses a serious problem not only for the distinction between topic drop and English and French null subjects but also for the syntactic analysis of topic drop as presented in Haegeman (1990), which incorrectly predicts that topic drop cannot target expletives. I revisit this issue in Section 3.1.5.

Finally, concerning (iii), Haegeman & Ihsane (1999) did find instances of null subjects in embedded clauses in diary contexts not only without overt complementizers such as (31) but also with overt complementizers (32).[30] This represents

[30]Note that while example (31) stems from the authentic diary of the American poet Allen Ginsberg, example in (32) is taken from a fictional diary, Helen Fielding's *Bridget Jones's diary*. Haegeman & Ihsane (1999: 130–131) note that embedded null subjects do occur more frequently in the fictional text by Fielding than in Ginsberg's real journals, but still, they do occur there as well. Haegeman & Ihsane (1999: 131) argue, referring to other fictional diary texts, that not all of them exhibit a particularly high rate of embedded null subjects. They speculate that *Bridget Jones's diary* is special because it originally appeared as a newspaper column and was subject to a word limit for print purposed (1 000 for *Bridget Jones's diary*) (Haegeman & Ihsane 1999: 131). At this point, the question remains as to how common subject omissions after overt complementizers are in English, and to what extent they are a (deliberately chosen) stylistic device in diaries, or whether they also occur in authentic (spoken) speech data.

a difference from topic drop in German, where the omission in embedded clauses with an overt complementizer is ungrammatical, as discussed in Section 2.1.2.

(31) also said Δ should make girls (Allen Ginsberg: *Journals 1954–1958*: 99; cited in Haegeman & Ihsane 1999: 128)

(32) was worried that Δ might split (Helen Fielding: *Bridget Jones's Diary*: 227; cited in Haegeman & Ihsane 1999: 128)

In sum, the similarities between null subjects in English and French and topic drop in the Germanic languages, in particular, the restriction to the left periphery and the typical occurrence in certain registers or text types, are striking. Therefore, a joint consideration of the two phenomena seems quite promising, despite the differences that also exist concerning the omission of objects or the co-occurrence with an overt complementizer.

2.4.2 Argument omission in Russian

While most other Slavic languages are *pro*-drop languages, Russian is discussed as a mixed type (Franks 1995: 300–301). It shares with *pro*-drop languages the possibility of null arguments, i.e., subjects and objects can optionally be omitted. However, in contrast to these languages, the overt realization of these arguments does not put a special emphasis on them. It is even the case that special pragmatic conditions are required to license null arguments in Russian (Gordishevsky & Avrutin 2004: 188–189, see also Franks 1995: 300–301). Franks (1995: 307–308) defines argument omissions in Russian (33) as discourse ellipsis because they only target elements whose reference is clear from the discourse context (see also Zdorenko 2010: 120–121).

(33) a. *A: My vstret-im-s'a?*
we meet-PRES.1PL-REFL?
A: 'Shall we meet?'

b. *B: Davajte, kak* Δ *i dogovariv-al-i-s'.*
let.us as we and agree-PAST-PL-REFL
B: 'Let's do it, as (we) already agreed.' (Zdorenko 2010: 120)

In her corpus study of null subjects in Russian,[31] Zdorenko (2010: 126) found that, similar to topic drop and to null subjects in English and French, Russian

[31] Zdorenko searched three written and three spoken subcorpora of the Russian National Corpus for sentences with a verb, excluding imperatives and impersonal constructions. She annotated the first 100 hits per subcorpus for whether the subject was a noun, a pronoun, or null (Zdorenko 2010: 124–125).

null subjects occur preferably in spoken colloquial speech where almost a third of the subject pronouns are omitted.[32] Unlike topic drop and the English and French null subjects, in Russian, there is no restriction of this phenomenon to the left periphery and null subjects can also occur in subordinate clauses (e.g., Franks 1995: 302, his example 30). In this respect, Russian is more similar to the "classical" *pro*-drop languages. It is primarily the typical occurrence in certain modalities, registers, or text types that brings Russian argument omission close to topic drop. However, due to the lack of restriction to the left periphery, it seems more questionable here whether a joint consideration of the two ellipsis types would be useful.

2.5 Summary: a typological view on topic drop

In Section 2.2, I showed that there are distributional differences and diverging syntactic properties between topic drop in German on the one hand and the similar phenomena *pro*-drop, discourse drop, dialectal null subjects, and V1 declaratives on the other hand that justify a distinction. While its restriction to the left periphery of declarative main clauses distinguishes it from the positionally unrestricted cases of *pro*-drop, discourse drop, and dialectal null subjects, it is also different from V1 declaratives. In V1 declaratives, all obligatory verb arguments are overtly realized, whereas in most topic drop cases either the subject or the object is absent from the sentence. It is therefore useful to consider topic drop as an ellipsis type in its own right.

This phenomenon is found not only in German but in almost all Germanic languages, where it is tied to the typical V2 positioning with the resulting prefield restriction. My brief overview of several Germanic languages suggests that the omission is always restricted to this position, that subjects can always be omitted and that (direct) objects can often be omitted as well. It also shows that omission occurs primarily in spoken language and conceptually spoken text types. Therefore, it seems reasonable to consider topic drop not only as an ellipsis type in its own right but as a common ellipsis type of the Germanic V2 languages. Consequently, many of the results obtained in this book for topic drop in German should be potentially transferable to topic drop in other languages. Accordingly, it would be promising to empirically investigate whether topic drop in these languages is also generally unacceptable in the middle field, whether certain conjunctions can precede it, and whether its positional restriction can be described as a position that is either the highest [Spec, CP] of a root clause or a [Spec, CP]

[32] In the other spoken text types, as well as in the written ones, it was at most 6%.

not c-commanded sentence-internally by a potential identifier (see the discussion in Chapter 3). In a second step, it could be examined whether the factors that, according to the second part of this book, play a role in the usage of topic drop (see Chapters 8 to 11 for corpus and experimental results), such as grammatical person and verb type, are also of relevance for topic drop usage in other languages. If this is the case, the information-theoretic explanation that I discuss in Chapter 6 might also be transferable, i.e., the usage of topic drop in the other Germanic languages might also be determined, at least in part, by distributing information in a sentence as uniformly as possible to reduce the processing effort for the hearer.

Similar considerations of transferability apply, at least in part, to the cases of argument omission in English, French, and Russian discussed in Section 2.4. In all three languages, the omission seems to be particularly frequent in certain modalities, registers, or text types, exactly as with topic drop. In addition, at least for English and French, a syntactic restriction of the null subjects to the left periphery can be observed, similar to the prefield restriction of topic drop. Since both languages are languages with a dominant SVO word order but not classic V2 languages, they do not have a prefield position. Therefore, one could speculate whether the different word order properties could lead to different manifestations of what is actually a similar ellipsis type. The impossibility of object omission in English and French may not be a categorical difference between topic drop and null subjects but merely the logical consequence of the fact that the immediate preverbal position in English and French is usually only occupied by the subject but not by the object. Future research is needed to explore the possibilities of a joint analysis at least for Germanic topic drop and English and French register-dependent subject omissions.

This summary has again shown the central position of the prefield restriction for topic drop in German since it functions both as a separating (from other ellipsis types) and as a unifying factor (within the group of Germanic languages). In Chapter 3, I take a closer look at this central syntactic property and specify it using theoretical discussions and four acceptability rating studies.

Part I

Licensing of topic drop

3 Syntactic licensing: prefield restriction of topic drop

The most cited licensing condition of topic drop in German is its structural restriction (Reich 2011: 1850) to the preverbal position of declarative verb-second (V2) clauses, as shown by the contrast in (1) where *ich* ('I') can only be omitted in (1a) but not in (1b).

(1) a. Δ *Habe das schon gemacht.*
I have that already made
'(I) have already made that.'
b. **Das habe* Δ *schon gemacht.*
that have I already made
'(I) have already made that.'

In what follows, I address this prefield restriction in detail. For this purpose, I discuss the positions from the theoretical literature and relate them to my empirical results from four acceptability rating experiments. The resulting findings motivate a specification of the prefield restriction of topic drop and form the basis for a critical examination of the generative-syntactic analyses proposed so far in the literature.

While the absolute majority of researchers agree on the prefield restriction of topic drop, there is, to the best of my knowledge, only one dissenting opinion on this matter, namely Helmer (2016), who argues that topic drop in German can also occur in the middle field. I discuss her argumentation in Section 3.2 and assume for this overview, based on my definition from Section 2.1, that the prefield restriction holds – an assumption that is supported by the experimental evidence presented in Section 3.2.2.

Pre-theoretically, the position to which topic drop is restricted can be defined as preverbal (Fries 1988, Sandig 2000, Frick 2017), i.e., as located immediately before the finite verb. In the literature, this position is most frequently referred to as *prefield* (e.g., Oppenrieder 1987, Auer 1993, Zifonun et al. 1997, Reis 2000, Reich 2011, Volodina 2011, Schalowski 2015, Trutkowski 2016, Frick 2017), following the terminology of the so-called *topological field model* (Drach 1937, Höhle

1986, Wöllstein 2018). This model has its origins in the 19th century (e.g., Herling 1821, Erdmann 1886 and later Drach 1937; see Höhle 1986) and is used to describe the structure of German clauses by dividing them into several linearly ordered "field" and "bracket" positions (Wöllstein 2018), in which specific elements occur. V2 declarative clauses are usually described as consisting of five positions, but not all of them need to be filled: the prefield, which usually holds exactly one constituent, the left bracket, which only contains the finite verb, the middle field, which can hold any number of constituents, the right bracket, which contains verb particles or infinite verbs, and the postfield, which contains extraposed constituents (see Wöllstein 2018). Table 3.1 illustrates the scheme for two example sentences, which contain *ich* ('I') and *der Mann* ('the man') as prefield constituents respectively.

Table 3.1: Analysis of two example sentences within the topological field model

Prefield	Left bracket	Middle field	Right bracket	Postfield
Ich	habe	das schon	gemacht.	
I	*have*	*that already*	*made*	
Der Mann	muss	ein Buch	abholen	in der Stadt.
the man	*must*	*a book*	*collect*	*in the city*

Wöllstein (2018: 159) illustrates how the topological field model can be brought into correspondence with a classic complementizer phrase / inflectional phrase (CP/IP) model of generative grammar. In this approach, the prefield position corresponds to [Spec, CP], as shown in Figure 3.1, so that, in generative terms, topic drop is restricted to [Spec, CP] (Ackema & Neeleman 2007, Trutkowski 2016).

Since [Spec, CP] is frequently considered one of several topic positions or even the only one in German (Pittner & Berman 2021: 154; Wöllstein 2018: 158), several authors assume that the restriction to [Spec, CP] is, in fact, a restriction to the topic position (Huang 1984, Auer 1993, Jaensch 2005, Volodina 2011) or, more explicitly, a restriction to topics (Sternefeld 1985, Erteschik-Shir 2007, Helmer 2016).

Rizzi (1994) approaches the positional restriction of topic drop from a generative point of view by stating that subject topic drop – similarly to early null subjects in child language – is restricted to the "specifier of the root" (Rizzi 1994: 155), a position that c-commands every other constituent in the sentence and is

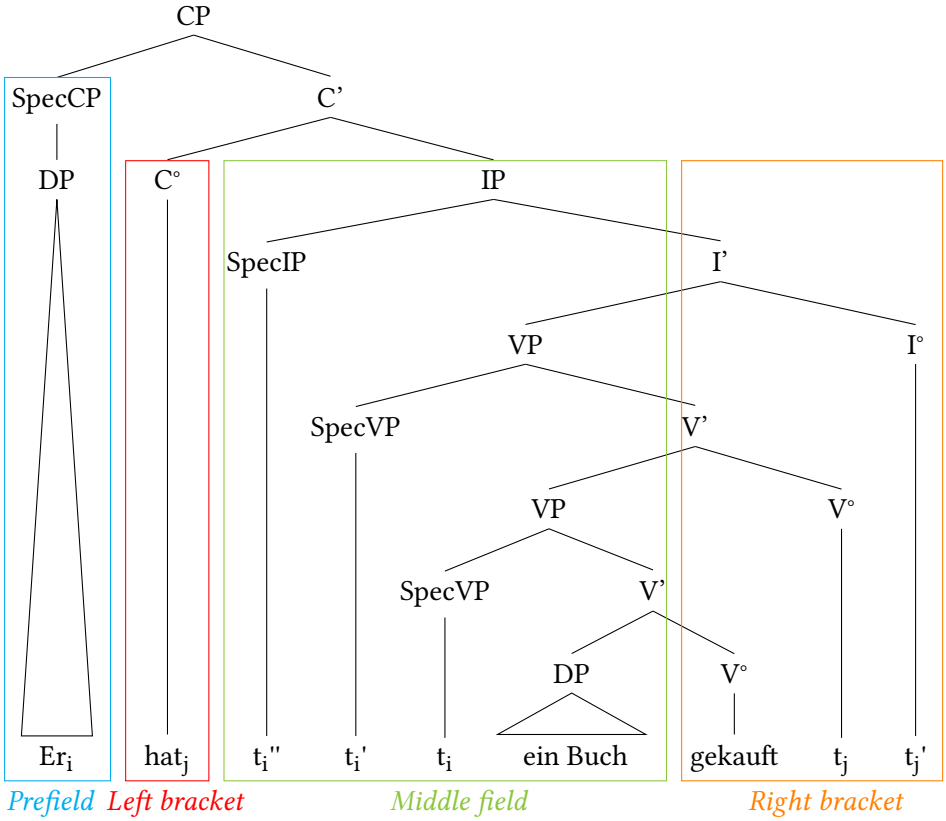


Figure 3.1: Correspondence between the classic CP/IP model of generative grammar and the positions of the topological field model, based on a figure by Wöllstein (2018: 159)

itself not c-commanded sentence-internally by a potential identifier.[1] Freywald (2020: 167) interprets the *specifier of the root*-restriction as a restriction to the highest syntactic position of an autonomous or root clause, which results in a restriction of topic drop to the highest [Spec, CP] of a sentence since Rizzi (1994) assumes that the root of V2 clauses is usually a CP.

In the literature, it is a matter of debate whether topic drop is possible in (potentially) embedded contexts.[2] While Cardinaletti (1990: 76) and Volodina (2011:

[1]The concept of c-command goes back to Reinhart (1976: 32), who defined it as follows: "Node A c(onstituent)-commands node B if neither A nor B dominates the other and the first branching node which dominates A dominates B."

[2]There is also an ongoing discussion about the status and the syntactic position of V2 clauses in complement function. I remain agnostic about this dispute, but I discuss it in Section 3.3.

272) deny this option, Jaensch (2005: 101) states that topic drop can occur in embedded clauses in specific cases like question-answer pairs, such as (2), and that its acceptability seems to be highly dependent on the speakers being asked.

(2) a. *Will Peter wirklich das neue Auto kaufen?*
wants Peter really the new car buy
'Does Peter really want to buy the new car?'
b. *Ich glaube [sic!] Δ hat er schon gekauft.*
I believe it has he already bought
'I think he's already bought (it).' (Jaensch 2005: 98, her judgment)

Trutkowski (2016: 224–225) likewise assumes that topic drop is generally possible in embedded clauses but states that there are stronger identity conditions between antecedent and gap for embedded than for unembedded cases. The question of whether topic drop can or cannot occur in potentially embedded clauses is related to the characterization of the position of topic drop as clause- or sentence-initial (Huang 1984: 547, Trutkowski 2016: 1), left-peripheral (Freywald 2020: 150), or being at the left edge of an utterance (Ackema & Neeleman 2007: 99). It remains unclear whether this also entails that topic drop is restricted to the syntactically highest position and/or to the very first position of the clause-forming sequence of elements.

From the above, four questions arise regarding the positional restriction of topic drop, which I address through theoretical discussions and experimental studies:
1) Is the positional restriction of topic drop at the same time a restriction to only topical constituents? – In Section 3.1, I argue that topic drop is not restricted to topics, that the prefield position is not a genuine topic position, and that topicality is neither a (strictly) sufficient nor a necessary condition for topic drop.
2) Is topic drop restricted to a preverbal, more specifically, prefield position or is it also possible postverbally in the middle field, as Helmer (2016) argues? – The results of a rating study that is presented in Section 3.2 suggest that argument omissions in the middle field are degraded and that topic drop is restricted to the prefield position of V2 clauses.
3) Can topic drop occur in (potentially) embedded V2 clauses, i.e., is it possible in any prefield position? – In Section 3.3, I argue based on the results of a further rating study that subject topic drop is restricted at least to a prefield position where it is not c-commanded by a potential identifier from within the sentence or even to the prefield of autonomous or root clauses.
4) Is topic drop restricted to a strictly linearly sentence-initial position or can

elements precede topic drop? – The results of a rating study that is presented in Section 3.4 suggest that topic drop is not restricted to an absolute sentence-initial position, but that it can follow conjunctions like *und* ('and') and *aber* ('but').

In Section 3.5, I summarize the answers to the four questions and thus the insights gained into the prefield restriction. Finally, in Section 3.6, I examine the extent to which they can be reconciled with previous syntactic analyses of topic drop.

3.1 Prefield restriction and topicality

As I mentioned above, the prefield restriction of topic drop is often implicitly or explicitly interpreted as a restriction to topics, i.e., in a way that only topical constituents can be targeted by topic drop. In this section, I investigate whether the equation of topic drop with the omission of topics is justified. To do so, I first discuss the concept of *topic*, which is essential to this book. Based on the central literature, I propose a topic definition that serves as the basis for the theoretical discussion in this chapter and the experimental manipulations of topicality in Section 9.2. Second, I show that the prefield position is not necessarily the topic position in German. In the third step, I discuss the role of topicality for topic drop and conclude that it is neither a (strictly) sufficient nor a necessary condition.

3.1.1 Defining *topic*

The term *topic* is one of the concepts that are used to describe the information structure (Halliday 1967) of an utterance, also called its information packaging (Chafe 1976), i.e., "the relation of what is being said to what has gone before in the discourse, and its internal organization into an act of communication" (Halliday 1967: 199).[3] As Lambrecht (1994: 118) and Musan (2017: 26) point out, topic as a concept has a long history, which goes back to Aristotle's distinction between

[3]According to Chafe (1976: 28) the concepts related to information packaging include but are not limited to givenness, focus and contrast, definiteness, subjecthood, topicality, and point of view. It is the multitude of these phenomena with their different, sometimes overlapping or contradictory definitions and their various relations to partly the same linguistic means of expression that have given information structure the reputation of being an opaque field of research. Musan (2002: 207) speaks here of a "lush thicket" of concepts, pairs of concepts, means of encoding, and functions that developed from the "seed" originally "sown" by the authors Weil (1844), von der Gabelentz (1868), and Paul (1919) and visualizes the complex relationships within the field of information structure with an intricately branching graphic. Similarly, Molnár (1991: 11) characterizes the research around the concept of topic as "chaotic" and diagnoses a terminological confusion and a theoretical mess.

(logical) subject and predicate, where the predicate says something about the subject. This basic idea of a dichotomy between the thing about which something is said and what is said about it has persisted into modern research.[4] Nowadays it is mostly discussed under the term *topic*, which goes back to Hockett (1958: 201): "The most general characterization of predicative constructions is suggested by the terms 'topic' and 'comment': the speaker announces a topic and then says something about it."[5] Hockett (1958: 201) illustrates this with the example sentence *John ran away*, in which *John* is announced as the topic and *ran away* is the comment that is said about *John*.[6] In this book, I assume the following topic definition:

Definition 2 *The topic of a sentence, represented through a topic expression, also called topic constituent, is the entity under which the information contained in the comment expression(s) should be stored in the common ground. Thereby, the topic can be determined both as the entity the speaker talks about or as the entity the addressee gains additional knowledge about. The topic and the corresponding proposition bear an aboutness relation relative to a particular discourse. Topic expressions need to be referential, but they do not need to be subjects nor given information, although they frequently are both.*

This definition is based on four influential topic definitions from the last 40 years by Reinhart (1981), Gundel (1988), Lambrecht (1994), and Krifka (2007), who all employ the aboutness criterion to define topic. It is worth noting that they implicitly or explicitly mean sentence topics, i.e., the topics of a single utterance,

[4] It was also discussed by von der Gabelentz (1868: 378) in the 19th century and by Paul (1919: 12) at the beginning of the 20th century. See Molnár (1991: 13–40) for a detailed history of the topic concept from a functional and formal perspective.

[5] The Prague School uses the similar conceptual pair *theme* and *rheme*. But while the theme is usually defined as given, old, or known information and the rheme as new or unknown information (e.g., Daneš 1970), such an equation with old and new information is rejected for topic and comment by Reinhart (1981), Molnár (1991), and Krifka (2007). They argue that while topics are often given or old information, they do not necessarily have to be. See Musan (2002: 204–205) and Molnár (1991: 12–13; 60–62) for a closer discussion and for more related terms.

[6] This topic concept is, like its historical ancestors, a categorical one, according to which a constituent is either topical or not. In contrast, some scholars of the Prague school assume a gradual topic-like notion. For instance, Firbas (1971, 1992) defines the related concepts theme and rheme (see Footnote 5) by means of so-called "communicative dynamism" as a part of his theory of functional sentence perspective. According to this view, the theme is the element with the lowest degree of communicative dynamism, i.e., the element that contributes least to pushing communication forward (Firbas 1971: 141, see also the discussions in Molnár 1991 and Erteschik-Shir 2007). In this book, I do not employ such a gradual topic notion but stick to the now "classic" categorical topic concept.

as opposed to discourse topics, i.e., often more abstract "topics of larger units" (Reinhart 1981: 54). In this book, I refer to sentence topics when I use the term *topic*; when I mean discourse topics, I make this explicit.

Gundel's (1988) definition focuses on the interaction between speaker and hearer: "An entity, E, is the topic of a sentence, S, iff in using S the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to E" (Gundel 1988: 210). The topic is the anchor point of this interaction, the entity about which some form of information is exchanged from the speaker to the hearer. According to Gundel, for an entity to be a topic, this entity has to be familiar or known to both speaker and hearer (Gundel 1988: 212). Additionally, to store information about the topic, the hearer must be able to uniquely identify it (Gundel 1988: 214). If we take Hockett's (1958) example sentence, *John ran away*, *John* is the anchor point, i.e., the entity about which the speaker wants to increase their addressee's knowledge. It is assumed that the speaker and the addressee both know John and that the addressee can uniquely identify John.

Lambrecht (1994) provides a similar topic definition that also takes the hearer into account, but he focuses on the aspect of increasing the hearer's knowledge about a referent: "A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent" (Lambrecht 1994: 131). He explicitly distinguishes between topic referent and topic expression. The former is the entity that serves as the topic, and the latter is the linguistic expression that is used to refer to that entity (Lambrecht 1994: 127–128). Lambrecht makes two statements about this linguistic expression. First, he argues, following Reinhart (1981: 67), "that only referential expressions can be topics" (Lambrecht 1994: 156), ruling out semantically empty proforms such as expletive *it* or *there*, as well as indefinite pronouns. Second, he stresses the fact that there is a cross-linguistically strong tendency for this expression to be the grammatical subject. "[S]ubjects are UNMARKED TOPICS" (Lambrecht 1994: 132, original emphasis), but there is no one-to-one mapping between topic and subject. Non-topics can be subjects and non-subjects can be topics (see also Reinhart 1981: 62). Finally, following Reinhart (1981), Lambrecht (1994: 127) describes the *topic-of*-relation as a pragmatic aboutness relation between a topic referent and a proposition relative to a particular discourse. In Hockett's (1958) example, *John ran away*, the string *John* is the topic expression, while the actual person John is the topic referent. The proper name is a referential expression and functions as the subject of the sentence *John ran away*, i.e.,

the subject is the topic. Relative to the discourse in which *John ran away* is uttered, an aboutness relation between the referent John and the proposition *John ran away* is established.

Reinhart (1981) provides a pragmatic topic definition using Stalnaker's (1978) notion of a *context set*, in fact, she uses the term *context set* but means Stalnaker's (1978) concept of *common ground*. Stalnaker (1974, 1978, 2002) defines the common ground[7] as a shared background between interlocutors in the form of presupposed propositions, i.e., the set of propositions that the speaker and the hearer assume, believe, or accept to be true. The context set, in turn, is the set of possible worlds that are compatible with the common ground. Reinhart (1981: 78–79) states that she disregards the worlds aspect and considers "the context set of a given discourse at a given point as the set of the propositions which we accept to be true at this point", which corresponds to the common ground. To illustrate the organization of the common ground and its relation to topicality, she makes use of the metaphor of a library catalog. The common ground is the library catalog, which is ordered in a useful way, and the propositions that are part of the common ground correspond to entries in this catalog (Reinhart 1981: 79–80). When a new book is entered into the catalog, this book is stored under a defining entry, like the subject it deals with. Similarly, when a new proposition is added to the common ground, this proposition needs to be stored under an entry. This entry corresponds to the topic of the respective sentence. We can illustrate this again with Hockett's (1958) example sentence *John ran away*. When a speaker utters this sentence and the hearer accepts it, the corresponding proposition is added to the common ground so that the comment *ran away* is stored under the entry of the topic *John*.

This metaphor is picked up by Krifka (2007), who talks of the common ground as a file card system and of topics as the titles of these file cards. He defines topic in the following way: "The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content" (Krifka 2007: 41). He states that utterances typically have exactly one topic, but, following Lambrecht (1994), there may also be sentences with more than one topic constituent (3a) or so-called *thetic* sentences with no topic constituent at all (3b) (Krifka 2007: 42–43). In Section 3.1.5, I come back to this issue when discussing whether sentences with expletive subjects should be considered thetic.

(3) a. As for John and Sue, they ran away.
b. It is raining.

[7]Stalnaker (1978, 2002) attributes the term to Paul Grice, who used it in the William James Lectures in the form of the common ground status of propositions (Grice 1989).

3.1.2 Relation between topic drop and topicality

With this definition at hand, we can now proceed to the question of the relationship between topic drop and topic, the main point of this section. In the research to date, two opposing views can be found. First, Trutkowski (2016: 19) and Frick (2017: 67) state that topic drop and topicality are independent notions and that the omission is by no means restricted to topic expressions. Frick (2017) clarifies: "The notion of topic on which this [Frick's, LS] investigation is based is hence a purely structural one and refers to the preverbal positioning of the elements in the prefield" (Frick 2017: 67).[8] The opposite opinion is held by Sternefeld (1985), Erteschik-Shir (2007), and Helmer (2016, 2017b), who explicitly state that topic drop is the non-realization of a topic expression. Helmer (2017b) defines topic drop as follows: "Analepses with topic drop are [...] sentence grammatically incomplete utterances [...] in which the topic of a previous utterance is not verbalized again" (Helmer 2017b: 2).[9] In her dissertation, she specifies that both the sentence topic and the discourse topic can be omitted (Helmer 2016: 168). Other researchers do not use the term *topic* but the related *theme* (see Footnote 5). They argue that only a thematic element can be omitted (Oppenrieder 1987, Poitou 1993, Zifonun et al. 1997, Sandig 2000, Günthner 2006), or that no rhematic expressions can be targeted by topic drop (Fries 1988, Zifonun et al. 1997). However, they do not explain which definition of theme they assume and do not relate it to topic. Even less clear is the position of authors such as Huang (1984), Fries (1988), Auer (1993), Jaensch (2005), and Volodina (2011). They do not define topic drop by using the information-structural topic concept but rather through the prefield restriction, which they term restriction to the topic position. As mentioned above, the German prefield position is frequently equated with a or even *the* topic position in German. When authors argue that topic drop is restricted to a syntactic position that they believe is reserved for topics, it is not unreasonable to assume that they at least indirectly consider topic drop to be restricted to topics.

From this literature review, two main issues arise: 1) Is the prefield position indeed the topic position in German so that an equation of prefield restriction and topic restriction is justified? 2) Is topic drop restricted to topics or, to put it more concretely, is being a topic both a sufficient and a necessary condition for a

[8] My translation, the original: "Die der [Fricks, LS] Untersuchung zugrunde gelegte Auffassung von Topik ist also eine rein strukturelle und bezieht sich auf die präverbale Positionierung der Elemente im Vorfeld" (Frick 2017: 67).

[9] My translation, the original: "Analepsen mit Topik-Drop sind [...] satzgrammatisch unvollständige Äußerungen, [...] in denen das Topik einer vorherigen Äußerung nicht erneut verbalisiert wird" (Helmer 2017b: 2).

constituent to be targeted by topic drop? I discuss these issues in the next three sections.

3.1.3 Contra prefield position as topic position

The positioning of a constituent in the prefield in German is frequently called topicalization (e.g., Cardinaletti 1986), which suggests a close relationship between topic and prefield (e.g., Pittner & Berman 2021). When authors such as Huang (1984) state that topic drop is restricted to "sentence-initial – namely, topic – position" (Huang 1984: 547; see also Auer 1993: 198, Jaensch 2005: 97, Volodina 2011: 272), this suggests that the prefield is exactly the position where the topic is placed in German.

However, the prefield position in German is neither restricted to topics nor can topics only occur in this position. The first point is easily illustrated by elements in the prefield that are not referential and therefore cannot function as topics. Examples are the expletive *es* ('it') (4a), the placeholder *es* (4b), the correlate *es* (4c), or indefinite pronouns such as *niemand* ('nobody') (4d).[10]

(4) a. *Es regnet heute schon den ganzen Tag.*
it rains today already the whole day
'It's been raining all day today.'

b. *Es spielen bekannte Rockbands und kleinere Indiegruppen.*
it play known rock.bands and smaller indie.groups
'Well-known rock bands and smaller indie groups play.'

c. *Es freut mich, dass es dir besser geht.*
it pleases me that it you better goes
'I'm pleased that you are feeling better.'

d. *Niemand mag Rosenkohl.*
nobody likes Brussels.sprouts
'Nobody likes Brussels sprouts.'

Second, while there seems to be a general cross-linguistic tendency for topics to occur at the left edge of an utterance (e.g., Gundel 2010: 471), this tendency is by no means universal. Lambrecht (1994: 200) shows this not only with the example of VOS or VSO languages but also with German and English, "in which topical non-subject constituents may appear in canonical argument position after the verb". Frey (2000: 140; 145) argues for a designated topic position in the middle field in German, which immediately precedes the position of sentence adverbials like *wahrscheinlich* ('probably') or *unglücklicherweise* ('unfortunately')

[10] In naming the various *es* forms, I follow the terminology in Eisenberg (2020: 190–193).

and is independent of whether the prefield is filled with a topic expression or not.[11] Finally, Speyer (2010) sketches an optimality-theoretic model of how the prefield is filled. He argues that constituents that function as scene-setters or that indicate contrast are ranked higher than those that represent the topic and are therefore more likely to occur in the prefield. In these cases, the topic must also occur in another topological position.

In summary, while the prefield position in German is a prototypical position for topics, it is not the only or exclusive topic position. Topics may occur in other positions and also non-topical constituents can be placed in the prefield. Consequently, the positional restriction of topic drop to the prefield position does not automatically entail a restriction of topic drop to topics.

3.1.4 Contra topicality as a sufficient condition

The question of whether topic drop is restricted to topic expressions has two aspects, namely whether topicality is a sufficient condition and whether it is a necessary condition for topic drop. When people argue in favor of a topicality restriction in the literature, they usually refer exclusively to the aspect of necessity and not to sufficiency. They assume that topicality is the precondition for omitting a constituent but not that topicality follows from being able to omit the constituent. Although necessity is thus the central point, I formalize and discuss both aspects in what follows.

For a condition P to be sufficient, it has to be the case that if condition P is met, the consequence Q inevitably follows from P. This means that whenever a prefield constituent[12] is topical, it must be omittable, as indicated in equation 3.1.

$$P \Rightarrow Q \text{ : prefield constituent topical} \Rightarrow \text{constituent omittable} \qquad (3.1)$$

[11] He illustrates the topic position in the middle field by means of the contrast in (i). The subject must precede the sentence adverbial and not follow it.

(i) 'I'll tell you something about Otto.'

a. *Nächstes Jahr wird Otto wahrscheinlich seine Kollegin heiraten.*
next year will Otto probably his colleague.FEM marry
'Next year, Otto will probably marry his colleague.'

b. #Nächstes Jahr wird wahrscheinlich Otto seine Kollegin heiraten. (Frey 2000: 141, his judgments)

[12] If one abandons the equation of prefield and topic position and assumes that topics in German can also occur in the middle field, the prefield restriction does no longer follow from the restriction of topic drop to topics. To account for the empirical facts (see Section 3.2.2), proponents of a topic restriction would have to additionally assume an independently motivated prefield restriction.

This can be refuted by cases in which a prefield constituent is a topic expression but still cannot be targeted by topic drop. A first counterexample are contrastive topics (Büring 1997), i.e., "topics with a rising accent" that "represent a combination of topic and focus" (Krifka 2007: 44) (5).

(5) A: What do your siblings do?
B: [My [SISter]$_{\text{Focus}}$]$_{\text{Topic}}$ [studies MEDicine]$_{\text{Focus}}$,
and [my [BROther]$_{\text{Focus}}$]$_{\text{Topic}}$ is [working on a FREIGHT ship]$_{\text{Focus}}$.
(Krifka 2007: 44)

According to Krifka (2007: 44), the function of the contrastive topic *my sister* is to indicate an alternative aboutness topic, namely *my brother*, about whom the second conjunct is. This way, B indicates that the first conjunct is an incomplete answer to A's question and that more information relevant to this question will come – Krifka calls this "a strategy of incremental answering in the CG [common ground, LS] management" (Krifka 2007: 44).[13]

What is relevant to the question of sufficiency is the fact that such contrastive topics cannot be targeted by topic drop, as Fries (1988: 28), Volodina (2011: 272), and Volodina & Onea (2012: 217) point out. Omitting *my sister* in (5) is impossible because this would result in the loss of the prosodic marking through the focus and thus in the loss of the function of incremental answering. Moreover, the omitted constituent *my sister* would be difficult to recover because the plural form *siblings* implies that there are at least two persons, to whom the topic drop could refer.

This leads us to the second counterexample against sufficiency: topics that cannot be omitted because the hearer is unable to link them to a previous mention in the discourse.[14] Krifka (2007) discusses example (6) uttered by speaker A to speaker B in a context where both A and B know John but have not talked about him in their recent conversation, and where John is also not present in the situation.

(6) Did you know? [John]$_{\text{Topic}}$ [married last week]$_{\text{Comment}}$. (Krifka 2007: 43, topic-comment structure added)

[13]Not all usages of contrastive topics have this function. In example (i), the contrastive topic indicates that B's assertion may not satisfy the expectations of speaker A (Krifka 2007: 45).

(i) A: Does your sister speak Portuguese?
B: [My [BROther]$_{\text{Focus}}$]$_{\text{Topic}}$ [[DOES]$_{\text{Focus}}$]$_{\text{Comment}}$.

[14]Such cases pose no problem for approaches that assume that only continuous or continued topics can be omitted (e.g., Schulz 2006, Freywald 2020).

A's second sentence clearly is about John. The information about the recent marriage will be stored under the entry for John in the common ground of A and B. This means that for John to be the topic of the utterance, he does not need to be part of the "recent state of the CG content", but it suffices that there is "a long established and known interest of B in John" (Krifka 2007: 43). However, for the topic expression *John* to be targeted by topic drop, the reference of the omitted constituent needs to be recoverable (see Chapter 4), like basically any type of omission. An omission out of the blue, as in (7), is not possible as B will not be able to identify and recover the omitted topic expression.

(7) Did you know? *Δ Has married last week.

The fact that contrastive topics and unrecoverable topics cannot be targeted by topic drop constitutes counterevidence to topicality being a sufficient condition for topic drop. These are topic expressions in the prefield that cannot be omitted so that we may conclude that not all topic expressions can be targeted by topic drop. However, one could restrictively object that in the cases of contrastive topics and non-recoverable topics with focus/contrast and lack of recoverability, there are, as it were, "confounding factors" R that can override the possibility to omit the constituent conditioned by topicality. Consequently, at most equation 3.2 applies but not equation 3.1.

$$\begin{aligned} P + \neg R \Rightarrow Q \ &: \text{ prefield constituent topical} \\ + \text{ no confounding factors} &\Rightarrow \text{constituent omittable} \end{aligned} \tag{3.2}$$

In this case, then, we can more conservatively conclude that topicality is at least not a *strictly* sufficient condition for topic drop.

3.1.5 Contra topicality as a necessary condition

The second aspect of the question of whether topic drop only targets topical constituents is whether topicality is a necessary condition for topic drop. A condition P is necessary if P must be true whenever the consequence Q is true, as shown in equation 3.3.

$$Q \Rightarrow P \ : \text{ constituent omittable} \Rightarrow \text{prefield constituent topical} \tag{3.3}$$

This means that for any constituent that can be omitted from the prefield position, it must hold that this constituent is topical. Only topic expressions should be targeted by topic drop, while non-topical constituents should not. Since there is

widespread agreement in the literature that non-referential constituents cannot be topics (see Section 3.1.1), it would be an argument against necessity if non-referential prefield constituents can be omitted. A first example of non-referential elements are quantifiers such as *jeder* ('everybody') or *keiner* ('nobody'), which apparently indeed cannot be targeted by topic drop.[15] The second example is the expletive *es* ('it'),[16] which is used as the subject of weather verbs like *regnen* ('to rain') (8a), of copulas in combination with adjectives like *kalt* ('cold') (8b), of verbs with further arguments in impersonal constructions like *geben* ('to give') (8c) and of verbs of perception like *stinken* ('to stink') (8d), which also allow for referential subjects (8e) (see Eisenberg 2020: 191–192).

(8) a. *Es regnet.*
it rains
'It is raining.'
b. *Es ist kalt.*
it is cold
'It is cold.'
c. *Es gibt Pizza.*
it gives pizza
'There is pizza.'
d. *Es stinkt.*
it stinks
'It stinks.'
e. *Der Müll stinkt.*
the garbage stinks
'The garbage stinks.'

In the theoretical literature, there is strong dissent about whether topic drop can or cannot target expletives. Fries (1988: 34), Cardinaletti (1990: 81), Kaiser (2003:

[15] It is probably the impossibility of reconstructing the omitted constituent that blocks topic drop of quantifiers. In example (i), both *jeder* and *keiner* could theoretically have been omitted, but it could also be a referential 3rd person singular pronoun that is unrealized. See also the related discussion on indefinite pronouns and the pronoun *man* in Footnote 9 on page 188.

(i) *Δ *hat die Klausur bestanden.*
everybody/nobody has the exam passed
'*(Everybody/nobody) passed the exam.'

[16] Recall that, unlike Frick (2017), for example, I do not consider the "omission" of other *es* types, such as the placeholder *es* and the correlative *es*, to be topic drop. See Sections 2.1.3 and 2.2.4 for details.

260), Haegeman (2007: 116), and Volodina & Onea (2012: 213) judge examples like (9) and (10) as ungrammatical.

(9) *A: Und wie ist das Wetter bei euch?*
and how is the weather at you.2PL
A: 'And how is the weather with you?'
a. *B: *Δ Ist warm. / *Δ Ist kalt.*
it is warm it is cold
B: '(It) is warm.' / '(It) is cold.'
b. *B: *Δ Regnet. / *Δ Schneit.*
it rains it snows
B: '(It) is raining.' / '(It) is snowing.' (Fries 1988: 34, his judgments)

(10) *Δ *Freut mich, dir zuzuhören.*
it pleases me you.DAT.2SG listen.to
'I'm glad to listen to you.' (Fries 1988: 34, his judgment)

In contrast, Poitou (1993: 116), Reis (2000: 218), Volodina (2011: 271), Trutkowski (2016: 120), Frick (2017: 67), and Ruppenhofer (2018: 220; 222–224) discuss corpus examples demonstrating that speakers do use topic drop to omit expletives in the prefield. Frick (2017) presents evidence from Swiss German text messages, such as (11). In her corpus data, she found 83 utterances with realized expletive *es* in the prefield and 179 utterances where an expletive *es* is omitted from this position. The resulting omission rate of 68.32% is significantly higher ($\chi^2(1) = 6.03$, $p < 0.05$)[17] than the rate of the referential personal pronoun of the 3rd person singular neuter of 57.32% (137 omitted instances vs. 102 realized instances) (Frick 2017: 140).

(11) a. *hüt chani s tel dänn abnäh, Δ rägnet nüme und mir*
today can.I the phone then pick.up it rains no.more and we
stecked nöd im morast
stick not in.the mud
'Then today I can pick up the phone, (it)'s no longer raining and we're not stuck in the mud.' (Frick 2017: 150)
b. Δ *Wird ehner spat, Δ bliibe na chli.*
it gets rather late I stay still a.little
'(It)'s getting rather late, (I)'ll stay a bit longer.' (Frick 2017: 151)

[17] I calculated a Pearson's chi-squared test with Yates's continuity correction in R (R Core Team 2021).

Ruppenhofer (2018) supports this finding for the Standard German of Germany. In a sample of German Twitter data, he found several cases of topic drop of expletives with weather verbs, although the possibility of omission seems to depend on the lexeme used (Ruppenhofer 2018: 222–223), see Table 3.2.

Table 3.2: Omission rates by weather verb in the twitter data by Ruppenhofer (2018), taken from Ruppenhofer (2018: 223)

Weather verb	Full form	Topic drop	Total	Omission rate
regnen ('to rain')	26	11	37	29.73%
pissen ('to rain hard')	63	1	64	1.56%
nieseln ('to drizzle')	41	1	42	2.38%
schiffen ('to rain hard')	35	10	45	22.22%
schneien ('to snow')	97	1	98	1.02%

Additionally, in the fragment corpus FraC (Horch & Reich 2017, see Section 7.2.1 for details), several cases of omitted expletives produced by different authors in different text types can be found, such as (12a) from a conversation about a hotel (the omitted referential constituent refers to this hotel), the tweet (12b), or the text message (12c).[18]

(12) a. *ja,* $\Delta_{expletive}$ *gibt auch Hallenbad,* $\Delta_{referential}$ *ist*
yes it gives also indoor.swimming.pool it is
zentral, $\Delta_{expletive}$ *gibt einen Fitneßraum zum Austoben.*
central it gives a fitness.room to.the work.out
'yes, (there) is also an indoor swimming pool, (it) is central, (there) is a fitness room to work out in.' [FraC D867]

b. *Oh...* Δ *ist ja fast schon warm draußen.*
oh it is PART almost already warm outside
'Oh... (It) is almost warm outside already.' [FraC T261]

c. Δ *WIRD ZEIT FÜR DEINE RÜCKKEHR!*
it becomes time for your.2SG return
'(It) is time for your return!' [FraC S748]

[18]Example (i) shows that the prefield restriction also applies to examples (12a), (12b), and (12c). Therefore, it seems legitimate to assume that the omission of non-referential constituents is just as much topic drop as that of referential ones.

(i) a. *Hallenbad gibt Δ auch.
b. *Zum Austoben gibt Δ einen Fitnessraum.
c. *Draußen ist Δ ja fast schon warm.

These examples indicate that expletives can be omitted, which challenges the assumption that topicality is a necessary condition for topic drop.

Trutkowski (2011, 2016) tries to reconcile, at least for weather verbs, the restriction of topic drop to topics and the possibility of omitting expletives. She suggests that an expletive subject can function as a topic expression that refers to an event or situation, and in these cases can be targeted by topic drop:

> [A]s situations are referential and representable by overt elements (also by expletive pronouns), nothing speaks against the fact that these overt elements can be dropped. I.e., I assume that the (dropped) expletive pronoun represents the event that is the current situation.
>
> (Trutkowski 2016: 120–121; see also Trutkowski 2011: 213)

She illustrates this with the contrast in (13). While she claims (13a) to be fine because the situation is present through the speaker looking out of the window, she marks (13b) as ungrammatical because the situation the expletive refers to is not present in the discourse context: "Whenever the event is actually present, the expletive *es* can be omitted – however, this is not possible when the event is not present in the given discourse" (Trutkowski 2016: 120).[19]

(13) a. *[While looking out of the window:] Δ Regnet grad'.*
it rains right.now
[While looking out of the window:] '(It) is raining right now.'

b. *Δ *Regnet bestimmt, wenn wir in Urlaub fahren.*
it rains definitely when we in vacation go
'(It) will definitely rain when we go on vacation.'

(Trutkowski 2016: 120, her judgments)

Trutkowski bases her argumentation on Falk (1993), who considers a topic drop interpretation for expletive drop in Swedish. Falk states that in this case "[t]he notion 'topic' must be given a somewhat wider interpretation, including situations that are often expressed in impersonal expressions as a topic of discourse, like the weather" (Falk 1993: 172). It is worth noting that Falk, unlike Trutkowski (2011, 2016), does not explicitly call the expletive *es* the topic expression but states that impersonal expressions serve the purpose of expressing a situation in the form of

[19] Note that there are also two types of usages of the expletive in example (13): an episodic (13a) and a generic usage (13b). It may well be that this is the relevant difference that leads Trutkowski to consider (13b) as ungrammatical. However, Trutkowski does not mention this difference. While in this book I only tested her proposal of the presence/absence of the situation, it may be beneficial in a future study to also contrast topic drop of different usages of expletives. I thank Ingo Reich (p.c.) for pointing out the two usage types to me.

a discourse topic. Even clearer is Firbas (1982), who terms sentences with weather verbs like *regnen* ('to rain') *thetic* sentences, following a philosophical tradition (see, e.g., Ulrich 1986). He states that these sentences express an event but do not have a rheme so the expletive subject is "a mere dummy element" (Firbas 1982: 105). Krifka (2007) argues that although thetic sentences do not have a topic expression, "they do have a topic denotation, typically a situation that is given in the context" (Krifka 2007: 43). Erteschik-Shir (1997, 2007) talks about *stage topics* "indicating the spatio-temporal parameters of the sentence (here-and-now of the discourse)" (Erteschik-Shir 2007: 16). To sum up, sentences with expletive subjects and weather verbs are regularly classified as thetic sentences without a topic expression. This means that the expletive subject *es* is not a topic expression. If *es* can be omitted, this argues against topicality being a necessary condition for topic drop.

This is different with Trutkowski's (2011, 2016) account, as she considers the possibility of identifying the topic with the expletive subject. In what follows, I present an experiment with which I tested her proposal empirically to exclude the possibility that the omission of expletives before weather verbs is the omission of a topic in Trutkowski's sense specified above. However, it is important to stress that Trutkowski's account was sketched for weather verbs and it remains unclear how other types of expletive subjects would be treated in cases where it may be not that easy to find a suitable situation.

3.1.6 Experiment 1: topic drop of expletives

Experiment 1[20] was an acceptability rating study that tested the following claim by Trutkowski (2011, 2016): Topic drop of the expletive subject *es* of weather verbs is said to be possible, provided that the subject represents a topic denotation in the form of the current weather situation, which must be present in the discourse.[21] Trutkowski's claim conflicts with the prevailing research opinion that non-referential expletive elements cannot serve as topics. In this view, it would speak against topicality as a necessary condition for topic drop if their omission is acceptable.

As mentioned above, Trutkowski (2011, 2016) argues that the situation that serves as the topic must be present in the current discourse. I operationalized this through a question that either asked for information about the current weather or not (see Section 3.1.6.1 below). Following Trutkowski (2011, 2016), the weather

[20] All materials and the analysis script can be found online: https://osf.io/zh7tr.

[21] At a poster presentation (Schäfer 2022), I presented a pilot study of this experiment with eight token sets, which yielded a comparable result.

question should make the weather, i.e., the current situation, present in the discourse and set it as the topic. With the other question proposing an activity, this was not the case. The weather as the current situation did not occur until the target sentence, so the situation was not set as the topic. Accordingly, the overt or covert expletive subject pronoun in the target sentence should once be a topical expression referring to the weather and once not.

The experiment had the form of a 2 × 2 design crossing Completeness (full form vs. topic drop) and Question Type (weather vs. other). If Trutkowski's (2011, 2016) approach is on the right track, there should be a significant interaction between both predictors so that topic drop is rated as more natural in the context of a weather question that sets *es* ('it') as topic expression than in the context of an unrelated question, whereas the full forms should be natural after both types of questions. If only referential expressions can be topics and if topicality is generally a necessary condition for topic drop, this predicts that topic drop should be degraded to the same extent compared to the full forms, regardless of the preceding question. On the contrary, if topicality is not a necessary condition for topic drop, there should be no difference between the utterances with topic drop and the full forms in both question conditions.

3.1.6.1 Materials

Items

I tested 24 items in the form of question-answer pairs, such as (14).

(14) a. (i) *A: Was macht das Wetter bei dir?*
what makes the weather at you.DAT.2SG
A: 'How's the weather with you?'
(ii) *B: (Es) regnet leider schon wieder ziemlich heftig*
it rains unfortunately already again pretty hard
B: 'Unfortunately (it) is raining pretty hard again' (weather)

b. (i) *A: Wolltest du nicht joggen gehen?*
wanted you.2SG not jog go
A: 'Didn't you want to go jogging?'
(ii) *B: (Es) regnet leider schon wieder ziemlich heftig*
it rains unfortunately already again pretty hard
B: 'Unfortunately (it) is raining pretty hard again' (other)

The question either asked for the weather in the current location of the addressee(s) (14a) or did not mention the weather but suggested an activity or asked

why an activity did not come about (14b). The answer was an utterance with a weather verb and the expletive subject *es* realized in or omitted from the prefield.[22]

Since weather verbs tend to be a closed class with a limited number of members, of which again only a few are very frequent, I did not use 24 different verbs but tested six frequent verbs in four different lexical environments respectively: *regnen* ('to rain'), *stürmen* ('to storm'), *schneien* ('to snow'), *donnern* ('to thunder'), *nieseln* ('to drizzle'), and *schütten* ('to pour (with rain)').[23] This means that there were four token sets respectively with each of the six verbs. Another token set with *regnen* ('to rain') is shown in (15).

(15) a. (i) *A: Wie ist das Wetter bei euch?*
how is the weather at you.DAT.2PL
A: 'How is the weather with you?' (weather)

(ii) *A: Und, was hältst du von einem Spaziergang im Park?*
and what hold you.2SG of a walk in.the park
A: 'So, what do you think about a walk in the park?' (other)

b. *B: (Es) regnet wohl noch den ganzen Tag*😒
it rains PART still the whole day
B: '(It) will probably rain all day😒'

Fillers

The 24 items were presented together with a total of 80 fillers. 24 of the fillers were also question-answer pairs that contained gapping or right node raising structures in the answers. Another 24 fillers were the items of experiment 9, which tested topic drop of the 1st and 2nd person singular and consisted of two

[22] I indicate the two levels of the predictor COMPLETENESS, full form and topic drop, by putting the prefield constituent into parentheses. Note that the target sentences were shown without a final period, as such points are often omitted in instant messages or replaced by emojis.

[23] For example, I did not consider (i) verbs like *graupeln* ('to sleet'), *winden* ('to be windy'), and *gewittern* ('to thunder', lit. 'to thunderstorm') because they are rarely used, (ii) synonyms for heavy raining like *pissen* ('to piss'), *gießen* ('to pour'), or *schiffen* ('to pee', literally 'to travel by ship') because they are very colloquial and partly vulgar, and (iii) verbs like *pladdern* ('to rain with heavy drops') or *drippeln* ('to drizzle') because they are regionally restricted. Furthermore, I excluded *blitzen* ('to flash', used for lightning) because it deviated remarkably from the other verbs already in the full forms in the pretest that I conducted (see Footnote 21). As a consequence, I also excluded *hageln* ('to hail'), which was also tested in the pilot study, to have an even number of verbs (and token sets).

turns by one speaker (see Section 10.5.4 for details).[24] The same structure of two turns produced by one speaker was used for 24 fillers with sluicing and sprouting structures in the target utterances. Additionally, I included eight ungrammatical catch trials with word order violations in the last utterance such as (16), which served as attention checks and exclusion criterion for the participants.

(16) a. *A: Hast du in letzter Zeit mal was von Ricardo gehört?*
have you.2SG in last time PART what about Ricardo heard
A: 'Have you heard from Ricardo lately?'

b. *B: Er in Köln wollte eine Ausbildung zum anfangen*
he in Cologne wanted a training to.the start
Mediengestalter
media.designer
B: 'He wanted to start training as a media designer in Cologne'
(Lit. 'He in Cologne wanted a training as a start media designer')
(ungrammatical)

3.1.6.2 Procedure

I used a web-based procedure to collect acceptability rating judgments. I implemented a survey using the LimeSurvey online survey tool (Limesurvey GmbH 2023). As participants, I recruited 47 German native speakers between the ages of 18 and 40 from the crowdsourcing platform Clickworker (Clickworker 2022)[25] who were compensated with €3.30.[26] Their task was to read the instant messaging dialogues and to rate how natural the last utterance is in the given context on a 7-point Likert scale with labeled end points: 1 was labeled "vollkommen unnatürlich" ('completely unnatural'), 7 was labeled "vollkommen natürlich" ('completely natural'). Participants were asked to base their judgments on the question of whether they or people with whom they communicate via WhatsApp or similar services would write this message in the corresponding dialogue in this way.

[24] Experiment 1 and experiment 9 were conducted as part of the same study. Their items both contained topic drop but were otherwise unrelated. Experiment 1 tested topic drop of 3rd person singular expletives before weather verbs, while in experiment 9 1st and 2nd person singular pronouns referring to sender and addressee were omitted. Having relatively many topic drop structures in the study should not be a problem since they are very common in instant messages (see the discussion in Section 2.1.5).

[25] Due to technical problems, I could only collect data from 47 instead of the planned 48 participants.

[26] For each study in this book, the compensation was based on the statutory minimum wage in Germany at the time of the study and a realistic estimate of the expected completion time.

They were told that there is no right or wrong but that all that matters is their intuition.

The 24 items were distributed across four lists according to a Latin square design so that each participant saw each token set exactly once and in only one condition. Items and fillers were presented in an individual pseudo-randomized order, ensuring that no two stimuli of the same type immediately followed each other. All materials were presented as instant messaging dialogues between two persons. I added a matching emoji to some of the materials, such as the item in (15), to increase their overall naturalness as instant messages.[27] The appearance of the chat resembled that of the world's largest instant messaging service WhatsApp (Statista 2022), as shown in Figure 3.2, and was created using CSS code by Rocha (2022).

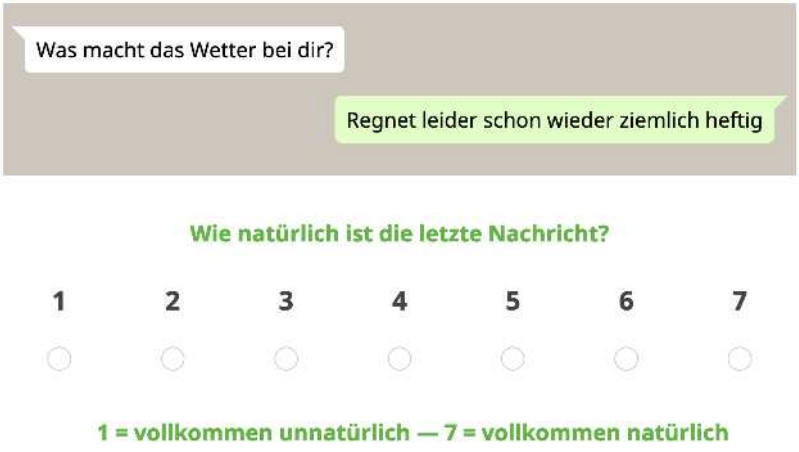


Figure 3.2: Screenshot of the instant messaging design of experiment 1

3.1.6.3 Data analysis

In what follows, I describe in detail how I analyzed the resulting acceptability rating data. Since the procedure for all rating experiments in this book is more or less identical, I am not as detailed in describing the further experiments but only outline the most important steps and any deviations from the following general scheme.

[27] All emojis shown in this book are the WhatsApp versions obtained from https://emojipedia.org/whatsapp/ (visited on 01/02/2025). In the experiments, equivalent versions were shown to participants depending on the browser they used to participate in the experiment.

Before the analyses, I excluded the data from presumably inattentive participants who gave too good ratings to the ungrammatical catch trials. For experiment 1, I defined "too good ratings" beforehand as rating four or more of the eight catch trials with 6 or 7 points. This is a liberal threshold, which allowed me to only exclude the data from subjects who repeatedly gave massively inappropriate ratings. For experiment 1, this resulted in the loss of the data from nine participants. I statistically analyzed the data from the remaining participants, 38 in experiment 1, using the programming language R (version 4.1.0, R Core Team 2021)[28] and the integrated development environment RStudio (version 2022.02.3, RStudio Team 2021).

The dependent variable in all of the rating experiments were numeric values between 1 and 7 indicating the perceived naturalness of the experimental stimuli assessed with a 7-point Likert scale. It is a matter of debate whether these data can be treated as interval data or need to be considered ordinal (Carifio & Perla 2008). While ordinal variables require a meaningful order, for interval variables furthermore the difference between two values needs to be equal (Rasinger 2013: 26). Likert scale data are clearly ordered, e.g., a rating of 5 is higher than a rating of 4, but it is unclear whether the distance between 4 and 5 is equal to the distance between a rating of 1 and a rating of 2 (Schütze & Sprouse 2014: 33–34). Such an equal distance between scale points would be the prerequisite to analyzing the data with parametric tests like linear mixed effect models. In this book, I pursue a rather conservative strategy by treating the rating data as ordinal (see, e.g., Jamieson 2004; but cf. Jaccard & Wan 1996, Carifio & Perla 2008) and analyze them with cumulative link mixed models (CLMMs) from the package ordinal for ordinal data (version 2019.12-10, Christensen 2019). These CLMMs take into account that the distances between the points of a scale may not be equal, but they also allow for an adjustment of this assumption. The threshold parameter of the CLMM function can be used to set constraints on the thresholds to simplify the corresponding model. It can be changed from the default flexible thresholds (scale points are ordered, but there are no further restrictions) to symmetric (distance between scale points is symmetric around the central point), symmetric2 (the latent mean of the reference group is 0) or equidistant (equal distance between scale points) (Christensen 2019). For the final model, I always started with the most liberal flexible thresholds and subsequently tried simpler models with more restricting symmetric, symmetric2, or equidistant thresholds. I compared the complex and the simpler final model with likelihood ratio tests calculated with R's anova function (R Core Team 2021) and kept the simpler model if the model fit was not significantly different between models.

[28] All analyses in this book used the same R-version to maximize their comparability.

CLMMs allow for the analysis of an ordinal response variable considering fixed effects of categorical or continuous independent variables, as well as random effects. The independent variables in my experiments were usually binary predictors coded with deviation coding[29] or, less often, numeric scores. The models contained the main effects of these independent variables, as well as the scaled and centered numeric position at which the critical trial was presented in the experiment, and all the two-way interactions between the predictors. To cover possible differences between the token sets and the participants, the models had random intercepts for items and subjects and by-item and by-subject random slopes.

In my analyses, I always started with a maximal model including all fixed and random effects (Barr et al. 2013). In case of convergence issues, I simplified the random effects structure starting with eliminating random slopes for interactions until the model converged. I systematically reduced the full model to the final model by using a backward model selection procedure for the fixed effects. In each step, I excluded a non-significant fixed effect and compared the models with and without the effect through likelihood ratio tests performed with the anova function in R (R Core Team 2021). If the more complex model did not have a significantly better model fit, I continued with the simpler model. The random effects structure remained unaffected by this procedure, i.e., constant for all models. The final model was always a model where all fixed effects were either significant or part of a significant higher-order interaction. To obtain the test statistics in the form of χ^2 values and the p-values for the effects in the final model, I used the same procedure as for model selection. Again, I compared a model with and a model without the effect in question with the anova function. For each analysis in this book, I present the model call of the full model and the model table showing the fixed effects in the final model.

[29] Deviation coding is a coding scheme for categorical variables that "compares the mean of the dependent variable for a given level to the overall mean of the dependent variable" (UCLA: Statistical Consulting Group 2022), the so-called grand mean. For a binary independent variable, the two levels are usually coded either as 1 and −1 respectively (this is also called sum coding) or as 0.5 or −0.5 (this is called deviation coding in the narrower sense). For such a binary variable coded with deviation coding, the intercept corresponds to the mean between conditions, while the slope indicates the deviation between the conditions and the mean (Alday 2022).

3.1.6.4 Results

Table 3.3 shows the mean ratings and standard deviations of the four conditions and Figure 3.3 the mean ratings and 95% confidence intervals.[30] They suggest that utterances following a weather question were preferred over utterances following another question, while there seems to be little difference between the full forms and the utterances with topic drop.

Table 3.3: Mean ratings and standard deviations per condition for experiment 1

Completeness	Question Type	Mean rating	Standard deviation
Full form	Weather question	5.84	1.35
Topic drop	Weather question	5.88	1.32
Full form	Other question	5.39	1.55
Topic drop	Other question	5.30	1.60

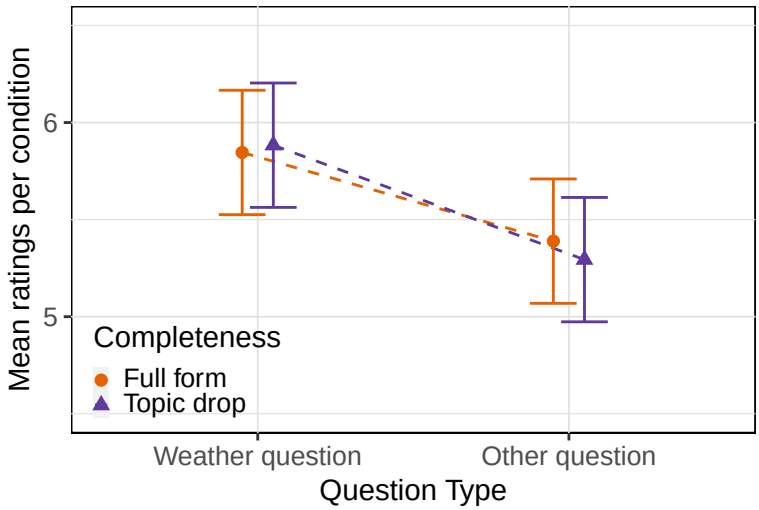


Figure 3.3: Mean ratings and 95% confidence intervals per condition for experiment 1

[30] All point-line plots in this book were created in R (R Core Team 2021) using the package ggplot2 (Wickham 2016).

I performed an inferential statistical analysis of the data, as described in the previous section. The full CLMM contained the ordinal ratings as the response variable. As fixed effects, I included the main effects of the independent variables Completeness (deviation coded, full form as 0.5, topic drop as −0.5), Question Type (deviation coded, weather questions as 0.5, other questions as −0.5), and of the numeric centered and scaled Position at which the trial was presented in the experiment, as well as all two-way interactions. The random effects structure consisted of random intercepts for subjects and items and of by-subjects and by-items random slopes for all three predictors and the two-way interactions between them.[31] By performing a backward model selection, I obtained the final model with symmetric thresholds and a significant main effect of Question Type, as shown in Table 3.4.

Table 3.4: Fixed effect in the final CLMM of experiment 1

Fixed effect	Est.	SE	χ^2	p-value	
Question Type	0.91	0.23	12.05	< 0.001	***

Regardless of topic drop, utterances in the context of a weather question were rated as more acceptable than utterances in the context of another question ($\chi^2(1) = 12.05$, $p < 0.001$). The interaction between Completeness and Question Type was not significant ($\chi^2(1) = 0.05$, $p > 0.8$), nor the main effect of Completeness ($\chi^2(1) = 0.11$, $p > 0.7$).

3.1.6.5 Discussion

The results of experiment 1 are inconsistent with an explanation according to which only referential expressions can be topical and according to which topicality is a necessary condition for topic drop. The utterances with non-referential and, thus, non-topical expletive subjects of weather verbs received comparable acceptability ratings regardless of whether they were syntactically complete or contained topic drop. There is also no evidence for the proposal by Trutkowski (2011, 2016), according to which topic drop of such expletive subjects is only possible if the situation they supposedly refer to is present in the current discourse. There was no interaction between the syntactic completeness of the utterance

[31] The formula of the full model was as follows: `Ratings ~ (Completeness + Question Type + Position)^2 + (1 + (Completeness + Question Type + Position)^2 | Items) + (1 + (Completeness + Question Type + Position)^2 | Subjects)`.

and the type of question used. Topic drop was not particularly degraded if the question did not ask about the current weather. The main effect of Question Type seems to be pragmatically motivated. A discourse where (in)complete utterances with weather verbs follow questions asking about the weather is more coherent than a discourse where the utterance with a weather verb is used as an excuse (not) to do something. In the latter case, the hearer needs to draw a relevance implicature[32] as there is a more indirect relation between question and answer.

In line with the corpus results by Frick (2017) and Ruppenhofer (2018) and the examples from the FraC, we can conclude that topic drop of expletives is possible and that it is as acceptable as the corresponding full forms. This indicates that a constituent does not need to be topical in order to be targeted by topic drop. Therefore, the results do not support the idea that topicality is a necessary condition for topic drop.

3.1.7 Summary: topic drop and topicality

In this section, I discussed in detail the potential role of topicality for topic drop. After having developed a topic definition drawing on the most central literature of the last 40 years, I showed that the prefield position is not to be equated with *the* topic position in German. The prefield can be filled with non-topical constituents, and topical constituents regularly appear in the middle field as well. In the second half of the section, I argued on both theoretical and empirical grounds that topicality is neither a (strictly) sufficient nor a necessary condition for topic drop. I showed that contrastive topics and non-recoverable topics cannot be omitted and that topic drop also targets elements that by definition cannot be topics, such as non-referential expletive subjects. This suggests that the term *topic drop* is at least misleading in German. In Chapter 9, I return to topicality and investigate whether it is a factor that favors topic drop. For the remainder of this chapter, I focus on the prefield restriction as a positional restriction.

3.2 Is topic drop restricted to the prefield?

3.2.1 Theoretical background

In light of the results about the relationship between topic drop and topicality gained in the previous section, it seems reasonable to interpret the prefield restriction of topic drop as structural (Reich 2011: 1850). As mentioned above, the

[32] I thank Robin Lemke (p.c.) for pointing this out.

assumption that topic drop in German is restricted to the prefield is the prevailing research opinion. It is, to the best of my knowledge, only contradicted by Helmer (2016), who argues that topic drop can also occur postverbally in the middle field.[33] She states that the widespread assumption of a prefield restriction is related to the basic tendency that given information precedes new information in sentences. Topics, which are usually given information, occur more often in the prefield and are omitted more often from there (Helmer 2016: 26–27). However, according to her, topic drop still regularly occurs in the middle field, which she proves with examples from her corpus study. In this study, Helmer examined a total of about 35 hours of spoken conversations from the Forschungs- und Lehrkorpus Gesprochenes Deutsch FOLK ('The research and teaching corpus of spoken German') (Schmidt 2014) and the corpus Gespräche im Fernsehen GIF ('conversations on TV') (Helmer 2016: 65). In these corpora, she found 541 instances of topic drop, which she defines as the non-realization of non-expletive[34] subject and object complements of fully uttered verbs (Helmer 2016: 70). Of the 541 instances, she classifies 55 (10.2%) as topic drop in the middle field (Helmer 2016: 214), but she mentions only five of them and provides little or no context. She claims that the majority of the 55 cases are responsive utterances with mental verbs like *nicht glauben* ('not believe') (Helmer 2016: 214) and cites (17), an instance from the FOLK, which I complemented with the relevant precontext. However, for this example and similar cases, an analysis as null complement anaphor (NCA), where "the understood sentential or VP complement of the verb must be interpreted from context" (Hankamer & Sag 1976: 411, see Klein 1993: 778 for German examples) seems to be more promising than the assumption of topic drop in the middle field.

(17) a. *A: Welche Methoden des Literaturunterrichts gibt es denn?*
which methods the.GEN literature.classes.GEN gives it PART
A: 'What are the methods of teaching literature?'

b. *B: Also es gibt natürlich Unterrichtsgespräche. Es gibt*
so it gives of.course class.discussions it gives
Textanalyse. Handlungs- und produktionsorientierte Methoden. Nja.
text.analysis action and production-oriented methods well
Ich glaub mir fehlt eine aber…
I believe me.DAT lacks one but

[33] But see Nygård (2018: 10) for a similar view on Norwegian topic drop.

[34] In this respect, she deviates from my definition of topic drop which includes expletive arguments, see Section 2.1.3.

B: 'So there is class discussion, of course. There is text analysis. Action- and production-oriented methods. Well. I believe I'm missing one but...'

c. *A: Ich glaube nicht, jedenfalls nicht nach dem Modell von*
I believe not at.least not following the model of
((AUTOR A)).
author A
A: 'I don't think so, at least not according to the model of ((AUTHOR A)).' [FOLK_E_00038], simplified

By considering *Ich glaube nicht* in (17c) a case of topic drop in the middle field, Helmer has to assume that the corresponding full form is (18a), i.e., a pronoun *das* is omitted in the middle field, which refers back to the last utterance of B. However, an alternative and more natural full form is the structure in (18b), in which the reference to the preceding utterance is made explicit in the subordinate *dass*-clause. This structure can be transformed to (17c) via NCA, i.e., by omitting the subordinate clause.

(18) a. *Ich glaube das nicht.*
I believe that not
'I don't believe this.'

b. *Ich glaube nicht, dass Ihnen eine Methode fehlt.*
I believe not that you.2SG.POL a method lack
'I don't believe that you are missing a method.'

Since *glauben* in the reading of 'considering something likely' most often takes a clause as its complement (Leibniz-Institut für Deutsche Sprache 2023a), the NCA analysis is a promising strategy to cover this most frequent type of Helmer's (2016) apparent counterexamples and to maintain the prefield restriction. Still, though, this explanation does not cover all the cases that Helmer discusses.[35] Therefore, I break away from Helmer's (2016) individual examples in the next section and address the question of whether topic drop is possible in the middle field more generally, using an acceptability rating study that compared pre- and postverbal argument omissions.

[35]Other examples of Helmer (2016) such as *Sagen Sie ruhig* ('Go ahead and say (it)') might be a case of argument omission possible only in imperatives as described in Külpmann & Symanczyk Joppe (2016) and Külpmann (2021). In the case of the question *kann ich da liegen lassen* ('Can I leave (it) there'), either the speaker might have produced an involuntary elision of a *s*, turning a demonstrative *das* ('that') referring to the omitted object into the adverb *da* ('there'), or the question could be a V2 question, in which the missing direct object is omitted from the prefield not from the middle field.

3.2.2 Experiment 2: prefield restriction

Experiment 2 tested the main licensing condition of topic drop, namely its restriction to the preverbal prefield position of declarative V2 clauses.[36] While almost all previous literature on topic drop in German lists this restriction, it was, to the best of my knowledge, only discussed with introspective examples.[37] The present acceptability rating study aimed to complement the introspection with empirical data. I systematically collected acceptability judgments for utterances with topic drop in the prefield and for utterances with corresponding argument omissions in the middle field, as well as for the corresponding full form, which served as a baseline. This resulted in a 2 × 2 design crossing Completeness (full form vs. topic drop) and Topological Position (prefield vs. middle field). If the majority of the previous theoretical literature is correct and topic drop is restricted to the prefield position, there should be an interaction between the predictors Completeness and Topological Position. Argument omissions in the middle field are expected to be degraded compared to omissions in the prefield. In contrast, if Helmer's (2016) assessment is correct, there should be no interaction between both independent variables. The omissions should be rated equally well in the prefield and the middle field.[38]

3.2.2.1 Materials

Items

To test these predictions, I created 24 items with a similar structure, such as (19). Person A asked a question to which person B responded with three utterances, the last of which was critical, i.e., (19d-i) or (19d-ii). In these utterances it is varied whether the subject, which is always the 1st person singular subject pronoun *ich* ('I'), is realized or omitted and whether it is positioned in the prefield (19d-i) or in the middle field (19d-ii). In the middle field conditions, the prefield position was occupied by another constituent of the clause, such as the temporal adverbial *jetzt* ('now') in (19), modal adverbials such as *leider* ('unfortunately'), conjunctive adverbs such as *deshalb* ('therefore'), or (prepositional) objects such as *auf das*

[36] All materials and the analysis script can be found online: https://osf.io/zh7tr.

[37] Helmer's (2016) corpus data, discussed in Section 3.2.1, are an exception that opposes not only the prevailing research opinion but also the introspective approach.

[38] It should be noted, however, that the cases that Helmer (2016) discusses as topic drop in the middle field were rarer in absolute numbers in her corpus than the cases of topic drop in the prefield. If this were confirmed by relative numbers with corresponding full forms as the baseline, this could indicate a preference for topic drop in the prefield and result in a weak interaction in the experiment.

Essen dort in *Auf das Essen dort steh (ich) aber total* ('But (I) am totally into the food offered there.').[39]

(19) a. *A: Was habt ihr heute Abend geplant?*
what have you.2PL today evening planned
A: 'What do you have planned for tonight?'

b. *B: Wir wollen uns den neuen Matrixfilm im Kino anschau-*
we want us the new matrix.movie in.the theater watch
en 😊
B: 'We want to watch the new Matrix movie in the theater 😊'

c. *B: Die ersten drei Filme gefallen mir total gut*
the first three movies please me.DAT totally well
B: 'I totally like the first three movies'

d. (i) *B: (Ich) bin jetzt richtig gespannt auf den neuen Teil*
I am now rightly keen on the new part
B: '(I am now really excited about the new part'
(full form / topic drop, prefield)

(ii) *B: Jetzt bin (ich) richtig gespannt auf den neuen Teil*
now am I rightly keen on the new part
B: 'Now (I) am really excited about the new part'
(full form / topic drop, middle field)

Fillers

The items were mixed with a total of 72 fillers. 24 fillers were embedded gapping structures and corresponding full forms with V2 or verb-final word order, which, like the items, consisted of four turns. 16 fillers functioned as items for another experiment. They were question-answer pairs, more specifically, dialogues with

[39]It cannot be excluded that the prefield constituent influences the (im)possibility of an ellipsis in the middle field, at least if the ellipsis targets a topical constituent. Recall that Frey (2000) argues that the topic denotation must precede the sentence adverbial, at least in the middle field. If this principle is expanded to include the prefield, it should not be possible to have a sentence adverbial in the prefield followed by a topic denotation in the middle field. If we still find such cases, it could be argued that the apparent topic is not a topic precisely because it is preceded by a sentence adverbial. Although it is unclear whether such an expansion of the ordering principle is reasonable, it suggests that the presence of a sentence adverbial could principally influence the acceptability of omitting a constituent in the middle field. However, since I tested a large number of different prefield constituents in my experiment, of which only a few were sentence adverbials, and the effects reported below showed up across all items, it seems unlikely that there was a systematic influence of some sort in my experiment.

two turns. The second turn was a prepositional phrase (PP) fragment answer exhibiting or not exhibiting preposition omission and/or article omission. Another 24 fillers were items from experiment 3 with two turns, which tested potentially embedded vs. unembedded topic drop structures and full forms with 3rd person singular subject pronouns (see Section 3.3.2 for details).[40] I also included eight catch trials with severe word order violations, similar to those used in experiment 1 (see Section 3.1.6.1), which served as attention checks for the participants. Half of them consisted of two turns, the other half of four turns.

3.2.2.2 Procedure

Like experiment 1, experiment 2 was conducted as a web-based acceptability rating study set up with LimeSurvey (Limesurvey GmbH 2023). I recruited another 48 participants from Clickworker (Clickworker 2022) who had not taken part in any of my other experiments on topic drop. They were German native speakers between the ages of 18 and 40 and received €3.20 for their participation. They should read the instant messaging dialogues and rate the naturalness of the last utterance on a 7-point Likert scale (1 = completely unnatural, 7 = completely natural) based on their intuitions (see Section 3.1.6.2 for details on the instructions). The 24 items were distributed across four lists according to a Latin square design to ensure that each token set was seen exactly once and in only one condition by each participant. To avoid two stimuli of the same type immediately following each other, items and fillers were mixed and presented in individual pseudo-randomized order. Like in experiment 1, the materials were presented as instant messaging dialogues between two persons and occasionally contained matching emojis, as shown in item (19).

3.2.2.3 Results

I excluded the data from three presumably inattentive participants who had rated four or more of the eight catch trials with 6 or 7. Table 3.5 shows the mean ratings and standard deviations for the four experimental conditions calculated on the data from the remaining 45 participants. In Figure 3.4 the mean ratings and 95% confidence intervals are plotted. While the ratings for full forms and topic drop

[40] Experiment 2 and experiment 3 were conducted as part of the same study. The items served as fillers for each other, which seems to be unproblematic because the manipulations were independent. In this experiment, I tested only topic drop of the 1st person singular manipulating the topological position of the covert constituent, while in experiment 3 the items only contained topic drop of the 3rd person singular and it was varied whether they occurred in a potentially embedded clause or not.

were comparable when the overt or covert constituent was in the prefield, there was a clear drop in acceptability of topic drop in the middle field compared to the corresponding full form. The mean rating of 4.94 was still relatively good, although participants differed more strongly in their judgment of this condition, as indicated by the increased standard deviation of 1.74.

Table 3.5: Mean ratings and standard deviations per condition for experiment 2

Completeness	Topological Position	Mean rating	Standard deviation
Full form	Prefield	6.13	1.25
Topic drop	Prefield	6.16	1.34
Full form	Middle field	6.03	1.29
Topic drop	Middle field	4.94	1.74

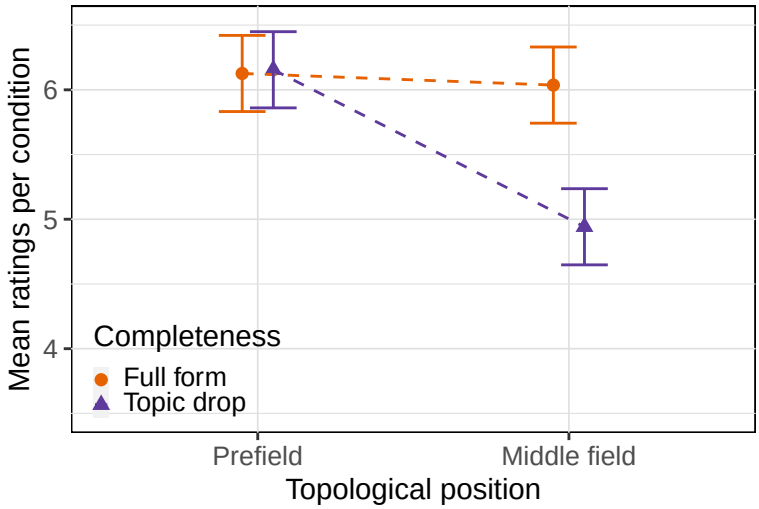


Figure 3.4: Mean ratings and 95% confidence intervals per condition for experiment 2

I analyzed the data with CLMMs (Christensen 2019), as described in Section 3.1.6.3. The full model contained fixed effects of the binary predictors Completeness and Topological Position, for which I used deviation coding (full form and prefield were coded as 0.5, topic drop and middle field as −0.5 respectively), and

of the numeric scaled and centered Position, at which the trial appeared in the experiment, as well as of all two-way interactions between these three variables. The random effects structure consisted of random intercepts for both items and subjects and of by-item and by-subject random slopes for Completeness, Topological Position, Position, and the interaction between Completeness and Topological Position.[41] The final model had symmetric thresholds and contained significant main effects of all three independent variables and a significant interaction between Completeness and Topological Position, as presented in Table 3.6.

The ratings for the full forms were significantly higher than for the utterances with topic drop ($\chi^2(1) = 13.06$, $p < 0.001$). Utterances with the overt or covert constituent in the prefield were rated significantly better than utterances where this constituent was in the middle field ($\chi^2(1) = 23.7$, $p < 0.001$). Additionally, utterances with an overt constituent in the middle field were particularly degraded ($\chi^2(1) = 28.18$, $p < 0.001$). Overall, the ratings improved in the course of the experiment ($\chi^2(1) = 7.62$, $p < 0.01$). Since the significant main effects of the three predictors are not of theoretical relevance, I do not discuss them further.[42]

Table 3.6: Fixed effects in the final CLMM of experiment 2

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	0.92	0.25	13.06	< 0.001	***
Topological Position	1.37	0.24	23.70	< 0.001	***
Position	0.27	0.10	7.62	< 0.01	**
Completeness × Topological Position	−2.46	0.42	28.18	< 0.001	***

3.2.2.4 Discussion

Experiment 2 tested argument omissions in the prefield and the middle field. The analysis revealed a significant interaction between Completeness and Topolog-

[41] The formula of the full model was as follows: `Ratings ~ (Completeness + Topological Position + Position)^2 + (1 + Completeness * Topological Position + Position | Items) + (1 + Completeness * Topological Position + Position | Subjects)`.

[42] Looking at Figure 3.4, it appears anyway that the main effects of Completeness and Topological Position are inconsistent, i.e., qualified by the interaction between Completeness and Topological Position (see Crump et al. 2019 for more details), which is the result of the deviation coding.

ical Position indicating that omissions in the middle field were particularly degraded. This result supports the assumption of a prefield restriction of topic drop. While topic drop was acceptable in the preverbal position with ratings comparable to the full forms, a comparable argument omission was significantly degraded in the middle field. This degradation is arguably the result of violating the positional restriction of topic drop. It is worth noting, however, that the mean rating of the omission in a postverbal position was still at 4.94 on a 7-point scale and, thus, relatively good. At first glance, this seems to question the validity of the conclusions drawn here, but there are at least two possible explanations for the relatively good ratings. First, the ratings for the omission in a postverbal position may have been driven up by the extremely poor catch trials, which, as described above, contained severe word order violations that are much more striking than the omission of a postverbal pronoun. These catch trials received a mean rating of 2.03 (SD = 1.59,) while the mean ratings for all other filler types were also around 5 points. Second, it could be that in some trials some participants simply overlooked the missing pronoun and did not realize that the sentence contained an omission. This seems plausible since previous research has established that readers skip around a third of all words during initial reading (Rayner 1998) and that the tendency to skip a word is stronger if the word is shorter and more predictable (e.g., Rayner & McConkie 1976, Brysbaert et al. 2005, Rayner et al. 2011). Consequently, readers should generally be likely to skip a predictable pronoun like *ich* ('I') anyway so that they may not realize that this pronoun is not present in the sentence. This line of reasoning seems to be supported by the ratings of up to a dozen participants who, based on visual inspection of the figure *Exp2_Barplot_Participants.pdf* in the online repository,[43] made no or only a little difference between the topic drop and the full form in the middle field.[44] It appears that the characteristics of the fillers and the way participants read the items may have caused the absolute ratings for the omission in the middle field to be higher than would have been expected. Nevertheless, they are clearly worse than those for the omission in the prefield, which argues for the prefield restriction of topic drop.

[43]The figure can be found at: https://osf.io/zh7tr.

[44]This hypothesis could be validated in further studies. Cases of topic drop in the middle field could be compared to cases where other short words are omitted from the middle field and the resulting structures cannot be interpreted as topic drop, such as the omission of prepositions, determiners, or even short nouns after determiners. If these clearly ungrammatical structures receive ratings comparable to those of topic drop in the middle field by the same participants, this might support the skipping hypothesis. This skipping hypothesis could also be addressed more directly using a lexical decision task where participants are asked whether or not they just read the pronoun.

3.3 Is topic drop possible in (potentially) embedded clauses?

In what follows, I refine the prefield restriction of topic drop by considering two further aspects. In this section, I investigate whether topic drop can occur in potentially embedded clauses (experiment 3), while in Section 3.4 I discuss whether the prefield restriction can be understood as a restriction to a sentence-initial position.

3.3.1 Theoretical background

There is dissent in the literature about whether topic drop is possible in embedded clauses, namely, in embedded V2 clauses that exhibit a prefield like main clauses.[45] Cardinaletti (1990: 76), Rizzi (1994: 169), and Volodina (2011: 272) claim that topic drop is not possible in clauses like (20), which stems from the items of experiment 3. In contrast, Jaensch (2005: 101) takes the view that topic drop is possible in embedded clauses, although its acceptability is said to vary stronger between speakers than for unembedded topic drop. Similarly, Trutkowski (2016: 224–225) considers embedded topic drop to be possible but states that it is subject to stronger identity conditions.

(20) ?*Am Freitag hat er mir gebeichtet,* Δ *hat seine neue Freundin betrogen.*
on Friday has he me confessed he has his new girlfriend cheated
'On Friday he confessed to me, (he) cheated on his new girlfriend.'

If topic drop was not possible in a potentially embedded clause, this would require modifying the prefield restriction by stating that topic drop is not possible in every prefield position but only in a "special" one. Rizzi (1994) and Freywald (2020) propose two different approaches on how to specify this special position. Based on an alleged asymmetry between subject and object topic drop postulated by Cardinaletti (1990), Rizzi (1994) analyzes cases of subject topic drop in German as so-called "root null subjects", while treating object topic drop separately as an "empty category bound by a discourse-identified null operator" (Rizzi 1994: 160) (see also Section 3.6). For the subjects, on which I focus here, he assumes an analysis as a null constant without an operator that is restricted to "the specifier of the root, the highest position of the structure, the position that c-commands

[45] There is an ongoing debate in the literature about the status and the syntactic position of these V2 clauses, to which I briefly turn below.

everything and is not c-commanded by anything" (Rizzi 1994: 162).[46] In this position, the null constant is exempted from a relaxed version of the empty category principle (in the sense of Rizzi 1990a) – according to which an empty category must be chain-connected to an antecedent, if it can – and is therefore available for "discourse identification", i.e., to "receive its referential value in discourse" (Rizzi 1994: 162).[47] Rizzi (1994) assumes that in the unmarked case, the root category is the CP so that subject topic drop occurs in [Spec, CP]. However, he argues that the alleged ban on topic drop in embedded clauses shows that topic drop cannot occur in every [Spec, CP] position but only in the *specifier of the root*: "A null element can be discourse-identified only if it is not c-commanded sentence-internally by a potential identifier" (Rizzi 1994: 169). In sum, Rizzi (1994) predicts that topic drop should not be possible in every prefield but only in the prefield of root clauses, a position that is not c-commanded clause-internally by a potential identifier.[48]

[46] It seems that the c-command relation plays a role in other types of null elements as well. For instance, Stowell (1991, 1996) observes an asymmetry for null articles in headlines. A DP cannot have a null article if it is c-commanded by another DP that has an overt article, as illustrated by the contrast in (i).

(i) a. Man bites a dog.
b. *A man bites dog.

Reich (2017) explains the licensing of null articles (and also of null copulas) with a strategy of discourse-orientation (in the sense of what Huang 1984 has proposed for languages that allow for null topics). He argues that discourse-orientation as a strategy is available in headlines but not in the standard registers of German or English, which are said to follow a strategy of sentence-orientation. See also the discussion of Reich's (2017) proposal in Section 3.6.

[47] In Trutkowski's (2016) interpretation, this means "that the prefield guarantees optimal access to the discourse, which provides the antecedent for the gap" (Trutkowski 2016: 19). She adds that this is particularly important for topic drop because, unlike other ellipsis types, it allows for semantic and syntactic mismatches between antecedent and target. Therefore, in her view, a close connection to the antecedent is of greater importance. Note, however, that other types of ellipses also allow for mismatches but are not positionally constrained. For instance, sluicing (e.g., Ross 1969, Merchant 2001) allows for mismatches in finiteness, tense, modality, and polarity, among others (see, e.g., Kroll & Rudin 2017, Anand et al. 2021). See also the discussions of the prefield restriction in several syntactic approaches in Section 3.6.

[48] Apparently independently of Rizzi (1994), Ackema & Neeleman (2007) develop a similar approach. To account for the distribution of *pro*-drop in Early Modern Dutch, they add a structural condition to the factor of locality in Ariel's (1990) list of factors determining accessibility (see Section 4.4 for a description of her concept of an accessibility hierarchy). Accordingly, an antecedent is (more) highly accessible if it "is part of the same finite CP as the anaphoric expression" (Ackema & Neeleman 2007: 97). To avoid that this condition excludes the possibility of topic drop, where the positioning in [Spec, CP] rules out that it finds an antecedent within the CP in which it is contained, they exempt null pronouns at the left edge from the structural condition, considering the left edge an " 'escape hatch' for anaphoric dependencies" (Ackema & Neeleman 2007: 99).

Freywald (2020) proposes to extend the topic drop analysis to further phenomena that do not occur in the prefield, e.g., object omissions in directive infinitives.[49] To do so, she refers to Rizzi (1994) but reinterprets the *specifier of the root*-account by stating "that topic drop can only occur in the top-most position of an autonomous sentence"[50] (Freywald 2020: 167). Thereby, she equates root clauses and autonomous[51] clauses (Freywald 2020: 150), a view that is questioned by Reich & Reis (2013). They point out that not only autonomous clauses[52] are root clauses[53] but also the conjuncts of an autonomous coordination or continuative relative clauses (see, e.g., Blümel et al. 2017 for this clause type in German), where the maximal sentence node is dominated by another node, but no integration relation exists between the two clauses. When describing Freywald's (2020) approach in the following, I preliminarily adopt her equation of root clause and autonomous clause. In sum, Freywald (2020) predicts topic drop not to be pos-

[49]Freywald (2020) argues that object omissions in directive infinitives such as (i) are not a syntactic phenomenon in their own right but can be subsumed under topic drop which she, unlike me, understands as the omission of topics. According to her, treating object omission as topic drop is justified because the omitted objects are continuous topics, they can refer to an extralinguistically present entity, and there is evidence of a topic position at the left periphery in directive infinitives.

(i) *Nach dem Öffnen* Δ *kühl lagern* *(6° bis 8°C) und* Δ *innerhalb weniger Tage*
after the opening it cool store.INF 6° to 8°C and it within few days
verbrauchen.
consume.INF
'Once opened, store (it) in a cool place (6° to 8°C) and consume (it) within a few days.'
(Freywald 2020: 148)

[50]My translation, the original: "dass Topikdrop nur in der höchsten Position eines selbstständigen Satzes stattfinden kann" (Freywald 2020: 167).

[51]I use the term *autonomous* to refer to what is called *selbständig* in German, while *dependent* is used as equivalent to the German *abhängig*. As Reich & Reis (2013: 541) point out, every dependent clause is non-autonomous, but independent clauses can also be non-autonomous, such as the conjuncts of a coordination.

[52]They define syntactic autonomy in the following way: "S ist genau dann syntaktisch selbständig, wenn der maximale Satzknoten von S von keinem (anderen) Knoten dominiert wird" ('S is syntactically autonomous if and only if the maximum clause node of S is not dominated by any (other) node') (Reich & Reis 2013: 541, my translation).

[53]According to Reich & Reis (2013), a root clause is defined as follows: "Ein Satz S ist genau dann ein Wurzelsatz, wenn S in keinen Satz S' im Sinne von (I) integriert ist." ('A clause S is a root clause if and only if S is not integrated into a clause S' in the sense of (I).') (Reich & Reis 2013: 542, my translation). For integration in the sense (I), they provide this definition: "S2 ist genau dann in S1 *integriert* (eingebettet), wenn der minimale Satzknoten von S1 den maximalen Satzknoten von S2 dominiert" ('S2 is *integrated* (embedded) in S1 if and only if the minimum clause node of S1 dominates the maximum clause node of S2.') (Reich & Reis 2013: 537, my translation).

sible in every prefield position but only in the highest syntactic position of an autonomous or root clause. In principle, Rizzi's (1994) and Freywald's (2020) approaches are very similar and only make a different prediction for the special case of initial V2 clauses if they are in fact interpreted as embedded (see below). In this book, it is neither possible nor my goal to distinguish between the two approaches. I remain agnostic here and discuss my predictions and results in terms of both approaches.

Before I turn to these predictions, it is necessary to briefly sketch the ongoing debate about the status and syntactic position of presumably embedded V2 clauses, which can appear both in sentence-final (21a) and in sentence-initial position (21b).

(21) a. *Am Freitag hat er mir gebeichtet,* Δ *hat seine neue Freundin*
on Friday has he me confessed he has his new girlfriend
betrogen.
cheated
'On Friday he confessed to me, (he) cheated on his new girlfriend.'

b. Δ *hat seine neue Freundin betrogen, hat er mir am Freitag*
he has his new girlfriend cheated has he me on Friday
gebeichtet.
confessed
'(He) cheated on his new girlfriend he confessed to me on Friday.'

Superficially, at least the full forms, i.e., the syntactically complete utterances, seem to be parallel to the corresponding embedded verb-last clauses with an overt complementizer in (22). Therefore, it is reasonable to assume that the V2 just like the verb-last structures appear in the prefield or postfield position of the matrix clause and are syntactically embedded.

(22) a. *Am Freitag hat er mir gebeichtet, dass er seine neue Freundin*
on Friday has he me confessed that he his new girlfriend
betrogen hat.
cheated has
'On Friday he confessed to me that he cheated on his new girlfriend.'

b. *Dass er seine neue Freundin betrogen hat, hat er mir am Freitag*
that he his new girlfriend cheated has has he me on Friday
gebeichtet.
confessed
'That he cheated on his new girlfriend he confessed to me on Friday.'

This seems to be the view of Cardinaletti (1990), Rizzi (1994), and Volodina (2011) because they, when arguing against the possibility of topic drop in embedded clauses, refer to them as such, i.e., as embedded, and do not discuss other options. Similarly, Grewendorf (1988: 253–254) assumes the possibility of having V2 complement clauses in the prefield. Reis (1997: 138) notes that, at least at the time of her 1997 paper, the default assumption is that these kinds of V2 sentences are syntactic complements in the strict sense. With this paper, however, Reis (1997) argued against this assumption and largely changed the prevailing opinion.

She terms final V2 clauses with argument function *syntactically relatively unintegrated dependent clauses*[54] (Reis 1997: 121) and assumes, based on several diagnostics, that they are adjoined to the VP (Reis 1997: 138, see also Truckenbrodt 2006).[55] Reis (1997: 139) furthermore claims that relatively unintegrated V2 clauses cannot be placed in the prefield (see also Frank 2000, Truckenbrodt 2006, Freywald 2009, 2013).[56] For apparent counterexamples, she proposes two alternative analyses, one as parenthetical structures and one as quasi-paratactic structures.

In her view, example (23a) is an autonomous V2 clause with a final integrated V1-parenthetical. *Er kommt* is considered to be the matrix clause in which the V1 clause *glaube ich* is syntactically integrated (Reis 1997: 139, footnote 23).

(23) a. *Er kommt, glaube ich.*
he comes believe I
'He's coming, I think.' (Reis 1997: 139, footnote 23)

Reis (1996) lists the following defining properties of V1 integrated parentheticals:

(i) Verb-first

(ii) Interpretational integration into the host clause

[54] My translation, the original: "syntaktisch 'relativ unintegrierte Nebensätze' " (Reis 1997: 121).

[55] Meinunger (2004, 2006) proposes an adjunct position next to the CP including so-called *double access reading*, where the V2 clause receives theta-roles in the base position as the complement of VP and the licensing of binding relations is permitted (Meinunger 2006: 481).

[56] In more recent work, Freywald (2016: 342, footnote 16) argues that speaker judgments regarding what she calls the "Vorfeldphobie" ('prefield phobia') of V2 clauses are not very robust and that in particular, the ratings for topical V2 clauses vary widely. However, the results of experiment 3 provide no evidence for this statement. While the mean ratings for utterances with initial embedding are generally slightly but significantly worse than for unembedded clauses, the standard deviations are comparable. Participants appear to be as confident in their judgments of utterances with initial embedded V2 clauses as they are for utterances without embedding (whereas there appears to be more variation for utterances with final embedded clauses, where the standard deviations are larger). See Section 3.3.2.3 for details.

(iii) Prosodic integration into the host clause

a) no focus-background-structure of their own

b) no stress/focus

c) no intonational breaks (i.e. no 'comma intonation') (Reis 1996: 48)

For longer examples like the item of experiment 3 in (24) (see Section 3.3.2.1 for details on the materials), an analysis as an autonomous V2 clause with integrated V1 parenthetical seems unlikely (see also Pauly 2013: 149–151 for a critical discussion of Reis's arguments for a parenthetical reading based on binding data).

(24) *B: (Er) hat seine neue Freundin betrogen, hat er mir am Freitag*
he has his new girlfriend cheated has he me on Friday
gebeichtet
confessed
B: '(He) cheated on his new girlfriend, he confessed to me on Friday'

The assumed V1 parenthetical *hat er mir am Freitag gebeichtet* is longer than a typical V1 parenthetical and *beichten* ('to confess') is less typical as a verb compared to classic "verbs of saying, thinking, and believing and verbs of asking" like *glauben* ('to believe'), *fragen* ('to ask'), *denken* ('to think'), *meinen* ('to think'), *sagen* ('to say'), and *wissen wollen* ('want to know') (Steinbach 2007: 56).[57] Additionally, despite a close interpretational relation to the V2 clause, the V1 clause provides additional information in the form of the temporal adverbial *am Freitag*.

[57] Note that Rapp (2015: 194, footnote 20) considers the very similar verb *gestehen* ('to confess'/'to admit') to be a factive verb that only allows for V2 embeddings in conjunctive mood (i).

(i) *Sie gesteht, sie könne/*kann diesen Mann nicht einschätzen.*
she confesses she can.CONJ/can.IND this man not estimate
'She confesses she cannot evaluate this man.' (Rapp 2015: 194, footnote 20, her judgment)

In the Deutsches Referenzkorpus DeReKo ('German reference corpus') (Leibniz-Institut für Deutsche Sprache 2022, see also Kupietz et al. 2010) there are, however, isolated examples of *gestehen* with V2 complement clauses in indicative mood, such as (ii). Thus, at least this verb should generally allow for embedded V2 clauses, regardless of the mood.

(ii) *In der "Bunten" gesteht er, er mag den harten Rap des Amerikaners*
in the Bunte confesses he he likes.IND the hard rap the.GEN American
Eminem [...].
Eminem
'In the 'Bunte' (tabloid, L.S.) he confesses that he likes the hard rap of the American Eminem [...].' [DeReKo, HMP11/APR.02340, Hamburger Morgenpost, 04/28/2011, p. 02]

In sum, there are good reasons not to treat examples like (23a) as autonomous V2 clauses with V1 parentheticals (though I come back to this option in the discussion in Section 3.3.2.4).[58]

The second alternative to the analysis as a syntactically embedded V2 clause is assuming a quasi-paratactic structure. The clause *hat er mir am Freitag gebeichtet* in (21b), repeated here as (25a), would then have to be interpreted as a V2 clause with topic drop (25b), as proposed in Reis (1995, 1996) for unintegrated parentheticals, similar to Reis's (1995: 66) example in (26).

(25) a. Δ *hat seine neue Freundin betrogen, hat er mir am Freitag*
he has his new girlfriend cheated has he me on Friday
gebeichtet.
confessed
'(He) cheated on his new girlfriend he confessed to me on Friday.'

b. (Er) hat seine neue Freundin betrogen, (das) hat er mir am Freitag gebeichtet.

(26) *Hans – (so/das) glaubt FRITZ – wird KOMMen.*
Hans so/that believes Fritz will come
'Hans – (so) Fritz believes – will come.' (Reis 1995: 66)

The corresponding full form would have a demonstrative *das* ('that') in the prefield referring back to the preceding clause (25b). The other V2 clause *Er hat seine neue Freundin betrogen* would not be in the prefield of the *gebeichtet*-clause, but both clauses would be considered root clauses and form together a quasi-paratactic structure.

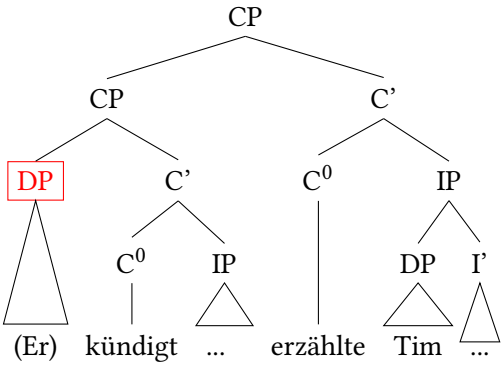
In sum, we end up with two proposals on how to analyze the final V2 clauses and with three for the initial V2 clauses. In the following, I briefly sketch how Rizzi's (1994) and how Freywald's (2020) refinement of the prefield restriction account for these proposals. For both analyses of the final V2 clauses, either as embedded or as an adjunct to the VP, Rizzi's (1994) and Freywald's (2020) theoretical accounts predict topic drop to be impossible because the [Spec, CP] of the final V2 clause is not the highest [Spec, CP] of the autonomous or root clause and it is c-commanded sentence-internally by a potential identifier, namely by

[58]To exclude the possibility of a reading as V1 parenthetical it might be an option to test the respective structures with and without an intonational break between the two clauses. Given Reis's (1996) defining properties, the variants with intonational break should not be interpretable as V1 parentheticals. Nevertheless, they are not automatically "true" complement clauses but could still be interpreted as unintegrated V2 parentheticals, as discussed in the following.

the subject. For the initial V2 clauses, there are the options to analyze them as parenthetical structures with a V1 and a V2 clause, or as quasi-paratactic structures with two independent V2 clauses. Under both analyses, Rizzi (1994) and Freywald (2020) would predict that topic drop should be possible since the corresponding prefield would be the highest of an autonomous or root clause and it would not be c-commanded sentence-internally by a potential identifier. Only if the initial V2 sentences were indeed analyzed as embedded, and if we assumed a similar structure to (27b) for (27a), would the predictions of the two approaches differ.

(27) a. *(Er) kündigt, erzählte Tim.*
he quits told Tim
'(He) quits, Tim told.'

b.



While, according to Freywald (2020), topic drop should be impossible because the [Spec, CP] with *er* is not the highest of the autonomous or root clause, it should be allowed according to Rizzi (1994) since the *er* is not c-commanded sentence-internally by a potential identifier.

While a distinction between the three analyses proposed for the initial V2 cases would be desirable in principle, it must be left to future research, just like a distinction between Freywald's (2020) and Rizzi's (1994) accounts. In the following experiment, I investigated more generally whether topic drop is possible at all in the potentially embedded initial or final V2 sentences. This is my main question here, as it allows for a refinement of the prefield restriction under investigation.

3.3.2 Experiment 3: topic drop in (potentially) embedded clauses

In experiment 3, I tested potentially embedded V2 clauses with topic drop and compared them to a simple baseline with the clause in isolation and to the cor-

responding full forms.[59] Moreover, I varied whether the potentially embedded V2 clause occurred sentence-initially or sentence-finally. The acceptability rating study had the form of a 2 × 3 design with the two factors of Completeness (full form vs. topic drop) and V2 Clause (initial vs. final vs. baseline). Table 3.7 summarizes the predictions for the analyses and approaches discussed in Section 3.3.1.

Table 3.7: Overview of the predictions for experiment 3: a checkmark indicates that topic drop is predicted to be possible.

	Final V2	Initial V2 embedded	Initial V2 parenthetical	Initial V2 paratactic
General prefield restriction	✓	✓	✓	✓
Rizzi (1994)	✗	✓	✓	✓
Freywald (2020)	✗	✗	✓	✓

If there is a general prefield restriction, it should not matter whether the clause is potentially embedded or not, i.e., the acceptability of topic drop should be comparable in all three V2 Clause conditions because the ellipsis always occurs in a prefield position. If Freywald's (2020) proposal of a restriction of topic drop to the highest [Spec, CP] of an autonomous or root clause is correct, topic drop in the final condition should be degraded compared to the baseline topic drop condition. For the initial condition, topic drop should be possible for Freywald (2020) if the V2 clause is analyzed as autonomous, combined either with a V1 parenthetical or with a further V2 clause with topic drop but not if analyzed as an embedded clause. According to Rizzi's (1994) approach, topic drop should be excluded in the final (possibly embedded) V2 clauses, while it should generally be possible in the initial V2 clauses, regardless of the analysis adopted. Although I present the predictions of Rizzi's (1994) and Freywald's (2020) approaches here, my main goal is not to distinguish between them but to answer the question of whether topic drop is possible in any prefield. Nevertheless, it seems reasonable to be able to discuss the implications of the results for these two proposals, as it may be the starting point for further research.

Note that experiment 2 and experiment 3 were part of the same study. As a consequence, the items served reciprocally as fillers (along with the additional fillers, see Section 3.3.2.1). Therefore, I only outline the main points of the experimental procedure, as the details were already sketched in Section 3.2.2.

[59] All experimental items, fillers, and the analysis script are available online at: https://osf.io/zh7tr.

3.3.2.1 Materials

Items

I constructed 24 items in the form of question-answer pairs, such as (28) . The question asked for news about a third person X, either in the form *Was gibt's Neues von X?* ('What's new from X?') or *Wie ist die Lage bei X?* ('What is the situation with X?'). The answer provided the requested information in the form of a V2 clause with an omitted or realized subject in the prefield. This clause was either presented on its own (28a), serving as a baseline, or in the initial (28b) or final (28c) position relative to another clause, i.e., in a potentially embedded position.[60] I only used the verbs *erzählen* ('to tell'), *beichten* ('to confess'), and *gestehen* ('to admit') (as opposed to typical V1 parenthetical verbs like *glauben* ('to believe'/'to think')), to make a parenthetical reading of the matrix clause less likely (see Section 3.3.1). In addition, I expanded the clause with additional information in the form of temporal adverbs. Both the verbs in the matrix clauses and the verbs in the embedded clauses were presented in perfect tense.

(28) *A: Was gibt's Neues von Tim?*
what gives.it new from Tim
A: 'What's new from Tim?'

a. *B: (Er) hat seine neue Freundin betrogen*
he has his new girlfriend cheated
B: '(He) cheated on his new girlfriend'
(topic drop / full form, baseline)

b. *B: (Er) hat seine neue Freundin betrogen, hat er mir am Freitag*
he has his new girlfriend cheated has he me on Friday
gebeichtet
confessed
B: '(He) cheated on his new girlfriend, he confessed to me on Friday'
(topic drop / full form, initial)

c. *B: Am Freitag hat er mir gebeichtet, (er) hat seine neue Freundin*
on Friday has he me confessed he has his new girlfriend
betrogen
cheated
B: 'On Friday he confessed to me, (he) cheated on his new girlfriend'
(topic drop / full form, final)

[60] For a discussion of the question of whether the V2 clauses should be considered embedded or not, see Section 3.3.1.

Fillers

As described in Section 3.2.2.1, 72 fillers were included in the experiment: 24 (potential) gappings, 16 PP fragments, 24 clauses with overt or covert constituents in the prefield or the middle field, and eight catch trials with word order violations.

3.3.2.2 Procedure

The procedure was described in detail in Section 3.2.2.2. As stated above, the materials were presented as WhatsApp dialogues using the design shown in Figure 3.2 on page 70 and partly contained emojis to increase the naturalness.

3.3.2.3 Results

As described in Section 3.2.2.3, I excluded three participants who did not pass the attention checks by rating four or more of the ungrammatical fillers with 6 or 7. For the remaining 45 participants, Table 3.8 shows the mean ratings and standard deviations per condition, and Figure 3.5 the mean ratings and 95% confidence intervals. The visual inspection indicates that there was a three-part degradation independently of Completeness. The baseline conditions were more acceptable than the initial conditions and the initial conditions, in turn, were more acceptable than the final conditions. It is striking that topic drop in the final condition was significantly worse, with a mean score of only 3.93.

Table 3.8: Mean ratings and standard deviations per condition for experiment 3

Completeness	V2 Clause	Mean rating	Standard deviation
Full form	Baseline	5.62	1.42
Topic drop	Baseline	5.40	1.53
Full form	Initial	5.31	1.44
Topic drop	Initial	5.13	1.58
Full form	Final	4.92	1.62
Topic drop	Final	3.93	1.80

The rating data were analyzed with CLMMs from the package ordinal (Christensen 2019) in R, generally following the procedure described in Section 3.1.6.3. For the three-level predictor V2 Clause, however, I used forward coding to compare every level to the subsequent one(s) using two contrasts: Baseline.Other

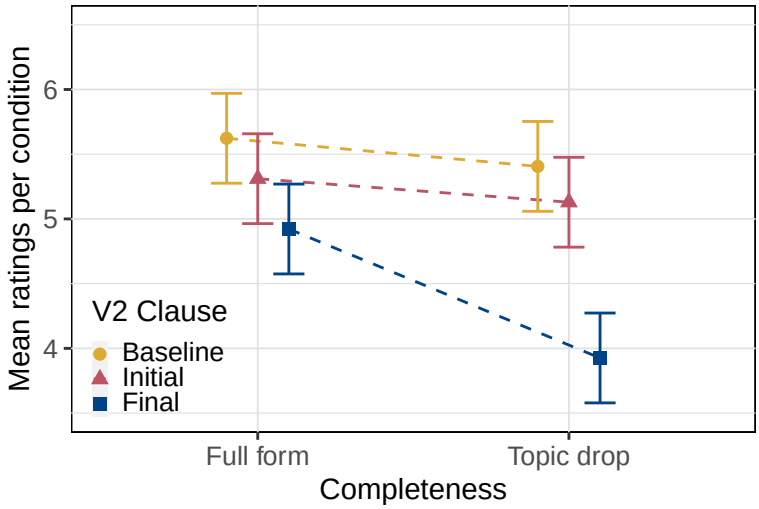


Figure 3.5: Mean ratings and 95% confidence intervals per condition for experiment 3

compares the level baseline, coded as $^2/_3$, to the other two levels, coded as $-^1/_3$. Other.Final compares the levels baseline and initial, coded as $^1/_3$, to the level final, coded as $-^2/_3$. Besides these two predictors that represent V2 Clause, I also included the deviation-coded predictor Completeness (full form coded as 0.5, topic drop coded as -0.5), the centered and scaled Position of the trial in the experiment, as well as the two-way interactions between the three predictors Completeness, Baseline.Other, and Position, and the two-way interactions between Completeness, Other.Final, and Position. The random effects structure consisted of random intercepts for subjects and items and of by-item and by-subject random slopes for the predictors Completeness, Baseline.Other, Other.Final, Position, and the two-way interactions of Completeness with Baseline.Other and Other.Final.[61] The fixed effects in the final model with symmetric thresholds are shown in Table 3.9.

The final model contained a significant interaction between Other.Final and Completeness ($\chi^2(1) = 8.96$, $p < 0.001$) according to which topic drop in a (potentially embedded) final V2 clause is significantly degraded compared to topic

[61] The formula of the full model was as follows: `Ratings ~ Completeness + Baseline.Other + Other.Final + Position + Completeness : Baseline.Other + Completeness : Other.Final + Position : Baseline.Other + Position : Other.Final + (1 + (Baseline.Other + Other.Final) * Completeness + Position | Items) + (1 + (Baseline.Other + Other.Final) * Completeness + Position | Subjects)`.

Table 3.9: Fixed effects in the final CLMM of experiment 3

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	0.76	0.15	20.57	< 0.001	***
Baseline.Other	0.64	0.26	5.65	< 0.05	*
Other.Final	1.33	0.22	26.62	< 0.001	***
Completeness × Other.Final	−1.25	0.41	8.96	< 0.001	***

drop in the initial V2 clause or the baseline condition. The significant main effects of Other.Final ($\chi^2(1) = 26.62$, $p < 0.001$) and of Baseline.Other ($\chi^2(1) = 5.65$, $p < 0.05$) show a gradation for V2 Clause independently of Completeness. Both full forms and utterances with topic drop were rated as more acceptable in the baseline condition than with an initial V2 clause. Additionally, utterances with an initial V2 clause were in turn more acceptable than utterances with a final V2 clause (see Figure 3.5). Finally, there was also a significant main effect of Completeness ($\chi^2(1) = 20.57$, $p < 0.001$), according to which utterances with topic drop received degraded ratings compared to full forms.

3.3.2.4 Discussion

In this acceptability rating study, I investigated whether topic drop is possible in every prefield position by testing utterances with sentence-initial or sentence-final (potentially embedded) V2 clauses. The results show that topic drop in a potentially embedded clause that is positioned sentence-finally was degraded, even though the constituent was omitted from a prefield position. This suggests that the prefield positioning of the omitted constituent is a necessary but not a sufficient condition for topic drop.

Utterances with an initial V2 clause were degraded compared to the simple baseline, but unlike for the final V2 clauses, this was not restricted to topic drop but also held for the full forms.[62] This means that topic drop in a (potentially embedded) V2 clause in sentence-initial position was not particularly degraded but can be considered to be possible. This result can be explained in two ways:

a) If the initial conditions indeed contain embeddings, it could be concluded that topic drop in embedded clauses is as acceptable as the corresponding full

[62]This drop in acceptability observed for the initial condition independently of Completeness compared to the baseline condition could be caused by the more complex structures being less natural in instant messages than simple sentences.

form, as long as the embedding is initial. This result would call into question Freywald's (2020) variant of the *specifier of the root*-account since in this case topic drop would not occur in the highest [Spec, CP] of an autonomous or root clause but in a subordinate position. However, it would not clash with Rizzi's (1994) variant because the [Spec, CP] at the left edge may not be the highest position, but it would be a position that is not c-commanded sentence-internally by a potential identifier and should in principle allow for topic drop.

b) If we consider the possibility that the initial conditions do not contain embeddings but are either autonomous V2 clauses with integrated V1 parentheticals or V2 root clauses in a paratactic relation to a second V2 clause with topic drop, Freywald's (2020) variant would not be at stake either. The [Spec, CP] of the autonomous or root V2 clause would be the highest [Spec, CP], from where topic drop should be possible.

These results do not allow for a decision between Rizzi's (1994) more liberal and Freywald's (2020) stricter variant of the *specifier of the root*-account. The important result is that topic drop is not just restricted to the prefield but only to a prefield that is at least not c-commanded sentence-internally by a potential identifier or even has to be the highest prefield of an autonomous or root clause. Furthermore, both accounts explain why topic drop is not possible in the middle field because positions in the middle field of V2 clauses are necessarily c-commanded sentence-internally by a lexical category and are not the highest position of the autonomous or root clause in which they occur.[63]

3.4 Is topic drop restricted to the very first element?

The results of experiment 3 on potentially embedded topic drop suggest that topic drop is not possible in every prefield position. It is restricted to at least a prefield that is not c-commanded sentence-internally by a potential identifier if not to the highest prefield of an autonomous or root clause. In any case, the position is usually located at the left edge of an utterance. Trutkowski (2016) argues that the restriction to this sentence-initial (Huang 1984: 547, Trutkowski 2016: 1), left-peripheral (Freywald 2020: 150), or left edge (Ackema & Neeleman 2007: 99) position is required to ensure that the antecedent in the previous discourse is optimally accessible. This brings up the question of whether the sentence-initial

[63]The fact that topic drop does not seem to be possible in left dislocations (see Volodina & Onea 2012 and the next section) suggests that it is not sufficient to restrict topic drop to a position that is not c-commanded sentence-internally by a potential identifier. The restriction to the prefield probably needs to be assumed independently.

position is necessary, i.e., whether topic drop must also be the linearly first word in the string of words that form the utterance.

3.4.1 Theoretical background

Examples like (29), taken from the FraC (see Section 7.2.1), suggest that topic drop does not generally need to be the very first element of an utterance. The insulting address *Vollidiot* ('complete idiot') (29a), the particle *ja* ('yes') (29b), the greeting *hey-ho* (29c), as well as the conjunctions *und* ('and') (29d), (29e) and *aber* ('but') (29f) all seem to be able to precede topic drop.[64]

(29) a. *Vollidiot Δ kannst mich mal!!!*
complete.idiot you.2SG can me PART
'Complete idiot, screw (you)!!!' [FraC S481]

b. *ja, Δ wäre mir lieber.*
yes that would.be me rather
'Yes, I would prefer (that).' [FraC D1416]

c. *HEY-HO, Δ WEIß NET, OB I NACH DER STRESSIGEN WOCHE SO LANGE DURCHHALTE,*
hey-ho I know not if I after the stressful week so long last
'Hey ho, (I) don't know if I'll last that long after the stressful week' [FraC S1691]

d. *Wir hängen sehr an ihr. Und Δ mögen uns gar nicht vorstellen, wie es irgendwann ohne sie ist.*
we hang very at her and we may us at.all not imagine how it sometime without her is
'We are very attached to her. And (we) don't even want to imagine what it will be like without her at some point.' [FraC B36–37]

e. *Da habe ich nun ein schönes neues Tablet. Und Δ kann's nicht mit dem Wlan verbinden.*
there have I now a pretty new tablet and I can.it not with the wifi connect
'Now I have a nice new tablet and can't connect it to the wifi.' [FraC B363]

[64] This is also discussed by Sigurðsson (2011: 287), who states that for topic drop "high discourse particles and left-dislocated elements" can precede topic drop.

f. *ich habe ja jetzt schon ferien [sic!]! aber* Δ *fahre erst in*
I have PART now already vacations but I go only in
2Wochen [sic!] weg!
2weeks away
'I am already on vacation now! But (I) am not leaving for another two weeks!' [FraC S417]

Apparently, topic drop can be preceded by elements that occupy a position to the left of the prefield, which is usually termed *Vorvorfeld* ('pre-prefield') or *linkes Außenfeld* ('left outfield') (Zifonun et al. 1997: 1577–1581, Auer 1997, Gallmann 2015, Speyer & Reich 2020) in the extended topological field model.[65] It is often divided into at least two subfields, a position for coordinating conjunctions (COOR(D)) and one for so-called *linksversetzte* ('left dislocated') constituents (e.g., Speyer & Reich 2020, Pittner & Berman 2021; see also Pafel (2011: 55), who calls the first *Anschlussposition* ('connection position') and the second *Topikfeld* ('topic field') and Eisenberg (2020: 410), who equates left dislocation and pre-prefield).

Zifonun et al. (1997: 1577) further assume that left of the coordination position, relatively independent interactive units such as interjections, answer particles, and forms of address can be placed. In the Grammis system by the IDS, it is pointed out that these units do not contribute to the construction of sentences and tend to be the units with the loosest connection to the following sentence (Leibniz-Institut für Deutsche Sprache 2019). Examples (29a)–(29c) fall into this category. Given that they are frequently considered units on their own, the fact that they precede topic drop does not necessarily question a sentence-initial position of topic drop because these elements can be considered to be outside of the actual sentence.

This is less plausible as an explanation for left dislocations. In (30a), the dislocated DP in the pre-prefield *den einen Schauspieler aus Breaking Bad* receives the accusative case from the verb *mögen* in the clause, which suggests a syntactic integration of the preposed element. However, Volodina & Onea (2012) argue that topic drop is neither possible for left dislocated constituents because these constituents usually function as contrastive topics (see Section 3.1.4), nor after left dislocations because the corresponding prefield constituent must be focal and focal constituents cannot be omitted (Volodina & Onea 2012: 230–231). The topic drop variant of (30a), (30b), indeed seems to be marked at least if not impossi-

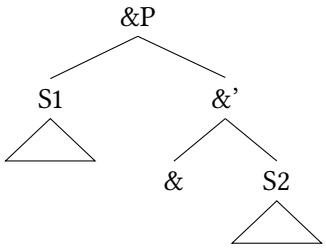
[65] Wöllstein (2014) distinguishes between a pre-prefield and a left outfield while Pasch et al. (2003) propose the term *Nullstelle* ('zero' or 'root') because they assume that this position exists independently of a prefield position, for example, for V1 sentences as well.

ble.[66] Moreover, since there do not appear to be instances of topic drop after a left dislocation in either the FraC or in Frick's (2017) corpus, left dislocations also do not seem to challenge the sentence-initial positioning of topic drop either.[67]

(30) a. *Den einen Schauspieler aus Breaking Bad, den mag ich total.*
the.ACC one.ACC actor from Breaking Bad the.ACC like I totally
'The one actor from Breaking Bad, I really like him.'
b. ?Den einen Schauspieler aus Breaking Bad, Δ mag ich total.

What remains are the conjunctions in the COORD position. Examples (29d)–(29f) show that they indeed occur before topic drop. Following Johannessen (1998), the phrase structure of coordinated structures can be modeled as in (31), taken from Reich (forthcoming). Johannessen assumes a conjunction phrase &P with the conjunction & as the (functional) head.[68] The conjunction is not a direct part of either conjunct but syntactically closer to the second conjunct (Reich forthcoming).

(31)



(Reich forthcoming)

Nevertheless, it is not clear whether the conjunctions in (29d)–(29f) do function as conjunctions linking two clauses or whether they should rather be consid-

[66] However, this is difficult to judge because, without a clear intonational pause, the dislocated constituent could also be understood as the prefield constituent of a full form without a pre-prefield.

[67] The potential impossibility of topic drop after left-dislocated constituents could be explained in Rizzi's (1994) logic by assuming that the left-dislocated constituent is a possible identifier for the prefield constituent. See also Frey (2005: 165–166), who assumes, based on binding and island effects, that the left-dislocated constituent and the resumptive pronoun, by which it is taken up again, together form one syntactic object and that the left-located constituent is in [Spec, CP], whereas the prefield is formed by [Spec, FinP].

[68] The fact that conjunctions are functional rather than lexical heads becomes important below. For a detailed justification of this assumption see Johannessen (1998: 96–105).

ered discourse particles or discourse markers. Proske (2014: 130) summarizes the properties of discourse markers in German as follows: They are not independent, cannot form a clause or a turn on their own, and can but need not be realized as separate intonation units (see also Auer 1997). When discussing the research on discourse particles along various dichotomies, Fischer (2006) summarizes that a considerable number of researchers consider conjunctions to be "items that constitute parts of utterances" (Fischer 2006: 8), i.e., they consider them to be syntactically integrated. In contrast, Gohl & Günthner (1999) and Imo (2012) argue that discourse markers are often not part of the following utterance and are only loosely syntactically connected to it. In summary, there seems to be no consensus on whether conjunctions are part of the subsequent utterance, regardless of whether they function as discourse markers or not. However, they seem to be less independent than the interjections, forms of address, and greetings discussed above. Therefore, they are the best candidate to test whether elements can precede topic drop.

Concerning the relationship between conjunctions and topic drop, it is important to note that the structures in (29d)–(29f) have clear parallels to subject gaps in coordination, also called subject deletion, as has been pointed out by, for example, Klein (1993) and Haegeman (2007). Without the sentence boundary, (29d) and (29f) could be analyzed as symmetric coordinations with a shared prefield (constituent), as in (32a) and (32c) (e.g., Wilder 1997).[69] This is not possible for example (29e), though, which would have to be analyzed as an asymmetric coordination (32b), more specifically, as a so-called *SLF construction* for *Subjektlücke in F-Sätzen* ('subject gaps in F-clauses')[70] (Höhle 1983, 1990; see also Reich 2007, 2009a,b, 2013, Bonitz 2013, Barnickel 2017, Oppermann 2021). SLF coordination is characterized by the fact that in the non-initial conjuncts, the finite verb is fronted and no overt subject is present, while in the initial conjunct, the subject is located in the middle field so that an analysis in terms of a shared prefield is impossible (Reich 2008, 2009b).

(32) a. Wir hängen sehr an ihr und Δ mögen uns gar nicht vorstellen, wie es irgendwann ohne sie ist.
b. Da habe ich nun ein schönes neues Tablet und Δ kann's nicht mit dem Wlan verbinden.
c. ich habe ja jetzt schon ferien, aber Δ fahre erst in 2Wochen [sic!] weg!

[69] See example (29) above for the glossing.
[70] F-clauses are clauses with a finite verb in the first or second position (Höhle 1983: 1).

It is important to note that not all conjunctions allow (to an equal degree) for a subject gap interpretation according to the literature. Such an interpretation in both symmetric and asymmetric coordination is possible for the basic case *und* ('and'), and also for *oder* ('or'), *sondern* ('but rather'), and *aber* ('but') (Höhle 1983, Wunderlich 1988). While Höhle (1983) argues that *aber* is special because in asymmetric coordination it must not precede the second conjunct like *und* does (33a) but follow the finite verb, as in (33b), Wunderlich (1988) does not mention a special status of *aber* but presents example (34), where *aber* precedes the second conjunct, as in (33a). As a mediating position, so to speak, van de Velde (1986: 504) argues that the use of the subjunctive adverb *aber* instead of the conjunction in subject gap constructions, while apparently preferred, is by no means obligatory.

(33) a. **da standen ein paar Leute rum, aber* Δ *rührten keinen*
there stood a few people around but they moved no
Finger
finger
'There were a few people standing around, but (they) didn't lift a finger'

b. da standen ein paar Leute rum, Δ rührten aber keinen Finger (Höhle 1983: 13, his judgments)

(34) *In den Wald ging der Jäger, aber* Δ *fing nichts.*
in the forest went the hunter but he caught nothing
'The hunter went into the forest, but (he) caught nothing.'
(Lit. 'Into the forest, the hunter went, but (he) caught nothing.') (Wunderlich 1988: 308, his judgment)

A subject gap interpretation is impossible for the empty conjunction (Reich 2013: 361), the two-part conjunctions *weder – noch* ('(n)either – (n)or') and *sowohl – als auch* ('both – and') (Höhle 1983: 14, Wunderlich 1988: 310, Reich 2013: 361), for the complex conjunctions *entweder – oder* ('either – or') and *einerseits – andererseits* ('on the one hand – on the other hand') (Reich 2009a: 93, Reich 2013: 361), and for *denn* (parordinating 'because'),[71] both in symmetric (35a) and asymmetric coordination (35b) (van de Velde 1986: 508, Bonitz 2013: 54).

[71]While, e.g., Zifonun et al. (1997: 60) and Eisenberg (2020: 216) classify *denn* as coordinating conjunction, the *Handbuch der deutscher Konnektoren* ('Handbook of German connectives') (Pasch et al. 2003: 584), Reich & Reis (2013: 549), and Breindl (2017: 16) list it as "syntaktischen Einzelgänger" ('syntactic loner'). According to Breindl (2017: 15), it deviates from the typical coordinating conjunctions *und* and *oder* in not allowing for ellipsis in coordination (i), nor for category variable internal conjuncts (ii), nor for embeddings (iii) (see also Reich & Reis 2013: 549).

(35) a. **Ich kann heute nicht kommen, denn* Δ *bin krank.*
I can today not come because I am sick
'I cannot come today because (I) am sick.'
b. **Heute kann ich nicht kommen, denn* Δ *bin krank.*
today can I not come because I am sick
'Today I cannot come because (I) am sick.' (van de Velde 1986: 508, adapted)

From these statements, it follows that topic drop with certain preceding conjunctions might be acceptable because speakers use it analogously to or hearers reanalyze it as subject gap in symmetric or asymmetric coordination.

Besides the fact that such an interpretation may not be possible for all preceding conjunctions, two further facts speak against the analogy to subject gaps as the only explanation. First, example (36) taken from an email by a student[72] suggests that topic drop after a conjunction is also possible if the subject of the preceding clause is distinct from the omitted subject and a subject gap analysis is not available.

(36) *Es tut mir wirklich sehr Leid [sic!] und* Δ *hoffe Sie haben*
it does me really very sorry and I hope you.2SG.POL have
Verständnis.
understanding
'I am really very sorry and (I) hope you understand.' (Lit. 'It does me very sorry and (I) hope you understand.') (email by a student, 05/19/2022)

(i) a. *Die Kinder sind heimgegangen* und *die Eltern (sind heimgegangen).*
the children are home.gone and the parents are home.gone
'The children have gone home and the parents (have gone home).'
b. Die Kinder sind heimgegangen *denn* die Eltern *(sind heimgegangen). (Breindl 2017: 15)

(ii) a. *Wir laden Uwe* und *Evi ein.*
we invite Uwe and Evi VPART
'We invite Uwe and Evi.'
b. *Wir laden Uwe *denn* Evi ein. (Breindl 2017: 15)

(iii) a. *{Dass Otto krank ist* und *heute nicht kam}, hat alle überrascht.*
that Otto sick is and today not came has all surprised
'That Otto is sick and did not come today surprised everyone.'
b. *{Dass Otto krank ist, *denn* heute nicht kam}, hat alle überrascht. (Breindl 2017: 15)

[72] I thank Nele Hartung for providing me with this example.

Second, a preceding conjunction in the form of a discourse particle apparently can occur with object topic drop, see example (37b) with *aber* ('but'), a modification of the attested example (37a), while object gaps in coordination are taken to be impossible (see Horch 2014 for experimental evidence against Fortmann 2005, who considers object gaps to be possible under certain circumstances).[73]

(37) a. *ach, Hotels habe ich. ja. zwei Möglichkeiten. [...] fünf Minuten*
INJ hotels have I yes two options five minutes
vom Bahnhof mit Garage. Δ *brauchen wir eh nicht,*
from.the station with garage that need we anyway not
so ein Quatsch.
such a nonsense
'Ah, hotels I have. Yes. Two options. [...] Five minutes from the station with garage. We don't need (that) anyway, such nonsense.'
[FraC D1974–1975]

b. [...] fünf Minuten vom Bahnhof mit Garage. Aber Δ brauchen wir eh nicht, so ein Quatsch.

To complement the corpus and introspective data just presented with experimental evidence, I conducted an experiment on topic drop with and without preceding conjunctions, which is presented in the next section.

3.4.2 Experiment 4: topic drop and preposed conjunctions

In experiment 4, I investigated whether topic drop is restricted to the left-most element of a clause using the test case of topic drop after conjunctions.[74] I tested utterances with topic drop and corresponding full forms that were or were not preceded by a conjunction and collected acceptability judgments. Since, as shown in Section 3.4.1, different conjunctions allow for a subject gap interpretation to varying degrees, I altered the type of conjunction between items for exploratory reasons. A third of the items contained *und* ('and') as a conjunction, a third *aber* ('but'), and another third *denn* (parordinating 'because').[75] As a control predictor, I also varied whether or not the sentence preceding the critical utterance theoretically allows for interpreting the target sentence as an SLF construction, i.e.,

[73] An informal survey of colleagues suggests that the variant with the conjunction is in fact even slightly preferred, although the general judgments for examples (37a) and (37b) vary widely, which calls for a systematic empirical investigation in a future study.

[74] All items, fillers, and the analysis scripts are available online: https://osf.io/zh7tr.

[75] The term *parordinating* was introduced by Höhle 1986: 329 for connectives that neither coordinate nor subordinate their conjuncts. See below for a more exhaustive discussion of *denn*.

as containing a subject gap. Based on the literature presented in the previous section, such an interpretation should be perfectly possible for *und*, impossible for *denn*, and possibly marked for *aber*, given that in my materials *aber* does not occur post- but pre-verbally in the second conjunct (see below). However, since the claims about *aber* and *denn* have not, to my knowledge, been empirically confirmed, it does not seem unreasonable to vary the possibility of a subject gap reading for all three conjunctions. In sum, the experiment had the form of a 2 × 2 × 2 design with Completeness (full form vs. topic drop), Presence of Conjunction (absent vs. present), and Subject Gap (possible vs. impossible).

3.4.2.1 Materials

Items

I constructed 24 items in the form of instant messaging dialogues between two persons, such as (38), adapting the materials of experiment 2. Person A always asked a question, to which person B responded with three utterances. The last of these utterances was the target utterance (38d), in which the 1st person singular subject pronoun *ich* was omitted from or realized in the prefield. The (empty or filled) prefield was (38d-i) or was not (38d-ii) preceded by one of the three conjunctions *und* ('and'), *aber* ('but'), or *denn* (parordinating 'because'). As mentioned above, the conjunction was varied between items, so that eight token sets were built with *und*, such as (38), another eight with *aber*, and another eight with *denn*.

(38) a. *A: Was habt ihr heute Abend geplant?*
what have you.2pl today evening planned
A: 'What do you have planned for tonight?'

b. *B: Wir wollen uns den neuen Matrixfilm im Kino anschauen.*
we want us the new matrix.movie in.the theater watch
B: 'We want to watch the new Matrix movie in the theater.'

c. (i) *B: Die ersten drei Filme mag ich total.*
the first three movies like I totally
B: 'I totally like the first three movies.' (subject gap possible)

(ii) *B: Die ersten drei Filme gefallen mir total gut.*
the first three movies please me.dat totally well
B: 'I totally like the first three movies.'
(Lit. 'The first three movies please me totally well.')
(subject gap impossible)

d. (i) *B: Und (ich) bin jetzt richtig gespannt auf den neuen Teil.*
and I am now rightly keen on the new part
B: 'And (I) am now really excited about the new part.'
(conjunction present)

(ii) *B: (Ich) bin jetzt richtig gespannt auf den neuen Teil.*
I am now rightly keen on the new part
B: '(I) am now really excited about the new part.'
(conjunction absent)

The penultimate utterance, (38c), was varied between whether it theoretically enabled an asymmetric coordination with a subject gap reading of the target utterance or not. In the condition that was intended to enable such a reading, (38c-i), the speaker appeared as a 1st person singular subject, just like in the target utterance. In the other condition (38c-ii), the speaker was referred to by means of an object pronoun in an impersonal construction with verbs such as *jdm. gefallen* ('to please sb.'), *jdm. gut gehen* ('to be well'), or *jdm. wohl sein* ('to feel comfortable'). Examples (39) and (40) show the relevant manipulation also for a token set with *aber* and *denn* respectively.[76]

(39) a. *Seit gestern kränkle ich ein bisschen* 🤧 *Aber* Δ *fühl mich*
since yesterday ail I a bit but I feel REFL
bestimmt nächste Woche besser.
surely next week better
'Since yesterday I have been a bit sick. But I'm sure I'll feel better next week.' (subject gap possible)

b. *Seit gestern geht's mir nicht so gut* 🤧 *Aber* Δ *fühl mich*
since yesterday goes.it me.DAT not so good but I feel REFL
bestimmt nächste Woche besser.
surely next week better
'Since yesterday I am not so well. But I'm sure I'll feel better next week.' (subject gap impossible)

(40) a. *So richtig wohl fühl ich mich dabei aber nicht. Denn* Δ *kenn*
so rightly well feel I REFL thereby but not because I know
den Typ ja gar nicht.
the guy PART at.all not
'But I don't really feel comfortable with it. Because I don't know the guy at all.' (subject gap possible)

[76] Note that the manipulation of SUBJECT GAP is lost in the translation to English, but it can be seen from the glosses.

b. *So richtig wohl ist mir dabei aber nicht. Denn* Δ *kenn den*
so rightly well is me.DAT thereby but not because I know the
Typ ja gar nicht.
guy PART at.all not
'But I don't really feel comfortable with it. Because I don't know the guy at all.' (subject gap impossible)

Fillers

The items were mixed with a total of 72 fillers: 16 items from experiment 6, which investigated the impact of topicality on topic drop (see Section 9.2.2) consisting of dialogues with four turns, 24 items on preposition omission in fragments with proper names vs. definite descriptions also consisting of four turns, 24 (potential) gapping structures with V2 vs. verb-final word order, half of which consisted of four and half of two turns, and eight slightly adjusted catch trials from experiment 1, which also equally consisted of two and four turns.

3.4.2.2 Procedure

Like the previously presented studies, experiment 4 was an acceptability rating study conducted over the Internet using LimeSurvey (Limesurvey GmbH 2023). The experiment involved 72 native German speakers (age 18–40) who were again recruited from the crowdsourcing platform Clickworker (Clickworker 2022) and who had not participated in any of my previous studies on topic drop. They received a compensation of €4.00. Their task was again to read the dialogues and to rate the naturalness of the last utterance using a 7-point Likert scale (1 = completely unnatural, 7 = completely natural). I distributed the materials across eight lists so that each subject saw each critical stimulus exactly once and in one condition. The items and the fillers were mixed and presented as WhatsApp-like dialogues in individual pseudo-randomized order so that no two stimuli of the same type immediately followed each other.

3.4.2.3 Results

I excluded the data from 14 participants who had exceeded the threshold of having rated four or more of the eight catch trials with 6 or 7. Table 3.10 shows the mean ratings and standard deviations per condition calculated on the data from the remaining 58 participants. In Figure 3.6, this information is also plotted.
While the ratings for full forms and utterances with topic drop were comparable when no conjunction was present, topic drop was degraded when it was preceded

Table 3.10: Mean ratings and standard deviations per condition for experiment 4

Completeness	Presence of Conjunction	Subject Gap	Mean rating	Standard deviation
Full form	Absent	Possible	6.03	1.13
Topic drop	Absent	Possible	6.01	1.20
Full form	Present	Possible	5.92	1.20
Topic drop	Present	Possible	5.56	1.54
Full form	Absent	Impossible	6.12	1.12
Topic drop	Absent	Impossible	6.06	1.24
Full form	Present	Impossible	5.90	1.11
Topic drop	Present	Impossible	5.38	1.72

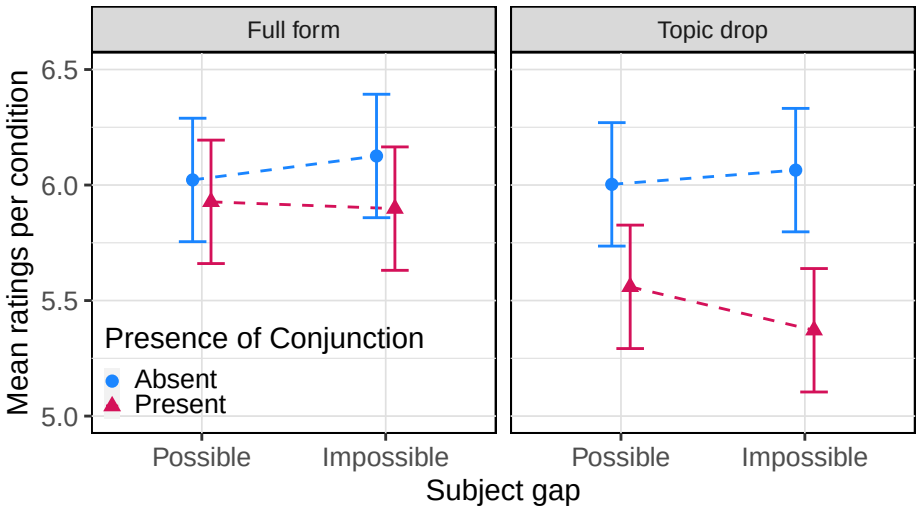


Figure 3.6: Mean ratings and 95% confidence intervals per condition for experiment 4

by a conjunction. From visual inspection, it seems that this effect is even more pronounced when a subject gap reading was impossible.

I analyzed the rating data with CLMMs from the package ordinal (Christensen 2019) in R, as described in Section 3.1.6.3. The full model contained the binary predictors Completeness, Presence of Conjunction, and Subject Gap coded using deviation coding (full form, conjunction absent, and subject gap possible were coded as 0.5, the other levels as −0.5 respectively), their three-way interaction and all two-way interactions between them, as well as the numeric scaled and centered Position of the trial in the experiment and all two-way interactions of the factors with Position. As random effects, I included random intercepts for both subjects and items and by-item and by-subject random slopes for Completeness, Presence of Conjunction, Subject Gap, and their interactions which each other, as well as for Position.[77] The fixed effects in the final model with flexible thresholds are shown in Table 3.11.

Table 3.11: Fixed effects in the final CLMM of experiment 4

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	0.32	0.22	2.02	> 0.15	
Presence of Conjunction	1.00	0.28	10.98	< 0.001	**
Position	0.08	0.08	1.00	> 0.31	
Completeness × Presence of Conjunction	−0.87	0.35	6.02	< 0.05	*
Completeness × Position	−0.27	0.13	4.63	< 0.05	*
Presence of Conjunction × Position	−0.32	0.13	6.45	< 0.05	*

There was a significant main effect of Presence of Conjunction ($\chi^2(1) = 10.98$, $p < 0.001$). It indicates that utterances without preceding conjunctions were preferred over utterances with conjunctions. The significant interaction between Completeness and Presence of Conjunction ($\chi^2(1) = 6.02$, $p < 0.05$) shows that utterances with topic drop and preceding conjunctions were particularly degraded. The Position of the trial in the experiment interacted significantly both

[77] The formula of the full model was as follows: `Ratings ~ Completeness : Presence of Conjunction : Subject Gap + (Completeness + Presence of Conjunction + Subject Gap + Position)^2 + (1 + Completeness + Presence of Conjunction + Subject Gap)^2 + Completeness : Presence of Conjunction : Subject Gap + Position | Items) + (1 + Completeness + Presence of Conjunction + Subject Gap)^2 + Completeness : Presence of Conjunction : Subject Gap + Position | Subjects)`.

with Completeness ($\chi^2 = 4.63$, $p < 0.05$) and with Presence of Conjunction ($\chi^2(1) = 6.45$, $p < 0.05$). Throughout the experiment, participants gave better ratings to both stimuli with topic drop and to utterances with preceding conjunctions, which points toward a habituation effect. The predictor Subject Gap was not involved in any significant effect.

A closer inspection of the data revealed large differences between token sets, in particular between the different conjunctions. In particular, the ratings for *denn* (parordinating 'because') deviated strongly from those for *und* ('and') and *aber* ('but'). Figure 3.7 shows the mean ratings and 95% confidence intervals per conjunction and indicates that the ratings for *denn* (parordinating 'because') strongly deviate from the ratings for *und* ('and') and *aber* ('but'). While there was little to no difference between conditions for *und* and *aber*, for *denn* already the full forms with preceding conjunction were rated worse but even more so the topic drop conditions.

This motivated me to conduct a post hoc analysis in which I split the data into three sets, an *und* ('and')-set, an *aber* ('but')-set, and a *denn* (parordinating 'because')-set and repeated the analysis for each conjunction separately. For *aber* ('but') and *denn* (parordinating 'because'), I had to reduce the random effects structure by removing the by-items and by-subjects random slopes for the three-way interaction between Completeness, Presence of Conjunction, and Subject Gap for the model to converge. Tables 3.12, 3.13, and 3.14 show the fixed effects in the final models of each subanalysis. The results confirm the visual inspection of Figure 3.7.

In the subanalysis of *und* ('and'), there was only a significant interaction between Presence of Conjunction and Position ($\chi^2(1) = 11.32$, $p < 0.001$). It indicates that utterances with a preceding *und* received better ratings when they occurred later in the experiment.

Table 3.12: Fixed effects in the final CLMM of the *und* ('and')-subanalysis of experiment 4

Fixed effect	Est.	SE	χ^2	p-value	
Presence of Conjunction	0.138	0.43	0.1096	> 0.74	
Position	0.001	0.21	0.0003	> 0.99	
Presence of Conjunction × Position	−1.015	0.35	11.3200	< 0.001	***

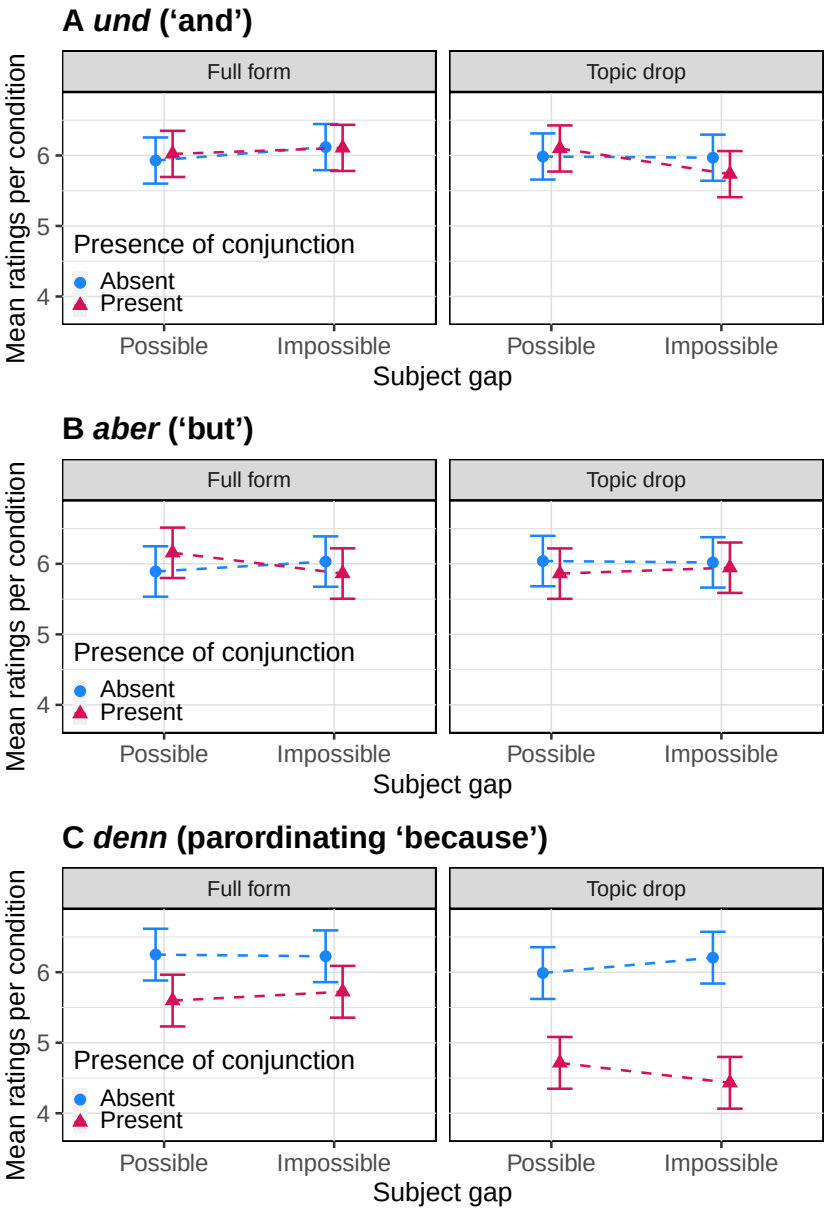


Figure 3.7: Mean ratings and 95% confidence intervals per condition for experiment 4 subdivided by conjunction

For *aber* ('but'), there was a significant Presence Of Conjunction × Subject Gap interaction ($\chi^2(1) = 3.92$, $p < 0.05$). Utterances (both full forms and utterances with topic drop) that did not allow for a subject gap reading and contained a conjunction were rated as particularly bad. Since this effect is independent of Completeness, it is, however, of no theoretical interest in this context.

Table 3.13: Fixed effects in the final CLMM of the *aber* ('but')-subanalysis of experiment 4

Fixed effect	Est.	SE	χ^2	p-value	
Presence of Conjunction	0.42	0.38	1.27	> 0.25	
Subject Gap	0.42	0.34	1.64	> 0.20	
Presence of Conjunction × Subject Gap	−1.16	0.61	3.92	< 0.05	*

In the subanalysis of *denn* (parordinating 'because'), there were significant main effects of Completeness ($\chi^2(1) = 8.6$, $p < 0.01$) and Presence of Conjunction ($\chi^2(1) = 17.4$, $p < 0.001$). They indicate that full forms were generally preferred over utterances with topic drop and that utterances with *denn* received worse ratings than utterances without. A significant interaction between both predictors ($\chi^2(1) = 4.31$, $p < 0.05$) suggests that the latter effect was particularly pronounced for topic drop. Utterances with topic drop and realized *denn* received particularly bad ratings. A significant Completeness × Position interaction ($\chi^2(1) = 6.73$, $p < 0.01$) shows that the ratings for topic drop improved throughout the experiment indicating a slight habituation effect for the *denn* materials. Subject Gap had no effect.

Table 3.14: Fixed effects in the final CLMM of the *denn* (parordinating 'because')-subanalysis of experiment 4

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.40	0.43	8.60	< 0.01	**
Presence of Conjunction	2.88	0.54	17.40	< 0.001	***
Position	0.36	0.22	2.53	> 0.11	
Completeness × Presence of Conjunction	−1.82	0.80	4.31	< 0.05	*
Completeness × Position	−0.73	0.30	6.73	< 0.01	**

3.4.2.4 Discussion

In experiment 4, I tested whether topic drop is restricted to a strictly sentence-initial position or whether elements, namely conjunctions, can be placed before topic drop. The results suggest that generally elements can precede topic drop, i.e., that it is not restricted to the very first position. The naturalness of utterances with *und* ('and') or *aber* ('but') followed by topic drop was comparable to the topic drop conditions without conjunctions and the corresponding full forms.

The fact that there were no effects of a possible subject gap reading on topic drop indicates that topic drop after *und* ('and') and *aber* ('but') is possible, irrespective of whether such a reading is licensed through a matching context sentence or not. This suggests that the acceptability of topic drop after conjunctions does not depend on the readers reinterpreting such structures as (asymmetric) coordination structures with subject gaps across sentence boundaries.[78]

The results for *denn* (parordinating 'because') strongly deviated from those for *und* ('and') and *aber* ('but') in that (i) utterances with *denn* were degraded across conditions and (ii) that in particular topic drop with *denn* was rated as less natural. A possible reason for (i) could be that the conjunction *denn* is too formal for instant messages. In spoken communication, *denn* has been largely replaced by *weil* (subordinating 'because') (Selting 1999, Wegener 2000) and text messages can be considered a conceptually spoken form of communication (Koch & Oesterreicher 1985). From (ii), we can conclude that while *und* and *aber* can precede topic drop, *denn* cannot do so equally well. It is obvious to attribute this difference to the difference between the conjunctions. As a starting point, we can note that they differ with respect to their semantics and to the discourse relations they encode. The *Handbuch der deutschen Konnektoren 2* ('The handbook of German connectives 2') classifies *und* as additive (Breindl et al. 2014: 1211), *aber* as adversative or concessive (Breindl et al. 2014: 1173), and *denn* as causal, marking the antecedent of the causal relation (Breindl et al. 2014: 1187). In the taxonomy of Kehler (2000, 2002), *und* usually either encodes a resemblance relation, specifically a parallel one, or a contiguity relation, specifically a narration. *Aber* encodes a contrast, a

[78] Recall that a subject gap interpretation would require analyzing the corresponding structure as asymmetric coordination and that according to the literature subject gaps in asymmetric coordination only function with *und* ('and'). They are impossible with *denn* (parordinating 'because') and presumably require the conjunction to be positioned postverbally with *aber* ('but'), which was not the case in my stimuli. However, even for only the token sets with *und*, which should allow for subject gaps in asymmetric coordination in any case, there was no effect of the predictor Subject Gap. That is, the conditions that permit a subject gap interpretation and the conditions that do not, according to the literature, did not differ significantly in their acceptability.

different subtype of a resemblance relation, whereas *denn* encodes an explanation, a subtype of a cause-effect relation (see Kehler 2000: 540–545, Kehler 2002: 15–23).

A first hypothesis derived from this observation is that the conjunctions differ in how likely they make the subject of the following sentence. In anticipation of the second part of this book, we can assume that topic drop is more felicitous if the omitted constituent is more predictable in context. If a certain conjunction makes *ich* a lot more likely, e.g., *and* as additive or parallel connective, this may boost the acceptability/appropriateness of topic drop. Coming back to *denn*, we can specify the hypothesis in the following way: Topic drop following *denn* received degraded ratings in the experiment because the subject *ich* is less likely to follow *denn* than *und* and *aber*. In a cloze study, where I collected continuations after the context sentence with or without the conjunction given, I found no evidence to support this hypothesis.[79] Whether participants produced a continuation with a 1st person singular subject differed more strongly between token sets than between conjunctions.

A second hypothesis can be built based on the semantic and syntactic properties of *denn*, which is considered to be a 'syntactic loner' (Pasch et al. 2003, Reich & Reis 2013, Breindl et al. 2014, Breindl 2017, see also Footnote 71). Reich & Reis (2013: 565) and Breindl et al. (2014: 866–867) state that a special property of *denn* is its asymmetric behavior. Semantically or discourse-structurally, *denn* functions as a subordinating connective because it encodes an explanation relation between the first and the second conjunct, with the second conjunct as explanation being subordinate to the first. Syntactically, however, *denn* is a "parordinating" connective (Höhle 1986: 329), that neither coordinates nor subordinates its conjuncts (Pasch et al. 2003: 590). Breindl et al. (2014: 870) state that the information expressed in the conjunct introduced by *denn* is always asserted

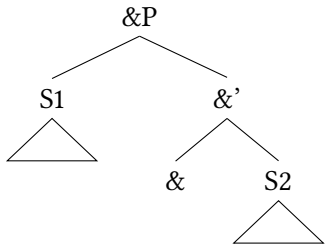
[79] In a 2 × 2 design (Presence of Conjunction × Subject Gap), 60 participants provided continuations for the 24 items of experiment 4 and 16 fillers. The items were cut off after the context sentence, as shown in (i), so that participants produced the target sentence. It was varied between conditions whether the conjunction was already determined as the first word of this sentence or not. I considered both the proportion of produced continuations with *ich* ('I') as the subject and the proportion of continuations with *ich* ('I') in the prefield. For both measures, the variation between token sets was higher than the variation between conjunctions.

(i) a. A: Was habt ihr heute Abend geplant?
b. B: Wir wollen uns den neuen Matrixfilm im Kino anschauen
c. B: Die ersten drei Filme (mag ich total | gefallen mir total gut)
d. B: (Und) _____

independently and therefore always in focus so that there is a tendency to introduce new information through a *denn* clause. The fact that *denn* seems to block topic drop may be a consequence of its hybrid status. Given the semantic subordination of the second conjunct, *denn* may indicate that the clause it introduces is not fully autonomous, therefore, its prefield is not available for ellipsis. To verify this speculation, further research on *denn* is needed.

In sum, experiment 4 indicates that topic drop is not restricted to the very first element of an utterance. Conjunctions like *und* and *aber* can precede it. The corresponding structures are not "hidden" subject gap structures but genuine topic drop structures. As discussed before the experiment, it is unclear whether the linear precedence translates itself into a syntactic superordination and whether this relation holds *within* a clause, which may potentially affect Rizzi's (1994) *specifier of the root*-account based on c-command or Freywald's (2020) highest clause position account. If we assume the coordination structure visualized in (31), repeated here as (41), then the covert constituent in [Spec, CP] of S2 would be c-commanded by the head &.

(41)



(Reich forthcoming)

At first glance, this is not compatible with Rizzi's c-command approach because the covert constituent in [Spec, CP] of S2 would be c-commanded sentence-internally and, therefore, should not be available for topic drop. There are two ways to counter this: (i) As discussed above, there is the assumption that the conjunction in & is not a genuine part of the following clause S2. Therefore, it could be argued that Rizzi's principle is not violated because the c-command relation between the conjunction in & and the covert constituent in [Spec, CP] of S2 does not hold sentence-internally but sentence-externally. (ii) Although there is a c-command relation between the conjunction in & and the covert constituent in [Spec, CP] of S2, the conjunction in & cannot function as an identifier for the covert constituent as it is a functional head (Johannessen 1998), which cannot provide an antecedent for a null pronoun. Thus, Rizzi's principle is not violated

because the covert constituent in [Spec, CP] of S2 may be c-commanded but not by a potential identifier.

Concerning Freywald's (2020) proposal, we have to come back to the fact mentioned already in Section 3.3.1 that she equates the concepts of autonomous clause and root clause. However, it is, in particular, the difference between both concepts, which Reich & Reis (2013) point out, that matters for the analysis of S2 in (41). According to Reich & Reis (2013: 542), the conjuncts of an autonomous coordination are considered to be root clauses but not autonomous clauses. This means that Freywald's account needs to be specified to allow for topic drop in S2. It must be determined that the relevant category is the root clause and not the autonomous clause, i.e., that topic drop is enabled in the highest [Spec, CP] of a root clause. This way, Freywald's approach becomes more general because root clauses are a superset of autonomous clauses. Consequently, topic drop is permitted in more configurations.

3.5 Summary: prefield restriction

Before I turn to the final section of this chapter and address the question of how a suitable syntactic analysis may look in light of what we learned so far about the prefield restriction of topic drop, I summarize these findings and specify the prefield restriction.

In this chapter, I investigated in detail the most prominent licensing condition of topic drop: its positional restriction to the prefield position. As a first step, I looked at the role of the information structural category topic for topic drop. I showed that the prefield in German is neither the only position in which topic expressions can be placed nor that all elements occurring in the prefield are topics. Based on theoretical arguments and experimental data, I argued that the topicality of the omitted constituent is neither a (strictly) sufficient nor a necessary condition for topic drop. Contrastive topics and topics that cannot be recovered question at least strict sufficiency indicating that there can be topic expressions in the prefield that cannot be omitted. More importantly, the fact that expletive subjects, i.e., non-referential and, thus, non-topical elements, can be omitted calls necessity into question. Apparently, elements can be targeted by topic drop without being topic expressions, which shows that the scope of topic drop goes beyond topics. If one does not want to abandon the relevance of the notion of topic for topic drop altogether, one could argue that topics are at least the prototypical candidates for omission since they are usually easy to recover

(see the discussions on recoverability in Chapter 4) and predictable in context (see Chapter 9). This means that the topic restriction, which is often regarded as a licensing condition in the literature, is only a favoring factor for topic drop. I investigate the degree to which topicality is indeed such a factor in Chapter 9 in the second part of this book.

In the second step, I showed experimentally that argument omissions in the middle field are degraded, which supports the validity of a structural prefield restriction of topic drop. With two further experiments, I refined the properties of this restriction. First, I showed that topic drop is not possible in every prefield position but only in a prefield position that is at least not c-commanded sentence-internally by a potential identifier, as proposed by Rizzi (1994), or even only in the highest prefield of an autonomous or root clause, according to Freywald (2020). Second, I was able to provide evidence that topic drop is not restricted to the very first element of a clause but that conjunctions like *und* ('and') and *aber* ('but') can precede it. I argued that this result does not question Rizzi's (1994) c-command-approach because the conjunction is not really part of the second utterance, nor can it be an antecedent for the covert constituent because it is a functional head. It makes it necessary, however, to specify Freywald's (2020) proposal in such a way that topic drop is restricted to the highest prefield of a root clause and not of an autonomous clause. In this book, I do not distinguish between Rizzi's (1994) and Freywald's (2020) accounts but include both as options in the following specified definition of topic drop:

Definition 3 *Topic drop is the omission of a constituent from the prefield of declarative verb-second (V2) clauses. This prefield must at least not be c-commanded sentence-internally by a potential identifier or it must even be the highest prefield of a root clause.*

3.6 Toward a syntactic analysis of topic drop

In light of this sharpened prefield restriction, I turn to a potential syntactic analysis of topic drop. Such an analysis faces the challenge of explaining the omission of both referential and non-referential constituents, as well as of subjects, objects, and adverbials, and of adequately modeling the prefield restriction as the central syntactic licensing condition. I discuss the three main syntactic approaches to topic drop, shown in Figure 3.8: null operator analysis, *pro*-analysis, and PF-deletion analysis.

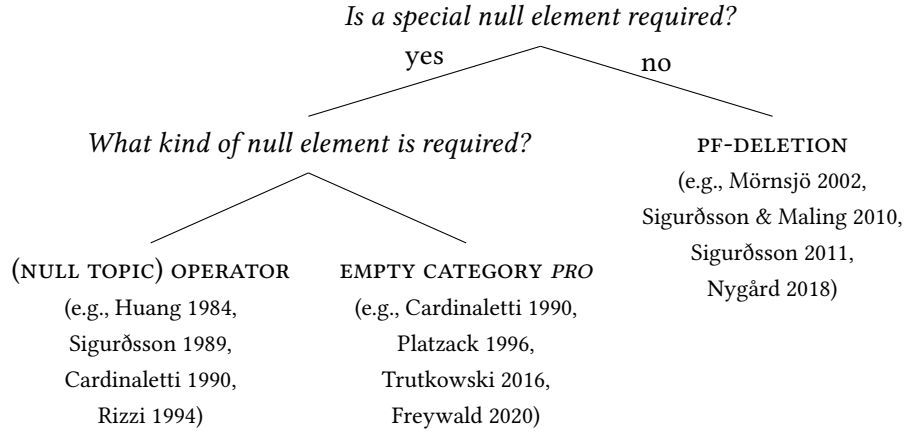


Figure 3.8: Schematic overview of the three main syntactic analyses of topic drop proposed in the literature and of their central representatives

I present their central representatives, as well as their respective advantages and disadvantages.[80] As Figure 3.8 illustrates, the null operator analysis, the analysis as null pronoun *pro*, and the analysis as PF-deletion differ, first, in whether they require a specific null element and, if so, second, in what type of null element is assumed.

3.6.1 Operator analysis

The first generative analysis proposed for topic drop that I am aware of is by Huang (1984). He suggests to analyze topic drop of both referential subjects and objects by means of a variable that is $\bar{A}$-bound by a null topic (or zero topic) operator in [Spec, CP], which in turn is linked to the discourse (Huang 1984: 543; 548; see also Sigurðsson 1989 and Haegeman 1990 for similar approaches).[81] It

[80] See Nygård (2018: 27–44) for a more detailed overview covering among others the suggestions by Huang (1984), Sigurðsson (1989), Cardinaletti (1990), Rizzi (1994), Haegeman (1990), Mörnsjö (2002), Sigurðsson & Maling (2010), and Sigurðsson (2011). I focus exclusively on generative approaches because they dominate the research. But cf. Helmer (2017a), who discusses the possibility of analyzing topic drop with construction grammar.

[81] Cardinaletti (1990) and, with slight modifications, also Rizzi (1994) adopt the operator analysis only for objects. They argue that it follows from the alleged restriction of object topic drop in German to the 3rd person (see Section 8.1 for a discussion) that the corresponding operator has intrinsic 3rd person properties (Cardinaletti 1990: 79, Rizzi 1994: 161). For subject topic drop, they both propose a different analysis. Cardinaletti (1990) advocates an analysis involving a

already hints at the main problem of the operator account that the operator is usually conceptualized as a null or zero topic operator (Huang 1984, Sigurðsson 1989, Cardinaletti 1990, Haegeman 1990). The concept seems to be closely tied to topicality and, therefore, cannot straightforwardly account for the omission of non-topics, in particular of non-referential expletives.

The operator analysis of topic drop is similar to (i) the analysis by Büring & Hartmann (1998) for subject gaps in asymmetric coordinations and (ii) Reich's (2017) proposal for null copulas and null articles.

3.6.1.1 Büring & Hartmann (1998: 176–177)

Büring & Hartmann (1998: 176–177) assume for cases of asymmetric coordination like (42) that there is an empty element in the subject position of the second conjunct [Spec, IP] that is bound and identified by an empty operator in [Spec, CP] and that this operator, in turn, is bound by the subject of the first clause.[82] In the appendix of their paper, they provide a semantic analysis of this idea (Büring & Hartmann 1998: 191–198).

(42) *Gestern ging ich aus und* Δ *traf Olaf Thon.*
yesterday went I out and I met Olaf Thon
'Yesterday, I went out and (I) met Olaf Thon.' (Büring & Hartmann 1998: 176)

Note that Büring & Hartmann's (1998) variant of the operator analysis also does not cover the omission of non-referential expletive subjects. For subject gaps, it probably does not have to because expletives cannot occur in subject gap configurations according to Reich (2009a: 53). For topic drop, on the other hand, it would be desirable, as mentioned above, if the omission of expletives were also

null pronoun *pro* (see below). Rizzi (1994) argues that in V2 languages [Spec, CP] can behave as an *A*- instead of an *Ā*-position if the subject is moved there, which enables the null element in [Spec, CP] to bind an NP-trace in [Spec, IP] (Rizzi 1994: 161). To prevent this null element, which is not identified sentence-internally, from violating the ECP, Rizzi (1994) proposes a relaxed ECP identification requirement that exempts the *specifier of the root*, [Spec, CP] in V2 languages (Rizzi 1994: 162), thereby providing a theory-driven justification for the prefield restriction of topic drop.

[82] It is worth noting that Hartmann (1994: 12) analyzes null subjects in coordinations as *pro* bound by an empty operator, following a proposal by Cinque (1990), and, earlier, Chomsky (1986), for parasitic gaps and extractions out of NP islands. She argues that the requirements imposed on the empty category involved in subject gaps are best fulfilled by *pro* since it bears its own theta role and can appear in any grammatical person (Hartmann 1994: 12–13). This analysis allows for combining operator and *pro* (see below), but it does not solve the problem that non-referential expletives are not accounted for.

explained by the syntactic account. Büring & Hartmann's (1998) approach could be adapted, though, for topic drop of referential constituents, with the following modifications: First, to account for all attested cases of topic drop, we must assume that the empty element can function not only as a subject but also as an object or an adjunct.[83] Second, the empty operator involved in topic drop constructions semantically works like a lambda abstraction, turning an open proposition of type t into a property of type $\langle e,t\rangle$, provided the bound element is of type e.[84] To become an interpretable proposition t, this unsaturated property $\langle e,t\rangle$ needs an argument of the object type e, on which it can operate. Since the utterance itself does not supply such an argument, it needs to come from the previous linguistic or extralinguistic context, as in example (43).

(43) a. 'What about Tino?'
b. Δ *Schläft.*
he sleeps.
'(He) is sleeping.'

Thus, the discourse orientation of [Spec, CP] that has often simply been postulated in previous research can be discourse-pragmatically motivated. For the utterance with topic drop to denote a proposition, it needs to access an entity from the discourse. This analysis has the severe drawback, though, that it leads semantically to the assumption that declaratives cannot only be of type t but also of type $\langle e,t\rangle$. That is, utterances with topic drop like (43b) are semantically properties of type $\langle e,t\rangle$, but they behave as if they were declaratives of type t in having a truth value ((43b) is true if Tino sleeps and false otherwise). While from a syntactic point of view, the operator analysis seems quite elegant, at least for referential constituents, it is therefore not very attractive from a semantic point of view.

3.6.1.2 Reich (2017)

Reich (2017) also proposes a null operator analysis of null copulas and null articles in newspaper headlines. He refers to the typological distinction between discourse-oriented and sentence-oriented languages (Tsao 1977, Huang 1984) and the property that, according to Huang (1984: 545), only discourse-oriented languages such as Chinese allow for a null or zero topic operator that binds a variable, while sentence-oriented languages such as English do not (Reich 2017: 192).

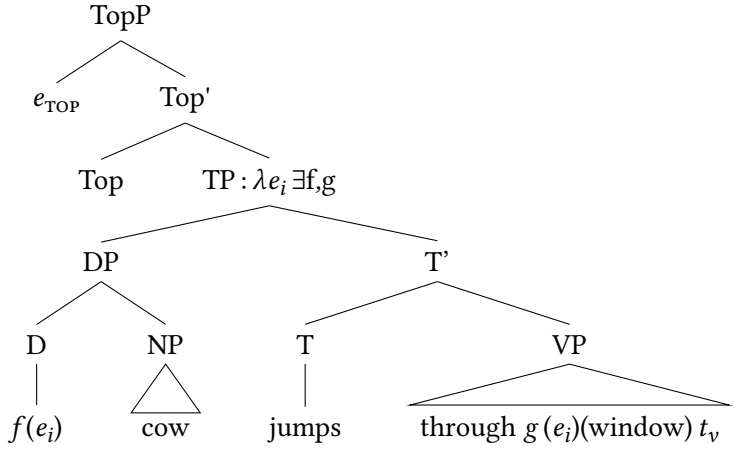
[83] Unlike the proposal by Cardinaletti (1990), the operator for the objects does not need to be restricted to the 3rd person. I argue that the apparent impossibility to drop 1st and 2nd person objects can be accounted for mainly by pragmatics (see Section 8.1).

[84] I thank Ingo Reich for the help with this semantic conceptualization.

Reich (2017: 193) argues that German allows for a switch from "normal" sentence orientation to discourse orientation in certain registers (such as headlines), thus enabling the phenomena of null copulas and null articles, which he investigates. He assumes in his analysis, exemplified for a null article in (44), that null copulas and null articles are represented by variables bound by a null operator located in the specifier of a TopP, which forms the topmost phrase of the headline. This null operator is a so-called *event topic*, *e*-topic for short, which is licensed in headlines because they are characterized by the fact that they report (newsworthy) events (Reich 2017: 193). The event variable e_i is existentially bound by λ-abstraction. Reich (2017: 194) argues that sentence-oriented and discourse-oriented registers differ in that "[i]n *sentence-oriented* registers covert variables are existentially quantified over at (some) propositional level of semantic interpretation. In *discourse-oriented* registers they are systematically bound by λ-abstraction" (original emphasis). It seems possible to extend this analysis also to topic drop by arguing that it is a further phenomenon that gets enabled by the switch to discourse orientation. It must be noted, however, that null copulas and null articles are much more restricted to certain text types, in particular to headlines, than topic drop. Topic drop is particularly common in text messages and spoken language, but it is not entirely excluded in many other text types and registers (see Section 2.1.5). For this reason, it is questionable whether one can argue for a shift from sentence- to discourse-orientation in every case and for the existence of the postulated *e*-topic.

(44) a. *Kuh springt durch Fenster*
cow jumps through window
'Cow jumps through window'

b.



(Reich 2017: 197)

In summary, an operator analysis of topic drop seems syntactically quite plausible and can motivate the prefield restriction discourse-pragmatically. However, it has the clear disadvantages that, at least in its present form, it can only explain referential topic drop and that a precise semantic analysis would lead to a type problem of declaratives.

3.6.2 *pro*-analysis

While I distinguished topic drop from *pro*-drop in Section 2.2.1, there are syntactic accounts of topic drop that analyze it in a very similar way. They assume an empty pronoun *pro* that is part of the lexicon, similar to the one that is assumed for *pro*-drop languages (e.g., Cardinaletti 1990, Platzack 1996, Trutkowski 2016, Freywald 2020). The traditional view is that Standard German only allows for a non-referential expletive 3rd person *pro* in subject function and in the nominative case (Sternefeld 2015: 170–171), such as (45) but, unlike *pro*-drop languages such as Italian (see Section 2.2.1), not for a referential *pro* with flexible person and case features and a theta role.

(45) *weil* pro *getanzt wird*
because pro danced gets
'because there is dancing' (Sternefeld 2015: 170)

To explain topic drop employing a *pro*, we have to assume either a second *pro* for German or that the properties of the German *pro* are more flexible than usually described.[85] Additionally, it is unclear whether this *pro* is generally part of the lexicon or only in special text types or registers (cf. the discussion of discourse-oriented vs. sentence-oriented registers in Section 3.6.1).

If we additionally assume a *pro* that can account for any occurrence of topic drop sketched in this book so far, it needs to be able to function at least as a referential and non-referential subject, an object, an adverbial, and a part of prepositional adverbs with *da*. The last two cases in particular could be problematic since Platzack (1996: 101) assumes that *pro* is an NP,[86] but a *pro* restricted to nominal

[85] In German dialects such as Bavarian, we find cases of referential null subjects, which are often analyzed as *pro* drop (see Section 2.2.3). Thus, at least for the dialects, a more flexible *pro* is to be assumed anyway.

[86] Platzack (1996: 101) considers it an advantage of the *pro*-analysis that it ensures that only nominal elements can be targeted by topic drop since *pro* can only represent such elements. What he considers an advantage, however, is criticized by Nygård (2018: 38), who points out, with reference to Mörnsjö (2002), that in Swedish and Norwegian also non-nominal adverbial elements can be omitted. The same is true for German, as discussed in Section 2.1.3.

elements does not apply to all cases of topic drop.[87]

In more recent literature, it is assumed that the *pro* that represents topic drop is underspecified with respect to its person, number, and gender phi-features and receives the values from a matching corresponding discourse referent (Freywald 2020: 172). Trutkowski (2016: 172) even assumes "an internally structured *pro*, i.e., an empty category that is of the category *pro* and that contains the category *pro*", which is sufficiently underspecified to account for strict and sloppy readings and mismatches in syntactic identity (Trutkowski 2016: 173). This conceptualization of *pro* does not apply to topic drop of expletives, which is disregarded by Freywald (2020: 150) and only marginally considered by Trutkowski (2016: 120–121) (see also Section 3.1.5). The expletive *es* is semantically empty, so it does not need a discourse referent to receive phi-features from. Therefore, if we wanted to apply the *pro*-analysis also to topic drop of non-referential constituents, we would need to assume two different *pro*s, one for referential and one for non-referential constituents. The latter, in turn, has similarities to the expletive *pro* assumed for standard German anyway. Unlike the latter, however, it is restricted to the prefield in German, like the referential *pro*. This results in the need for at least three different *pro* types in German.

To account for the prefield restriction of topic drop, Platzack (1996: 94) stipulates a feature A. He argues that a marker [Repel A] can be assigned any phrase of a clause to force the correspondingly marked phrase to "move to the first available node outside the highest A-position." According to his approach, the *pro* representing the covert constituent is always marked [Repel A] and, thus, has to move to this position before spell out (Platzack 1996: 101). The prefield restriction is hard-coded in the lexicon, so to speak. In short, Platzack's (1996) account requires an additional feature that, stated in terms of Minimalism, can only be checked in [Spec, CP]. However, his solution does not explain the prefield restriction, i.e., there is no explanation for why this feature should exist.[88]

Trutkowski (2016) and Freywald (2020), on the other hand, assume that the prefield restriction of *pro* results from it being only discourse accessible in this

[87] In order to clarify this issue, it would be necessary to examine in detail, first, whether the *pro* is in fact an NP, or whether Chomsky's original definition of a "pure pronominal", an empty category with the features [+pronominal, -anaphor] (Chomsky 1982: 81–82), can be extended to other proforms that act as substitutes for adverbials, prepositional phrases, or prepositional adverbs. Second, it needs to be clarified whether it is reasonable to assume that these categories have a nominal or at least pronominal part. Cf. on the last point the remarks of Oppenrieder (1991: 162, footnote 6), who contradicts Wunderlich's (1984: 88) claim according to which the pronominal part of prepositional adverbs is an NP.

[88] See also Sternefeld (2009: 762–763) for a critical discussion of topicalization as a feature-driven movement.

position and from the restriction of topic drop to topics that they assume (see Section 3.1.5 for counter-evidence). Accordingly, Freywald (2020: 170–172) states that the "topic drop *pro*" is restricted to the specifier of a ConTopP reserved for continuous topics, where it receives its corresponding features from discourse, as already outlined above. Similarly, Trutkowski (2016: 176) formulates an identification condition for topic drop according to which "[o]nly in the spec of the root-position an empty element is available for the identification by a discourse antecedent." She argues that the prefield restriction of topic drop is necessary to ensure that the antecedent can be optimally accessed and that this is particularly important "because a topic drop gap does not constitute a blind copy of its antecedent but can accommodate semantic and syntactic variation (different interpretations, case mismatches)" (Trutkowski 2016: 19). However, this argumentation is weakened by the fact that other (more "classic") types of ellipsis such as sluicing also allow for mismatches (Kroll & Rudin 2017, Anand et al. 2021) but are not positionally restricted to the left edge of an utterance. Moreover, Trutkowski (2016) and Freywald (2020) can only explain the prefield restriction of referential and topical constituents. Even in these cases, the question remains open as to why it should be precisely and exclusively the root or [Spec, ContTopP] position that enables discourse accessibility. Overt pronouns, for instance, can also occur in the middle field and hearers can still establish the corresponding discourse reference (as shown in the trivial example (46)).

(46) a. *Was ist mit Tino$_i$?*
what is with Tino
'What about Tino$_i$?'
b. *Am Montag hat er$_i$ gesagt, dass er$_i$ kommt.*
on Monday has he said that he comes
'On Monday he$_i$ said that he$_i$ was coming.'

In summary, the *pro*-analysis of topic drop follows similar existing analyses of *pro*-drop languages by assuming a null pronoun *pro* in the lexicon representing the omitted constituent. However, since usually only a non-referential *pro* is assumed for Standard German, either another *pro* must be postulated in the lexicon or the properties of the German *pro* must be adjusted. Moreover, if we were to adopt Platzack's (1996) characterization of the *pro* as an NP, the consequence would be that not all cases of topic drop could be explained by a *pro*. Finally, the prefield restriction can only be accounted for by postulated features or a discourse accessibility restriction that remains vague. Thus, several additional assumptions are required for the *pro*-analysis and still its explanatory force remains unsatisfactory.

3.6.3 PF-deletion analysis

The last major class of syntactic accounts to topic drop can be subsumed under the term *PF-deletion*. Instead of postulating a special null element, those accounts assume that utterances with topic drop only lack the spell-out of the prefield constituent while being otherwise (mainly) identical to the full form (Mörnsjö 2002, Sigurðsson & Maling 2010, Sigurðsson 2011, Nygård 2018): "[T]he semantic and grammatical features of the phonetically non-realized element are present in Spec-CP, in order to feed the interpretation process. Lacking phonological features, this element cannot be spelled out. Consequently, the syntactic licensing of a phonetically non-realized element is identical to its visible alternative" (Mörnsjö 2002: 133–134).[89] This approach nicely captures Reis's (2000) claim that utterances with topic drop are functionally parallel to their overt counterparts. It furthermore can straightforwardly account for the cases of omitted non-referential expletives that cannot be (easily) explained by the operator or *pro*-analyses.

Mörnsjö (2002) does not develop a detailed account of PF-deletion but suggests that it could be implemented using Distributed Morphology (Halle & Marantz 1993) and its concept of the late insertion of phonological features (Mörnsjö 2002: 133). Nygård (2018) adopts Mörnsjö's (2002) account of phonological non-realization, assuming that overt and covert elements are identical with respect to the syntax (Nygård 2018: 47) but embeds it into a generative exoskeletal frame-based syntactic model.[90]

Sigurðsson & Maling (2010) propose a sophisticated cross-linguistic approach to null arguments (see also Sigurðsson 2011), arguing that any apparent differences between different types of null arguments result exclusively from restrictions at PF (Sigurðsson & Maling 2010: 66). Their approach also has similarities to the *pro* accounts discussed in Section 3.6.2 because they consider null arguments to be pronouns and, in turn, pronominal arguments to be bundles of morpho-syntactic features for which spell-out at PF is principally optional (Sigurðsson & Maling 2010: 68). In their *Context-Linking Generalization* (Sigurðsson & Maling 2010: 61) (*C/Edge-Linking Generalization* in Sigurðsson (2011: 282)), they argue

[89] Note that Sigurðsson & Maling (2010: 66) qualify this for their approach stating that "[o]vert pronouns tend to be more specific or 'bigger' than null-arguments in the sense that they express some additional properties like Focus or Shifted Topic, not present in corresponding null-argument constructions."

[90] The idea of an exoskeletal frame-based syntax is that, unlike in Minimalism, the syntactic structure consists of abstract sentence frames which are "generated independently of lexical items" (Nygård 2018: 78). This means that the lexical insertion takes place after the syntax and that the interpretation of the inserted lexical items with respect to their theta roles is determined by this frame (see Nygård's (2018) chapters 3 and 4, for details).

that any overt or covert pronominal argument must positively match a context-linking C-feature, which comprise at least topic, speaker, or hearer features located in the C-domain (Sigurðsson & Maling 2010: 61).[91] To account for the prefield restriction of Germanic topic drop, they propose a clause-internal operative *Empty Left Edge Condition* (Sigurðsson & Maling 2010: 62): "The left edge of a clause containing a silent referential argument must be phonetically empty." That is, topic drop is only possible if [Spec, CP] is phonetically empty. This is explained as an intervention effect on feature matching at PF. Overt material in [Spec, CP] blocks the null argument to raise and match the context-linking feature in the C-domain, as illustrated in (47).

However, it remains unclear why an occupied [Spec, CP] blocks the context-linking of null arguments, given that overt pronouns can be easily context-linked in the middle field. Even if one explains this restriction to be operative at PF rather than in the syntactic system, still the origin of the PF rule remains unclear. Sigurðsson (2011: 294) states that it "seems more promising to assume that it is movement (internal Merge) of more than one constituent across the finite verb in C that is blocked (for reasons that remain to be explicated [...])." Additionally, the account by Sigurðsson & Maling (2010) in its outlined form seems to be restricted to referential pronominal arguments, which can be context-linked at all to the discussed topic, speaker, and hearer features. To account for non-referential expletive subjects, linking them with a topic feature does not seem to be sensible (see again Section 3.1.5). We would probably either have to assume a new C-feature for expletives or, and this seems to be intuitively more plausible, to exempt expletives from context-linking since they are in fact not linked to the context in any form.

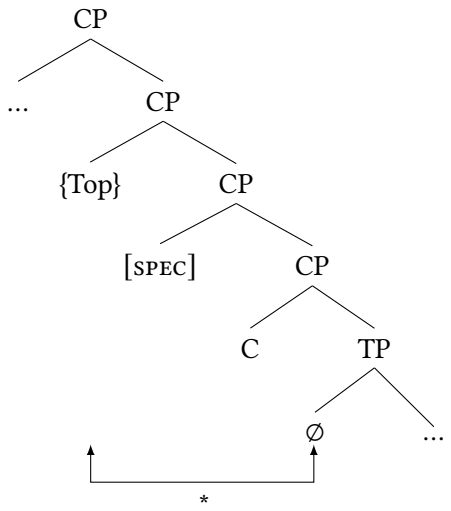
[91] Sigurðsson & Maling (2010: 61) refer back to Sigurðsson (2004a,b) by stating that "the speaker/hearer features are referred to as the logophoric agent (speaker) and the logophoric patient (hearer), ΛA and ΛP for short." They sketch the features of the C-domain as follows:

(i) $[_{CP} \ldots$ Top $\ldots$ ΛA $\ldots$ ΛP $\ldots [_{TP} \ldots$ (Sigurðsson & Maling 2010: 61)

The context-linking is defined as a " 'transitive' matching relation (where A ↔ B reads 'A is matched by B' or 'B is interpreted in relation to A'):

(ii) Context ↔ C-features ↔ TP-internal elements" (Sigurðsson & Maling 2010: 61)

(47)



(Sigurðsson & Maling 2010: 63)

This shows that the main challenge for the approach remains to adequately explain the prefield restriction of topic drop. Either we could adopt Sigurðsson & Maling's (2010) approach just described, but this implies that topicality has a central role to play, or we could solve it more technically, similarly to the *pro* approach. Then, we would have to assume a feature that sits in a [Spec, CP] position that is either not c-commanded sentence-internally by a potential identifier or the highest [Spec, CP] of a root clause, and that allows for the phonological non-realization of the corresponding constituent only in this position. In the latter case, however, as discussed above, the question of what would be an adequate motivation for such a feature still remains open.

3.6.4 Summary: syntactic analysis

In the following, I summarize the three syntactic approaches to a generative description of topic drop. In the operator approach, the operator can be semantically conceptualized as a lambda abstraction that triggers the retrieval of a salient discourse referent in order to pragmatically enrich the lambda abstraction to a proposition. In this way, the operator approach provides a discourse-pragmatic explanation for the prefield restriction of topic drop. The two major drawbacks of such an analysis are the following: First, the assumption of an operator as a lambda abstraction leads to a type problem of declaratives. It would have to be assumed that in addition to declaratives of type t, there are also declaratives of type $\langle e, t \rangle$. This conflicts with standard assumptions concerning the semantics

of declaratives (though it recalls cases of pragmatic or conceptual enrichment as argued for in the case of fragments by Stainton 2006 and others). Second, an operator approach cannot be easily extended to non-referential constituents targeted by topic drop, which precisely have no linguistic antecedent or extralinguistic referent. Besides these two obvious disadvantages, a potential advantage of the operator account is that it provides a similar analysis of the related ellipsis types topic drop, subject gaps, null copulas, and null articles. However, there are also obvious differences between these phenomena, which call into question whether a joint consideration is desirable at all. Subject gaps, as the name suggests, are restricted to subjects, unlike topic drop, and null copulas and null articles are limited to headlines as a text type.

The *pro* approach assumes the existence of a null element *pro* that is underspecified with respect to its features and fixed in the lexicon. This *pro* is argued to receive its feature values from a discourse antecedent, a mechanism that, to put it simply, syntactically anchors the felicity condition of recoverability (see Chapter 4). Again, the disadvantage arises that the cases of omitted non-referential constituents cannot be explained. However, this could be solved by assuming another *pro* for these cases, but this again complicates the lexicon and the grammatical system. Another drawback of the *pro*-analysis is that the prefield restriction can only be integrated into the proposal by assuming a special feature, i.e., mainly by stipulation.

Finally, the PF-deletion approach has the advantage that relatively few additional assumptions have to be made. We could say that, other things being equal, it is the null hypothesis for analyzing topic drop. The lexicon and the grammatical system are not complicated, because this approach does without special null elements. Since topic drop utterances are grammatically and functionally parallel to their full counterparts, it can also easily account for expletives. However, it also has the disadvantage that the prefield restriction (at least of non-referential topic drop cases) does not follow directly from the theory but requires postulated features.

Despite this drawback, the PF-deletion approach seems to harmonize best with the empirical results I obtained. Since it can explain as much as the other two approaches (or even more, i.e., the omission of expletives) without the assumption of additional null elements, I argue for giving it preference following Occam's Razor.

In the next chapter, I turn to recoverability, which has also been discussed as a factor impacting the licensing of topic drop.

4 Recoverability as a prerequisite for topic drop

Chapter 3 identified the prefield restriction, or more precisely the restriction to a prefield that is not c-commanded sentence-internally by a potential identifier or to the highest prefield of a root clause, as the main licensing condition of topic drop. This chapter focuses on recoverability, which is often discussed in the literature as a further condition for topic drop (e.g., Fries 1988, Cardinaletti 1990, Helmer 2016, Freywald 2020). For example, Fries (1988: 27) lists it as a discourse condition, while Cardinaletti (1990: 75) states that topic drop is "only possible with contextually salient elements, i.e. the reference of the null argument must be recoverable either from the linguistic or the extralinguistic context." In this chapter, I argue that recoverability should not be understood as a syntactic licensing condition, i.e., not analogous to the prefield restriction but as a felicity or usage condition.

Recoverability, roughly speaking, denotes the process of identifying the omitted constituent and, in the case of referential constituents, its reference. I argue that a hearer is able to recover an omitted constituent (and its reference) if they can determine, based on several factors, which constituent (and reference) the speaker intended. Possible referents are either directly available through the linguistic or extralinguistic context or can be inferred indirectly from the linguistic context. It is important to note that recoverability is not an exclusive feature of topic drop but a prerequisite for any type of ellipsis or omission. Presumably, this mechanism is even at work in other related phenomena, such as pragmatic enrichment, where something not explicitly verbalized must be co-understood, or the resolution of pronouns or anaphors in general, where the reference of a placeholder must be established.

I assume that the process of recovery requires cognitive effort on the part of the hearer and that this effort varies depending on how easy or difficult this process is. The ease of recoverability, in turn, may be influenced by several factors such as the type of context that is used for the recovery, the distance between the ellipsis site and the linguistic antecedent, and how direct the link between the ellipsis site and the linguistic antecedent is. It may be possible to operational-

ize recoverability, at least of referential constituents, by recourse to the concept of givenness, as suggested in the literature (e.g., Fries 1988, Helmer 2016, Trutkowski 2016). In this context, categorical givenness is often seen as a proxy for licensing the omission of a constituent or its recoverability, while gradual givenness is used to argue how well or poorly something can be omitted or recovered. Hence, it seems useful to also discuss givenness as a concept and, in particular, to relate it to recoverability.

This chapter is structured as follows: In the first two sections, I look at the type of context used to recover the reference of topic drop distinguishing between linguistic and extralinguistic context. The third section briefly sketches the role of recoverability for expletives before I turn to givenness in the fourth section and discuss whether and to what extent (concepts related to) givenness can be used to describe recoverability. Thereby, I also establish a link between recoverability and processing effort. In the concluding section of this chapter, I summarize the results and come back to the above-given definition of recoverability. There, I also already hint at a link between recoverability, processing effort, and the information-theoretic account of the usage of topic drop, which I outline in detail in Chapter 6.

4.1 Linguistic antecedents

There is consensus in the literature that the reference of topic drop of referential constituents can be recovered via linguistic antecedents. For example, Günthner (2006: 100) and Schwitalla (2012: 102) state that the omitted constituents can be reconstructed based on a previous mention (see also Fries 1988, Cardinaletti 1990, Zifonun et al. 1997, Volodina 2011, Volodina & Onea 2012). Trutkowski (2016: 22) argues that recovery is not only possible through an antecedent but also through a so-called *postcedent*. That means that the antecedent can also follow topic drop, as in the discourse-initial example (1), where the subject of the first utterance is omitted (1a) but can be recovered based on the co-referent overt subject of the second utterance *Ortschronist Günter Bergner* (1b) (and based on an additional photograph).

(1) a. Δ *Kennt die Straßen im Karlshorster Kiez wie seine*
he knows the streets in.the Karlshorst's neighborhood like his
Westentasche:
vest.pocket
'(He) knows the streets in the Karlshorst neighborhood like the back of his hand:'

b. *Ortschronist Günter Bergner ist häufig auf Achse.*
local.chronicler Günter Bergner is often on axis
'Local chronicler Günter Bergner is often on the road.' [DWDS: Corpus, Berliner Zeitung, 05/04/1995, p. 23] (Trutkowski 2016: 22)

In example (1), the antecedent and the omitted constituent are both a subject in the nominative case and, thus, formally identical. Fries (1988: 27) points out, however, that such a formal identity is not necessary, but that referential identity is sufficient to allow for recoverability. This means that mismatches in grammatical function and case between the antecedent and the omitted constituent are possible, as shown in example (2). In the context sentence (2a), the linguistic antecedent *Sofia* is a DP in the dative case that is part of a PP, but the non-realized prefield constituent in (2b) is required to be the subject in the nominative case. However, recoverability works because *Sofia* is a highly salient potential antecedent (probably even the only available antecedent in this context), so it immediately suggests itself that both, the antecedent and topic drop, refer to the same person.

(2) a. *A: Was ist denn mit Sofia?*
what is PART with Sofia
A: 'What's the matter with Sofia?'
b. *B:* Δ *Ist mir fremd gegangen!*
she is me strange gone
B: '(She) cheated on me!' (Fries 1988: 20)

Helmer (2016: 76) terms topic drop with referentially identical antecedents "direct analepsis" and distinguishes it from "indirect analepsis", similar to the distinction between direct and indirect anaphora (e.g., Consten & Schwarz-Friesel 2007: 284–286).[1] Indirect analepsis, which amounts to 20% of the 541 instances of topic drop in Helmer's corpus (see Section 3.2) (Helmer 2016: 74), is characterized by the fact that the referent of the antecedent and the referent of the omitted constituent are not identical but have a certain semantic relationship to each other, e.g., a metonymic one. To this end, Helmer (2016) discusses example (3), taken from a conversation of a couple that wants to quit smoking. She argues that no immediate antecedent is available for the topic drop in (3d) and, based on the wider discourse context (see Helmer 2016: 88–89, for details), that the referent of the covert expression would conceptually be something like *the health (of all*

[1]Consten & Schwarz-Friesel (2007: 284) refer to similar concepts such as Clark's (1975) *bridging* and Prince's (1981) *inferrable* category (see below). They likewise assume that a hearer establishes a relation between an expression and its "anchor" through an inference process.

smokers) endangered by the many smoking.[2] According to Helmer, it is linked to the discourse via a so-called *linguistic anchor*, namely *the many smoking* implicitly contained in A's ironic utterance (3c). Put simply, Helmer argues that the recovery of the omitted constituent works because there is a cause-effect relationship between the referent of the antecedent, i.e., the many smoking, and the referent of the omitted constituent, the endangered health.

(3) a. *A: Na, wollen wa noch eene roochen?*
INJ want we still one smoke
A: 'Well, do we want to smoke another one?'
b. *B: Joa*
yep
B: 'Yep.'
c. *A: Weil wa heut schon so wenig jeraucht ham?*
because we today already so little smoked have
A: 'Because we've already smoked so little today?'
d. *B: Δ is ja unsre jesundheit.*
yes it is PART our health
B: '(It) is our health.' [FOLK_E_00039_SE_01_T_03_jesundheit] (cited in Helmer 2016: 88, simplified)

Example (4) from the fragment corpus FraC (see Section 7.2.1) can serve as a less complex example of this type of indirect topic drop. The topic drop in (4b) cannot be recovered directly via a concrete linguistic antecedent but only indirectly by assuming what I informally term a process-result relationship. The reference of the topic drop is the result of the mixing process described in (4a).

(4) a. *Gleiche Mengen Giersch, Zitronenmelisse und*
equal amounts goutweed lemon.balm and
Gundermannblättchen in Apfelsaft ziehen lassen, abseihen und
ground.ivy.leaves.DIM in apple.juice steep let strain and
mit Sprudelwasser (oder Prosecco?) verdünnen.
with sparkling.water or prosecco dilute
'Steep equal quantities of goutweed, lemon balm, and ground ivy leaves in apple juice, strain and dilute with sparkling water (or Prosecco?).'
b. Δ *Schmeckt prima an heißen Tagen...!*
that tastes great on hot days
'(That) tastes great on hot days...!' [FraC B23–B24]

[2] Importantly, Helmer (2016: 89, footnote 71) clarifies that this is not to be understood as a lexical reconstruction but as a "conceptual structure".

However, in both examples, the procedure of determining the reference via a cause-effect or process-result relationship would be analogous if there was an overt pronoun in the prefield. This observation is supported by the fact that Helmer (2016: 211) classified 26.5% of the *das* ('that')-occurrences in her reference data (see Section 11.2 for details) as indirect anaphora, a figure that is even larger in absolute terms than the 20% indirect cases of topic drop.

Despite the indirect relationship between the antecedent and the covert constituent in these examples, the respective utterances are nevertheless directly adjacent to each other. Volodina & Onea (2012: 217) argue that this is not coincidental but the result of a restriction. They state that the referent of the omitted constituent needs to be named immediately before the omission without intervening utterances because only this would allow for what they call sufficient activation.[3] Consequently, they judge example (5) as ungrammatical.[4]

(5) *A: Hast du Marcus gesehen? Es regnet und es blitzt und es donnert*
have you.2SG Marcus seen it rains and it flashes and it thunders
*so sehr. – B: *(Er) ist im Haus.*
so very he is in.the house
A: 'Have you seen Marcus? It is raining and it is flashing, and it is thundering so much.' – B: '(He) is in the house.' (Volodina & Onea 2012: 217, their judgment)

I do not share this judgment but consider the utterance with topic drop to be acceptable even with the intervening utterance. This is supported by example (6) from the dialogue subcorpus of the FraC,[5] where several utterances lie between the antecedent *das Maritim-Hotel* in (6a) and the topic drop in (6e).[6]

[3] They seem to make use of the concept of activation, discussed, for example, by Chafe (1994) (see below), although they do not explicitly refer to him.

[4] Sandig (2000: 299) claims that topic drop is restricted to the utterance following the antecedent.

[5] The data are transcripts of spoken dialogues about organizing a business trip collected as part of the Verbmobil project (Burger et al. 2000) and stem from the Tübinger Baumbank des Deutschen/Spontansprache (TüBa-D/S, 'Tübingen treebank of spoken German', Hinrichs et al. 2000). Unfortunately, the information about which speaker produced which utterance is not available.

[6] Theoretically, the antecedent of topic drop could also be the price per room from utterance (6c). However, the reading that the price is "good middle class" is very uncommon and, therefore, unlikely. A more likely option is to assume that "the price" is needed for the statement that the hotel is middle class and that "the hotel" is reactivated by the immediately adjacent "price" from utterance (6c) in a kind of bridging (see above). To determine the role of the distance independently of such alternative analyses, a controlled experiment is required.

(6) a. *tja, dann nehmen wir vielleicht gleich das Maritim-Hotel, das*
well then take we maybe right.away the Maritim-hotel that
ist ziemlich zentral gelegen.
is pretty centrally located
'Well, then maybe we'll just take the Maritim Hotel, it is quite centrally located.'

b. *Einzelzimmer hundert einundfünfzig Mark.*
single.room hundred fifty.one German.mark
'Single room one hundred fifty-one German mark.'

c. *also, ich denke, das kann ich schon vertreten, ja, diesen Preis können*
well I think that can I PART defend yes this price can
wir schon noch nehmen.
we PART still take
'Well, I think I can defend that, yes, we can still take this price.'

d. *ja.*
yes
'Yes.'

e. *gut, ja, Δ ist so gute Mittelklasse.*
good yes it is so good middle.class
'Well, yes, (it) is good middle class.' [FraC D508–D512]

Examples like (6) suggest that the statement by Volodina & Onea (2012) according to which topic drop with intervening utterances is ungrammatical may be too strong. Nevertheless, there seems to be a general tendency for the antecedent and the covert constituent to be close to each other. Helmer (2016: 90) reports that for the majority of topic drop cases in her data set, the utterance with the antecedent and the utterance with topic drop are immediately adjacent, or there are only two to three turns in between. Furthermore, the intervening utterances are often incomplete through restarts or function merely as backchannelling, such as (6d) above. In these cases, they are argued not to disturb the linking between the antecedent and the covert constituent.

This observation of a short distance between the antecedent and topic drop coincides with the facts in the FraC, as shown in Figure 4.1. The figure depicts the distance between the antecedent and the target for 322 of the 873 occurrences of topic drop in the FraC. I measured this distance in the number of intervening utterances, i.e., 0 means the antecedent and the target occur in the same utterance, 1 means the utterances with the antecedent and the target are directly adjacent

to each other, 2 means there is one utterance in between, etc.[7]

For the vast majority of the cases, the linguistic antecedent of topic drop occurs in the precontext. This is in line with the central concept of my information-theoretic account of topic drop usage (see Chapter 6), namely predictability. If an antecedent is present in the preceding discourse, the coreferential constituent becomes predictable and is more likely to be omitted (see Section 6.2.1).[8]

In most of these cases (211, 65.53%), the antecedent does not only occur somewhere in the preceding context but in the utterance that immediately precedes the utterance with topic drop. A distance of two utterances is already less frequent with only 49 instances (15.22%) and includes many cases where topic drop occurs in two subsequent utterances after the antecedent. Instances where the antecedent (or postcedent) occurs in the same utterance as topic drop (13, 4.04%) and instances where the distance is larger than two (in total 42, 13.04%) are even rarer.[9] This result supports the assumption that topic drop needs some proximity to its antecedent, which, in turn, is in line with the concept of predictability. If a referent has been mentioned recently, that referent is most likely still (more)

[7] A distance could not be determined for the remaining 551 cases because there was either no precontext available or no antecedent could be found in the precontext (or no postcedent could be found in the postcontext). These were usually cases where the antecedent was known through the communicative situation, as in example (i) from the beginning of a blog post. In this case, the text type blog post is such a strong hint toward the blogger writing about themselves that the pronoun referring to them can be omitted.

(i) *15.05.13* Δ *Bin im Garten*
15.05.13 I am in.the garden
'05/15/13 (I) am in the garden' [FraC B9–B10]

[8] There are seven apparent counterexamples with a postcedent. In three cases each, the postcedent is in the following or the same utterance, in one case in the utterance after the next. They can be interpreted as the result of stylistic considerations rather than redundancy avoidance, e.g., in (i) the intention is probably to first introduce the slogan and then the car.

(i) a. Δ *MACHT DIE KURVE ZUR GERADEN*
he makes the curve to.the straight
'(It) turns the curve into a straight line.'
b. *DER NEUE NISSAN QASHQAI MIT CHASSIS CONTROL TECHNOLOGIE.*
the new Nissan Qashqai with chassis control technology
'The new Nissan Qashqai with chassis control technology.' [FraC P143–P144]

[9] See the file *Examples_Topic_Drop_Distances.pdf*, provided online at https://osf.io/zh7tr, for examples of different distances. Most instances with a distance of more than five between the antecedent and the target stem from chat room conversations with numerous participants. In these conversations, multiple threads are often followed in parallel, and multiple utterances from different people can be understood as a response to an assertion. These responses are then displayed one after the other, even if they have equal status as responses to the same antecedent utterance. Therefore, these cases are probably also cases with distance one.

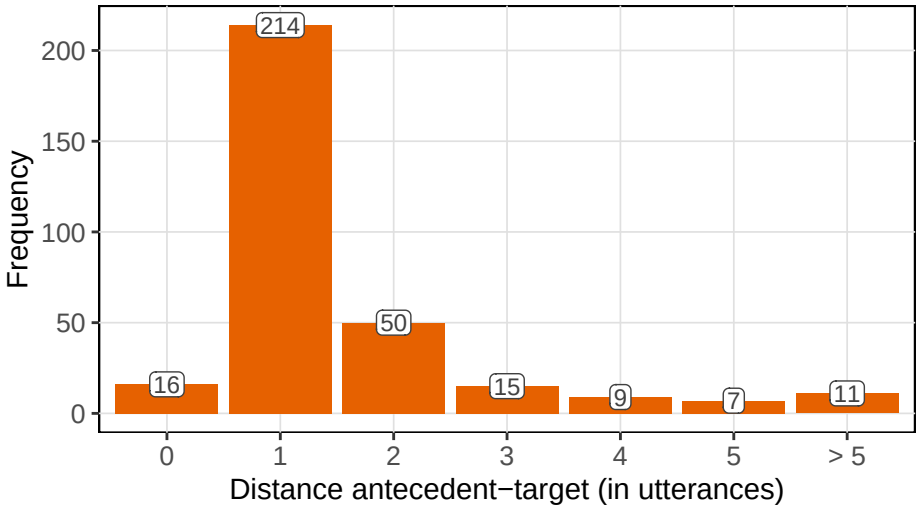


Figure 4.1: Distance between the antecedent or postcedent and topic drop in number of utterances in the FraC

present in the hearer's mind, so any constituent that refers to that referent is also predictable and can be better omitted (see also the similar discussion on recoverability and givenness in Section 4.4).

At the same time, the proximity between the antecedent and the target does not seem to be a unique property of topic drop. Quite similar tendencies are observed for overt personal pronouns by Portele & Bader (2016). They used the German DeWaC corpus (Baroni et al. 2009) to analyze two random samples of about 500 overt p- and d-subject pronouns in sentence-initial position respectively, based on a total search result of more than 900 000 instances. They considered a precontext of five sentences and determined the last DP that was coreferent with the target pronoun as the antecedent (Portele & Bader 2016: 11–13). Table 4.1 shows the proportion of sentences with a certain distance between the antecedent and the target pronoun, subdivided for pronoun type, using the same measurement of the number of utterances as in Figure 4.1 for topic drop.

For their sample at least, the effect of recency seems to be even more pronounced for the overt pronouns than for topic drop. For over 90% of the sentence-initial subject pronouns, the antecedent occurs in the immediately preceding utterance. Portele & Bader (2016: 19) argue that unlike lexical NPs, which often allow for a longer distance to their antecedent according to Arnold (2010), both pronoun types "seem to require an antecedent that occurred recently and is there-

Table 4.1: Proportions of distances between an overt pronoun and its antecedent in Portele & Bader (2016: 16), original distance measure adapted to fit Figure 4.1

Pronoun type	Distance antecedent-target				
	1	2	3	4	5
P-pronoun	91.1%	6.7%	2.6%	0.0%	0.0%
D-pronoun	95.8%	3.5%	0.5%	0.2%	0.0%

fore in an activated state in working memory." Consequently, not only the necessary proximity to the antecedent seems to be another parallel between topic drop and overt pronouns but also the relation to associated cognitive activity as the underlying explanation.

4.2 Extralinguistic antecedents

Besides the postulated restriction with respect to the distance between the antecedent and topic drop, Volodina & Onea (2012: 217) additionally claim that for topic drop to be felicitous the referent needs to be explicitly (verbally) named. By presenting the contrast in examples (7) and (8), they deny that topic drop can be recovered from the extralinguistic context. However, the variant with the recovery from the extralinguistic context (8) is highly artificial because it is completely unclear why B would point to a portrait in this situation.

(7) a. *A: Hast du Marcus gesehen?*
have you.2SG Marcus seen
A: 'Have you seen Marucs?'
b. *B:* Δ *Ist im Haus.*
he is in.the house
B: '(He) is in the house.'

(8) [A zeigt auf Marcus Portrait.] *B: *Δ Ist im Haus.*
A points to Marcus's portrait he is in.the house
[*A points to Marcus's portrait.*] B: '(He) is in the house.' (Volodina & Onea 2012: 217, their judgments)

More natural examples like (9) indicate that topic drop can be acceptable although the referent is not explicitly named but only present in the current utterance situation.

(9) [A, B, and C are in a restaurant. When C barks aggressively at the waiter, A asks B in a whisper:]
   a. *A: Was ist los?*
      what is loose
      A: 'What's wrong?'
   b. *B:* Δ *Hat schlecht geschlafen.*
      he has badly slept
      B: '(He) slept badly.'

This view is shared by Fries (1988), Cardinaletti (1990), Zifonun et al. (1997),[10] and Volodina (2011), who state that the referent of the omitted constituent can be recovered both from the linguistic and the extralinguistic context so that topic drop can also refer to objects or persons that are present in the current situation like the speaker, the hearer, or a movie in the form of a poster (10).

(10) a. [A and B standing in front of a poster for the new Downton Abbey movie.] (situational antecedent)
   b. *B:* Δ *Muss ich unbedingt sehen!*
      that must I desperately see
      B: 'I really have to see (it)!'

Frick (2017: 73) specifies that the recovery of topic drop from the extralinguistic context requires that the speaker and the hearer(s) share a mutual context in the concrete communicative situation, a context that is shaped by an orientation at the origo or deictic center (I, here, now) (Bühler 1965: 102) (see also Section 10.1.1). Accordingly, elements that can be recovered based on such a context orientation can be targeted by topic drop.

4.3 Non-referential elements

While resolving topic drop of referential constituents requires recovering both the omitted constituent and its referent, the situation is different for non-referential constituents like expletive subjects. Since expletives are not referential but

[10] Zifonun et al. (1997) assume two different processes here depending on the nature of the context from which the ellipsis is recovered. For linguistic contexts, they talk about *Analepse* ('analepsis'), presumably a composition of *Anapher* ('anaphora') and *Ellipse* ('ellipsis') borrowed from Blatz (1896) (Zifonun et al. 1997: 569–571). If the context is extralinguistic, they use the term *Ellipse* ('ellipsis') (Zifonun et al. 1997: 413). For topic drop, they explicitly discuss both possibilities, i.e., that its reference is recovered from the linguistic or the extralinguistic context (Zifonun et al. 1997: 636).

semantically empty, there is no reference to recover. The question, then, is how hearers handle utterances in which an expletive is omitted. There are principally two options. Hearers could reconstruct this expletive as a semantically empty element, i.e., insert it into the prefield during sentence processing. Or they could process the utterance directly without a prefield constituent so that in this case no reconstruction takes place at all. At this point, I cannot decide between the two variants, and it is also not clear to me which empirical method should succeed in such a differentiation. In any case, a hearer will understand an elliptical utterance like (11) in the intended way, either because they can unambiguously recover the omitted *es* ('it') as the prefield constituent, or because they do not need to recover anything, since all semantically relevant information is given. Therefore, we can conclude with Frick (2017: 137) that expletives are uniquely recoverable and identifiable, which favors their omission.

(11) Δ *Regnet.*
it rains
'(It) is raining.'

4.4 Recoverability and givenness

I now turn to the question of what kind of relationship exists between the omitted referential constituent and its linguistic antecedent and/or extralinguistic referent that allows for recoverability. In the literature, it is common to define this relationship as some form of categorical givenness relation. For example, Fries (1988: 27) uses givenness as a proxy to recoverability when he demands that the constituent targeted by topic drop is coreferential to a "known quantity"[11] in the discourse. An extension of the givenness relation is proposed by Trutkowski (2016: 19), who states in her "[s]alience/givenness condition" that "[t]he referent of a topic dropped element must be inferrable/contextually given (via some antecedent expression)." Similarly, Freywald (2020: 150) postulates: "The dropped element always refers to an entity that is highly salient[12] in the discourse (i.e.,

[11] My translation, the original: "bekannte[] Größe" (Fries 1988: 27).

[12] Both authors use the concept of salience differently from Ariel (1990), who equates it with topicality, see below. While Freywald (2020) simply states that salient means being inferrable, Trutkowski (2016) refers to Chiarcos et al. (2011) and adopts their definition of salience: "Salience defines the degree of relative prominence of a unit of information, at a specific point in time, in comparison to the other units of information" (Chiarcos et al. 2011: 2). Trutkowski (2016: 20) admits that this definition is rather vague, and clarifies that for her a salient referent is (highly) accessible in a given discourse without having to be a topic. See also the review article by Boswijk & Coler (2020) on different concepts of salience in linguistics.

that can be inferred from the discourse context)."[13] From these two statements, it follows that the referent of topic drop need not necessarily be given (in the sense of, e.g., Prince (1981)); it is sufficient that it is inferrable, a term that neither Trutkowski (2016) nor Freywald (2020) explicitly defines.

Inferrable is often used as a kind of middle point of givenness, located between given and new. In this way, it was introduced as a linguistic term by Prince (1981). Prince created a taxonomy distinguishing a total of seven different states of givenness, which she terms "assumed familiarity" (Prince 1981: 233; 237), shown in Figure 4.2. She distinguishes between new, given – which she terms evoked – and inferrable discourse entities and defines the latter in the following way: "A discourse entity is Inferrable if the speaker assumes that the hearer can infer it, via logical — or, more commonly, plausible — reasoning, from discourse entities already Evoked or from other Inferrables" (Prince 1981: 236). She gives the example (12a), where the discourse entity *the driver* is a *(noncontaining) inferrable* because it can be inferred from the textually evoked entity *a bus* and from the world knowledge that buses usually have a driver. In example (12b), *one of these eggs* is a so-called *containing inferrable* since the hearer can infer it by a set-member inference from the situationally evoked discourse entity *these eggs*.

(12) a. I got on a bus yesterday and the driver was drunk.
b. Hey, one of these eggs is broken! (Prince 1981: 233)

So, if we combine the statements by Trutkowski (2016) and Freywald (2020) with Prince's (1981) taxonomy, recoverability in their sense means that the constituent targeted by topic drop must not be new, i.e., it must be either textually or situationally evoked or a noncontaining or containing inferrable. This way, the concept of inferrable may also be suitable to describe the cases that I discussed as indirect topic drop above. For example, in (4b) above, the discourse entity to which the topic drop refers can be inferred from the ingredients mentioned in the previous utterance.

While Prince's (1981) taxonomy is based on a categorical notion of givenness, there are also gradual approaches to givenness. For example, Chafe (1994) defines it with recourse to activation, which is a cognitive concept used, roughly speaking, to describe the focus of a speaker's consciousness (see also Chafe 1974). According to Chafe (1994), there are three activation states, i.e., *active*, *semiactive*,

[13] My translation, the original: "[D]as gedroppte Element referiert stets auf eine Entität, die im Diskurs hoch salient ist (d.h. aus dem Diskurskontext inferiert werden kann)" (Freywald 2020: 150).

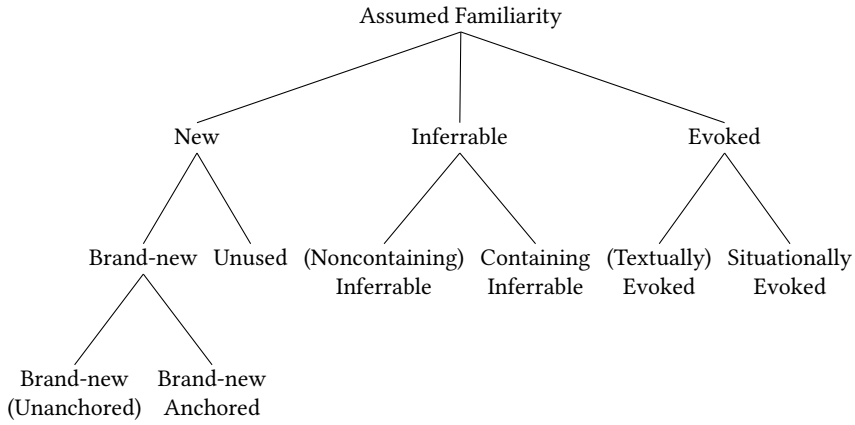


Figure 4.2: Prince's (1981) levels of assumed familiarity, recreated from Prince (1981: 237)

and *inactive*,[14] which he links to three levels of givenness – given, accessible, and new – using the concept of *activation cost*:

> Suppose that at a certain time, t_1, a particular idea is active, semiactive, or inactive. Suppose that at a later time, t_2, whatever its earlier state may have been, this idea is now active. If it was already active at t_1, we can say that at t_2 it is given information. If it was semiactive at t_1, it is accessible at t_2. If it was inactive at t_1, it is new at t_2. It is helpful to think of these three processes in terms of cognitive cost: given information is least costly in the transition from t_1 to t_2 because it was already active at t_1. Accessible information is somewhat more costly, and new information is the most costly of all, presumably because more mental effort is involved in converting an idea from the inactive to the active state. (Chafe 1994: 72–73)

Chafe (1994: 75) claims that usually speakers use accented full noun phrases to express new and accessible information while given information is expressed using weakly accented pronouns or null pronouns. The link between activation and referential form is picked up by Volodina & Onea (2012) (without reference to Chafe) and by Helmer (2016) (with reference to Chafe). They argue that a constituent can usually only be targeted by topic drop if the entity it refers to

[14] For convenience, Chafe argues only with these three states, but basically, he assumes a continuum of activation states (Chafe 1994: 55). Given the direct mapping between activation and givenness, he seems to also assume a gradual givenness concept.

is given, i.e., if it was already active before the utterance with topic drop in the sense of Chafe (1994). According to Volodina & Onea (2012) and Helmer (2016), recoverability corresponds to having been active in the speaker's consciousness, so the cognitive cost in accessing the referent of the omitted constituent is low. This seems like a more concrete formulation of Schwitalla (2012: 102), who claims that constituents can be omitted because their referent is still present in memory.

The link between activation cost and referential form has parallels to Ariel's (1990) continuous *accessibility marking scale*.[15] Along this scale, Ariel (1990: 73) arranges referential expressions as so-called "accessibility markers" from very explicit markers (e.g., full names plus modifiers) at the top to very implicit markers like gaps, i.e., null elements, at the bottom. She argues that the closer a marker is toward the bottom, the more accessible its antecedent needs to be and vice versa. A lowly accessible antecedent is rather picked up with a full name plus a modifier (or a definite description, etc.), while a highly accessible antecedent can be referred to using a gap (or at least a cliticized or unstressed pronoun). According to Ariel (1990), accessibility[16] is determined by at least the following four factors:

a Distance: The distance between the antecedent and the anaphor (relevant to subsequent mentions only).

b Competition: The number of competitors on the role of antecedent.

c Saliency: The antecedent being a salient referent, mainly whether it is a topic or a non-topic.

d Unity: The antecedent being within vs. without the same frame/world/point of view/segment or paragraph as the anaphor. (Ariel 1990: 28–29)

Ariel (1990: 29) argues that an antecedent is "more likely to be in a highly activated state in memory" and, therefore, more likely to be picked up with an accessibility marker from the lower pole of the scale, e.g., with topic drop, if it is spatially close to this marker, if it has few competitors, if it is salient, which Ariel mainly defines as being a topic, and if it is more strongly connected in terms of coherence to the marker. In sum, Ariel (1990) considers recoverability to be a form of activation or presence in memory or consciousness, similar to Chafe

[15] See also Ackema & Neeleman's (2007) application of Ariel's (1990) accessibility theory to *pro* drop and topic drop in Early Modern Dutch, which I discussed briefly in Footnote 48.

[16] Ariel's (1990) term *accessibility* is distinct from Chafe's (1994) *accessible*. For Ariel, *accessibility* is comparable to Chafe's (1994) concept of activation, while Chafe uses *accessible* to denote an intermediate givenness level.

(1994). She argues that the degree of activation does not only depend on givenness but also on the four factors introduced: distance, competition, saliency, and unity.

4.5 Recoverability as felicity or usage condition

In this section, I come back to the definition of recoverability presented at the beginning of this chapter and expand it in light of the further insights gained. I argued that recoverability means that the hearer must be able to uniquely identify the (reference of the) covert constituent. For referential constituents, the recovery requires some kind of antecedent or "anchor" in the extralinguistic or linguistic context, as shown in Figure 4.3.[17] In the case of linguistic context, topic drop and the antecedent tend to be close to each other, often in directly adjacent utterances.

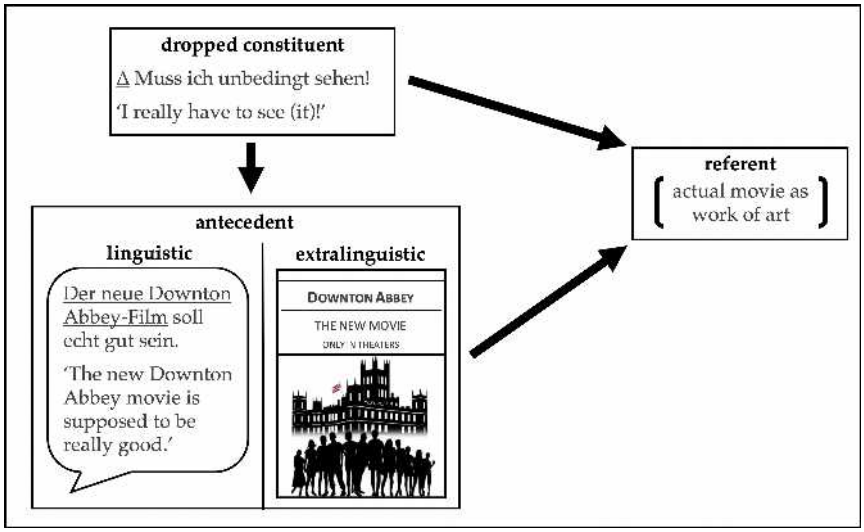


Figure 4.3: Relation between the omitted constituent, the antecedent, and the referent (loosely based on a figure by Helmer 2016: 76)

Following Helmer (2016), I distinguish between two "modes of recovery" for referential topic drop: direct and indirect topic drop. For direct topic drop, the antecedent and the covert constituent do not have to be formally identical but are

[17] I created this figure in Microsoft Word using pictures from pixabay. Source for the castle: https://pixabay.com/de/vectors/highclere-castle-downton-abbey-4515425/ (visited on 05/09/2022). Source for the group of people: https://pixabay.com/de/vectors/junge-kind-zusammenarbeit-vati-2026064/ (visited on 01/02/2024).

typically coreferential. This means that the hearer can directly determine the reference of the omitted constituent when they identify the coreferential antecedent since the referent of this antecedent is already known to them. This is illustrated in the example of Figure 4.3, where *the new Downton Abbey movie* is the referent of both the linguistic antecedent and the omitted constituent. For indirect topic drop, however, the two referents are distinct but have a certain semantic or pragmatic relationship to each other that the hearer must determine. The mechanisms sketched for recovering topic drop are parallel to the commonly assumed mechanisms that underlie the use of (personal) pronouns and, more generally, anaphora (see, e.g., Huang 2000, Bhat 2004).

In this book, I additionally draw on the idea that the recoverability of referential topic drop can be described by concepts of givenness. While givenness can be understood as categorical, i.e., a constituent can be targeted by topic drop if its referent is not new, i.e., if it is given in or inferrable from the current discourse context in Prince's (1981) taxonomy, I pursue a gradual givenness concept. Following Chafe (1994) and Ariel (1990), I assume that the recoverability of a covert constituent depends on how accessible its referent is from memory or how strongly activated this referent is in consciousness. Such a gradual givenness concept includes cognitive processes and is argumentatively closer to the information-theoretic approach that I present and discuss in the second part of this book. More specifically, in Section 6.3, I argue that the verb following topic drop may play a role in facilitating the recovery of the omitted constituent and reducing the overall processing effort for the hearer.

A gradual notion of givenness almost inevitably leads to a gradual notion of recoverability, as I assume in this book. The scale of recoverability ranges from (i) cases at the upper end, which can be recovered unambiguously, to (ii) cases at the lower end, where topic drop could be merely recovered by guessing. Cases like (i) are the above-mentioned omission of the expletive *es*, which can be uniquely recovered but also the omission of a 1st person singular subject before a verb that is distinctly marked for inflection. An example for cases like (ii) is (7) from Chapter 3, repeated here as (13). If it is uttered out of the blue, i.e., without a previous discourse in which John was mentioned or without John being present in the current situation, the hearer cannot determine the referent of topic drop but can only guess wildly.

(13) Did you know? *(John) married last week.

I argue that in such cases, topic drop is not ungrammatical but simply infelicitous. From this argumentation, it follows that recoverability is not a licensing condition for topic drop, as it is the prefield restriction, but rather a felicity or usage

condition. While in example (13), the speaker, unless completely lost in thought, should realize that their hearer has no chance of retrieving the referent of topic drop, there may be cases where this is not so obvious. That is, it could be that the referent is sufficiently given to the speaker but not given enough to the hearer. This idea is supported by Helmer (2016). She states that topic drop is generally well understood in her data despite instances of indirect topic drop and existing mismatches between the antecedent and the target (Helmer 2016: 181). However, she also found instances in her corpus where the hearer asked a comprehension question after an utterance with topic drop. This indicates that they were unable to recover the reference of the ellipsis or were at least not completely sure if their reconstruction was correct (Helmer 2016: 190–191). Such a case is shown in example (14) from a conversation during a soccer manager game. The player Simon produces a topic drop that another player, Jan, apparently cannot recover. This leads first to Jan asking *what?* and then Simon making the reference of the topic drop explicit by verbalizing the proposition the ellipsis referred to (*dass du auch noch einsteigen willst*).

(14) [*Simon, Maik, and Jan are playing a soccer manager game with other friends and are currently bidding on a player. Maik is in the lead with 14 million.*]

a. *Simon: Vierzehn fürn Maik.*
fourteen for.the Maik
Simon: 'Fourteen for Maik.'

b. *Simon: Jan.*
Jan
Simon: 'Jan.'

c. *[...] Simon:* Δ *Hätt ja sein können.*
it had.CONJ PART be can
Simon: '(It) could have been.'

d. *[...] Jan: Was?*
what
[...] Jan: 'What?'

e. *Simon:* Δ *Hätt ja sein können, dass du auch noch einsteigen willst.*
it had.CONJ PART be can that you also still get.in want
Simon: '(It) could have been that you also wanted to get in.' (Helmer 2016: 190, adapted)

In other cases, the speaker apparently preempts such a comprehension question by making the reference of the omitted constituent explicit and verbalizing it

in the postfield or a subsequent utterance (Helmer 2016: 191–192). This suggests that in many cases it depends not only on the referent's givenness but also on the hearers and their general and situational cognitive resources whether and how well an omitted referent can be recovered and, consequently, omitted.

In sum, I consider recoverability to be a felicity or usage condition for topic drop. Topic drop may be syntactically licensed, i.e., the omission occurs in the prefield as required, but it may still fail if the hearer cannot retrieve the omitted constituent or its reference, as in example (13) above. I assume that recoverability is gradual in nature, i.e., that topic drop can succeed more or less well depending on how easy it is to resolve the ellipsis, which in turn depends on how much cognitive effort is required to do so. Such a link between the usage of topic drop and assumed underlying cognitive processes anticipates the second part of this book. In Chapter 6, I argue that the usage of topic drop is guided by the intention to distribute processing effort efficiently across the utterance and that this is also associated with recovering the omitted constituent based on cues in the following context. Therefore, I come back to recoverability in Section 6.3.

Part II

Usage of topic drop

5 Previous accounts of topic drop usage

While the first part of this book was concerned with the licensing of topic drop and its syntactic properties, in the second part, I address the question of when speakers *use* topic drop, assuming it is licensed. Occasionally, I still cover aspects concerning licensing or syntax in this part since many authors do not sharply separate licensing and usage, and such a separation is not always easy either.

Before I propose an information-theoretic account of topic drop usage in Chapter 6, in this first chapter, I present existing approaches to topic drop usage, approaches that try to explain when and why topic drop is used, and what function it serves. First, I discuss a classic explanatory approach to ellipsis in general, namely avoiding redundancy and achieving linguistic economy. Then, I focus on the function of topic drop in creating cohesion, as well as the function of topic avoidance. Afterward, I discuss what I termed socio-pragmatic functions, which include expressiveness, condensation, action orientation, and the usage in completions, elaborations, and responses, as well as rhetorical and social functions. Finally, I deal with the potential role of prosody.

5.1 Redundancy avoidance and economy

The first explanation of why speakers use topic drop is a classic one, namely ellipsis as a means of economy, which allows speakers to avoid redundancy. Merchant (2001: 1) puts it as follows: "Elliptical processes capitalize on the redundancy of certain kinds of information in certain contexts, and permit an economy of expression by omitting the linguistic structures that would otherwise be required to express this information." According to Schwitalla (2012: 103), a speaker does not need to say more than the hearer needs to understand the intended message.

Helmer (2016: 167) applies this reasoning to topic drop and distinguishes two cases related to linguistic economy: (i) Speakers can omit a constituent if it is redundant because all interlocutors know based on the previous discourse how to understand the elliptical utterance anyway. (ii) Speakers can refrain from explicating a topic if it would be uneconomic to boil down a complex, global discourse

topic. By (ii), Helmer (2016) primarily explains the usage of what she terms *indirect analepsis*, and what I termed *indirect topic drop* in Section 4.1, that is, cases where determining the reference requires more complex inference processes. In (i), one can argue that topic drop is economical both for the speaker and the hearer because the speaker does not need to produce redundancy and the hearer does not need to process it. However, Helmer (2016: 45) argues that from the view of redundancy avoidance, topic drop is at first glance not as easy as possible to understand for the hearer (she quotes Sperber & Wilson's (1986) relevance theory here) since the hearer must at least think about what the antecedent of the ellipsis is.[1] As there is often one very salient antecedent in the discourse, this will most of the time not result in increased processing effort. But this is different for (ii): (ii) is mainly economical for the speaker as they do not need to explicate a potentially complex concept. In contrast, the hearer may even be confronted with additional processing effort as they need to infer this intended concept (Helmer 2016: 169–170).

Poitou (1993) argues that redundancy avoidance is in particular a useful explanation for the omission of the expletive *es*, which is semantically empty and, as he puts it, in some sense superfluous since it simply serves the purpose of filling the prefield. However, he also points out that there are limits to the explanation of redundancy avoidance. Often even shorter expressions would be possible, for example, telegraphese (Poitou 1993: 123), but speakers still do not use them. For instance, in example (1), topic drop allows the speaker to omit the prefield constituent *ich*, but in telegraphese and potentially also in spoken language the even shorter form *Schlüssel gefunden*, i.e., verbalizing only the object as bare noun plus the verb participle, would be possible. This shows that for many instances of topic drop, not all constituents that are redundant in some sense are omitted.

(1) *[(Ich) hab den] Schlüssel gefunden.*
I have the key found
'[(I) have] found [the] key.'

However, the limitations of redundancy avoidance as the (only) explanation for the use of topic drop lie not only in the competition with shorter forms but also, conversely, in the presence of full forms. As Helmer (2016: 198) points out, relevance theory, for instance, predicts that when topic drop is licensed, it should always be used, with overt pronouns being avoided. But still, we find cases where

[1] Helmer (2016: 45) points out that it is the cohesive function of topic drop that facilitates processing for the hearer. See Section 5.2 for details.

speakers or writers use the full form when topic drop would be possible. This shows that redundancy avoidance as an explanation for the use of topic drop is plausible but falls short as the only explanation.[2]

5.2 Cohesion

Auer (1993), Sandig (2000), Helmer (2016), and Frick (2017) consider topic drop to be a means of building cohesion. In their seminal book *Cohesion in English*, Halliday & Hasan (1976) state that "[c]ohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it" (Halliday & Hasan 1976: 4, original emphasis). Halliday and Hasan consider ellipsis to be one of the grammatical resources that establish cohesion in texts.

Auer (1993) frames the relationship between cohesion and topic drop in two slightly different ways. On the one hand, he states that positioning the verb in sentence-initial position can be an alternative to using an anaphoric pronoun with the same function of establishing cohesion to the previous discourse (Auer 1993: 199). On the other hand, he argues that topic drop is used in particular when the corresponding utterance exhibits what he considers to be an extremely strong cohesion to the previous linguistic context or the context of the communicative situation (Auer 1993: 203). The subtle difference between both statements is that in the first case, topic drop is a means to establish cohesion, whereas in the second case, topic drop results from the already existing strong cohesion relation. Since Auer (1993) does not distinguish between the two variants, he might assume that both play a role. That is, that topic drop is used when there is already cohesion between two utterances and then makes this relationship explicit and reinforces it if necessary.

The latter point is also made by Frick (2017: 128–129) in the context of object topic drop. She states that it is precisely the omission of the prefield constituent that can reinforce the cohesion to the previous discourse because it signals that

[2] As mentioned above, while one can imagine that the most reduced form of example (1) can not only occur in telegraphese but also in spoken language, it will most likely be rather marginal there. Other reductions in telegrams are probably even impossible in spoken language. Tesak & Dittmann (1991: 1124) have therefore considered telegraphese as a register of its own. Heidolph (1992: 404) argues that it has, like headlines, potentially its own grammar enabling special forms of omission. From this point of view, the argument against redundancy avoidance that has just been put forward can be qualified to some extent. If a shorter form is not available in a register/grammar, it cannot be used as an alternative to topic drop.

the referent of this constituent is known to the hearer. Similarly, Sandig (2000: 300–301) argues that the omission of a verb argument is meant to indicate that the current utterance is still about the referent of that omitted argument, and, thus, indicates local cohesion. According to Helmer (2016: 216), utterances with topic drop are strongly cohesive to a speaker's preturn, frequently to the immediately adjacent one.[3] She states that they add information to the precontext instead of establishing a new topic or a new "communicative project" and, thus, depend on this precontext. Helmer (2016: 45) argues, again with recourse to Sperber & Wilson's (1986) relevance theory, that the cohesive function of topic drop results in a "positive cognitive effect" for the hearer by indicating that the current discourse sequence continues.

In sum, topic drop can be seen as a means to enforce an already existing cohesion relation between two utterances by indicating thematic continuity. I already mentioned that Auer (1993) argues that topic drop is similar to the use of an overt anaphoric pronoun in this case. However, he does not explain when speakers use one and when they use the other. If topic drop strengthens cohesion and facilitates processing for the hearer, it should always be preferred to the overt pronoun, but, as discussed in Section 5.1, this is not the case. Eckert (1998) explains the variation between utterances with overt (demonstrative) anaphoric pronouns and utterances with topic drop as a division of tasks. She states that while "demonstratives are used when reference is made to the previous utterance as a whole and the predicate supplies new information", topic drop occurs "when the semantic information of the verb is somehow given (explicitly or implicitly) and the participants expect a statement to be made about the previous utterance (eg [sic!] agreement or disagreement with it)" (Eckert 1998: 217). I come back to the potential role of the verb for topic drop below in Section 5.4. What emerges from Eckert's functional description, based on an analysis of authentic spoken data, is that both utterances with anaphoric pronouns and utterances with topic drop fulfill cohesive functions, but that utterances with topic drop often link more strongly to the previous discourse by commenting on it evaluatively. I revisit this aspect in Section 5.4, too.

5.3 Topic avoidance

Oppenrieder (1987), Auer (1993), Günthner (2006), Schwitalla (2012), and Helmer (2016) discuss that speakers use topic drop to avoid having to designate a con-

[3] See also the discussion on the usual small distance between antecedent and target in Section 4.1.

stituent in the utterance as the topic and to focus on the comment or rheme. Oppenrieder (1987: 179) speculates that topic drop might be a means to indicate that all remaining constituents are equally rhematic and/or equally important. This claim is called into question by Auer (1993: 204). He argues that the remaining constituents in an utterance with topic drop also vary in terms of how rhematic they are, e.g., pronouns are less rhematic than full NPs, etc. However, he agrees with Oppenrieder in assuming that in utterances with topic drop no constituent is designated as the topic because either the topic is known from the linguistic or extralinguistic context or none of the constituents is "un-rhematic" or thematic enough to function as a topic (Auer 1993: 204). Günthner (2006: 105) states that topic drop as the omission of thematically given information increases the "rhemacity" because the semantically more important information comes first (see also Helmer 2016: 168–169). Similarly, Schwitalla (2012: 103) argues that the omission of constant themes (or topics) allows the hearer to focus their attention on the rhemes. While Helmer (2016) links this strategy to redundancy avoidance and thematic progression, Oppenrieder (1987) connects it to increased expressiveness (see Section 5.4).

However, as shown in Section 3.1.5, topic drop does not only affect topics but also non-topics, such as non-referential expletives. For these elements, the explanation of topic avoidance cannot be valid, since the omitted prefield element is not a topic. However, it could be argued that the omission of a semantically empty element also directs the focus more strongly to the rest of the utterance. But here, too, the question remains of why this strategy is not pursued more often or, more precisely, whenever possible, i.e., why there are still utterances with topical elements or expletives in the prefield at all. Topic avoidance or rheme focus can thus presumably only be a part of a broader explanation for the usage of topic drop.

5.4 Socio-pragmatic functions

In this section, I discuss jointly those functions of topic drop that it fulfills primarily in spoken or narrative texts according to the relevant literature and which may be termed socio-pragmatic functions.

5.4.1 Expressiveness, condensation, and action orientation

Auer (1993) assumes that the function of topic drop varies according to the type of text it is used in and the linguistic acts as which the corresponding utterances

occur. He argues that topic drop can indicate a switch to narration and states that in the narrative genre, it stresses what he terms the action character of the proposition (Auer 1993: 219). More specifically, he suggests that it highlights the semantics of the finite verb in the left bracket, marking action orientation in a general sense. While topic drop does not necessarily draw the hearer's attention to a concrete action, since at least for analytic verb forms, the semantic content of an action is expressed through the infinite verb part in the right bracket (Auer 1993: 218), it often does so in the present tense, as Günthner (2006: 101) points out. Accordingly, the frequent combination of the narrative present tense and topic drop allows the speaker to focus on the concrete action, which leads to the plot advancing. According to Günthner (2006: 104), the combination of the present tense and topic drop, which she considers a technique of verb foregrounding, has the effect of "liveliness, condensation and expressiveness of the utterance."[4] Similarly, Sandig (2000: 302) interprets topic drop as a conventionalized form of expressing dramatics and pace.[5]

In the last two statements, several claims about the effect of topic drop seem to be condensed and partly intermingled. First, the focus on the verb and the action, including the concurrent omission of the agent or patient of the action, can indeed result in the effect of an increased pace of the narration – one event follows the other. In this respect, there may also be a similarity to the likewise action-oriented drama, which can justify the denotation as "expressing dramatics". The term "liveliness", which Günthner (2006: 104) uses, may be a more subjectively biased word for this, which already anticipates the aspect of expressiveness discussed below.

Second, Günthner (2006: 104) argues that topic drop has the effect of condensation, which is not surprising given that omitting something always results in a shorter and denser structure. Auer (1993) describes this process in more detail by stating that topic drop leads to a closer connection between two adjacent utterances that does not endanger their semantic or syntactic autonomy. Furthermore, he links topic drop to the global strategies of condensation (analogous to syntactic subordination) and fragmentation (analogous to syntactic parataxis), which are at work in spoken language (Auer 1993: 219).

Third, Günthner's (2006) statement about the expressiveness of topic drop is probably the vaguest one. Alongside her, Oppenrieder (1987: 179–180) uses the term in a similarly unspecific way by stating that the strategy of "rhematizing"

[4] My translation, the original: "Lebendigkeit, Dichte und Expressivität der Äußerung" (Günthner 2006: 104).

[5] All three authors, Auer (1993), Sandig (2000), and Günthner (2006), furthermore emphasize the role of prosody and rhythm for topic drop (see Section 5.5).

the utterances leads to "a strongly expressive coloring", which is enforced by the omission of a functionless prefield constituent. What is meant by this expressiveness and why exactly it is supposed to be evoked or enhanced by topic drop remains unclear in both authors. Here Poitou (1993: 127–128) is somewhat clearer, at least specifying what he means by expressiveness, namely that the speaker also communicates their own (emotional) attitude. According to him, this function of topic drop would be expressed with modal particles or intonational patterns in the complete V2 sentence.

Auer (1993) lists five so-called conversational environments in which topic drop occurs. The last one is the already discussed narration while the first and second ones are directly related to expressiveness. First, topic drop is said to occur in modalizations. According to Auer (1993: 207), these are statements in which the speaker expresses their attitude to a proposition by judging the truth, the probability, the reliability, etc. of the information contained in these statements. Second, Auer (1993: 208) lists evaluations as a further environment. They also express the speaker's attitude but not in terms of truth values or probability but in terms of aesthetics and morality. Here, we seem to run into a chicken-and-egg problem. Is topic drop a means of expressiveness, as proposed by Oppenrieder (1987), Poitou (1993), and Günthner (2006), or does topic drop preferably occur in expressive environments, i.e., in statements with expressive function, as proposed by Auer (1993)?

5.4.2 Elaborations/reformulations and responses

The last two conversational environments discussed by Auer (1993) in which topic drop occurs preferentially also establish a relation to the previous utterance. They are elaborations or reformulations and responses. According to Auer (1993: 209), the former are "non-corrective (mostly self-)repairs."[6] Here, the speaker specifies, exemplifies, or reformulates their own previous utterance or (more rarely) the utterance of their interlocutor. Auer (1993) illustrates this function with example (2), an excerpt from an investment consultancy.

(2) a. *A: Was nich schlecht is s Siemens*
what not bad is is Siemens
A: 'What's not bad is Siemens'
b. B: m'm
c. *A: Also* Δ *isn, isn, isn starkes Papier*
so that is.a is.a is.a strong paper

[6] My translation, the original: "nicht-korrigierende (meist Selbst-)Reparaturen" (Auer 1993: 209).

A: 'So (this) is a strong security' (Auer 1993: 209, simplified)

A's second utterance *isn starkes Papier* is an addition or specification to their first utterance *was nich schlecht is s Siemens*. Auer (1993: 210) emphasizes that elaborations or reformulations can be grouped with the modalizations and evaluations mentioned above. In all three types of conversational environments, he argues, there is a particularly strong sequential link between the precontext and the actual utterance, without the actual utterance starting a new linguistic action or being more relevant. Therefore, Auer characterizes them as backward-looking.

According to Auer (1993: 212), this is somewhat different for the group of responses because the responses are at least as relevant as the question or, more generally, the preceding utterance to which they refer – since they do not need to answer an explicit question. This is shown in example (3), where the response follows a request. Auer (1993) argues that the type of conversational responses with topic drop that he refers to are similar to the modalizations in that the same verba sentiendi are used. He further speculates that the responses could be captured as modalizations or elaborations of an implicit or explicit direct yes-or-no response and thereby connects four of the five mentioned conversational environments (Auer 1993: 212).

(3) a. *A: Grüß alle!*
greet all
A: 'Say hello to all!'
b. *B:* Δ *Mach ich, du auch!*
that make I you too
B: 'I will do (that), you too!' (Auer 1993: 212, simplified)

5.4.3 Completion or continuation

Poitou (1993) discusses a further function of topic drop that can be termed completion. He, too, makes the observation, discussed earlier in this book, that topic drop frequently occurs in utterances that are immediately adjacent to the utterance with the antecedent (Section 4.1). He adds, however, that this often takes place in dialogues and in such a way that the speaker who utters the utterance with topic drop would, in a sense, continue the utterance of the previous speaker (Poitou 1993: 125). About half of his instances of subject topic drop, he says, are structured this way, and the vast majority of instances of object topic drop. He states that in these cases, the utterance with topic drop continues and completes the preceding utterance or simply follows on from it. Or in more flowery terms,

he puts it: "So it's as if speaker B's sentence had actually begun before B starts speaking"[7] (Poitou 1993: 125).

Auer (1993: 202–203), on the other hand, disputes that the cases of collaborative turn constructions, as he terms them, occur frequently. He describes them as cases in which a speaker does directly continue or complete the utterance of the previous speaker with a topic drop so that both utterances together form one syntactic structure. Auer even goes so far as to consider examples like (4), from Fries (1988), as irrelevant to an analysis of topic drop. He argues that in such cases, both utterances F and A need to be considered in combination, as one syntactic structure, and that there is no topic drop because the verb would be in the second position in the complete structure, as usual.

(4) a. *F: Mir las sie immer Simone de Beauvoir vor und dem*
me.DAT read she always Simone de Beauvoir VPART and the.DAT
Fritz?
Fritz
F: 'She used to read Simone de Beauvoir to me, and to Fritz?'

b. *A: Δ Liest sie zur Zeit "Wie kommt das Salz ins Meer?"*
him.DAT reads she to.the time how comes the salt in.the sea
vor.
VPART
A: 'She is currently reading 'Why Is There Salt in the Sea?' (to him).'
(Fries 1988: 31, cited in Auer 1993: 202)

While I do not agree with this interpretation and also analyze examples like (4) as structures with topic drop,[8] I would like to support Auer's view that such question-answer sequences with topic drop, as they are much discussed especially in (mainly) introspectively oriented works like those of Fries (1988) or also Trutkowski (2016), often seem constructed and occur rather rarely in corpora of natural language.[9] Thus, it seems at least unlikely that the function of topic drop is to complete previous utterances.

[7] My translation, the original: "Es ist also so, als hätte der Satz des Sprechers B eigentlich schon begonnen, bevor B zu sprechen beginnt" (Poitou 1993: 125).

[8] Although topic drop of a dative object is a rare case anyway.

[9] Remember that Poitou's (1993) corpus contains an undefined portion of literary texts, in which examples like (4) may be deliberately used as stylistic devices.

5.4.4 Rhetorical and social function

Besides the functions redundancy avoidance and cohesion, Helmer (2016) also discusses a rhetorical and a social function of topic drop. She states that topic drop can be used to allow the hearer to continue the individually preferred topic, i.e., a certain vagueness lets the hearer continue in their desired way (Helmer 2016: 174). She exemplifies this with (5), taken from an interview, where she argues that the question with topic drop is so vague that the interviewee has a large freedom in answering.

(5) a. *Nun sagt man ja oft Frauen nach, also man, Männer sagen das oft Frauen nach, dass Frauen sehr oft an sich zweifeln. Kann ich das überhaupt?* [...] *Da wächst mir was aufn* [sic!] *Kopf. Nein, lieber nicht. Lass ich die Finger davon.*
'Now it is often said of women, well, men often say of women that women very often doubt themselves. Can I do that at all? [...] Something is becoming too much to handle for me. No, better not. I'll leave it alone. '

b. Δ *Hat bei Ihnen keine Rolle gespielt?*
that has at you.2SG.POL no role played
'(That) didn't matter to you?' (Helmer 2016: 175, adapted)

However, it seems to me that an utterance with an overt pronoun *das* in the prefield would be likewise vague, i.e., offering several possible interpretations as it can refer back to different referents. Therefore, it is dubious whether the rhetorical function of the utterance does in fact hinge on topic drop.

Furthermore, Helmer (2016: 177) assumes that certain utterances with topic drop function more or less as formulas or idioms, or more precisely, she treats them as social action formats following, e.g., Thompson & Couper-Kuhlen (2005) and Fox (2007). Social action formats are linguistic expressions frequently used to perform certain social actions such as requests, suggestions, or demands, with their propositional content receding into the background (see also Deppermann 2021). Helmer (2016: 177) argues that in her data set, social action formats with topic drop like those in (6) are often a means to express empathy while at the same time terminating the current discourse topic. Again, she argues that topic drop allows for semantic vagueness, which in these cases has the function of expressing understanding.

(6) a. Δ *Is ja auch nich so einfach immer.*
it is PART also not so easy always

'(It) is not always that easy.' (Helmer 2016: 178, shortened and simplified)

b. Δ *Wird schon*
that will already
'(It) will be fine' (Helmer 2016: 179, shortened and simplified)

While this analysis seems reasonable, again it is not clear to me why only or mainly topic drop and not the corresponding full forms with the vague anaphoric pronoun *das* ('that') should have this function.

5.5 Prosody

Related to the effect of liveliness discussed in Section 5.4.1, Auer (1993: 218) argues that this effect is enhanced in spoken language through a special rhythmic structure, which is also discussed in Günthner (2006: 104). They state that utterances with topic drop differ from the full forms in that they lack an anacrusis. They do not begin with an unstressed or weakly stressed syllable, as the full forms with pronouns do, but directly with a more or less strongly stressed finite verb. Sandig (2000: 300) agrees with this observation and states that the finite verb is more strongly stressed than it would be with a preceding pronoun. She argues that the resulting structure is then more strongly intonationally marked and that a two-peaked tone pattern is created; however, she does not explain what the two peaks consist of.

Helmer (2016) investigated prosodic differences between utterances with topic drop and utterances with an anaphoric pronoun using a data set based on the research and teaching corpus of spoken German – FOLK (Schmidt 2014). It consists of 321 instances of topic drop and 200 instances that contain an overt anaphoric pronoun (Helmer 2016: 217). In her rather concise presentation, which leaves some methodological and terminological questions unanswered, Helmer concentrates, among other aspects,[10] on the position of the focus accents in both types of utterances. She notes that in utterances with an overt anaphoric pronoun, the focus accent is often (in about 19.5% of the instances) on the subject in predicative constructions of the form *das Kopula NP* ('that copula NP') but is also placed frequently (here, Helmer does not provide a proportion) on predicative adjectives in constructions of the form *das Kopula ADJP* ('that copula ADJP'). From

[10] See Helmer (2016: 216–227) for her complete comparison between utterances with topic drop and utterances with an anaphoric pronoun with respect to prosody, which also includes the rhythm and the length of the utterances, as well as the relation between prosody and social interaction.

this, she concludes that in utterances with anaphoric pronouns often evaluative elements or elements in comments are stressed. In turn, in utterances with topic drop,[11] the focus accent is placed most often on the lexical verb, in about 36.5% of the cases, as Helmer (2016: 217) reports. Furthermore, in many cases (again Helmer does not report a proportion), there is a verum focus (e.g., Höhle 1992) on the full verb or adverbs and adjectives are focused (Helmer 2016: 217–218). At this point, it remains unclear from Helmer's (2016) statements whether all cases where the lexical verb is accented are cases of verum focus and if not, what the other cases are. Helmer (2016: 218) points out that in the utterances with topic drop, mostly elements are stressed that mark the responsivity of the utterance, as in the verum focus cases, where the prosodic structure emphasizes that the speaker considers the previously uttered proposition to be true.

In a second step, Helmer (2016: 218) investigated the position of the focus accent and differentiated initial, medial, and final position, as shown in Table 5.1.

Table 5.1: Position of the focus accent in utterances with topic drop vs. utterances with an anaphoric pronoun in Helmer's (2016) data set, taken from Helmer (2016: 218)

Position of	Full form		Topic drop	
focus accent	Frequency	Proportion	Frequency	Proportion
Initial	17	8.5%	96	29.9%
Medial	103	51.5%	98	30.5%
Final	80	40%	127	39.6%

The initial cases, according to her, are those in which the first element of an utterance bears the focus accent. For final and medial, she is less clear, but it can be assumed that the final cases are accordingly those in which the very last element of an utterance is focused, and medial are all cases in which it is an element somewhere in between. While both in utterances with topic drop and with an anaphoric pronoun the focus accent is final in about 40% of the cases, the values for initial and medial positions differ significantly between elliptical

[11] Recall that Helmer (2016) considers topic drop to be possible also in the middle field (see Section 3.2 of this book for a critical discussion). This means that her statements about the position of focus accents do not only refer to utterances with a V1 word order. Since in her complete data set, which contains not only 321 instances from the FOLK but also 220 instances from the corpus Gespräche im Fernsehen ('conversations on TV'), there are 55 instances of what she describes as topic drop in the middle field, there are at most 55 instances of topic drop in the middle field in the data set discussed here but presumably less.

and non-elliptical utterances, as Helmer (2016: 218) points out. In utterances with topic drop, the focus accent is significantly more often on the utterance-initial element than in utterances with an anaphoric pronoun, where it is put significantly more often on elements in the utterance-medial position. This partly supports the claims by Auer (1993), Sandig (2000), and Günthner (2006). In comparison to their non-elliptical counterparts, utterances with topic drop more frequently start with a stressed element, which is usually the finite verb (except for Helmer's alleged cases of topic drop in the middle field). However, the fact that still in about 70% of the cases, the focus accent is not on the initial element but later in the utterance does not indicate a very strong tendency to stress the initial verb.

It can be concluded that while topic drop allows the focus accent to be on the initial verb, this possibility is not always used. If it is, the initial focus accent draws attention to the verb and the action, which brings us back to the effects of topic drop outlined in Section 5.4. Thus, prosody seems less able to explain why topic drop is used than to be an additional factor that may, under certain circumstances, enhance the actual functions of topic drop, such as action emphasis. In addition, it is at least not immediately evident to what extent prosody is relevant to topic drop in written (if also conceptually spoken) text types. Féry (2006), for instance, argues that the human parser not only takes syntax into account but also builds up an unmarked prosodic structure when processing silently read sentences, which defaults to wide focus.

5.6 Summary: previous accounts of topic drop usage

I started this chapter on previous accounts of topic drop usage with the most classic explanation that can probably be applied to any elliptical phenomenon: linguistic economy through the avoidance of redundancy. This explanation is also a main component of my information-theoretic account of topic drop usage, sketched in Chapter 6. Following Levy & Jaeger's (2007) *uniform information density hypothesis*, I argue that predictable expressions are more likely to be omitted by speakers to achieve communicative efficiency. This probability-driven operationalization of economy may contribute to explaining the limits of the redundancy avoidance approach sketched above.

What my information-theoretic account does not explicitly consider are the potential socio-pragmatic functions of topic drop, as well as its function as a cohesive device. As described above, they partly involve a kind of chicken-and-egg problem. Is topic drop indeed deliberately used to create cohesion or to fulfill a certain socio-pragmatic function, or does, in turn, a higher cohesion or a certain

socio-pragmatic function lead to more uses of topic drop? In any case, it seems to be reasonable that topic drop is associated with these functions, functions that play a more important role in spoken and narrative texts. Since such texts are not the focus of the present book, it does not contribute to this issue other than encourage future research. The question of whether and how socio-pragmatic factors can be integrated into my proposed information-theoretic account must, thus, remain equally open.

The same applies to prosody, which is naturally restricted to spoken language to an even stronger degree. For this factor, Helmer (2016) evidenced that topic drop is at least not consistently used only as a device to stress the sentence-initial verb.

In this chapter, I furthermore argued that topic avoidance and completion/continuation only play a marginal role, if any, in motivating the usage of topic drop since non-topical constituents are also omitted and it is rare in authentic speech data that one interlocutor completes the utterance of another one.

For the rhetorical and social function that Helmer (2016) discusses as well as for the use in certain conversational environments that Auer (1993) lists, it remains unclear whether topic drop is really needed to fulfill the respective functions or whether the corresponding full forms would work equally well. Here, more research in the vein of Eckert (1998) and Helmer (2016), who focus on usage differences between utterances with topic drop and corresponding full forms, is needed. In the following chapters, I turn to my information-theoretic approach, its implications, and my empirical investigations.

6 Information-theoretic account of topic drop usage

To explain the usage of topic drop, I propose an account that builds on concepts from information theory.[1] According to this account, the choice between topic drop and the corresponding full form is guided by the predictability and recoverability of the preverbal constituent. If the constituent is predictable from the preceding context or if it can be easily recovered, e.g., through the following verb, speakers are more likely to omit it because this reduces the processing effort for the hearer.

In order to be able to expound my account, I first provide an overview of information theory and its most important concepts. In the second step, I turn to the influential *uniform information density hypothesis* (*UID*) proposed by Levy & Jaeger (2007), which is the basis for my information-theoretic account. I discuss two central predictions that follow from this hypothesis and their implications for topic drop: the principle of avoiding troughs, i.e., regions of low information, and the principle of avoiding peaks, i.e., regions of high information. In addition to the *UID hypothesis*, my information-theoretic account considers recoverability as a further factor that impacts the usage of topic drop. More specifically, I build upon the definition of recoverability developed in Chapter 4 to further expand the link to processing effort.

6.1 Information theory: basic concepts

Information theory is a mathematical theory of communication based on Claude Shannon's seminal 1948 paper *A mathematical theory of communication* (Shannon 1948). Shannon addresses the question of how signals of information can

[1] Parts of this chapter appeared in a similar form in Schäfer (2021), such as the presentation of the *UID hypothesis* with the *avoid troughs* and *avoid peaks* principles, its application to topic drop, and the additionally assumed concept of the *facilitate recovery* principle. However, they were specified and extended.

be accurately and efficiently transmitted from a technical, engineering perspective (see also Weaver 1949). He explicitly states that "semantic aspects of communication are irrelevant to the engineering problem" (Shannon 1948: 379). At first glance, this statement seems to call into question whether information theory can be applied at all to linguistics in general and semantics, pragmatics, and psycholinguistics in particular, where the focus is on the meaning of messages exchanged in communication. While the concept of information that underlies Shannon's information theory is indeed distinct from that used in linguistics (see below), already Weaver (1949: 24–28) notes that the technical problem of communication overlaps with semantic and pragmatic issues and that information theory could also be applied to them. For example, he proposes the application of Markov processes, a special kind of stochastic process (Shannon 1948: 385), to handle the impact of context on meaning (Weaver 1949: 28), and suggests understanding the channel capacity, the amount of information that can be successfully sent over a channel (see below), not only as a technical limitation but also as a cognitive limitation on the part of the audience (Weaver 1949: 26–27).[2] Consequently, there has been a fruitful application of information-theoretic concepts in linguistic and psycholinguistic research, with beginnings going back even before Shannon (1948) and his formalization, such as Zipf's (1935) famous law about the inverse relationship between word length and word frequency (see also Mandelbrot 1965). Before discussing the application of information theory to topic drop, I first introduce the most central concepts of information theory.

6.1.1 Information

In linguistics or everyday language use, the concept of *information* is usually closely related to meaning. For example, the dictionary Merriam-Webster defines information as "knowledge that you get about someone or something; facts or details about a subject" (Merriam-Webster.com Dictionary 2021). In linguistic terms, "[t]o inform a person of something" can be considered "to induce a change in that person's knowledge state by adding one or more propositions" (Lambrecht 1994: 44) (this view is related to the concept of common ground by Stalnaker (1974, 1978, 2002), see Section 3.1.1 for a discussion). Lambrecht (1994: 43) points out that whether the meaning expressed by an utterance is informative depends on the communicative situation, i.e., on whether it changes the hearer's mental representation of the world or not. In this sense, then, information is related to meaning because for a proposition to change the hearer's knowledge, it needs to

[2]This anticipates the reinterpretation of the channel capacity as a limit to the hearer's cognitive resources, which I assume in this book. I come back to this below.

be in one way or another "about" the world. In information theory, however, *information* is a purely probabilistic concept, independent of facts about the world. An event x, in particular an occurrence of an expression, is more informative the less probable it is.

Following Shannon (1948), the information content (I), also termed Shannon information or surprisal (S) (Samson 1953, Attneave 1959, Hale 2001),[3] of an event x is mathematically defined as the negative logarithm to the base 2 of the probability of x:

$$I(x) = S(x) = \log_2 \frac{1}{p(x)} = -\log_2 p(x) \tag{6.1}$$

Information or surprisal is measured in the unit *bits*. 1 *bit* corresponds to the amount of information that is contained in a choice between two equally possible options (Shannon 1948: 380), like, e.g., a coin flip with the two options heads and tails, as shown in equation 6.2.

$$I(\text{coin flip}) = -\log_2 \frac{1}{2} = 1 \textit{ bit} \tag{6.2}$$

In particular in linguistics, the formula is often modified in the following way to express that the probability of x usually depends on its context (see also Levy 2008: 1130):

$$I(x) = S(x) = -\log_2 p(x \mid context) \tag{6.3}$$

According to Shannon (1948: 379–380), using the logarithmic measure instead of bare probabilities is "practically more useful", more intuitive, and simplifies mathematical operations. The inverted polarity in (6.1) and (6.3) has two logical implications as Lemke (2021: 156) points out. First, this way, the surprisal never becomes negative and, second, it ensures that the amount of surprisal decreases with the likelihood of the event x.

This event x can be any linguistic unit, from very basic units like a phoneme, a morpheme, or a syllable, to larger units like a word or a phrase, to complete utterances or messages, which I use as examples below. By message, I mean an abstract preverbal, nonlinear, and propositional representation of what a speaker wants to communicate (Konopka & Brown-Schmidt 2014: 8). For example, Levelt (1989: 101) expresses the message underlying the question (1a) formally as (1b),

[3]The term *surprisal* was reintroduced into linguistic research by Hale (2001: 159). While Hale attributes it to Attneave (1959: 6), Attneave himself points to Samson (1953: 293), who seems to be the creator of the term in this usage. In this book, I use the terms (Shannon) information and surprisal interchangeably.

using a bracket structure to indicate thematic roles, tense, and sentence mood (see also Konopka & Brown-Schmidt 2014: 9, for a review on further approaches to how messages are structured).

(1) a. Did John fall?
b. ?(PAST(FALL(JOHN))) (Levelt 1989: 101)

The concrete lexicalization of a message is termed a *signal* in information-theoretic terms and corresponds to an utterance in practice, e.g., in the experimental studies of the following chapters. In the following, I surround utterances or signals with quotation marks and set messages in capitals.

I illustrate the concept of surprisal with the following toy example: Imagine after several years of relationship, person A decides to propose to person B and asks, "B, will you marry me?". Person B now has several options to answer, of which YES and NO are probably the most prominent messages in this highly ritualized conversational situation.[4] Normally, one can assume that A will propose to B only if A is relatively sure that B will say "yes". Therefore, YES is the more likely and less informative message, we could say:

$$\begin{aligned} & p(\text{YES} \mid \text{proposal}) > p(\text{NO} \mid \text{proposal}) \\ \Rightarrow & -\log_2\ p(\text{YES} \mid \text{proposal}) < -\log_2\ p(\text{NO} \mid \text{proposal}) \\ & \Rightarrow I(\text{YES}) < I(\text{NO}) \end{aligned} \tag{6.4}$$

By contrast, the message NO would be the less likely answer in this situation, having greater Shannon information or higher surprisal.

Compare the proposal situation to a situation where the couple wants to have dinner at a diner. A waiter comes to the table and asks: "What can I get you?" Assume for the sake of simplicity that it is the couple's favorite diner, where they like all 20 dishes on the menu equally, and that the waiter knows this. Then all messages related to dishes on the menu are about equally likely (1/20 in this case) and thus equally informative. Suppose that person A is not very hungry and

[4] As Levelt (1989: 90) points out, it is not straightforward how the messages underlying the utterances "yes" and "no" would actually look like in a formal representation such as the one proposed by him. Reich (2003) formalizes them by means of two rhetorical relations `agree` and `disagree`, which function as binary operators and include a context variable Γ, which anaphorically takes up the meaning of the preceding question. An answer "yes" to the question "B, will you marry me?" would then be formalized as `agree`(Γ, that B will marry A), additionally presupposing that *p = that B will marry A* is already given. See Reich (2003: 189–195) for more details. However, for the sake of simplicity, I refrain from using this formalization here and represent the answer particles just like the corresponding utterances.

therefore orders only one serving of fries with the message SPEAKER(WANT (FRIES)). The information content of this message would be:

$$\begin{aligned} &-\log_2\ p(\text{SPEAKER(WANT(FRIES))} \mid \text{main dish order}) \\ &= -\log_2\ \frac{1}{20} \approx 4.3\ \textit{bits} \end{aligned} \tag{6.5}$$

Assume further that person B orders a cheeseburger and is asked which side dish should be served with it, fries or coleslaw. Suppose that B takes fries half the time and coleslaw the other half. Then the messages SPEAKER(WANT(FRIES)) and SPEAKER (WANT(COLESLAW)) would both have a probability of about ½ in the context of ordering the side dish. If B responds with basically the same message as A, i.e., SPEAKER(WANT(FRIES)), the surprisal of this message would be:

$$-\log_2\ p(\text{SPEAKER(WANT(FRIES)} \mid \text{side dish order}) = -\log_2\ \frac{1}{2} = 1\ \textit{bit} \tag{6.6}$$

Thus, the same message is less informative in the side dish situation than in the main dish situation because the number of (equally likely) competing messages is smaller, and the probability that this message will be selected is higher.

In summary, the surprisal of a message is higher the less likely the message is. This means that in the special case where all possible messages are equally likely, the surprisal increases with the number of competing messages. If there is only one possible message, its probability is 1, thus, the surprisal in this case is 0 – the message conveys no information.

6.1.2 Communication and channel

According to Shannon, "[t]he fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon 1948: 379). He illustrates this process schematically by presenting the general communication system in Figure 6.1 (Shannon 1948: 381).[5]

In communication, an information source first selects a message to be communicated from a set of possible messages. This message is then converted into a signal in some way by the transmitter and sent over a channel with a limited capacity. The signal is received and decoded back into a message by the receiver. Finally, the corresponding message gets to its destination, i.e., the intended addressee.

[5] I recreated the figure in Microsoft Word.

Referring to Weaver (1949: 7), I illustrate Shannon's (1948) model using (spoken) communication. Suppose we are in the restaurant situation described above, and the speaker wants to tell the waiter their order. The information source is the speaker's brain, which chooses a message encoding the desired dish from a set of possible messages, e.g., (SPEAKER(WANT(FRIES)):

$$m = \{(\text{SPEAKER (WANT(FRIES))}, (\text{SPEAKER(WANT(BURGER))}, \\ (\text{SPEAKER(WANT(PIZZA))}, ...\} \tag{6.7}$$

This message is then transferred by the articulatory system, which functions as the transmitter, into a concrete signal consisting of sound waves, such as "Fries, please".[6] The sound waves are sent through the air, the channel, and get to the waiter's ear, the receiver. Here the sound signals are decoded by the auditory system and then reach the waiter's brain, the destination.

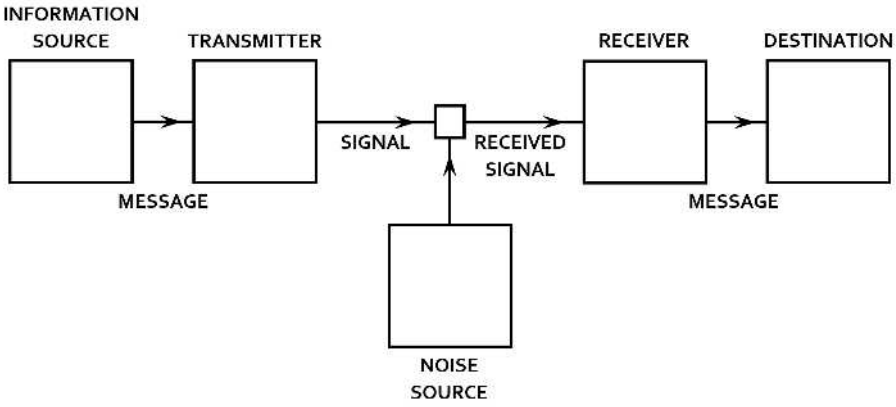


Figure 6.1: Scheme of a general communication system according to Shannon (1948), recreated from Shannon (1948: 381)

There are two properties of the communication channel that need to be discussed as they are important for the successful and efficient transmission of information. First, a communication channel can be noisy. This means that during

[6]This is of course an extremely simplified description of the complex process of speech production. It disregards important aspects of speech production like how a message is conceptualized, how the message is actually transferred into linguistic units, or how these units are arranged relative to each other. This is, however, not the place to discuss them in detail. See, e.g., Goldrick et al. (2014) for detailed overviews of these topics.

transmission, the signal may get corrupted by noise so that the received signal is distorted or erroneous compared to the original signal (Shannon 1948: 406). In the ordering example, the restaurant may be crowded with many people chattering and clattering dishes and cutlery. This noise can distort the signal as it is transmitted from the speaker to the hearer, so that instead of "Fries, please", the waiter might understand the signal as something else, such as "Rice, please". To avoid such a misunderstanding, speakers can mitigate the effect of a noisy channel by adding redundancy to the signal (Shannon 1948: 410), e.g., they could say "I would like to have a serving of your golden, crispy French fries please". The additional adjectives increase the chance that the waiter will understand the intended message correctly.

The second relevant property of a communication channel is that it has a certain capacity, i.e., "the maximum possible rate of transmission" (Shannon 1948: 410) at which signals can be sent over the channel. This capacity is crucial for Shannon's (1948) theorem of communication over a noisy channel. If the source sends bits per second at a rate that is equal to or lower than the channel capacity, the error rate in the signal can become arbitrarily small with suitable coding systems (Shannon 1948: 411; see also Weaver 1949: 21). This, however, is not possible if the bits per second sent by the source exceed the channel capacity. Any attempt to transmit at a rate higher than the channel capacity will result in an even greater increase in the error rate (Shannon 1948: 410). It follows that communication over a noisy channel is guided by a trade-off between communicating efficiently, i.e., sending as much information as possible, and being understood, i.e., preventing as many errors as possible. Therefore, the optimal transmission rate should be close to the channel capacity so that as many bits per second as possible can be transmitted, but it should not exceed it to prevent errors in the signal.

Above, I already briefly mentioned Weaver's (1949) idea of a reinterpretation of this technical concept of channel capacity. Weaver assumes that also the audience, i.e., the hearers or addressees in communication, have a certain capacity and that "if you overcrowd the capacity of the audience you force a general and inescapable error and confusion" (Weaver 1949: 27). This sensible transfer, which Weaver proposed just one year after the publication of Shannon's (1948) seminal paper, can be summarized as follows in light of the information-theoretic concepts already discussed: Given that surprisal is known to index processing effort (Hale 2001, Levy 2008, Demberg & Keller 2008), the channel capacity can be interpreted as an upper bound to the processing resources that are available to the hearer for language understanding (Lemke 2021: 254; see also Fenk & Fenk 1980). That is, there is some sort of hearer- and probably also situation-dependent

threshold that determines the amount of surprisal a hearer can process with the cognitive resources available to them. If the amount of surprisal contained in an utterance is too high, this capacity is exceeded, i.e., the hearer's cognitive resources are no longer capable of processing all of the input, which results in, e.g., longer processing times and additional comprehension difficulties. While an utterance such as "Fries, please" is not likely to exceed the hearer's channel capacity in the ordering situation, the utterance "Can I also have sauce for home?" might. In the context of the utterance situation, this question is less predictable than an utterance that contains an order. It could be that it leads to an overload of the waiter's processing capacities resulting in a delayed response or a clarification question such as "So, you are asking whether you can buy or get some of our burger sauce to use at home?". Speakers can minimize such processing difficulties if they construct their utterances to conform to the so-called *uniform information density hypothesis*.

6.2 *Uniform information density* (*UID*)

The central idea of the *uniform information density hypothesis* (*UID*) proposed by Levy & Jaeger (2007) and Jaeger (2010) is that rational speakers strive to optimize their production and transmit surprisal as uniformly as possible to their hearers. This is made concrete in Jaeger's (2010: 25) definition of *UID*:

> Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus)

Thereby, information density (ID) corresponds to Shannon information or surprisal per linguistic unit (Levy & Jaeger 2007: 849) or per time (Jaeger 2010: 25), as illustrated in Figure 6.2, which shows three hypothetical ID profiles.[7]

The strategy described by *UID* of distributing surprisal uniformly, or more precisely efficiently (Jaeger 2010), across an utterance is a direct consequence of the trade-off described above between sending as much surprisal as possible over the available channel and minimizing the error rate. Both the profile in (A) and the profile in (B) are uniform in that they avoid an alternation between areas of low surprisal, so-called *troughs*, and areas of high surprisal exceeding the channel

[7] All figures with ID profiles in this book were created using the package ggplot2 (Wickham 2016) in R.

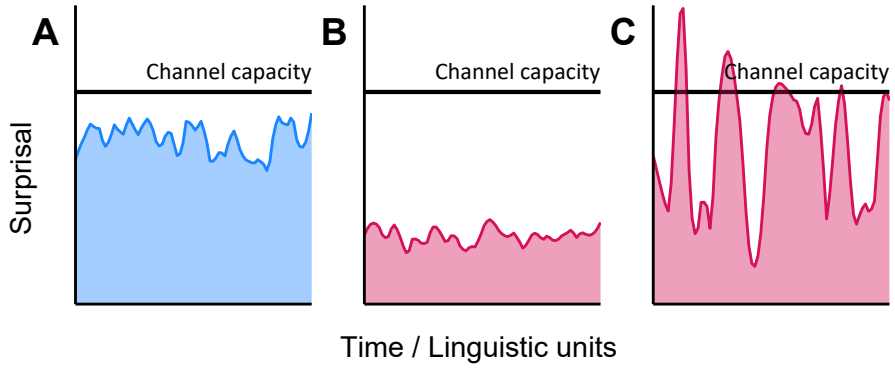


Figure 6.2: Hypothetical information density profiles showing (A) a uniform distribution of surprisal close to channel capacity, (B) a uniform but underchallenging distribution of surprisal, and (C) a non-uniform distribution of surprisal

capacity, so-called *peaks* (Levy & Jaeger 2007: 849). This way, they contrast with the profile in (C), which is much more variable and has visible troughs and peaks, some of which even exceed the hypothetical channel capacity, i.e., the hearer's processing capacities (see above). But there is also a clear difference between profile (A) and profile (B): (A) is more efficient because the transmission rate is much closer to the channel capacity than in (B). The profile in (A) follows the principle of exploiting the channel capacity by sending as much surprisal as possible, whereas in (B) available resources are wasted and the hearer is underchallenged.

With respect to Jaeger's (2010) *UID* definition, another point must be stressed. *UID* is said to operate only "within the bounds defined by grammar". This means that utterances are constrained not only by the distribution of surprisal but also by grammatical rules that a priori exclude variants from the set of possible utterances that are more optimal concerning the distribution of surprisal but are ungrammatical. For topic drop, this means that topic drop is only a valid alternative if it is grammatically licensed, i.e., if it fulfills the licensing conditions sketched in the first part of this book.

It is important to emphasize that while the work by Levy & Jaeger (2007) was highly influential during the rise of information-theoretic approaches in linguistics in the last 20 years, they were not the first ones who developed the idea behind *UID*. Already in 1980, Fenk & Fenk (1980) proposed what was later discussed as the *principle of constant information flow* (Fenk-Oczlon 1989: 536). According to this principle, the lossless transmission of messages requires first an average redundancy level that does not exceed the hearer's short-term memory

capacity and, second, that the information is distributed as uniformly as possible over short time intervals (Fenk & Fenk 1980: 402). Fenk & Fenk (1980: 402) illustrate the latter point by stating that areas with maximal information exceed the hearer's memory capacity while shortfalls of information would not use the available capacity (see also Fenk-Oczlon 1990: 38). This is very similar to *UID*. Both principles say that speakers should communicate efficiently by distributing information at a constant rate without peaks and troughs. That is, such a uniform distribution is "an optimal solution to the problem of *low-effort comprehension*" (Levy & Jaeger 2007: 851, original emphasis) because it minimizes the comprehension difficulty for the hearer. There is, however, a difference between *UID* and the *principle of constant information flow* for peaks and the capacity limit. According to *UID*, speakers should by all means avoid exceeding the channel capacity and smooth surprisal peaks by adding redundancy before the expression that causes the peak. For Fenk & Fenk (1980: 403), in contrast, short-term peaks of surprisal are basically acceptable, as long as they are followed by redundant passages, which allow the hearer to catch up on processing the information excess. This means that according to *UID*, speakers should communicate *below* channel capacity because any exceedance would overuse the processing capacities of the hearer. In contrast, according to the *principle of constant information flow*, speakers can communicate *around* the capacity because short-term peaks can be compensated for by catch-up processing.[8]

While *UID* is explicitly defined to operate on all levels of linguistic representation (Jaeger 2010: 24), there are related hypotheses that have been applied to certain linguistic units only. Genzel & Charniak (2002, 2003) propose the *entropy rate constancy principle* for whole texts and their constituting sentences, according to which it is optimal to communicate at a constant rate that is equal to the channel capacity. They evidence the validity of this principle for written texts in several languages. In the area of phonetics, Aylett and Turk suggest the *smooth signal redundancy hypothesis*, according to which redundancy is distributed uniformly across speech (Aylett & Turk 2004, Turk 2010). They argue that what they term language redundancy and acoustic redundancy are inversely related to each other. Predictable linguistic units such as syllables are articulated less strongly than less predictable syllables in terms of phonetic features such as duration and prosodic prominence.

[8]This also leads to different predictions with respect to topic drop: While, according to Fenk & Fenk (1980), topic drop before a verb with a high surprisal should be easier to process if it is followed by a predictable expression that causes a trough, *UID* does not predict such an effect. In this book, I do not test such predictions motivated by Fenk & Fenk's (1980) approach, but it may be beneficial to do so in future research.

What these presented hypotheses have in common is that they are based on the information-theoretic principle that information, surprisal, or redundancy is best distributed uniformly across utterances, avoiding too much variation in the information density (ID) profile, i.e., the overall distribution of surprisal across an utterance. In what follows, I refer to the *UID hypothesis* for this idea since it has been applied decidedly to other phenomena involving optional omissions in various languages, such as the omission of relativizers (Levy & Jaeger 2007), complementizers (Jaeger 2010, Kaatari 2016), discourse connectives (Asr & Demberg 2015, Yung et al. 2016), and verb phrases (Schäfer et al. 2021) in English, the use of contractions (Frank & Jaeger 2008) and short forms of nouns (Mahowald et al. 2013) in English, null subjects in Russian (Kravtchenko 2014), optional object case-marking in Japanese (Kurumada & Jaeger 2015), complementizer omission in Quebec French (Liang et al. 2021), the use of article omissions (Lemke et al. 2017), fragments (Lemke 2021), and sluicing (Lemke et al. 2022) in German, as well as the reduction of compounds in German (Zarcone & Demberg 2021). Since I am concerned with topic drop in this book, i.e., the omission of constituents, mostly of one-word proforms, I explain the predictions of *UID* at the word level in the following, but they can be easily applied to other levels as well.

There are two general principles concerning the ID profile that can be derived from *UID* and that speakers should follow to achieve efficient communication: *avoid troughs* and *avoid peaks*. To avoid troughs on the utterance level, speakers can omit predictable words, i.e., words with low surprisal. Such an omission is beneficial because redundant words cause undesirable minima of surprisal, i.e., regions way below channel capacity. Therefore, the omission increases efficiency because it avoids that processing resources available to the hearer are wasted. Peaks in the ID profile, i.e., regions exceeding channel capacity, can be avoided by inserting additional redundancy before the peak. A speaker can place a word or several words before the highly informative word that causes the peak to make that word more predictable. This lowers the information maximum, i.e., it reduces the processing cost, ideally to the point where the peak falls below the channel capacity and the hearer's processing capacity is not overloaded. In what follows, I look more closely at both principles and discuss them for topic drop.

6.2.1 *Avoid troughs* principle

Troughs in the ID profile are caused by predictable linguistic units, e.g., by words that are likely given the previous context and, thus, have low surprisal. If speakers include a lot of such predictable words in their utterances, they communicate inefficiently because in these regions of low surprisal, the channel capacity, i.e.,

the hearer's processing capacity, is heavily underused. Speakers can solve this issue and increase the efficiency of their utterances by omitting the predictable expressions where grammar permits it. One way to do this is to use topic drop, if licensed, because it allows speakers to omit prefield constituents with low surprisal. In the optimal case, an utterance with topic drop distributes surprisal more uniformly than the corresponding full form. This is illustrated in Figure 6.3, which shows the hypothetical ID profiles for an utterance with and without topic drop.[9]

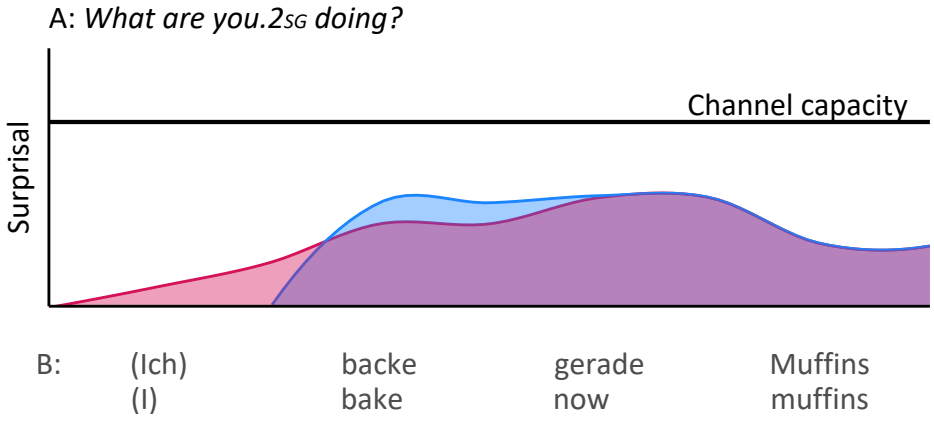


Figure 6.3: Hypothetical ID profile: Since *ich* creates a surprisal trough, the utterance with topic drop is more uniform.

On the x-axis, the words of the utterance *(Ich) backe gerade Muffins* ('(I) am baking muffins right now') produced as an answer to the question *Was machst du gerade?* ('What are you.2SG doing?') are plotted. The y-axis shows hypothetical surprisal values for these words which together form the ID profile of the utterance. The 1st person singular pronoun *ich*, which refers to the speaker, is very predictable because the speaker is prominent both linguistically (through the question 'What are you doing?') and extralinguistically (through the communication situation with speaker and hearer as default roles). Consequently, *ich* creates a trough in the ID profile of the full form, as indicated by the reddish curve. In other words, the surprisal of *ich* is far below the (hypothetical) channel capacity. To avoid this suboptimal trough in the ID profile, speakers can use topic

[9]Since the plots here serve to illustrate the difference in processing effort between predictable and unpredictable words, hypothetical values suffice. Therefore, I also refrain from giving concrete numeric values, which would furthermore depend on the corpus and model used.

drop and omit *ich*, as shown in the blue curve. In this way, surprisal is distributed more uniformly across the utterance, so that the hearer's processing capacity is used at a constant level.

The *avoid troughs* principle can be considered a greedy strategy because it prevents the waste of cognitive capacities and increases efficiency. That is, a speaker should tend to avoid not only severe troughs but every possible trough, as long as (i) grammar permits it and (ii) the omission does not cause a surprisal peak on the subsequent material. For topic drop, this predicts that the prefield constituent is omitted whenever (i) topic drop is licensed and (ii) this does not lead to a peak on the following verb (see Section 6.2.2). As I discuss in the next chapters, the tendency to avoid troughs in the ID profile can explain the impact of several grammatical factors on topic drop observed in the literature.

6.2.2 *Avoid peaks* principle

In addition to the *avoid trough* principle, *UID* also predicts that speakers should avoid peaks of surprisal when communicating. A surprisal peak is caused by an "overinformative" expression whose surprisal exceeds the channel capacity. The information content of this particular expression is so high that the hearer has difficulties processing the expression. Such a peak can be smoothed by adding additional redundancy before the expression that causes the peak. This additional redundancy serves the purpose of making this expression more predictable. For topic drop, the *avoid peaks* principle has the consequence that the omission of the prefield constituent should be depreciated if a verb with high surprisal would move to the sentence-initial position. This is shown in Figure 6.4.

Suppose that A asks *Was macht ihr gerade?* ('What are you.2PL doing?') and that instead of writing *Backe gerade Muffins*, B would use an utterance with topic drop of a 3rd person singular subject such as *Karamellisiert gerade Nüsse* ('Is caramelizing nuts'). First, the less frequent verb *karamellisieren* should generally have a higher surprisal than *backen*. Second, we can assume that the surprisal of *karamellisiert* is even high enough in the context of A's very general question to cause a peak in the ID profile that exceeds the channel capacity, as shown in the reddish curve. This region of high information is expected to challenge A's processing resources.

Assume that although A asks the general question *What are you doing?*, they do know that B and their flatmates C and D are just preparing a birthday cake for a common friend. Assume further that C is known to top every cake they bake with nuts caramelized by themselves. In this situation, it would be more beneficial for B not to use topic drop but to overtly realize C's name *Kim* as the

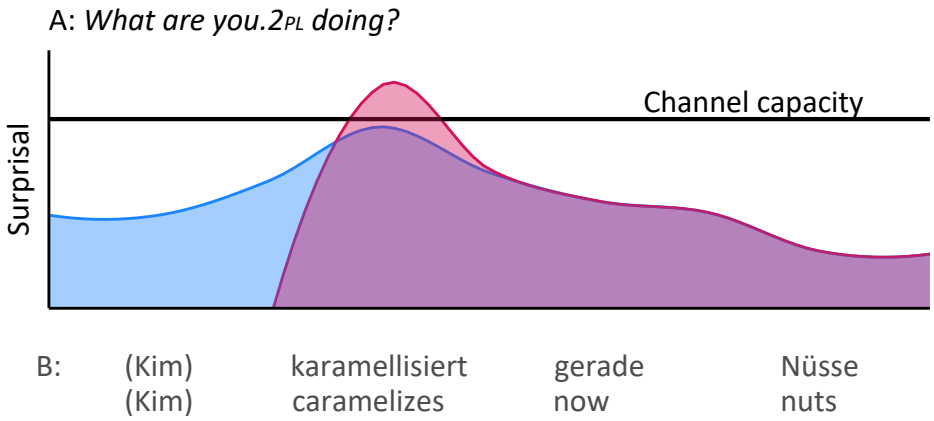


Figure 6.4: Hypothetical ID profile: Since *karamellisiert* creates a surprisal peak, the full form is more uniform.

prefield constituent, as indicated in the blue curve. This increases the likelihood of *karamellisiert* for (i) structural reasons and (ii) because of A's knowledge about C's behavior. (i) The proper name *Kim* in the prefield position makes it almost certain that it will be followed by a congruent finite verb.[10] (ii) The fact that C is passionate about caramelizing nuts increases the probability of *karamellisiert* following C's name. Therefore, inserting *Kim* lowers the surprisal of *karamellisiert* and the effort required to process the verb. This effort is even reduced in a third way. Using the full form instead of topic drop avoids the processing effort that would result from having to resolve the ellipsis. In the case of an omitted 3rd person singular subject, this effort would even be higher than for the 1st or the 2nd person because the ellipsis could potentially refer to any 3rd person referent known or present in the discourse or the situation, in our example, at least to both C and D (see also Section 6.3).

Finally, using the full form has another advantage for processing the verb. If the speaker does not use an ellipsis, the hearer does not have to resolve it. Thus, the effort associated with this process of ellipsis resolution can be saved. Since under an assumption of an incremental parser (see Section 6.3), this effort would presumably also be incurred directly on the verb, the overt realization of the prefield constituent should in any case facilitate the processing of this verb.

[10] Of course, this is also true for the utterance *Backe gerade*, but it is of more relevance here because the surprisal of *karamellisiert* is higher and likewise the need to lower it. This means that indicating that a congruent verb will follow by inserting the subject into the prefield is more beneficial if processing this verb requires more effort.

Ideally, by inserting an overt prefield constituent, the surprisal associated with the verb falls below channel capacity. However, a reduction of the peak and the processing effort is desirable in any case. From this reasoning, it follows that the surprisal of the following verb is a predictor of topic drop and that an effect of this predictor would provide genuine support for the information-theoretic account of topic drop usage. Such support comes from my corpus study of text messages in Section 11.4.4 but could not be obtained in experiments 11 and 12, which are discussed in Sections 11.5.1 and 11.5.2.

6.3 *Facilitate recovery* principle

The *avoid peaks* and the *avoid troughs* principles are determined by predictability, as I illustrated for topic drop. The prefield constituent is predictable from the linguistic or extralinguistic precontext. The constituent itself, in turn, impacts the likelihood of the following verb. However, the processing effort on this verb is not only associated with its predictability but also with the process of ellipsis resolution, which can only take place after the ellipsis is identified. Assuming an incremental parser, which uses any incoming data immediately to make a parsing decision (Marslen-Wilson 1973, 1975, Altmann & Kamide 1999), it is reasonable that topic drop is at least initially resolved directly on the subsequent verb (2b). I say at least initially because in written discourse, due to the lack of prosodic cues,[11] subsequent material may require a reanalysis of the apparent topic drop structure, e.g., as a polar question, when the parser finds a postverbal subject (2c).[12]

(2) a. A: 'What are you doing?'
b. *B:* Δ *Backe gerade Muffins.*
I bake now muffins
B: '(I) am baking muffins right now.'

[11] Zifonun et al. (1997: 415) state that in spoken discourse an empty prefield in declarative mood is indicated prosodically by a falling tone, among other things.

[12] The likelihood that such a reanalysis is necessary depends heavily on the context. Also, it may generally be more likely for a verb inflected for the 2nd person to introduce a question than for a verb inflected for the 1st person. In the case of a suitable syncretic verb form, a reanalysis as imperative may also be possible (i) (see Section 10.1.2).

(i) *Bring* *Muffins mit!*
bring.1SG.IND.PRS/IMP.SG muffins with
'(I) bring muffins!' or 'Bring muffins!'

c. *B: Backe ich Muffins oder nicht? Rate mal!*
bake I muffins or not guess.IMP.SG PART
B: 'Do I bake muffins or not? Guess!'

On the verb, the processing load increases with the difficulty of resolving topic drop at this point. This resolution process can be facilitated by cues that make it easier to identify the omitted element: the *facilitate recovery* principle. In German, for instance, a distinct inflectional marking on the verb provides information about the person and number of the congruent subject. Sigurðsson & Maling (2010: 70–71) argue that "by reducing ambiguity, agreement morphology both facilitates and constrains interpretation or identification." A distinct inflectional marking can thus facilitate recovering an omitted subject, prevent an information peak, and reduce the overall processing costs for the verb. This idea is illustrated graphically in Figure 6.5.

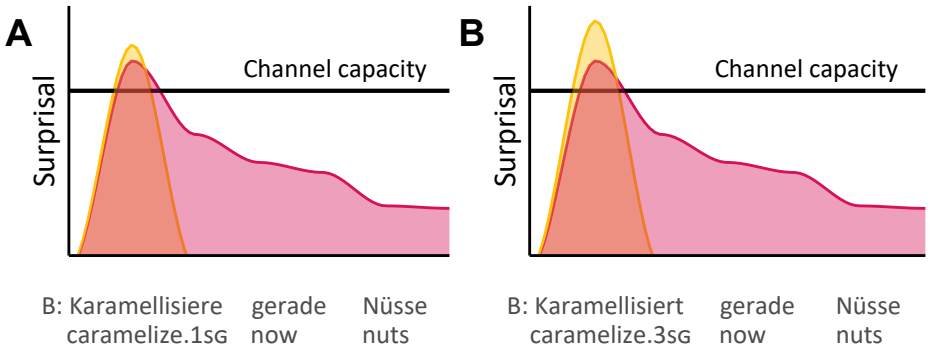


Figure 6.5: Hypothetical ID profiles indicating the additional processing effort on the verb for resolving topic drop in yellow. Resolving subject topic drop is usually easier from an inflectional ending for the 1st person singular (A) than for the 3rd person singular (B).

The yellow curve indicates the hypothetical additional processing effort associated with ellipsis resolution on the verb. For the verb form that is inflected for the 1st person singular present tense, this effort is lower (A) than for the 3rd person singular verb form (B). This is because upon finding a verb with an inflectional ending indicating the 1st person singular, the hearer can be certain that if they are just processing a topic drop structure with an omitted subject,[13] the

[13]It is also possible that an object is omitted from the prefield. However, in text messages, for instance, object topic drop is less frequent than subject drop. Thus, it is reasonable for the hearer to first assume that the subject is omitted and only consider an alternative analysis as object topic drop when they encounter a postverbal subject.

omitted subject must be a 1st person singular pronoun referring to the speaker.

In contrast, if they encounter a verb form inflected for the 3rd person singular, they usually cannot equally easily determine the referent of topic drop. Unlike for the speaker, who is uniquely determinable in an utterance situation, there are usually several potential 3rd person singular antecedents. This means that the set of potential referents for topic drop is also reduced by a 3rd person singular inflectional ending but to a lesser extent than by a 1st person singular inflectional ending. Therefore, the 1st person singular ending lowers the processing effort on the corresponding verb compared to the same verb with a 3rd person singular ending.

To sum up, the idea is that ellipsis resolution causes processing effort on the subsequent verb and that this processing effort is not just impacted by the predictability of the verb but can also be modulated by its inflectional ending. From this, it follows that an information-theoretic account of topic drop usage must also take recoverability-driven processing costs into account. This fits naturally with my interpretation of *UID* as a hypothesis of distributing processing effort uniformly and efficiently. Recoverability and the *facilitate recovery* principle are further factors that impact this processing effort besides the predictability, which is central to the original *UID hypothesis*. Note that this concept of recoverability is underlyingly identical to the concept of recoverability as a felicity or usage condition for topic drop that I presented in Chapter 4. Here, I simply integrate it explicitly into my information-theoretic approach.

6.4 Summary: information-theoretic approach to topic drop usage

I presented an information-theoretic account of the usage of topic drop according to which, in environments where it is licensed, topic drop is used or not used to distribute processing effort uniformly and efficiently across utterances. The use of topic drop is subject to a trade-off between two tendencies. The first tendency is to omit a prefield constituent whenever it creates a trough to avoid underchallenging the hearer. The second tendency is to realize the prefield constituent even if it creates a trough, provided that it smooths the peak on the subsequent verb because this prevents an overload of the hearer's processing capacities. Besides the processing effort caused by the predictability of the omitted constituent in context, there is an additional source of effort, namely the recoverability of the omitted constituent given the context and the following verb. Cues on the verb,

in particular the inflectional marking, can lower the overall processing effort on the verb and, thus, increase the likelihood or acceptability of topic drop.

The information-theoretic approach that I propose here is based on the *UID hypothesis*, which I interpret as a hypothesis about the efficient distribution of surprisal. In principle, *UID* is a theory about production, but it decidedly also considers perception because the efficiency of communication is increased by facilitating processing for the hearer. This means that to accomplish this facilitation the speaker is taken to perform audience design (Bell 1984), i.e., they adapt their utterances to the hearer and the situation. Consequently, insights into whether and how *UID* and its underlying principles shape language production can be gained from two sides. First, corpus or production studies can be used to directly investigate how speakers structure their utterances to increase their communicative efficiency. Second, perception experiments like acceptability rating studies can indirectly test whether utterances in line with *UID* are perceived as more well-formed by hearers. In the next chapter, I outline the relationship between production, perception, and processing effort in more detail.

7 Methodology of the empirical studies

The following four chapters each address one or more closely related factors that potentially influence the usage of topic drop: Chapter 8 syntactic function, Chapter 9 topicality, Chapter 10 grammatical person, verbal inflection, and ambiguity avoidance, and Chapter 11 verb type and verb surprisal. These chapters all have the same basic structure. First, I outline the central claims from the theoretical literature and the most important results from previous empirical studies. From these results, several hypotheses for the usage of topic drop emerge. Second, I discuss the information-theoretic predictions concerning the factor(s) and their relation to the hypotheses from the literature. Third and finally, I present my empirical studies in the form of a corpus study, acceptability rating studies, or both. The current chapter provides the methodological background for these empirical studies.

It is important to note that most of these studies examined multiple factors at once to determine if they collectively / in combination have an impact on the usage of topic drop. In my corpus study, for example, I considered grammatical person, verbal inflection, verb type, and verb surprisal, while in experiments 5, 7, and 8 I looked at grammatical person and topicality in combination. Since the following chapters are built around the factors rather than the individual studies, the results of the studies are discussed in several places, e.g., experiment 5 is addressed in both Chapter 9 on topicality and Chapter 10 on grammatical person. I give the details on the corpus study in a bundled way in Section 7.2 in this chapter, while for each experiment, I present the basic information concerning design, stimuli, and analysis at the first mention.

This chapter is structured as follows: First, I give an overview of the connection between the corpus study and the acceptability rating experiments and, more generally, the connection between production and perception. Then, I present the details of the fragment corpus FraC and three data sets derived from that corpus, which I used for my corpus investigations. I conclude this chapter with some brief remarks on the methodology of the experiments.

7.1 Production, perception, and processing

For my empirical studies, I employed a combination of methods to look both at the production, i.e., the frequency, and the perception, i.e., the acceptability, of topic drop. This combination was both practically and conceptually motivated. The practical motivation was that I could not investigate all factors of interest in the fragment corpus FraC. First, the FraC is not suitable for studying topicality because it is not annotated for information structure and for many utterances, the precontext required for such an annotation is not available. Second, it is a rather small corpus and certain factor combinations occur too rarely in there to investigate their impact on topic drop systematically. Such a systematic investigation was possible with my rating studies, which could at least partly offset the data sparsity problem. The conceptual reason for using both corpus and experiments was to investigate whether the effects found in the corpus study also show up in the experiments and vice versa. The intention was to provide a first indication of whether or not audience design (Bell 1984) is a reasonable assumption as a linking hypothesis for the usage of topic drop, as discussed in Section 6.4.

The corpus data allowed me to focus on the production of topic drop and to determine its frequency relative to certain influencing factors. Looking at the production is a natural first step to test the information-theoretic predictions for topic drop usage since *UID*, a core component of my approach, is in the first place a hypothesis about language production, i.e., how speakers optimize their utterances (see Section 6.2). With the authentic corpus data, I could directly investigate whether produced utterances with and without topic drop differ significantly with respect to relevant parameters, such as the grammatical person of the overt or covert prefield constituent or the surprisal of the following verb.

I examined the perception of topic drop through experiments. The linking hypothesis that connects production and perception is audience design (Bell 1984), i.e., the idea that a speaker shapes their linguistic production in such a way that the processing of the hearer is facilitated. If an utterance with topic drop is easier to understand for a hearer, a speaker should be more inclined to use it.[1] In this way, I could indirectly investigate the usage of topic drop via perception experiments. Since the acceptability rating experiments are an offline method, they do

[1]It could be argued that processing difficulties on the part of the hearer could always be avoided by the speaker being as explicit as possible and omitting nothing. However, according to *UID*, such overly explicit and thus partly redundant utterances waste resources from the speaker's perspective and may underchallenge the hearer (see Section 6.2.1). Furthermore, they may lead to undesired implicatures for violating the maxim of relevance or quantity (see the research on atypicality inferences, e.g., Kravtchenko & Demberg (2022)).

not allow for directly observing the processing of an utterance. However, it is possible to indirectly draw conclusions about the processing via the perceived acceptability (see, e.g., Fanselow & Frisch 2006, Hofmeister et al. 2014). The general idea is that the more difficult it is to process an utterance, the less acceptable or natural should it be perceived, and the worse should it be rated.[2]

Recall that in the information-theoretic logic, processing difficulties are caused by a suboptimal distribution of information, i.e., by peaks and troughs in the information density profile resulting from too predictable or too unpredictable linguistic expressions (see Section 6.2). This reasoning allows for connecting frequency as the dependent variable in the corpus study and acceptability as the dependent variable in the experiments. An expression that generally occurs very frequently (in a certain position) has a high predictability (in that position), i.e., a low surprisal, and is likely to create an information trough. According to the *avoid troughs* principle (Section 6.2.1), hearers should prefer an utterance where this expression is omitted (topic drop) over an utterance where it is realized (full form), provided that grammar permits the former. In contrast, a very rare expression should have a low predictability and a high surprisal, which might cause a processing overload for the hearer. In this situation, it is to be expected from the *avoid peaks* principle (Section 6.2.2) that hearers prefer an utterance where an additional prefield constituent is inserted before the expression with the high surprisal to reduce the peak of information and the processing effort (full form) over one where nothing is inserted (topic drop).

In summary, I used corpus data to determine frequencies of authentic utterances with and without topic drop. My experiments provided acceptability judgments of systematically constructed sentences with and without topic drop and, thus, also indirect evidence for possible processing difficulties. When the same factors are investigated with both methods, I expect comparable results according to audience design as a linking hypothesis, i.e., that rarer topic drop constructions are also less acceptable and vice versa.

7.2 Methodology: corpus study

In this section, I introduce the fragment corpus FraC, which I used for my corpus study. I outline how I annotated the instances of topic drop and how I extracted the corresponding full forms. Finally, I briefly present the three data sets that are based on the FraC and in which I investigated topic drop in this book.

[2]Support for this idea of a correlation between acceptability and processing comes, for example, from Hofmeister et al. (2013) for superiority effects in multiple *wh*-questions and from Hofmeister et al. (2015) for "frozen" extraposed constituents.

7.2.1 The fragment corpus FraC

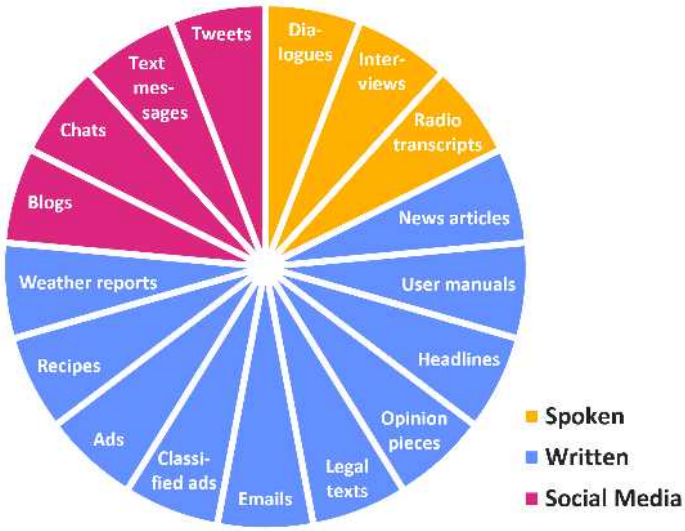


Figure 7.1: The 17 text types in the fragment corpus FraC, divided into spoken, written, and social media text types

My corpus study, the results of which I present in the following chapters, was conducted on the fragment corpus FraC (Horch & Reich 2017). The FraC is a Standard German corpus dedicated to the investigation of fragments, i.e., non-sentential utterances (e.g., Morgan 1973, Merchant 2004, Reich 2007, Lemke 2021). The corpus consists of 17 different text types,[3] which range from prototypically written (e.g., news articles, legal texts, etc.) to prototypically spoken (e.g., dialogues, interviews, etc.) to so-called "social media texts" (e.g., online chats, text messages, etc.). Figure 7.1 shows an overview of the different text types.[4]

There are about 2 000 utterances of each text type in the corpus, resulting in a total of about 34 000 utterances and 380 000 tokens (Horch & Reich 2017). The corpus was automatically POS-tagged and lemmatized using the TreeTagger (Schmid 1994, 1995) and is manually annotated for a variety of categories. Relevant to my work is the annotation of omission types. 919 utterances were classified as containing topic drop. However, the annotation of the FraC is based on a different definition of topic drop than the one that I assume and also includes V1 declaratives. Therefore, I manually reviewed all 919 utterances and excluded

[3]See Section 2.1.4 for a discussion of the term *text type* and how I use it in this book.
[4]I created the figure myself in Microsoft Excel.

those that did not match the definition of topic drop developed in Section 2.1 and that were neither covert subjects nor objects.[5] This resulted in a total of 873 instances of topic drop, which provide the basis for my corpus investigations. Note that the number of instances is not equal to the number of utterances with topic drop, since some utterances contained two or more cases of topic drop (separated by commas and not split into individual sentences in the corpus).

7.2.2 Annotation procedure

In the following, I describe first how I annotated the instances of topic drop in the FraC and then how I obtained and annotated the corresponding full forms that serve as reference data.

7.2.2.1 Annotating the instances of topic drop

I annotated each of the 873 instances of topic drop in the FraC manually for several categories. In what follows, I list these categories and illustrate the annotation process with example (1).

(1) a. *Sein fast 70jähriges Herz?*
his almost 70-year-old heart
'His almost 70-year-old heart?'
b. Δ *Schlägt gut und regelmäßig.*
it beats well and regularly
'(It) beats well and regularly.' [FraC B47–B48]

For each instance, I provided a subjective intuitive reconstruction of the omitted element based on the context and the morphological information in the form of a personal pronoun, a demonstrative pronoun, or a proper name (if this name cannot be naturally replaced by a pronoun) and also annotated the category of this reconstruction (personal pronoun, demonstrative pronoun, proper name, etc.). In the case of (1), I reconstructed the personal pronoun *es* ('it').

I annotated the grammatical function, person, number, and (if applicable) gender of the omitted constituent. In (1), it is a 3rd person singular neuter subject.

[5] Furthermore, I did not consider variants of the abbreviation *HD(GD)L* (*hab dich* (*ganz doll*) *lieb* ('love you (very much)'), which occur around 30 times in the text message subcorpus. Since the abbreviation is conventionalized in German chatspeak without the subject, i.e., usually there is no *IHDL* (*ich hab dich lieb* ('I love you')), i.e., there is actually no real full form that the speaker could use alternatively.

Next, I determined whether there is an antecedent or a postcedent present in the linguistic context and annotated its category. By default, I only considered postcedents if the precontext provided no suitable antecedent at all. I also determined the distance of the antecedent or postcedent in the number of utterances (0 means that the antecedent occurs in the same utterance as topic drop).[6] In example (1), there is a linguistic antecedent in the last utterance (distance 1), namely the determiner phrase *sein fast 70jähriges Herz* ('his almost 70-year-old heart'). See Section 4.1 for a discussion of the distance annotation.

For the verb following topic drop, I annotated the verb type distinguishing copular verbs (*sein* ('to be') and *werden* ('will') when used without another verb form), auxiliaries (*sein* ('to be'), *haben* ('to have'), and *werden* ('will') when used to form periphrastic verb forms), modal verbs (*dürfen* ('may'), *können* ('can'), *mögen* ('may'), *müssen* ('must'), *sollen* ('shall'), and *wollen* ('to want')), reflexive verbs, and lexical verbs. Additionally, I annotated whether the verb form was syncretic or distinct. In example (1), topic drop is followed by a lexical verb in present tense indicative, which constitutes a distinct verb form.

7.2.2.2 Obtaining and annotating the reference data

To obtain relative numbers and omission rates based on usage factors, I extracted full forms as reference data. This means that I considered those utterances in the FraC that could potentially be targeted by topic drop. For the FraC-TD-Comp data set (Section 7.2.3.1), which is based on the complete FraC, I used a semi-automatic approach. For the smaller FraC-TD-SMS (Section 7.2.3.2) and FraC-TD-SMS-Part (Section 7.2.3.3) data sets, which are based on only the text message subcorpus, I manually reviewed all 1 961 utterances in the text message subcorpus of the FraC.

To decide which utterances could be targeted by topic drop, it is necessary to verify that the licensing and the felicity condition of topic drop discussed in Chapters 3 and 4 are met: prefield position and recoverability of the corresponding constituent.[7] Since no linguistic context is available for several utterances

[6] Note that for some utterances, in particular for a part of the text messages, the precontext is not available. In these cases I could often infer the antecedent but not determine whether it was linguistically present and in what distance.

[7] Recall that I exempted non-referential expletives from the recoverability condition (see Section 4.3). However, I ensured that an overt constituent would be recoverable, if targeted by topic drop, by searching for all utterances with a substitutive pronoun in the prefield, as explained below. This way, I also captured the cases with overt expletive *es* in the preverbal position, as desired.

in the FraC and since the extralinguistic context is not accessible, it is partially impossible to determine which realized referential prefield constituent would be recoverable and which would not if omitted. To circumvent this problem, I restrict my reference data to those utterances where the prefield constituent is a substitutive pronoun. I assume that the use of such pronouns is restricted to cases where their reference can be recovered from the linguistic or extralinguistic context. That means that they have the same felicity condition of recoverability as referential topic drop. In practice, I included every utterance in which one of the object (accusative or dative)[8] or subject pronouns, listed in Table 7.1, occurs in the prefield.

I searched not only for the demonstratives with the stem *dies** but also for those with *jen**, but there were no occurrences in the preverbal position. I also included the dialectal variant *i* of *ich* ('I'), which is common in southern Germany and occurs several times in the text message subcorpus. Furthermore, I considered full forms with *man* in the prefield because I found instances where it is omitted

[8]There were no preverbal genitive object pronouns in the corpus. I did find 20 instances of PPs with pronouns in the prefield, of which some were adverbial adjuncts such as (i) and some prepositional objects such as (ii). However, I argue that none of them can be targeted by topic drop (as indicated by the asterisks) because the prepositions *mit* ('with') and *von* ('from') cannot be recovered from context.

(i) a. *Schon beim zweiten Tagesordnungspunkt, der Wahl eines Vorsitzenden für das neue Gremium, wurde der Gewerkschafter Rappe ungeduldig – und beantragte den Rückzug in Sitzungssaal Nummer zwei.*
'Already at the second item on the agenda, the election of a chairman for the new committee, the unionist Rappe became impatient – and requested to retire to meeting room number two.'

b. **(Mit mir) geht sowas nicht, entrüstete er sich über die*
with me goes something.like.this not outraged he himself about the
Gegenrede einer Studentin, ich mach so'n Zirkus nicht mit.
contradiction a.GEN student.FEM I make such.a circus not with
'That's not possible with me, he said indignantly in response to a counter-speech from a student, I don't take part in that kind of circus.' [FraC A2010–A2011]

(ii) a. *Ich bin Dozent für Computerlinguistik / Computerlexikographie in T-Stadt am Seminar für Sprachwissenschaft*
'I am a lecturer in computational linguistics / computational lexicography in T-city at the Seminar for Linguistics'

b. **(von mir) stammen die meisten Skripte (oder Skripts?)...*
from me stem the most scripts or scripts
'Most of the scripts (or scripts?) are from me...' [FraC C64–C65]

Table 7.1: Subject and object pronouns considered in the corpus search for the full forms

Grammatical person	Nominative	Accusative	Dative
1SG	*i*(*ch*)	*mich*	*mir*
2SG	*du*	*dich*	*dir*
3SG	*er*, *sie*, *es*, *man*, *der*, *die*, *das* *dieser*, *diese*, *dieses*	*ihn*, *sie*, *es*, *den*, *die*, *das* *diesen*, *diese*, *dieses*	*ihm*, *ihr*, *dem*, *der* *diesem*, *dieser*
1PL	*wir*	*uns*	*uns*
2PL	*ihr*	*euch*	*euch*
3PL	*sie*, *die*	*sie*, *die*	*ihnen*, *denen*

in the FraC.[9]

Since the FraC is not annotated topologically, I operationalized the licensing condition of topic drop, i.e., the prefield restriction, in the semi-automatic approach that yielded FraC-TD-Comp, as follows: I considered a pronoun to be in the prefield if it occurs in the first position of an utterance and is immediately followed by a finite verb.[10] In Section 3.4, I discussed that certain elements

[9] One such example of a covert *man* is (i).

(i) *Man kämpft mit den erstaunlich kräftigen Kindern,* Δ *presst sie in Klamotten.*
one fights with the astonishingly strong children one presses them in clothes
Δ *Stopft die Füße in die Schuhe.*
one stuffs the feet in the shoes
'You fight with the astonishingly strong children, (you) press them into clothes. (You) stuff their feet into shoes.' [FraC B1640–B1641]

At first glance, the possibility of omitting *man* seems to be problematic since *man* is frequently classified as an indefinite pronoun (e.g., Zifonun et al. 1997: 43, Gallmann & Sitta 2007: 66–67, Imo 2016: 87, Thielmann 2021: 140–141), and Volodina & Onea (2012) explicitly deny that indefinite pronouns can be targeted by topic drop. However, Eisenberg (2020: 187) characterizes *man* not as an indefinite pronoun but as an "impersonal personal pronoun", while the Leibniz Institut für Deutsche Sprache (IDS) terms it a "generalizing personal pronoun" (Leibniz-Institut für Deutsche Sprache 2018a). The IDS stresses that *man* differs from indefinite pronouns in that consecutive occurrences can refer to the same person, as the example above demonstrates. It seems that it is precisely this property that enables recoverability and, thus, topic drop.

[10] It has to be noted that the automatic extraction procedure that I used for FraC-TD-Comp

such as conjunctions, particles, and interjections can precede topic drop. Consequently, I also considered corresponding full forms with both the semi-automatic and the manual approach, i.e., utterances where a word, separated by a comma, precedes the pronoun in the prefield. For each full form, I annotated the prefield constituent, its category, its number, person, gender, and case, as well as the verb type of the following verb.

7.2.3 FraC data sets

I created three data sets: (i) The FraC-TD-Comp data set contains all instances of topic drop in the FraC and (semi-automatically retrieved) corresponding reference data, i.e., full forms where the prefield constituent could potentially be omitted. (ii) The FraC-TD-SMS data set consists of only the utterances with topic drop and the corresponding (manually extracted) full forms of the text message subcorpus of the FraC. (iii) The FraC-TD-SMS-Part data set is a reduced variant of FraC-TD-SMS containing only the 1st and 3rd person singular subjects.[11] As discussed in Section 2.1.4, topic drop is not (fully) licensed in certain text types contained in the FraC, such as news articles or legal texts. This means that these text types contribute full forms to FraC-TD-Comp but not instances of topic drop, making the full forms overrepresented in the data set. This distorts relative measures such as the omission rate. To circumvent this problem, I focused on the text message subcorpus in my further investigations because it contains the highest number of topic drop of all subcorpora and also has the highest omission rate of almost 64%. The corresponding FraC-TD-SMS data set allowed me to statistically compare relative values as a function of syntactic function, grammatical person, and verb type. I used the third data set, FraC-TD-SMS-Part, to look at distinct inflectional marking and verb surprisal and to perform an inferential statistical analysis. It simultaneously considered these two factors as well as grammatical person and verb type and evidenced an impact of them on the frequency of topic drop.

might have yielded false positive and false negative instances, as it is dependent both on the POS tagging and the tokenization in sentence units. The accuracy of both varies in particular with text type, e.g., for news articles it is higher than for dialogues.

[11] In Schäfer (2021), I also presented a corpus study of such a subset of the FraC, which yielded similar but partly different results. For this book, I was able to increase the amount and quality of the data I analyzed by not only using automatically extracted instances of topic drop and full forms but by manually reviewing the entire text message subcorpus. Additionally, I revised the criterion to choose the reference data from pronouns and full forms to only pronouns to ensure recoverability (see Section 7.2.2.2), refined the predictor that encodes whether the verb in the left bracket is syncretic or not and also the annotation of the corresponding verbs, and additionally included a predictor for the verb type following topic drop.

7.2.3.1 FraC-TD-Comp

The data set FraC-TD-Comp (*Fra*gment *C*orpus *T*opic *D*rop *Comp*lete) contains all 873 instances of topic drop in the FraC and reference data in the form of 3 211 full forms that could potentially be targeted by topic drop (omission rate of 21.38%), resulting in a total of 4 084 instances, as shown in Table 7.2.

Table 7.2: Overview of the FraC-TD-Comp data set

Full form	Topic drop	Total	Omission rate
3 211	873	4 084	21.38%

Each instance is annotated for the following categories: Completeness (Does the instance contain topic drop or is it a full form?), Syntactic Function and Case of the omitted or realized prefield constituent, Grammatical Person, Number, and Gender of the omitted or realized prefield constituent, Verb Type of the verb following the omitted or realized prefield constituent and Text Type in which the instance occurs.

In the following chapters, I present the results of a descriptive look at the FraC-TD-Comp data set with respect to syntactic function, grammatical person, number, gender, and type of the following verb, as well as their potential role for topic drop.

7.2.3.2 FraC-TD-SMS

As a result of the typical occurrence of topic drop in certain text types (see Section 2.1.4), I decided to reduce the corpus to a text type that allows for topic drop. I created the data set FraC-TD-SMS (*FraC T*opic *D*rop *SMS*), which contains the 353 instances of topic drop and the 201 full forms of the text message subcorpus of the FraC (see Table 7.3).

Table 7.3: Overview of the text message data set FraC-TD-SMS

Full form	Topic drop	Total	Omission rate
201	353	553	63.83%

The restriction to the text message subcorpus seems appropriate for two reasons: (i) By far the largest number of cases of topic drop in the FraC stem from text

messages, 353 out of 873 cases, or 40.44%, indicating that it is very natural for topic drop to occur in this text type. (ii) Analyses of this data set can be better compared to the existing corpus studies by Androutsopoulos & Schmidt (2002) and Frick (2017) (see Section 10.2.1 for details on their studies), which have also been conducted on text messages. In the following chapters, I present analyses of the FraC-TD-SMS data set considering syntactic function, grammatical person, and following verb, which complement the analysis of FraC-TD-Comp.

7.2.3.3 FraC-TD-SMS-Part

The final data set FraC-TD-SMS-Part is a further reduction that limits the data of FraC-TD-SMS to only the overt and covert subjects of the 1st and the 3rd person singular. This is motivated by the observation that overall mainly subjects are omitted in the FraC-TD-SMS data set, above all subjects of the 1st person singular, followed with some distance by the 3rd person singular (see Section 8.4.2). Since additionally for subject topic drop, the following congruent verb is predicted to play a role in the recovery according to my information-theoretic account, at least if this verb is distinctly marked for inflection, it seems reasonable to focus on the subjects, in particular subjects of the 1st and the 3rd person singular. The contrast between these two grammatical persons furthermore allowed me to investigate a potential effect of syncretic inflectional endings, which exist for the 1st and 3rd person singular in the present and past tense (see Section 10.1.1), and the overall influence of grammatical person on the likelihood of topic drop. Table 7.4 shows the distribution of full forms and instances of topic drop as a function of the grammatical person in the resulting data set FraC-TD-SMS-Part. In Section 10.4.4, I present the logistic regression analysis that I conducted on this data set and discuss its results there and in Section 11.4.4.

Table 7.4: Full forms, instances of topic drop, and omission rates as a function of grammatical person in the FraC-TD-SMS-Part data set

Grammatical person	Full form	Topic drop	Total	Omission rate
1SG	131	264	395	66.84%
3SG	29	41	70	58.57%
Total	160	305	465	65.59%

7.3 Methodology: experiments

In the following chapters, I present a total of eight acceptability rating experiments that tested various factors potentially influencing the usage of topic drop. I tested the impact of topicality, grammatical person, verb surprisal, and, more indirectly, verb type and ambiguity avoidance. These factors were selected based on the discussions in the literature, as well as based on my information-theoretic account of topic drop usage.

There were two goals that I pursued with the experiments. First, I tested several isolated hypotheses from the theoretical literature for the first time. That is, I investigated whether grammatical person, topicality, and verb type impact the usage of topic drop, as hypothesized by authors such as Auer (1993), Zifonun et al. (1997), Imo (2013), and Helmer (2016). Second, I related these results to the predictions of my information-theoretic account, more specifically to the *avoid troughs* and the *facilitate recovery* principles. The predictions of the *avoid peaks* principle were then decidedly tested with the two experiments on verb surprisal.

The methodology of the experiments is basically identical to that of the experiments in the first part of this book. I compared utterances with topic drop to a non-elliptical baseline to see whether ellipsis-specific effects of the influencing factors emerge in the form of an interaction. The statistical analysis followed the procedure described in Section 3.1.6.3. As discussed above, I interpret a degraded acceptability judgment as a display of processing difficulties caused by a suboptimal distribution of information.

8 Syntactic function

As discussed in Section 2.1.3, the focus of this book lies on topic drop of verb arguments, i.e., subjects and objects. In this chapter, I discuss whether and to what extent the syntactic function of the preverbal constituent impacts its omission. First, I provide a review of the literature on topic drop of subjects, objects, and predicatives and of previous empirical studies and their results. In the second step, I turn to the information-theoretic predictions for syntactic function. Finally, I present my empirical results in the form of a descriptive overview of the role of syntactic function in the fragment corpus FraC (see Section 7.2.1).

8.1 Theoretical overview

8.1.1 Subjects

In the literature, it is uncontroversial that topic drop can target subjects,[1] but there is dissent as to which kind of subjects can be omitted. In Section 3.1.5, I discussed in detail the controversy over whether only referential subjects (Fries 1988, Cardinaletti 1990) or both referential and non-referential subjects (Reis 2000, Frick 2017, Ruppenhofer 2018) can be targeted by topic drop. My experiment 1 showed that non-referential topics, more precisely *es* expletives, can very well be omitted (see Section 3.1.6). This result is in line with previous results from corpus studies, which I discuss in Section 8.2.

8.1.2 Objects

The literature, starting with Ross (1982) and Huang (1984), agrees that topic drop in German can also target objects. However, there are three (possible) limitations of object topic drop that are frequently discussed. First, it is debated whether object topic drop is as common and as acceptable as subject topic drop. Second,

[1] In an early paper that mentions topic drop as telegraphese, Reis (1982: 190) even suggested that only nominative DPs, i.e., subjects, can be omitted in telegrams. However, a broader notion of topic drop, not limited to the use in telegrams, also includes (at least) object drop (see Ross 1982, Reis 2000 and the next paragraph).

the omission of direct objects in the accusative case seems to be more felicitous than the omission of prepositional objects and indirect objects in the dative or the genitive case, a difference explained by various approaches in the literature. Third, there seems to be an asymmetry concerning grammatical person, according to which 1st and 2nd person objects cannot be omitted or are difficult to omit while omitting 3rd person objects is perfectly fine.[2] In what follows, I present the relevant contributions to the discussion on each of the three issues.

8.1.2.1 Subject vs. object topic drop

Concerning the first issue, Trutkowski (2016: 2) states that subjects and objects can be equally well omitted, whereas Volodina (2011) claims that object topic drop is rarer than subject topic drop, and Fries (1988) suggests that objects are less likely to be omitted. Fries refers to Klein (1985: 15), who postulates that for context ellipsis,[3] how well an expression can be omitted tends to be determined by a hierarchy of syntactic functions in the form of subject > direct object > indirect object > prepositional complement.[4] It seems that most authors discussing this issue take the view that object topic drop has a disadvantage over subject topic drop.

8.1.2.2 Direct vs. indirect objects

Regarding the second point, namely a possible difference between types of objects, i.e., direct vs. indirect objects or objects with a structural vs. objects with a lexical case, several authors argue for an asymmetry concerning how well they can be omitted. I discuss the different arguments and approaches in detail in what follows.

First, there is the view that indirect objects can in principle be targeted by topic drop but less well or less often than direct objects. For example, Schalowski (2015: 3) claims that topic drop of indirect objects is possible but less acceptable. He notes that he did not find any such instances in a previous corpus

[2] In this respect, syntactic function overlaps with grammatical person, to which Chapter 10 is devoted. I discuss the object asymmetry in relation to grammatical person in the present chapter and focus on the grammatical person of omitted subjects rather than objects in the next chapter.

[3] Klein (1985: 5) uses this term to subsume coordination ellipsis and what he terms "Adjazenzellipse" ('ellipsis under adjacency'), e.g., question-answer pairs, corrections, or continuations. Fries (1988) extends the scope of this tendency by applying it to topic drop.

[4] Similar hierarchies are proposed in centering theory (Grosz et al. 1995) for English (Walker et al. 1998) and German (Speyer 2007) to determine the reference of anaphora (see Section 9.2.3.1).

study (Schalowski 2009, cited in Schalowski 2015). According to Ruppenhofer (2018), the omission of indirect objects is only possible to a very limited extent, in "generic statements, especially ones about the efficacy of means and instruments" (Ruppenhofer 2018: 224, footnote 14) such as (1). There, however, the full form may not even obligatorily require an indirect object.[5]

(1) Δ *Hilft beim Abnehmen.*
it helps with.the losing.weight
'(It) helps to lose weight.' (Ruppenhofer 2018: 224, footnote 14)

In contrast to the assessment of Schalowski (2015) and Ruppenhofer (2018), who assume that indirect objects can be omitted to a limited extent, other authors hold the view that topic drop cannot affect indirect objects at all. Zifonun et al. (1997: 418) state that topic drop of dative objects and prepositional objects is not possible, regardless of the presence or absence of a linguistic antecedent.[6] This view seems to be supported by Jaensch (2005) and the Duden (2016, 2022) who do not list topic drop of indirect objects as an option (although they also do not explicitly deny it).

[5]While in any case, a subject referring to what promotes weight loss is omitted, the simplest full form is the generic statement (i) with *es* ('it') in the prefield but without an indirect object.

(i) *Es hilft beim Abnehmen.*
it helps with.the losing.weight
'It helps to lose weight.'

This is supported by the verb valence dictionary E-VALBU where for *helfen* ('to help') in the sense of causing health improvement the dative argument is optional (Leibniz-Institut für Deutsche Sprache 2023b) and for the sense of enabling one to be successful at something, an example without a dative object is listed (Leibniz-Institut für Deutsche Sprache 2023c). To consider (1) as topic drop of an indirect object, Ruppenhofer (2018) has to assume a full form like (iia), where a beneficiary like *Ihnen* ('you.POL') is included in the prefield (or a full form like (iib) with *Ihnen* in the middle field).

(ii) a. *Ihnen hilft es beim Abnehmen.*
you.2SG.DAT.POL helps it with.the losing.weight
'It helps you to lose weight.'
b. Es hilft Ihnen beim Abnehmen.

However, assuming the full forms in (ii) would result in (1) violating the restriction that topic drop can target only one constituent per clause. In short, the dubiousness of Ruppenhofer's example does not allow me to validate the restriction to "generic statements" that he formulates.

[6]Zifonun et al. (1997) seem to be one of the few researchers who mention topic drop of prepositional objects at all. But see also Trutkowski (2016: 25).

It is Sternefeld (1985) who is the first to propose that how well an object can be omitted depends on the distinction between *structural case* and *lexical case*. While for verb arguments, structural case depends on the sentence structure itself (e.g., the accusative as the case of the direct object), lexical case is determined by properties of the head that governs the corresponding DP (e.g., the German verb *helfen* ('to help') requires an object in the dative case) (see, e.g., Haspelmath 2009).[7] Sternefeld (1985) claims that only arguments with a structural case can be omitted, whereas arguments with a lexical case cannot. However, he points out that this is not equivalent to distinguishing between direct objects in the accusative case on the one hand and indirect objects in the dative or the genitive case on the other. He presents examples such as (2), where an accusative argument cannot be omitted either, according to his judgment, and contrasts them with the examples in (3), where topic drop shall be possible.

(2) a. **(Sie)* *Hat die Sache nicht interessiert.*
she.ACC has the thing not interested
'(She) was not interested in the matter.'

b. **(Ihn)* *Hat bedrückt, daß Ede krank ist.*
him.ACC has depressed that Ede ill is
'(He) was depressed that Ede is ill.' (Sternefeld 1985: 407, adapted, his judgments)

(3) a. *(Das)* *Kenn ich schon.*
that.ACC know I already
'I know (that) already.'

b. *(Den)* *Haben wir gestern erst gesehen.*
him.ACC have we yesterday only seen
'We saw (it) only yesterday.' (Sternefeld 1985: 407, adapted, his judgments)

Sternefeld (1985) argues that in (2), the verbs *interessieren* ('to interest') and *bedrücken* ('to depress') are so-called "flip verbs". In a classic generative analysis, these verbs are considered not to have a logical subject but two internal objects in the

[7] In the literature, structural case is also frequently contrasted with the term *inherent case*. Inherent case is used either for case assignment that is related to theta roles (Chomsky 1981: 171), later also called *thematic case* (e.g., Reinhart & Siloni 2005), or as an umbrella term for lexical case assignment and assignment by theta roles (Haspelmath 2009). While the umbrella term would make lexical case a hyponym of inherent case, Woolford (2006) argues for a distinction between inherent and lexical case. Since I restrict topic drop, as explained in Section 2.1.3, to the omission of arguments of the verb, it seems reasonable to focus on the concept of lexical case in this book.

deep structure, of which one becomes the subject by receiving the nominative, whereas the other becomes the object by receiving the accusative (Sternefeld 1985: 400; 427). Sternefeld (1985: 427) argues that this case assignment does not result in a structural but in a lexical accusative required by the flip verb, and in the impossibility to omit this argument. Sternefeld (1985) supports this claim only with the introspective judgments in (2) and (3), but even if his predictions were confirmed in empirical studies yet to be conducted,[8] it would remain unclear why lexical case should block topic drop.[9] This also applies, first, to Fries's (1988: 31) hypothesis that DPs with a lexical case and dative DPs seem to require structural and semantic constancy between antecedent and target utterances to be able to be omitted, an idea that partly anticipates Trutkowski's (2016) argumentation that I discuss below. Second, this open issue also concerns Haider's (2010: 269, footnote 22) remark that topic drop is one of several phenomena that indicate a dichotomy between the nominative and the accusative on the one hand and the dative on the other.

Trutkowski (2016) presents a more elaborate account of the role of the case for (object) topic drop. She differentiates between two types of topic drop: *non-verbatim* (NVTD) and *verbatim topic drop* (VTD). NVTD can only target arguments with a structural case, under which she subsumes the nominative and the accusative but allows for a change of the predicate. This change can be accompanied by mismatches in case and theta role between the antecedent and the omitted constituent (Trutkowski 2016: 3). VTD does not allow for such mismatches, since it requires semantic identity, but it can target both arguments with a structural and a lexical case (Trutkowski 2016: 3).[10] Trutkowski illustrates the apparent contrast with topic drop of a genitive object, i.e., an argument with a lexical case, which should only allow for VTD and, thus, block mismatches. Therefore, the matching VTD in (4) should be grammatical, while the NVTD involving a change in predicate from *gedenken* ('to commemorate') to *schämen* ('to

[8] But see Trutkowski's (2018) rating study discussed below, in which a part of Sternefeld's (1985) predictions are tested.

[9] As a starting point for an explanation, one could argue that lexical case is capable of blocking certain syntactic processes. For instance, passivization only targets structural but not lexical case in German. Topic drop could then be another such process.

[10] Trutkowski (2016: 31) specifies that "the crucial condition for the well-formedness of VTD is the identity of finely granulated (micro) theta roles assigned to antecedent and gap, respectively, which is ensured when context and target predicates are 'highly' synonymous (i.e. semantically equivalent)." By "micro theta roles" she means very fine-grained semantically enriched theta roles (Trutkowski 2016: 31). I do not go further into the details of her proposal but refer to her detailed account.

be ashamed') in (5) should not.[11] According to Trutkowski (2016), this is only possible if the antecedent and the omitted constituent have a structural case, as example (6) shall illustrate, where the omitted constituent is an accusative object. Therefore, the change from *gedenken* to *kennen* ('to know') should be fine.

(4) a. *A: Der Hans gedenkt der Rosa Luxemberg [sic!].*
the Hans commemorates the.GEN Rosa Luxemberg
A: 'Hans commemorates Rosa Luxemberg [sic!].'
b. *B:* Δ *Gedenkt der Otto auch.*
her.GEN commemorates the Otto too
B: 'Otto commemorates (her), too.' (VTD, Trutkowski 2016: 4, her judgment)

(5) a. *A: Der Hans gedenkt der Rosa Luxemberg [sic!].*
the Hans commemorates the.GEN Rosa Luxemberg
A: 'Hans commemorates Rosa Luxemberg [sic!].'
b. *B:* *Δ *Schämt sich manch ein Politiker.*
her.GEN is.ashamed REFL some a politician
B: 'Many a politician is ashamed (of her).' (NVTD impossible, Trutkowski 2016: 4, her judgment)

(6) a. *A: Der Hans gedenkt der Rosa Luxemberg [sic!].*
the Hans commemorates the.GEN Rosa Luxemberg
A: 'Hans commemorates Rosa Luxemberg [sic!].'
b. *B:* Δ *Kennt der Otto gar nicht.*
her.ACC knows the Otto at.all not
B: 'Otto doesn't know (her) at all.' (NVTD possible, Trutkowski 2016: 4, her judgment)

As Trutkowski herself notes, utterances with covert genitive objects generally seem odd because of a register clash between topic drop as a phenomenon of informal speech and genitive objects, which are rarely used in contemporary German and when they are, it is mainly in formal registers (Trutkowski 2016: 26). Nevertheless, she claims that there are acceptable instances of topic drop of genitive objects and tested respective utterances in a survey. Her stimuli with unrealized genitive objects such as (7), however, still seem highly marked in an

[11] What is omitted in (5) could also be the demonstrative pronoun *dessen* referring back to the proposition that Hans commemorates Rosa Luxemburg. Under this reading, the utterance would mean that many a politician is not ashamed of Luxemburg but of Hans commemorating her.

informal setting[12] and received correspondingly low ratings.[13]

(7) a. *Mir gefällt Martin so gut.*
me.DAT pleases Martin so well
'I like Martin so well.'
b. Δ *Konnte ich mich bislang auch nicht erwehren.*
him.GEN could I me yet also not resist
'So far, I have not been able to resist (him) either.' (Trutkowski 2016: 46, her judgment)

I come back to Trutkowski's account in Section 8.2, where I discuss an experiment she conducted in Trutkowski (2018) to test her hypotheses.

In sum, several authors argue for an asymmetry between topic drop of direct and topic drop of indirect. Sternefeld (1985) and Trutkowski (2016) attribute this asymmetry to case differences (structural vs. lexical case or verbatim and non-verbatim topic drop, respectively). However, it remains unclear from Sternefeld's (1985) account why such case differences should block topic drop and Trutkowski's (2016) account is based on judgments, which are at least dubious.

8.1.2.3 1st and 2nd person objects vs. 3rd person objects

Besides the difference between direct and indirect objects, there is a second asymmetry within object topic drop, which is the third issue concerning object topic

[12]This is supported by the fact that the Duden classifies the verb *erwehren* ('to resist') from example (7) as "elevated" (Dudenredaktion 2021).

[13]The mean rating for utterances with omitted genitive objects was at -0.775 on a scale from -2 (worst) to 2 (best) and consists of ratings for 4 items with different case combinations, as shown in the following table taken from Trutkowski (2016: 48):

Case combination	Mean rating	Standard deviation
Nominative–genitive	−1.15	1.21
Accusative–genitive	−1.25	1.07
Dative–genitive	0.12	1.39
Prepositional phrase–genitive	−0.82	1.37

In particular, the mean ratings for the items with nominative–genitive and with accusative–genitive are heavily degraded given that the best mean rating of the inquiry was 1.88 (0.32) for topic drop of a nominative DP with a PP antecedent. As Trutkowski (2016: 39) admits, her survey does not meet the requirements of a linguistic experiment, since she tested 30 different conditions with only one item each. This makes a clean statistical analysis impossible and prevents the generalizability of the results since possible effects could also depend on the lexicalization and not on the case combinations.

drop. Topic drop of objects is mainly restricted to the 3rd person, whereas the 1st and 2nd person object pronouns can hardly be omitted, if at all (Fries 1988, Auer 1993, Jaensch 2005, Erteschik-Shir 2007, Volodina & Onea 2012, Duden 2016, 2022, Trutkowski 2016). This is illustrated by the contrast between (8) and (9).[14]

(8) a. *A: Wo warst du so lange?*
where were you.2SG so long
A: 'Where have you been for so long?'
b. *B: *Δ Hat die Chefin gesucht.*
you.ACC.2SG has the boss searched
B: 'The boss has been looking for (you).' (Duden 2016: §1378, adapted)

(9) a. *A: Wo ist bloß wieder mein Schlüssel?*
where is PART again my key
A: 'Where is my key again?'
b. *B:* Δ *Hast du in die oberste Schublade getan.*
that.ACC has you.2SG in the top drawer put
B: 'You put (it) in the top drawer.'

There are different explanations for this pattern. Fries (1988) traces it back to the fact that in German, the reflexive and non-reflexive forms of the 1st and 2nd person object pronouns are identical, while with *sich*, there is a reflexive form of its own for the 3rd person. He argues that in contexts where the antecedent is neither distinctly marked for reflexivity nor non-reflexivity, this syncretism blocks topic drop (Fries 1988: 33), even for the 3rd person. (10) is such an example where according to Fries (1988), the 3rd person reflexive *sich* cannot be omitted, as the antecedent *ihm* ('him') does not entail reflexivity. In (11), in turn, speakers should even be able to omit the 1st person reflexive because the *selbst* ('self') in A's question entails a reflexive usage.

(10) a. *A: Gisbert hat den Sascha rasiert, und was ist mit ihm?*
Gisbert has the Sascha shaved and what is with him
A: 'Gisbert shaved Sascha, and what about him?'
b. *B: *Δ Rasiert er auch.*
himself shaves he too
B: 'He shaves too.' (Lit. '(Himself,) he shaves too.') (Fries 1988: 33, his judgment)

[14]Note that I glossed the omitted constituent in (9) as a demonstrative rather than a personal pronoun. Below, I come back to the difference between demonstratives and personal object pronouns.

(11) a. *A: Mich hast du rasiert, und was ist mit dir selbst?*
me has you shaved and what is with you.DAT self
A: 'You shaved me, and what about yourself?'
b. *B: Δ Rasier' ich auch.*
me shave I too
B: 'I shave too.' (Lit. '(Myself,) I shave too') (Fries 1988: 33, his judgment)

These subtle judgments require an empirical investigation, which cannot be accomplished in this book.[15] However, even if they were correct, it remains unclear why such a syncretism should block topic drop when other ambiguities are generally possible with topic drop, like the one caused by syncretic 1st and 3rd person singular forms of modal verbs tested in experiment 8 (see Section 10.5.3.3).

A further explanation for why object topic drop seems to be restricted to the 3rd person is given by Schulz (2006). She claims that only recoverable constituents can be targeted by topic drop and that a necessary prerequisite for recoverability is that the omitted constituent is a continued topic (Schulz 2006: 8). Schulz (2006: 11) argues that topics in subject function and topics in object function with the 3rd person are more easily recoverable than 1st and 2nd person object topics because the latter are less typical, i.e., more marked continued topics. She bases this claim on universal prominence hierarchies. More specifically, she uses the concept of harmonic alignment from optimality theory (Prince & Smolensky 2004) and applies it to the two subhierarchies for syntactic function and grammatical person (Schulz 2006: 11–12).[16] As a result, 1st and 2nd person objects are ranked lowest and, therefore, should not be able to be omitted (Schulz 2006: 13; 17).[17] While Schulz's approach is more elaborate than Fries's (1988) proposal, two aspects seem problematic. First, it presupposes that all omitted constituents must be continuous topics, i.e., that topicality is a necessary condition for topic drop. In Section 3.1.5, I argued against this view and showed that non-topical constituents can also be omitted. Second, the relationship between salience, topicality, and the possibility of omitting a constituent remains

[15] Note that *ihm* ('him') in (10) is also ambiguous as to whether it refers to Gisbert or to Sascha (with a preference for Gisbert in interpreting the discourse as a whole), further complicating matters.

[16] For the hierarchy of syntactic function, she refers to Keenan & Comrie (1977: 66) but omits the genitive: Subject > direct object > indirect object > major oblique case NP (> genitive) > object of comparison (">" means "more accessible"). For grammatical person she follows Aissen (1999: 674), who in turn derived the hierarchy from Silverstein (1976): local person (1st & 2nd) > 3rd.

[17] Based on this result, Schulz (2006) derives an optimality theoretic model with several constraints for the omission pattern in German, which can also be applied to Japanese null topics.

vague. While it is plausible to assume, based on the two hierarchies discussed, that 1st or 2nd person objects are more marked, it is not clear why this makes them necessarily less good continuous topics, thus, blocking omission.

As a related explanation, Sigurðsson (2011: 290) introduces the so-called *relative specificity constraint* for Germanic topic drop, according to which an omitted object cannot be more specific than the overt subject. He argues that the 1st and 2nd person are more specific than the 3rd person and that [+human] is more specific than [-human]. Thus, a 1st or 2nd person object cannot be omitted if a 3rd person subject is simultaneously realized overtly, as shown in the Swedish example (12a) with a clitic subject pronoun and the German equivalent (12b) (slightly adapted) with a full subject pronoun.

(12) [Context: 'That is Johnson over there, the new manager. We should say hello to him.']

a. *Δ *Vill'an säkert inte prata med nu.*
us/me wants'(h)e certainly not talk with now
'He certainly doesn't want to talk to (us/me) now.' (Sigurðsson 2011: 290, his judgment, translation added)

b. *Δ *will er jetzt bestimmt nicht sehen.*
us/me wants he now certainly not see
'He probably doesn't want to see (us/me) now.'

He states that this restriction is an intervention effect related to his proposed C/edge-linking generalization (see Section 3.6) (Sigurðsson 2011: 291). Sigurðsson (2011: 292) argues that the subject must also be C/edge-linked and that "the dropped object cannot be featurally 'bigger' " than the subject, as it would otherwise violate Relativized Feature Minimality (Rizzi 2001). Since he limits his argument to cases with clitic subject pronouns in Swedish, Norwegian, and Icelandic (Sigurðsson 2011: 291, footnote 36), it is unclear whether and how it could be applied to German and Dutch, where objects can also be omitted in the presence of full subject pronouns.[18]

Mörnsjö (2002), as well as Volodina & Onea (2012) and the Duden (2016, 2022), propose a fourth explanation. Mörnsjö (2002: 70) considers example (13), which Cardinaletti (1990) discusses to illustrate the apparent ban on omitting 1st and 2nd person objects, a "pragmatically inappropriate" answer to the question *Did I disturb you?*, independently of the ellipsis. She argues that, at least in Swedish,

[18]In the FraC, there seems to be no instance that violates the relative specificity constraint.

the full form would be acceptable if the direct object *mich* ('me') is stressed to indicate contrastiveness (see the contrastive topics discussed in Section 3.1.4), but in this case, it cannot be omitted (Mörnsjö 2002: 71). Similarly, Volodina & Onea (2012: 219) argue that when 1st and 2nd person object pronouns are placed in the prefield, this positioning is marked and involves stress, which, in turn, blocks topic drop.

(13) **(Mich) hast du sehr gestört.*
me have you.2SG very disturbed
'You disturbed (me) a lot.' (Cardinaletti 1990: 79, cited in Mörnsjö 2002: 70)

A related argument comes from the Duden (2016, 2022). There, it is stated that weakly stressed object pronouns such as *mich* and *dich* are rarely placed in the prefield and, therefore, rarely omitted from this position (Duden 2016: §1378; Duden 2022: §35). The fact that 3rd person objects can be better omitted may be explained by the fact that in this case, the omitted constituent is not a weak personal pronoun as in (14a) but a demonstrative pronoun as in (14b).

(14) *Was ist mit Tino?*
what is with Tino
'What's with Tino?'
a. *(Ihn) Hab ich gesehen.*
him.ACC have I seen
'I have seen (him).'
b. *(Den) Hab ich gesehen.*
him.ACC.DEM have I seen
'I have seen (him).'

This is supported by the fact that demonstrative pronouns in object function of the sort *den*, *die*, and *das* are a lot more frequent in the prefield than the corresponding personal pronouns, as evidenced by Bosch et al. (2007). In a corpus study of the NEGRA newspaper corpus,[19] they found that less than 0.5% of the 3rd person singular personal pronouns in object function occurred in the prefield, whereas for 3rd person singular demonstrative object pronouns, it were

[19]Bosch et al. (2007: 148, footnote 4) state that the NEGRA corpus, a POS-tagged and syntactically annotated subset of the Frankfurter Rundschau corpus, consists of 355 000 words from news articles.

around 20% (Bosch et al. 2007: 149–150).[20] Since these demonstratives are regularly placed in the prefield without a special marking or stress, they can be omitted from there more easily. No demonstrative counterparts are available for the 1st and 2nd person singular, since they are deictic by themselves. Thus, their positioning in the prefield almost always involves stress, which blocks topic drop. Overall, this last approach to the object asymmetry of topic drop in German seems to me to be the most promising one to explain the pattern. It is based on the incompatibility of prosodic prominence and ellipsis and does not rely on only introspective data or additional assumptions concerning topic continuity or specificity, like the other three approaches.

8.1.3 Predicatives

For completeness, I also mention topic drop of predicatives, even though it is generally a rare phenomenon and, therefore, has not received as much attention in the literature as the omission of subjects and objects. Sternefeld (1985: 406) gives example (15) to prove that topic drop cannot target predicatives. Fries (1988: 29) counters this with example (16), from which he concludes that predicatives can be omitted after all, provided they are used referentially. The Duden (2016: §1378) (also Duden 2022: §35) also argues that predicatives can be omitted in the form of weakly stressed demonstratives. I endorse this view.

(15) a. *A: Weißt du was über Säugetiere?*
know you something about mammals
A: 'Do you know anything about mammals?'

b. *B: *Δ Sind z.B. Wale.*
that are for.example whales
B: '(These) are, for example, whales.' (Sternefeld 1985: 406, his judgment)

[20] Zifonun (2001: 116–117) points out that the object personal pronoun of the 3rd person singular neuter *es* cannot be placed in the prefield at all (i).

(i) *Was ist mit dem Kind?*
what is with the child
'What's with the child?'

a. *(*Es) sehe ich nicht.*
it.ACC see I not
'I don't see (it).'

b. *(Das) sehe ich nicht.*
it.ACC.DEM see I not
'I don't see (it).'

(16) a. *A: Wale sind Säugetiere.*
whales are mammals
A: 'Whales are mammals.'
b. *B:* Δ *Sind Hunde auch.*
that are dogs too
B: 'So are dogs.' (Fries 1988: 29)

8.2 Previous empirical evidence

In this section, I discuss the results of previous empirical studies that investigated the syntactic function of the omitted prefield constituent. The goal is to validate the claims from the theoretical literature regarding the above-mentioned restrictions on subject and object topic drop. The results of three corpus studies (Poitou 1993, Frick 2017, Ruppenhofer 2018) (i) suggest that topic drop of referential and non-referential subjects as well as of direct objects occurs in natural speech data, (ii) do not evidence that subjects are more often omitted than objects, and (iii) provide no evidence that indirect objects are targeted by topic drop. However, an acceptability rating study (Trutkowski 2018) suggests that topic drop of an indirect object can be as acceptable as topic drop of a direct object under certain circumstances.

8.2.1 Poitou (1993)

Poitou (1993) presents the results of a corpus study that he conducted on 200 instances of topic drop. He states that the omission of subjects is most frequent in his corpus with about half of the instances, followed by object omissions (Poitou 1993: 119). The rarest is a so-called "rest group" with cases where an adverbial could occupy the preverbal position in the full form (Poitou 1993: 117).[21] Poitou presents a subdivision for the omitted subjects by animacy. Mostly inanimate subjects are omitted, whereas animate subjects only make up a fifth of the instances he found (Poitou 1993: 116).[22] He further notes that there are cases of non-anaphoric *es* ('it'), i.e., non-referential subject topic drop. For objects, he states

[21] As discussed in Section 2.1.3, I disregard such cases in this book.

[22] Poitou (1993) does not discuss what the animate or inanimate subjects refer to. However, in example (i), which he provides for the inanimate case, taken from Dürrenmatt's 1956 play *Der Besuch der alten Dame [The visit]*, the reference of the omitted constituent would be a proposition. A paraphrase could be something like ⟨*Why nobody pays taxes*⟩ *needs to be investigated.* Based on this example, it can be speculated that the high number of omitted inanimate subjects is the result of many instances of propositional reference. Interestingly, in the English translation by Patrick Bowles, the utterance is translated with overt subject as "It'll have to be investigated".

that only direct but no indirect objects are omitted in his corpus (Poitou 1993: 116, footnote 1).

However, in interpreting Poitou's results, we must take into account that his evidence comes from very diverse sources, i.e., interviews, phone calls, plays, and comics (Poitou 1993: 111). In particular, the unclear proportion of fictional and potentially dialectal speech data extracted from, among others, older plays by the Austrian author Nestroy (19th century) and the Swiss author Dürrenmatt (mid of 20th century) represents a problem for the significance of his results, as it is not clear how authentic and representative the data are. Furthermore, the results are presented without relative or absolute numbers and, thus, can only provide a tentative impression of how topic drop of different syntactic functions is distributed.

In sum, Poitou (1993) confirms that referential and non-referential subjects can be targeted by topic drop and his results tentatively suggest that indirect objects are omitted at least less frequently, if at all. Since he does not provide or did not determine relative numbers, it cannot be concluded that subject topic drop is more frequent than object topic drop.

8.2.2 Frick (2017)

Frick (2017: 38, 42) investigated topic drop and other ellipsis types in a corpus of 3 999 Swiss German text messages. This corpus is a part of the Swiss SMS corpus (Ueberwasser 2015–2022), which was created in the project sms4science.ch between 2009 and 2010. She presents relative numbers for omitted subjects and objects.

Of 4 385 subject pronouns in the prefield, 2 059 (46.96%) were realized and 2 326 (53.04%) were omitted,[23] with large differences in terms of grammatical person. For instance, the omission rate of the 1st person singular (59.32%) is higher than that of the 2nd person singular (47.01%), which in turn is considerably higher than

(i) a. *Der Bürgermeister: Unsere Kassen sind leer. Kein Mensch bezahlt Steuern.*
the mayor our coffers are empty no human pays taxes
The mayor: 'Our coffers are empty. No one pays taxes.'

b. *Der Pfarrer [sic!]:* Δ *Muß untersucht werden.*
the priest that must investigated be
The priest [sic!]: '(That) must be investigated.' (Poitou 1993: 116; in fact the bailiff [orig. Pfändungsbeamte] not the priest utters the topic drop)

[23] Note that the number of 2 326 omitted subjects is calculated based on Frick's (2017) figure 8 on page 88. It slightly deviates from the sum of all omitted subject pronouns, 2 294, which Frick provides in figure 7 on page 86, but it is identical to the sum of the numbers given in figure 10 on page 93.

that of the 2nd person plural (13.79%) (Frick 2017: 88, see Section 10.2.1 for details). Besides referential subjects, she also discusses the omission of non-referential expletives. Of 262 expletives in her data set, 179 (68.32%) were omitted (Frick 2017: 140). This provides further support that non-referential subjects can be targeted by topic drop and, in fact, are in natural speech data.

For objects, Frick (2017: 125–126) lists 68 (53.13%) realized and 60 (46.88%) omitted instances. First, the numbers indicate that subjects are a lot more frequent in the prefield of Swiss German text messages than objects. Second, the omission rate for subjects is overall larger than for objects in absolute terms. However, a Pearson's chi-squared test with Yates's continuity correction, which I calculated in R (R Core Team 2021), reveals that this difference is not significant ($\chi^2(1) = 1.66, p > 0.19$). Frick (2017) does not distinguish between topic drop of direct and indirect objects. Rather than defining object topic drop syntactically, she follows Auer (1993) and Zifonun et al. (1997) by stating that an object can be omitted from the prefield if it is present in the context, which functions as orientation for the interlocutors (Frick 2017: 125). However, the examples that she discusses are only those that contain an overt *das* or those with topic drop where a *das* could naturally be inserted into the prefield (Frick 2017: 126–127), i.e., direct objects in the accusative case Accusative case. This suggests that there are at least mainly direct objects in her data set.

In summary, Frick's (2017) corpus study suggests that topic drop can target referential and non-referential subjects and direct objects, whereas there is no evidence that indirect objects are omitted. Even though subjects seem to be a lot more frequent in the prefield than objects, subject topic drop is not significantly more frequent than object topic drop in her data set, as shown above.

8.2.3 Ruppenhofer (2018)

Ruppenhofer (2018) investigated the frequency of V1 constructions, which in addition to topic drop include conditional inversion, imperatives, yes–no questions, etc.[24] He examined 5 subcorpora containing news articles, websites, tweets, parliamentary speeches, and telephone conversations, as listed in Table 8.1.
He randomly chose 100 V1 constructions[25] from each subcorpus and determined their type. Of the 17 types that Ruppenhofer (2018) considers, four are of rele-

[24]The entire list of the 17 constructions that he investigated contains conditional/concessive inversion, exclamative, apodosis stranding, formulas, reporting inversion, *da*-drop, presentational inversion, contrast inversion, yes-no questions, formal imperative, infinitive imperative, hortative/optative, informal imperative, subject topic drop, subject expletive drop, cataphoric subject drop, and object topic drop. See Ruppenhofer (2018: 218) for details.

[25]These were all sentences where the first token was a verbal form (Ruppenhofer 2018: 217).

Table 8.1: Composition of Rupppenhofer's corpus, taken from Ruppenhofer (2018: 209), adapted

Corpus name	Type	N(Tokens)
Huge German Corpus (HGC)	Newspaper articles	204 813 118
German web corpus (deWaC)	Websites	1 627 169 557
Twitter corpus	Tweets	105 074 399
Bundestag corpus	Parliamentary speeches	5 756 188
CALLHOME German speech	Telephone conversations	202 964

vance here, which he terms subject topic drop, cataphoric subject drop, subject expletive drop, and object topic drop (Ruppenhofer 2018: 218).

Among the sample of 500 V1 constructions, he found 83 omitted referential subjects with an antecedent and 4 with a postcedent (i.e., the cataphoric cases), 13 omitted expletives, and 22 instances of object topic drop. He notes that the different V1 constructions vary clearly between the corpora he used. For instance, the tweets contained the highest proportion of covert arguments. This is in line with the result of my corpus study according to which tweets are the text type in my corpus that exhibits the fourth-highest omission rate with about 35% (see Section 2.1.5 for details).

While Ruppenhofer (2018) provides relative numbers, these are only relative to other V1 utterances but not relative to the corresponding full forms. Therefore, it is also impossible to determine from his corpus whether subject topic drop is more frequent than object topic drop. Ruppenhofer (2018) found no instances where indirect objects were omitted, even when he explicitly searched for verbs that require an indirect object, such as *helfen* ('to help'), *schenken* ('to give as a gift'), or *spenden* ('to donate') in Twitter data (Ruppenhofer 2018: 224–225).

8.2.4 Trutkowski (2018)

Trutkowski (2018) presents an acceptability rating study to test her account of verbatim and non-verbatim topic drop. An unknown number of participants rated the acceptability of 6 utterances[26] on a 7-point scale (7 = absolutely ac-

[26]This seems to be a rather low number of items for a study with 6 experimental conditions and results in each participant seeing each condition only once. When presenting the statistical results, Trutkowski (2016) also mentions herself that instead of testing 6 items, she should have actually tested 30 in order to be able to find significant results in the ANOVA item analysis (F2).

ceptable). In a 2 × 3 design, she crossed Case Antecedent (structural vs. lexical)[27] and Case Gap + Predicate (identical predicate, identical case vs. different predicate, identical case vs. different predicate, different case).[28] In fact, it was a reduced 2 × 2 × 2 design with Case Antecedent (structural vs. lexical), Case Gap (structural vs. lexical), and Predicate (identical vs. different), as exemplified in (17) and (18). The combination of a mismatch between Case Antecedent and Case Gap with an identical predicate is absent because a predicate usually takes arguments with the same case in German.[29] This rules out the *lexical case antecedent, structural case gap, identical predicate* condition and the *structural case antecedent, lexical case gap, identical predicate* condition.

(17) *A: Ich treffe den Peter.*
I meet the.acc Peter
A: 'I am meeting Peter.' (structural antecedent)

a. *B:* Δ *Treffe ich auch.*
him.acc meet I too
B: 'I am meeting (him), too.' (structural gap, identical predicate)

b. *B:* Δ *Mag ich wirklich gern.*
him.acc like I really a.lot
B: 'I really like (him).' (structural gap, different predicate)

c. *B:* Δ *Vertraue ich nicht.*
him.dat trust I not
B: 'I do not trust (him).' (lexical gap, different predicate)

(18) *A: Ich vertraue dem Peter.*
I trust the.dat Peter
A: 'I trust Peter.' (lexical antecedent)

a. *B:* Δ *Vertraue ich auch.*
him.dat trust I too
B: 'I trust (him) too.' (lexical gap, identical predicate)

[27] Trutkowski (2018) uses the term *oblique case*, which traditionally meant any case besides the nominative , but which is synonymous to inherent case in generative grammar and in her usage. For reasons of consistency, I use *lexical case* as the umbrella term for the dative and the genitive. However, it is unclear whether Trutkowski tested only objects in the dative or also objects in the genitive case.

[28] I renamed Trutkowski's variable Case Gap to Case Gap + Predicate to indicate that she considered not only the case of the omitted constituent but also whether the predicate was identical or different between the antecedent and the target utterance.

[29] The *same predicate but different case* condition could theoretically be created by exploiting case alternations, e.g., by passivization, i.e., the change from the nominative to the accusative in the *bekommen*-passive and from the dative to the nominative in the *werden*-passive (see, e.g., Wegener 1990).

b. *B:* Δ *Helfe ich auch gern.*
him.DAT help I too gladly
B: 'I am also happy to help (him).' (lexical gap, different predicate)

c. *B:* Δ *Kenne ich kaum.*
him.ACC know I hardly
B: 'I hardly know (him).' (structural gap, different predicate)
(Trutkowski 2018: 3, adapted)

Table 8.2 shows the mean ratings per condition (Trutkowski provides no standard deviations). Without knowing the corresponding fillers it is hard to judge the descriptive statistics, but the ratings seem relatively low on the 7-point scale.[30] This could already indicate that the omission of (pronouns referring to) 3rd person singular proper names is somehow marked (see also experiments 3, 5, 6, 7, and 8 in this book that tested topic drop of the 3rd person singular referring to persons).[31]

Table 8.2: Mean ratings per condition of Trutkowski's (2018) acceptability rating study, taken from Trutkowski (2018: 3), adapted

Case Antecedent	Case Gap	Predicate	Mean rating	Example
Structural	Structural	Identical	3.90	(17a)
Structural	Structural	Different	3.67	(17b)
Structural	Lexical	Different	2.60	(17c)
Lexical	Lexical	Identical	3.57	(18a)
Lexical	Lexical	Different	2.53	(18b)
Lexical	Structural	Different	3.43	(18c)

Trutkowski (2018) reports a significant main effect of Case Gap + Predicate and a significant interaction between Case Gap + Predicate and Case Antecedent in the two-way ANOVA subject analysis (F1) but not in the item analysis (F2). Items in the *identical case and identical predicate* condition were rated higher than items in the condition *identical case but different predicate* condition, which, in turn, were rated higher than items with *different case and different predicate.*

[30] Trutkowski (2018) does not provide any information whether she included any fillers and if so, which kind of fillers. If she had not used any fillers or marked or ungrammatical fillers, this would support the claim that the ratings for the materials with topic drop were relatively low. However, if she had included grammatical and unmarked structures as fillers, it is likely that these structures would have received very high ratings and that, in turn, the more marked items would have been perceived as degraded relative to these "good" fillers.

[31] A further factor could be that for the identical predicate conditions, a simpler and more natural form is available, which may compete with topic drop: the fragment *ich auch* ('me too').

However, this gradation is only present in the conditions where the antecedent has a structural case , i.e., those in (17). For antecedents with a lexical case, the *identical case but different predicate* condition (18b) received the worst mean rating. The *different case and different predicate* condition (18c) was more acceptable, with a mean close to the *identical case and identical predicate* condition (18a). I assume that this difference is caused by the fact that in the *different case and different predicate* condition with an antecedent with a lexical case, a constituent in the accusative case is omitted (18c), which should generally work better than omitting a constituent in the dative or the genitive case.

Table 8.2 shows that topic drop of a constituent with a lexical case is only as acceptable as topic drop of a constituent with a structural case if both the case and the predicate are identical in the context and the target utterance (18a). In contrast, the other conditions where a constituent with a lexical case is omitted, i.e., when there is either a case mismatch between antecedent and target (17c) or a change in predicate between antecedent utterance and target utterance (18b), are degraded.

From this, Trutkowski concludes that "[t]opic drop of obliquely [i.e., lexically, LS] cased arguments is fine when context and target predicates are the same" and that "case identity is not a sufficient condition for obliquely [i.e., lexically, LS] cased NPs to be dropped" (Trutkowski 2018: 4). However, it remains questionable, first, whether a rating of 3.57 on a 7-point scale for the *identical case and identical predicate* condition with a lexical case can really be called "fine". Second, it is also dubious whether a difference in acceptability of one point to the other two conditions, in which the covert constituent is a constituent with a lexical case, justifies the assumption that topic drop is grammatical in the first case and ungrammatical in the second. This more gradual acceptability cline provides first support for treating the syntactic function not as a licensing condition but as a usage factor.

8.3 Information-theoretic predictions for syntactic function

As I indicated in Section 6.2.1, the *avoid troughs* principle derived from *UID* can potentially explain the influence of several factors on the speaker's choice between the full form and topic drop. The syntactic function of the prefield constituent is one of these factors that have been discussed in the previous literature, partly even as licensing factors for topic drop. In Section 8.1, I reviewed three issues concerning syntactic function. (i) Topic drop of subjects is argued to

be more frequent than topic drop of direct objects, (ii) topic drop of direct objects or objects with a structural case shall be more frequent than topic drop of indirect objects or objects with a lexical case or the latter should not be possible at all, and (iii) topic drop of 1st and 2nd person objects is said to be largely impossible in contrast to topic drop of 3rd person objects. These effects could be explained by the relative frequency of the corresponding full forms.

8.3.1 Subject vs. object topic drop

While the relatively free word order in German allows almost any constituent to occupy the prefield position, in practice, the sentence-initial constituent in V2 sentences is often the subject. Engel (1972: 44) claims that it occupies the prefield in about 60% of the cases, while arguments in the dative or the accusative seem to be very rare in this position. This is supported by the results of several empirical studies. Hansen-Schirra & Gutermuth (2017) analyzed the German-English CroCo corpus, which contains original English and German texts, as well as translations from English into German and vice versa (Hansen-Schirra et al. 2012). What is of interest here are the 121 original German texts with about 288 000 tokens (Hansen-Schirra et al. 2012: 31). There, the prefield position is occupied by the subject in 50.25% of the declarative sentences, by an adverbial in 31%, by an object in 8.46%, and by another element in 10.29% of the sentences (Hansen-Schirra & Gutermuth 2017: 311). Zybatow (2014) classified prefield constituents by syntactic category in 20 textbook texts each, for the subjects German, biology, and history for grades 5 to 9 (Zybatow 2014: 97). She finds that in about 47% of the cases, the prefield constituent is the subject, in about 45% of the cases it is an adverbial, and in about 5% it is an object (Zybatow 2014: 98).[32] Further empirical evidence comes from Bader & Häussler (2010), who compared, among other things, subject vs. object prefield positioning in a data set based on newspaper articles. They restricted themselves to cases where the object was a DP with the masculine definite accusative singular or the dative plural article *den* (see for details Bader & Häussler 2010: 722–723).[33] They found that in their subcorpus

[32] Zybatow presents a subdivision of her data by school subject and grade that reveals that objects in the prefield are most frequent in history books for grade 6, where about 10% of the prefield constituents are objects. See Zybatow (2014: 98) for more details.

[33] They provide the following examples for illustration:

(i) a. *Wir haben den Studenten unterstützt.*
we have the student.ACC.SG supported
'We supported the student.'

b. *Wir haben den Studenten geholfen.*
we have the students.DAT.PL helped
'We helped the students.' (Bader & Häussler 2010: 722)

prefield SO/OS, where they looked at cases with a *den* object at any position, the prefield was occupied by the subject in 58.5% of the cases, by the *den* object in 12.4%, and by another phrase like an adverbial phrase or a prepositional phrase in 29.1% of the cases (Bader & Häussler 2010: 725).

Despite the variation across empirical studies in terms of the specific proportion of objects in the prefield and without neglecting the diverse factors that determine the word order in German (see, e.g., Hoberg 1981, Kempen & Harbusch 2005, Rauth 2020 for word/object order in the middle field and Speyer 2007, Filippova & Strube 2007, Bader et al. 2017 for how the prefield is filled), it seems to be a priori, and if we neglect the precontext, more likely that the subject of the clause will appear in the prefield than another element, such as an object or even a predicative (which seems to be contained in the rest category of Hansen-Schirra & Gutermuth 2017 and which apparently did not appear in the texts investigated by Zybatow 2014). This increased frequency also raises the chance that the subject will be targeted by topic drop in suitable text types[34] because according to the information-theoretic approach, the likelihood of a constituent being omitted increases with its predictability in context. Therefore, the frequency of the full form may explain issue (i), the postulated higher relative frequency of subject topic drop compared to object topic drop. As shown in Section 8.2, this frequency difference between subjects and objects has not yet been empirically proven. Due to the small size of the FraC, I am not able to provide this evidence either (see Section 8.4). This question must remain a research desideratum until a sufficiently large syntactically annotated corpus of a text type that allows for topic drop is available.

8.3.2 Direct vs. indirect objects

Issue (ii), the potential difference between direct and indirect objects or between objects with a structural vs. a lexical case, could be explained similarly as the potential difference between subject and object topic drop. The data from Bader & Häussler (2010) indicate that dative objects occur only half as often in the prefield than accusative objects . Consequently, the likelihood of their omission should also be reduced in relative terms. However, not only did the corpus studies discussed in Section 8.2 find fewer omissions of indirect objects, but they found none at all. On the one hand, this could be because the data sets used were too small to find a sufficient number of the rare indirect objects. On the other

[34] Note that the studies that I just discussed investigated (conceptually) written text types, where topic drop does not or only rarely occur. This explains why the subjects in the prefield are realized despite their high frequency.

hand, the text types studied may also play a role. For example, Bader & Häussler's (2010) examined news articles, whereas topic drop studies usually focus on text messages or spoken language. In the latter two text types, the proportion of indirect objects in the prefield could in turn be lower. This would also result in a lower proportion of topic drop of indirect objects, possibly a proportion that is so small that it cannot be measured in the data sets used. Here, corpus studies of more extensive data sets are pending. As mentioned above, my corpus study was conducted on a rather small, mostly hand-annotated corpus and, thus, cannot provide insights into this issue either. The results of Trutkowski's (2018) acceptability rating study suggest that topic drop of indirect objects can be as acceptable as topic drop of direct objects provided that the antecedent and the omitted constituent have the same case and the verb of the antecedent and the target utterance is the same. However, recall that the acceptability ratings were rather mediocre and that the study has some methodological drawbacks. In information-theoretic terms, it could be that the previous mention of the prefield constituent in the local context with an identical verb could make it highly predictable, even if it is not generally predictable in this position. Such contexts as the ones used by Trutkowski (2018), two subsequent assertions about the same person with the same verb like (18), repeated as (19), however, are special and probably not very frequent in natural language, not least because the fragment *ich auch* ('me too') is the more natural alternative for (19b).

(19) a. *A: Ich vertraue dem Peter.*
I trust the.DAT Peter
A: 'I trust Peter.'

b. *B:* Δ *Vertraue ich auch.*
him.DAT trust I too
B: 'I trust (him) too.' (Trutkowski 2018: 3, adapted)

8.3.3 1st and 2nd person objects vs. 3rd person objects

This leaves us with issue (iii), which is also related to grammatical person: the nonexistence or at least extremely low frequency of object topic drop with the 1st and the 2nd person. As discussed in Section 8.1, the Duden (2016, 2022) argues that weakly stressed 1st and 2nd person object pronouns rarely occur in the prefield and, thus, are rarely omitted from there (Duden 2016: §1378, Duden 2022: §35). Similarly, Volodina & Onea (2012: 219) state that when 1st and 2nd person object pronouns occur in the prefield, this involves stress, which in turn blocks topic drop. I conducted a corpus study to investigate the information-theoretic

frequency hypothesis for issue (iii). Specifically, I investigated whether the low number of topic drop of 1st and 2nd person object pronouns can be explained by a likewise low number of corresponding full forms with overt 1st and 2nd person object pronouns in the prefield.

I searched the German reference corpus DeReKo (Leibniz-Institut für Deutsche Sprache 2022) to assess the general frequency of subjects and objects, as well as their frequency in the prefield. Specifically, I searched for subject and object pronouns with different cases and grammatical persons in all corpora of the archive TAGGED-T of DeReKo. This archive is large enough that one can expect to find a substantial amount of utterances with object pronouns in the prefield. It contains almost 4.5 million texts (newspaper articles, Wikipedia articles, speeches, and interviews) with more than 1 billion words from 1997 to 2010, which were part-of-speech tagged using the TreeTagger (Schmid 1994, 1995). I restricted myself to subject pronouns in the nominative case and object pronouns in the accusative and the dative case of the 1st and 2nd person singular and the 3rd person singular masculine to have as few syncretic forms as possible[35] to ensure that the matches for one form correspond as good as possible to the matches for one function. For each form, I assessed (a) how frequent it is generally in the data set and (b) how frequent it is in the prefield. I approximated the positioning in the prefield by searching for sentence-initial occurrences that are followed by a finite verb.[36] Table 8.3 shows the results including the proportion of prefield occurrences in the total number of occurrences.

I assume that two quantities are relevant to the argument here. First, the total number of occurrences in the prefield is an estimate of how likely it is that any given utterance begins with the corresponding pronoun as the prefield con-

[35]While the 3rd person masculine object pronoun in the dative case, *ihm*, is identical to the form of the 3rd person neuter object pronoun in the dative, the paradigms of the plural persons and of the 3rd person singular of different gender exhibit even more syncretic forms (marked in bold):

	3SG F	3SG N	1PL	2PL	3PL
Nominative	***sie***	***es***	*wir*	***ihr***	***sie***
Accusative	***sie***	***es***	***uns***	***euch***	***sie***
Dative	***ihr***	***ihm***	***uns***	***euch***	*ihnen*

[36]I used the following query: `mich /w0 <sa> /+w1 MORPH(VRB fin)`. For the demonstrative 3rd person pronouns, I additionally included an expression that ensured that the *der*, *den*, or *dem* was used as a demonstrative and not as a determiner or relative pronoun: `den /w0 MORPH(PRON dem sub) /w0 <sa> /+w1 MORPH(VRB fin)`.

Table 8.3: Frequency of personal and demonstrative subject and object pronouns of the 1st person singular, the 2nd person singular, and the 3rd person singular masculine in the DeReKo TAGGED-T archive

Pronoun	Person, number	Gender	Case	Total occurrences	Occurrences in prefield	Proportion in prefield
ich	1SG	–	NOM	2 068 332	570 809	28.60%
du	2SG	–	NOM	130 166	10 230	7.86%
er	3SG	M	NOM	3 808 561	721 514	18.94%
*der**	3SG	M	NOM	7 035	3 882	55.18%
mich	1SG	–	ACC	380 986	4 534	1.19%
dich	2SG	–	ACC	33 224	110	0.33%
ihn	3SG	M	ACC	410 037	3 510	0.86%
*den**	3SG	M	ACC	10 172	2 795	27.48%
mir	1SG	–	DAT	346 165	18 192	5.26%
dir	2SG	–	DAT	28 869	121	0.42%
ihm	3SG	M/N	DAT	442 615	15 465	3.49%
*dem**	3SG	M/N	DAT	11 985	9 952	83.04%

*demonstrative pronoun

stituent.[37] Second, the proportion of occurrences in the prefield to the total number of occurrences serves as an estimate for the probability that an utterance begins with the corresponding pronoun as the prefield constituent, provided that the corresponding pronoun occurs in the utterance at all.

The data confirm the observations just made about the difference in frequency between subjects and objects in the prefield. The personal pronouns of all grammatical persons are more frequent as subjects than as objects, both generally and in the prefield (but not so the demonstrative pronouns, which we come to below).

In addition, we see that the 1st person singular object pronouns *mich* and *mir* are rare in the prefield according to both measures. The number of occurrences in the prefield of *mir* (about 18 000) and *mich* (about 4 500) is remarkably lower than the number of occurrences of the 1st person singular subject pronoun *ich*

[37] This number could be converted into a proportion by dividing it by the total number of (declarative) utterances in the corpus. However, since this total number is the same for all pronouns, the ratios can also be read from the absolute numbers.

(about 570 000). This suggests that it is generally less likely for an utterance to start with either of the two object pronouns than with the corresponding subject pronoun. The same result arises from the relative numbers. While almost 30% of all *ich* occurrences are in the prefield position, for the dative object pronoun *mir* it is just over 5%, and for the accusative object pronoun *mich* even only around 1%. The 2nd person singular shows a similar tendency, with even lower absolute and relative values. At first glance, also the pattern for the 3rd person singular masculine seems to be the same. The subject pronoun *er* appears more than 720 000 times in the prefield and in almost 20% of the utterances. In turn, the dative object pronoun *ihm* occurs about 15 000 times in the prefield and only in about 3.5% of the utterances, and the accusative object pronoun *ihn* only about 3 500 times and in less than 1% of the utterances. The picture partly changes when we look not only at the 3rd person personal pronouns but also at the demonstratives. The likelihood that an utterance starts with a 3rd person demonstrative pronoun is relatively low, about 2 800 occurrences of *den* occur in the prefield, about 10 000 for *dem*, and about 4 000 for *der*. However, if they occur in an utterance, they do so particularly frequently in the prefield, in over 80% of the cases for the dative, in more than 55% for the nominative, and still in around 27% for the accusative. This confirms the results of Bosch et al. (2007) mentioned in Section 8.1, according to which demonstrative pronouns are more frequent in the prefield than personal pronouns. Furthermore, it suggests that a more or less equivalent full form for a 3rd person singular topic drop should be formed with a demonstrative rather than a personal pronoun in the prefield.

The results provide evidence for the frequency-based information-theoretic explanation. Full forms with 1st or 2nd person singular object pronouns in the prefield are rare in absolute and relative terms, the latter in particular in comparison to full forms with 3rd person demonstrative object pronouns in the prefield. Consequently, *mich*, *dich*, *mir*, and *dir* are less predictable in this position and cannot be omitted equally well.

For the hypothesis that the positioning of 1st and 2nd person object pronouns in the prefield requires contrast or special stress, the corpus data from the DeReKo provide mixed evidence. There are cases such as (20), where *dich* cannot be omitted according to my judgment and where the object pronoun in the prefield is indeed contrastive or would require stress.

(20) a. *Ihre Goldene Hochzeit feiern heute Renate und Helmut Urschel. Seit 50 Jahren geht das Bad Münsterer Ehepaar gemeinsam durchs Leben.*
'Renate and Helmut Urschel are celebrating their golden wedding anniversary today. The couple from Bad Münster has been going through life together for 50 years.'

b. *„*(Dich) heirate ich einmal", sagte Helmut Urschel seiner*
you.ACC.2SG marry I once said Helmut Urschel his
späteren Frau gleich am ersten Abend beim Tanz
eventual wife immediately on.the first evening at.the dance
im Hotel Kaiserhof.
in.the hotel Kaiserhof
' 'You're the one I'm going to marry one day,' Helmut Urschel told his future wife on the very first evening at the dance in the Hotel Kaiserhof.' [DeReKo, RHZ07/AUG.29363, Rhein-Zeitung, 08/30/2007], my judgment

For other examples such as (21), I would likewise argue that *mir* cannot be omitted, but the object pronoun in the prefield does not seem to be used contrastively nor does it necessarily require special stress. The hypothesis that stress blocks omission cannot explain why topic drop is not possible in this case.

(21) a. *Ich spiele gerne Fußball, das ist mein liebstes Hobby. Ich spiele beim VfB Fallersleben in der E-Jugend und mache das schon seit fünf Jahren. Ich bin in meiner Mannschaft der rechte Stürmer.*
'I like playing soccer, it's my favorite hobby. I play for VfB Fallersleben in the under-11 team and have been doing so for five years. I'm the right striker on my team.'

b. **(Mir) macht das einfach Spaß, und Tore schießen ist natürlich*
me.DAT makes that simply fun and goals scoring is of.course
am schönsten.
at.the most.beautifully
'I just enjoy it, and scoring goals is the best thing, of course.' [DeReKo, BRZ08/JAN.12343, Braunschweiger Zeitung, 01/28/2008], my judgment

Also in information-theoretic terms, there seems to be no clear explanation for why topic drop is not possible or at least very marked. The speaker, to whom the pronoun refers, is given and occurs linguistically as the subject in the previous utterances. Therefore, it is predictable and its recovery in the critical utterance should be perfectly fine. Still, though, topic drop seems to be blocked. So while the general pattern of object topic drop with different grammatical persons is accounted for by the information-theoretic explanation based on frequencies, examples such as (21) suggest that it may not be sufficient to explain the whole picture.

At this point, it may be fruitful to briefly come back to the other proposed explanations for the potential ban on 1st and 2nd person object topic drop from the literature, discussed in Section 8.1. Schulz (2006) proposes that 1st and 2nd person objects are harder to omit because they are more marked as continued topics. Recall that she considers being a continued topic a prerequisite for topic drop. In example (21), however, the 1st person singular object pronoun must be considered a rather optimal continued topic. In the previous utterances, the speaker is clearly the topic denotation. Thus, *mir* can be considered an ideal continued topic expression, which should be easily recoverable. The fact that it still cannot be omitted, therefore, also poses a problem for Schulz's (2006) account. In contrast, Fries's (1988) approach according to which the syncretism between 1st and 2nd person object and reflexive pronouns blocks their omission if the context is not explicitly marked for (non-)reflexivity can account for example (21). The form of the preverbal object pronoun *mir* is syncretic with the corresponding reflexive *mir* (e.g., *Ich merke mir das* (lit. 'I memorize myself that')), the context does not explicitly provide a non-reflexive reading, and, therefore, topic drop is blocked, as predicted by Fries (1988). Furthermore, also Sigurðsson's (2011) relative specificity constraint according to which an omitted object cannot be more specific than the overt subject, correctly predicts topic drop to be impossible in (21). The omitted object *mir* is a 1st person singular element and, thus, more specific than the overt subject *das*, referring to the 3rd person referent *Fußball spielen*, which is furthermore [-human]. Thus, also the higher specificity of the object compared to the subject may block its omission.

In sum, neither the information-theoretic account, nor the account based on stress and contrast, nor Schulz's (2006) continued topic account can explain why topic drop seems to be impossible in example (21). By contrast, it is Fries's (1988) reflexivity account and Sigurðsson's (2011) relative specificity constraint that adequately describe the introspective judgment. Here, then, the information-theoretic approach motivated purely by frequencies seems to reach its limits for the first time. It seems that an additional factor to explain the choice between full forms and topic drop is needed. Determining what that factor is must be left to future research.

8.4 Corpus study of syntactic function

In the following, I present the results concerning syntactic function from two of the three data sets that I analyzed in my corpus study. I looked at the frequencies and omission rates of referential and non-referential subjects, as well as of

accusative and dative objects. Due to data sparsity, I cannot provide an inferential statistical analysis, but I present a descriptive overview that may inform in-depth research on this factor in the future.[38]

8.4.1 Syntactic function in FraC-TD-Comp

As sketched in Section 7.2.3.1, the FraC-TD-Comp data set contains 873 instances of topic drop and 3 211 full forms that principally allow for topic drop (neglecting the typical occurrence of topic drop in certain text types, see Section 2.1.4). Table 8.4 shows the frequency per syntactic function and the omission rate in FraC-TD-Comp.

Table 8.4: Full forms, instances of topic drop, and omission rates as a function of syntactic function in the FraC-TD-Comp data set

Syntactic function	Full form	Topic drop	Total	Omission rate
Subject	3 051	807	3 858	20.92%
– referential	2 780	778	3 558	21.87%
– non-referential	271	29	300	9.67%
Object	160	66	226	29.20%
– accusative	152	66	218	32.14%
– dative	8	0	8	0.00%

In the prefield position of the utterances in FraC-TD-Comp, there are either overt or covert subjects in the nominative case, overt or covert objects in the accusative case, or overt objects in the dative case. There is no instance where topic drop targets an indirect object in the dative case. Objects in the genitive case or prepositional objects never occur in the prefield, neither overtly as pronouns nor covertly.

For subjects, it can be differentiated between referential and non-referential subjects (see the theoretical discussion in Section 8.1), namely the expletive *es* in subject function.[39] There are 271 instances with an overt expletive pronoun in the prefield and 29 cases where an expletive is omitted from this position. This

[38] The corpus data and the analysis scripts can be accessed online: https://osf.io/zh7tr. For copyright reasons, I cannot provide the actual linguistic material but only the IDs and the annotations of each instance.

[39] For the reasons discussed in Footnote 9 and for consistency, I classified all occurrences of *man* as referential.

result is consistent with the corpus data discussed in Section 3.1.5, as well as with the result of experiment 1. Expletives can be targeted by topic drop.

What stands out is the fact that objects in the prefield generally seem to be much rarer than subjects. This is also true for accusative objects but especially for dative objects, of which only 8 appear in pronominalized form in the prefield and none as topic drop.[40] Due to the small size of the corpus, no definitive conclusions can be drawn about the omission of indirect objects or objects with a lexical case, but this observation is consistent with the assumption that topic drop of dative objects (and of genitive and prepositional objects) occurs rarely or not at all in corpus data because the corresponding full forms are already very infrequent.

8.4.2 Syntactic function in FraC-TD-SMS

Table 8.5 shows the frequency of topic drop and the omission rate by syntactic function in the FraC-TD-SMS data set, a subset of FraC-TD-Comp that is restricted to only the 353 occurrences of topic drop and the 201 full forms in the text message subcorpus of the FraC (see Section 7.2.3.2).

Table 8.5: Full forms, instances of topic drop, and omission rates as a function of syntactic function in the FraC-TD-SMS data set

Syntactic function	Full form	Topic drop	Total	Omission rate
Subject	199	348	547	63.62%
– referential	194	337	531	63.47%
– non-referential	5	11	16	68.75%
Object	2	5	7	71.43%
– accusative	2	5	7	71.43%
– dative	0	0	0	0.00%

Like in FraC-TD-Comp, most instances of topic drop in this data set are referential subjects. They are significantly more often omitted than realized ($\chi^2(1) = 38.51$, $p < 0.001$), according to a chi-squared test for given probabilities calculated in R (R Core Team 2021) against a 50:50 baseline. Their omission rate of 63.47% is even 10% higher than the 53% that Frick (2017) attested in her larger data set of Swiss German text messages, a significant difference according to

[40]The eight dative objects occur in seven different text types: two in opinion pieces and one each in a news article, blog, online chat, email, radio transcript, and tweet. The diversity of these text types does not hint at a systematic influence of the text type, as one might have assumed.

a Pearson's chi-squared test with Yates's continuity correction calculated in R ($\chi^2(1) = 20.3, p < 0.001$).

For both objects and non-referential subjects, there are fewer than 20 instances each in the FraC-TD-SMS data set. Therefore, the resulting omission rates (non-referential subjects: 68.75%, direct objects: 71.43%) are hardly reliable. A much larger corpus of text messages would be needed to find enough corresponding instances of object topic drop and non-referential subject topic drop to compare with Frick's (2017) Swiss German data. This is even more necessary for indirect objects and predicates, for which I did not find any instances neither in the FraC-TD-SMS nor in the larger FraC-TD-Comp data set.

8.5 Summary: syntactic function

In this chapter, I bundled the relevant theoretical and empirical literature on the syntactic function of the prefield constituent as a factor for topic drop. I presented my qualitative empirical results and discussed them in light of my information-theoretic predictions. For subjects, I already showed in Section 3.1.5 in the first part of this book that topic drop is not restricted to referential subjects but can also target non-referential expletives. This position was strengthened through the results of the corpus studies reported and through my own corpus results. For object topic drop, I presented three issues in the theoretical overview. (i) Object topic drop may be rarer than subject topic drop. (ii) There may be a difference in how well direct vs. indirect objects or objects with a structural vs. a lexical case can be omitted. (iii) 3rd person objects may be better omitted than 1st and 2nd person objects (if the latter can be omitted at all).

The three corpus studies from the previous literature that I discussed and my corpus investigations provided no support for claim (i) that topic drop more frequently targets subjects than objects. It must be noted, however, that it was only possible to determine and compare relative frequencies in Frick's (2017) study and my own since Poitou (1993) and Ruppenhofer (2018) did not determine full forms as baselines, i.e., complete utterances with subjects or objects in the prefield that could be omitted but were not. Since Frick's study is based on Swiss German data and since my study suffered from data sparsity, the determination of possible frequency differences between subject and object topic drop in the Standard German of Germany remains a research desideratum.

The corpus studies also do not reveal how frequently indirect objects or objects with a lexical case are targeted by topic drop. Neither Poitou (1993), Frick (2017), Ruppenhofer (2018), nor myself found instances of topic drop of dative or genitive

objects, not even with identical predicates.[41] As a possible information-theoretic, frequency-based explanation, I discussed that they do not occur because the corresponding full forms with an overt indirect object in the prefield are already very infrequent. Despite some methodological shortcomings, Trutkowski's (2018) acceptability rating study suggests at least that topic drop of an indirect object or an object with a lexical case can be as acceptable as topic drop of a direct object or an object with a structural case if the antecedent and the omitted constituent have the same case and the verbs or predicates of the antecedent and the target utterance are the same.

For (iii), the object asymmetry with respect to grammatical person, I argued initially that it is most reasonable to adopt Mörnsjö's (2002) and Volodina & Onea's (2012) explanation according to which 1st and 2nd person objects, unlike 3rd person objects, are usually stressed when they occur in the prefield and that this prosodic prominence blocks ellipsis. The empirical studies from the previous literature did not report the grammatical person of the omitted objects, nor do the studies presented in Chapter 10 on grammatical person (they focus on subjects). Based on the results of my corpus study of the DeReKo, I argued that the presumed impossibility or very low frequency of topic drop of objects of the 1st and 2nd person at least partly hinges on the rarity of the corresponding full forms. At the same time, however, it became evident that neither the frequency nor the incompatibility of stress and omission can explain the whole picture. An authentic example with an unstressed 1st person dative object pronoun in the prefield that cannot be omitted suggests that the impossibility of 1st and 2nd person object topic drop is impacted by further factors such as the syncretism with reflexive forms (Fries 1988) or the relative specificity constraint (Sigurðsson 2011).

A final remark concerns predicatives. They are not only rarely discussed in the theoretical literature but were also not considered in the empirical studies. Thus, there is no corpus or rating data that allows me to judge how often or well they can be omitted. Also, there were no cases of covert predicatives in the FraC data sets. I leave it to future research to address this research desideratum.

In the next chapter, I revisit the factor of topicality from the first part of this book, this time considering it as a usage factor.

[41] Note that in neither of the corpus studies from the literature, nor in my own study there was a distinction between a lexical and a structural accusative case as proposed by Sternefeld's (1985). Thus, there is no empirical basis to verify whether they behave differently.

9 Topicality

In Section 3.1, I argued that topicality is not a (strictly) sufficient condition for topic drop, and, more importantly, that it is also not a necessary one. From this, I concluded that topicality is not a licensing condition for topic drop. Nevertheless, it remains an open issue whether topicality still plays a role in topic drop as a factor that impacts its usage. The guiding question of the present chapter is whether topical constituents are more often targeted by topic drop than non-topical constituents. This would be expected from an information-theoretic point of view because topics are often held constant across multiple utterances, and thus should be predictable and more likely to be targeted by topic drop.

Since I already provided an overview of the relevant theoretical literature on topicality in Section 3.1, I refrain from further theoretical discussion in this chapter. Instead, I first present the predictions of my information-theoretic account and then turn to the four acceptability ratings experiments on topicality, which, to the best of my knowledge, are the first empirical studies that investigate the role of topicality for topic drop. It was not possible to examine topicality with the corpus analysis because the FraC is not annotated for information structure. Since for many utterances, no precontext is available, it would not be feasible to perform such an annotation myself.

9.1 Information-theoretic predictions for topicality

The information-theoretic predictions regarding topicality follow from the *avoid troughs* principle (Section 6.2.1) according to which predictable constituents are more likely to be targeted by topic drop. To describe the relationship between topicality and predictability, we can start with Krifka (2007). He points out, referring to Givón (1983), that "[t]here is a well-documented tendency to keep the topic constant over longer stretches of discourse" (Krifka 2007: 43), i.e., to form so-called *topic chains*. Consequently, a hearer should expect that in an utterance following such a chain, the topic will remain constant instead of changing. If we further realize that topics tend to occur in the prefield (but cf. Section 3.1.3), this means that a constituent referring to the topic of the previous utterance is

not only likely to generally occur in the current utterance, but is particularly likely to occur in the prefield. This means that the topic of an utterance in the prefield should generally be more predictable , and therefore more likely to be omitted if the speaker adheres to *UID* and avoids troughs by omitting predictable constituents, where grammar allows it.

9.2 Experimental investigations of topicality

In the following, I present four acceptability rating studies (experiments 5 to 8), which investigated the impact of topicality on topic drop in a controlled setting using minimal pairs. These studies used three different methods to manipulate the topic status of the prefield constituent: a simple question method, a complex question method, and a method using the subject function. At the same time, in three of these studies, I also investigated the role of grammatical person experimentally. The aim was to test the claim from the literature that topic drop of the 1st person singular is particularly frequent or acceptable by comparing it and the corresponding full form to 3rd person singular pronouns referring to humans. I do not discuss the corresponding results in this chapter but in the next one, specifically in Section 10.5.

9.2.1 Experiment 5: topicality (simple question method)

In Section 3.1 in the first part of this book, I argued that topicality is not a (strictly) sufficient condition for topic drop and showed by means of corpus and experimental data that it is not a necessary condition either. In this experiment,[1] I investigate whether it is at least a factor that impacts the usage of topic drop. I examined whether topic drop is more acceptable when the omitted prefield constituent is coreferential with the discourse topic, which is set through a question beforehand. As I discussed in Section 9.1, my information-theoretic account can explain such an effect. The discourse topic is what the corresponding discourse is about and is often held constant across multiple utterances. This makes it predictable and a good candidate for omitting it to avoid a surprisal minimum.

In addition to the effect of topicality, this experiment also investigated the influence of grammatical person on topic drop by comparing the 1st and the 3rd person singular (see Section 10.5.1). It crossed the three binary predictors

[1] This experiment as well as experiments 7 and 8 were part of my bachelor thesis (Schäfer 2019). For this book, I reanalyzed the data considering further effects and using a different coding schema. Additionally, I expanded and modified the conclusions drawn. The materials and the analysis script can be accessed online: https://osf.io/zh7tr.

Topicality, grammatical Person, and Completeness, which resulted in a 2 × 2 × 2 within-subjects design. For topicality, an interaction between Topicality and Completeness is predicted. Topic drop should be more acceptable when the omitted prefield constituent is coreferential with the current discourse topic.

9.2.1.1 Materials

Items

The items were short instant messaging dialogues between two persons, A and B. They all exhibited the same overall structure, illustrated in (1).

(1) a. (i) *A: Na, wie läuft's bei dir?*
well how runs.it at you.DAT.2SG
A: 'Well, how's it going with you?'
(ii) *B: Vorhin haben Anna und ich den Vertrag unterschrieben.*
earlier have Anna and I the contract signed
B: 'Earlier Anna and I signed the contract.'
(iii) *B: (Ich) befördere sie zur Abteilungsleiterin.*
I promote her to.the head.of.the.department.FEM
B: '(I) promote her to head of the department.'
(topic drop / full form, 1SG, topic)

b. (i) *A: Na, wie läuft's bei Anna?*
well how runs.it at Anna
A: 'Well, how's it going with Anna?'
(ii) *B: Vorhin haben Anna und ich den Vertrag unterschrieben.*
earlier have Anna and I the contract signed
B: 'Earlier Anna and I signed the contract.'
(iii) *B: (Ich) befördere sie zur Abteilungsleiterin.*
I promote her to.the head.of.the.department.FEM
B: '(I) promote her to head of the department.'
(topic drop / full form, 1SG, no topic)

c. (i) *A: Na, wie läuft's bei Anna?*
well how runs.it at Anna
A: 'Well, how's it going with Anna?'
(ii) *B: Vorhin haben Anna und ich den Vertrag unterschrieben.*
earlier have Anna and I the contract signed
B: 'Earlier Anna and I signed the contract.'
(iii) *B: (Sie) befördert mich zur Abteilungsleiterin.*
she promotes me to.the head.of.the.department.FEM

B: '(She) promotes me to head of the department.'
(topic drop / full form, 3SG, topic)

d. (i) *A: Na, wie läuft's bei dir?*
well how runs.it at you.DAT.2SG
A: 'Well, how's it going with you?'

(ii) *B: Vorhin haben Anna und ich den Vertrag unterschrieben.*
earlier have Anna and I the contract signed
B: 'Earlier Anna and I signed the contract.'

(iii) *B: (Sie) befördert mich zur Abteilungsleiterin.*
she promotes me to.the head.of.the.department.FEM
B: '(She) promotes me to head of the department.'
(topic drop / full form, 3SG, no topic)

The conversation started with a question or a request by person A, who wants to get news about either person B, as in (1a-i) and (1c-i), or a third person, e.g., Anna, as in (1b-i) and (1d-i). This served the purpose of setting either person B or Anna as the (discourse) topic. Next, person B provided the requested news by replying with two utterances. The first utterance, namely (1a-ii), (1b-ii), (1c-ii), or (1d-ii), took up the topic, which was either the speaker, i.e., person B, or the third person, Anna. Additionally, it introduced the other person, i.e., Anna or person B, by naming an event in which both are involved such as signing a contract. In this utterance, Anna and person B always appeared as the coordinated subject in the middle field to keep the topic prominent: *Anna und ich* ('Anna and I'). This ordering was held constant across conditions for reasons of politeness (it is considered impolite in German if the speaker names themselves first). The prefield was always filled with a temporal adverbial, such as *vorhin* ('earlier').

The second utterance of person B was the critical utterance to be rated by the participants. There, either person B, as in (1a-iii) and (1b-iii), or Anna, as in (1c-iii) and (1d-iii), was picked up as the realized or omitted preverbal subject. If the person about whom person A wants to get news was identical to the preverbal constituent, this was a *topic* condition, otherwise, it was a *not topic* condition. The other referent appeared in a subordinate syntactic position as an object or adverbial. Topic drop was always followed by the inflected form of a lexical verb in present indicative, which was distinctly marked in terms of grammatical person.

Fillers

The 24 items were mixed with 84 fillers, which were also dialogues between two persons. 28 of these fillers were items for an experiment on gapping and right node raising consisting of three turns. Thus, they also contained (the possibility of) optional omissions. This was to prevent the items with topic drop from

being the only elliptical materials in the experiment. Seven of the fillers were dialogues with an ungrammatical target utterance containing two finite verbs in a single clause. I included them as attention checks to be able to exclude the data from inattentive participants from the analysis. The remaining 49 fillers were dialogues with three or four turns and were designed to represent a broad acceptability spectrum ranging from perfectly acceptable target utterances to marked ones, which should only be slightly more acceptable than the ungrammatical catch trials. I ensured that none of the filler utterances started with a DP, which could potentially be targeted by topic drop. This was to prevent participants from being primed by seeing more full forms than utterances with topic drop.

9.2.1.2 Procedure

48 native speakers of German between the ages of 18 and 50 who had not taken part in any of my other studies of topic drop participated in the experiment. They were recruited from the crowdsourcing platform Clickworker (Clickworker 2022) and received a compensation of €2.50. The study was implemented with the LimeSurvey survey presentation software (Limesurvey GmbH 2023). The participants' task was to rate the naturalness of the last italicized utterance of a small instant messaging dialogue on a 7-point Likert scale with labeled endpoints (1 was "vollkommen unnatürlich" ('completely unnatural') and 7 "vollkommen natürlich" ('completely natural')). The items were distributed across eight lists according to a Latin square design so that each participant rated each of the 24 critical token sets exactly once and only in one of the eight conditions. They were mixed with the fillers and shown to the participants in individual pseudo-randomized order ensuring that no two items or fillers of the same category immediately followed each other. The utterances were presented as instant messages (similar to those in the previous experiments, see Section 3.1.6.2) and were fully lowercased to keep the conditions with and without topic drop parallel in terms of the spelling of the verb.[2]

[2]While writing everything in lowercase is a common stylistic device in text messages (Schnitzer 2012: 152), the majority of writers tries to adhere to the rules of capitalization. Dittmann et al. (2007: 15–16) report that in a corpus of 847 text messages by 115 persons collected in 2002, 60.1% of the text messages exhibit correct capitalization. The majority of capitalization violations were due to messages that were either completely capitalized or completely lowercased. Therefore, I abandoned the lowercase spelling in later experiments, such as those presented in the first part of this book, in favor of greater naturalness.

9.2.1.3 Results

I excluded six participants because they had rated four or more of the ungrammatical control utterances with a 6 or 7 on the Likert scale, which suggests that they had been inattentive. Table 9.1 shows the mean ratings and standard deviations per condition for the remaining 42 participants. In Figure 9.1, the mean ratings and 95% confidence intervals are plotted.

Table 9.1: Mean ratings and standard deviations per condition for experiment 5

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.48	1.61
Topic drop	1SG	Topic	4.76	1.87
Full form	3SG	Topic	5.71	1.45
Topic drop	3SG	Topic	4.13	1.88
Full form	1SG	No topic	5.45	1.69
Topic drop	1SG	No topic	4.90	1.69
Full form	3SG	No topic	5.60	1.58
Topic drop	3SG	No topic	3.71	1.88

The data were analyzed with cumulative link mixed models (CLMMs) (Christensen 2019) in R (R Core Team 2021), following the procedure described in Section 3.1.6.3. The full model contained the ordinal ratings as the dependent variable and as independent variables the binary predictors Completeness, Person, and Topicality, which were coded using deviation coding (full form, 1SG, and topic were coded as 0.5, the other levels as −0.5), as well as the numeric scaled and centered Position at which the trial appeared in the experiment, the three-way interaction between Completeness, Person, and Topicality[3] and all two-way interactions between the predictors. The random effects structure consisted of random intercepts for participants and items and of by-item and by-subject random slopes for all three binary predictors and their two-way interactions, as

[3] Although I have no concrete predictions about a joint influence of topicality and grammatical person on the acceptability of topic drop, I included it in my analysis because it could be of theoretical interest, unlike the other possible three-way interactions with Position.

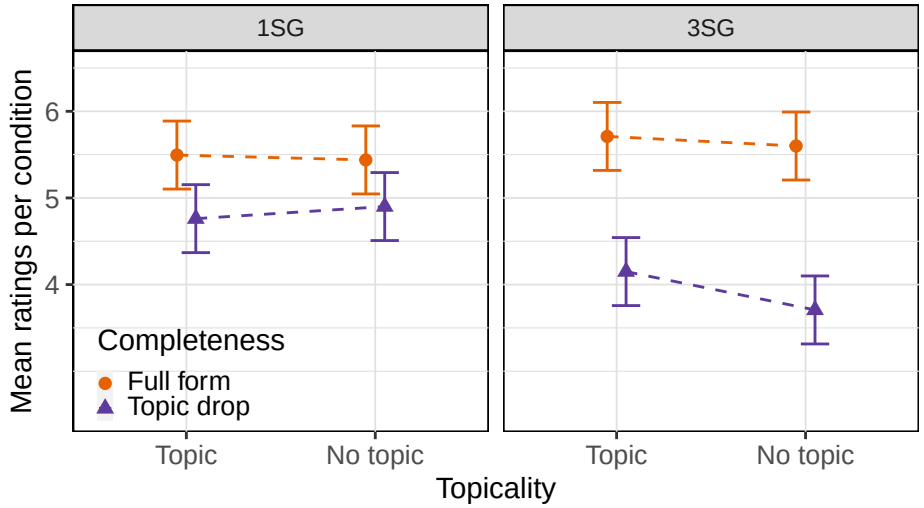


Figure 9.1: Mean ratings and 95% confidence intervals per condition for experiment 5

well as for Position.[4] I performed a backward model selection to obtain the final model, which in this analysis had symmetric thresholds. Table 9.2 shows the fixed effects in this model.

Table 9.2: Fixed effects in the final CLMM of experiment 5

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	2.03	0.34	26.67	< 0.001	***
Person	0.55	0.23	5.30	< 0.05	*
Completeness × Person	−1.57	0.34	15.07	< 0.001	***

There were significant main effects of Completeness ($\chi^2(1) = 26.67$, $p < 0.001$) and Person ($\chi^2(1) = 5.30$, $p < 0.05$), as well as a significant interaction between them ($\chi^2(1) = 15.07$, $p < 0.001$). Full forms received significantly better ratings than utterances with topic drop. Utterances with the 1st person singular were preferred over utterances with the 3rd person singular. Topic drop of the 3rd

[4]The formula of the full model was as follows: `Ratings ~ Completeness : Topicality : Person + (Completeness + Topicality + Person + Position)^2 + (1 + (Completeness + Topicality + Person)^2 + Position | Subjects) + (1 + (Completeness + Topicality + Person)^2 + Position | Items)`.

person was particularly degraded. The main effect of Topicality was not significant ($\chi^2(1) = 1.47, p > 0.2$), nor was the interaction of Topicality with Completeness ($\chi^2(1) = 0.0001, p > 0.9$), nor the three-way interaction between Completeness, Topicality, and Person ($\chi^2(1) = 2.61, p > 0.1$), nor any of the other effects.

9.2.1.4 Discussion

Experiment 5 intended to test for effects of topicality and grammatical person on the acceptability of topic drop. Whereas the results for grammatical person are discussed in Section 10.5.1, its results for topicality provide no evidence that this factor plays a role in the acceptability of topic drop. The predictor Topicality was not involved in any significant effect, i.e., topic drop was not more acceptable when the omitted constituent was coreferential with the current discourse topic. A possible concern with the design of the experiment is the context sentence, i.e., person B's second utterance. There, the previously set discourse topic appeared as part of the coordinated subject *Anna und ich*. It could be that this subject overwrote the previous topic so that no longer the speaker *or* Anna was the topic but both (see, e.g., Krifka 2007: 42–43 for the possibility of having more than one topic in a sentence). This would cancel out the topicality manipulation because a part of the topic would be omitted in the target sentence in both Topicality conditions. To circumvent this problem, I conducted a further experiment, namely experiment 6.

9.2.2 Experiment 6: topicality (complex question method)

Experiment 6 was an acceptability rating study that again focused on topicality, leaving aside grammatical person,[5] and, thus, had the form of a 2 × 2 design (Topicality (topic vs. no topic) × Completeness (full form vs. topic drop)).[6] Like in experiment 5, I set the (discourse) topic with a question, but this time I used two questions instead of one: the first to set the topic and the second to reinforce it after having introduced a competing referent. This should reverse any overwriting of the topic as may have occurred in experiment 5. Even if a context sentence with a different subject (e.g., a competing referent or a coordination of competing referent and current topic) were to overwrite the discourse topic set by the first question, the second question would reestablish the original discourse

[5] I restricted myself to topic drop of the 3rd person singular here because this allowed me to have two referents, one topical, the other non-topical, with the same grammatical person.

[6] All items, fillers, and the analysis script can be accessed online: https://osf.io/zh7tr.

topic. Like in experiment 5, the prediction is that if topicality is a favoring factor for topic drop, the conditions with topic drop should be preferred relative to the full forms if the omitted constituent is coreferential with the current discourse topic. Experiment 6 was part of the same study as experiment 4.

9.2.2.1 Materials

Items

I constructed 16 token sets, such as (2), which were always short dialogues between two interlocutors A and B talking about two other persons of different gender, e.g., Jennifer and Marcel in (2). These dialogues took the form of two consecutive question-answer pairs:

(2) a. (i) *A: Was gibt's Neues von Jennifer?*
what gives.it new from Jennifer
A: 'What's new from Jennifer?'
(ii) *B: Marcel ist inzwischen wieder Single.*
Marcel is meanwhile again single
B: 'Marcel is now single again.'
(iii) *A: Und was hat das mit Jennifer zu tun?*
and what has that with Jennifer to do
A: 'And what does that have to do with Jennifer?'
(iv) *B: (Sie) hat ihn nach einem Date gefragt.*
she has him for a date asked
B: '(She) asked him out on a date.'
(topic drop / full form, topic)

b. (i) *A: Was gibt's Neues von Marcel?*
what gives.it new from Marcel
A: 'What's new from Marcel?'
(ii) *B: Jennifer ist inzwischen wieder Single.*
Jennifer is meanwhile again single
B: 'Jennifer is now single again.'
(iii) *A: Und was hat das mit Marcel zu tun?*
and what has that with Marcel to do
A: 'And what does that have to do with Marcel?'
(iv) *B: (Sie) hat ihn nach einem Date gefragt.*
she has him for a date asked
B: '(She) asked him out on a date.'
(topic drop / full form, no topic)

First, speaker A asked for information about either Jennifer in (2a-i) or Marcel in (2b-i). Similar to experiment 5, this should set one of them as the (discourse) topic. Then, speaker B answered only indirectly by giving information about the other person, i.e., Marcel in (2a-ii) or Jennifer in (2b-ii). This way, Marcel or Jennifer were introduced into the discourse, presumably without changing the (discourse) topic. In the third turn, A asked for the connection of B's answer to Jennifer in (2a-iii) or Marcel in (2b-iii) and, in this way, reinforced Jennifer or Marcel as the discourse topic. Finally, B's answer, which established this connection, was the target utterance, i.e., (2a-iv) and (2b-iv). The omitted or realized feminine subject pronoun *sie* ('she') was always placed in the prefield position, whereas the masculine referent was referred to by an object pronoun in the accusative or the dative case *ihn*/*ihm* ('him').

This means that in the conditions where A set Jennifer as the discourse topic, the prefield constituent and the discourse topic were coreferential, whereas in the conditions where Marcel was set as the discourse topic they were not (predictor Topicality). The target utterances were always in the perfect tense with the 3rd person singular form of the auxiliary *haben* ('have') in the left bracket. Similar to the 3rd person conditions in experiment 5, the target utterances in the topic drop conditions were locally ambiguous until the participants encountered the gender-marked object pronoun, which allowed them to disambiguate between the two persons of different gender. Since the structure of the stimuli with four turns including two questions was quite conspicuous and not as natural as the items used in my previous studies, I constructed and tested only 16 token sets to prevent a stronger habituation effect.

Fillers

The 80 fillers consisted of the 24 items of experiment 4 on topic drop after conjunctions and, as described in detail in Section 3.4.2.1, of 24 items on preposition omission, 24 (potential) gapping structures, and eight ungrammatical catch trials.

9.2.2.2 Procedure

The procedure has been described in detail in Section 3.4.2.2.

9.2.2.3 Results

As described already in Section 3.4.2.3, I excluded the data from 14 inattentive participants who had rated four or more of the eight catch trials with 6 or 7. Table 9.3 shows the mean ratings and standard deviations per condition for the

remaining 58 participants. In Figure 9.2 the mean ratings and 95% confidence intervals are plotted.

Table 9.3: Mean ratings and standard deviations per condition for experiment 6

Completeness	Topicality	Mean rating	Standard deviation
Full form	Topic	4.29	1.89
Topic drop	Topic	4.06	1.80
Full form	No topic	4.21	1.81
Topic drop	No topic	3.99	1.79

From visual inspection, it does not seem that topic drop of a non-topic was particularly degraded compared to the full forms. The mean ratings were generally lower than the mean ratings for the items with topic drop after conjunctions (all above 5.3), which may indicate that the topicality items were indeed perceived as marked and less natural because of their unusual structure. Still, they were rated considerably better than the ungrammatical catch trials (mean = 2.95, SD = 2.16).

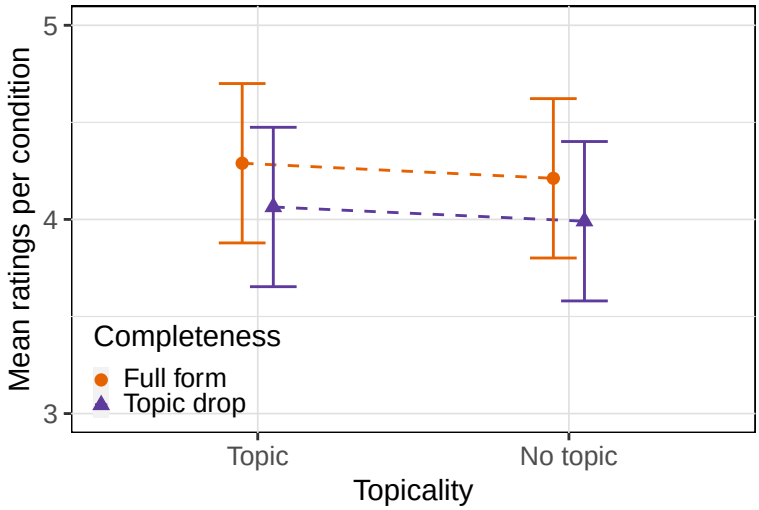


Figure 9.2: Mean ratings and 95% confidence intervals per condition for experiment 6

I analyzed the data again with CLMMs from the package ordinal (Christensen 2019) in R, as described in Section 3.1.6.3. The full model contained the ordinal ratings as the dependent variable and as independent variables the two binary predictors Completeness and Topicality, which were again coded using deviation coding (full form and topic as 0.5, the other levels as −0.5), as well as the numeric centered and scaled Position of the trial in the experiment. As random effects, I included random intercepts for subjects and items, as well as by-subject and by-item random slopes for Completeness, Topicality, and their interactions.[7] Table 9.4 shows the fixed effect in the final model, which had symmetric2 thresholds.

Table 9.4: Fixed effect in the final CLMM of experiment 6

Fixed effect	Est.	SE	χ^2	p-value	
Position	−0.38	0.07	27.78	< 0.001	***

There was only a theoretically uninteresting significant main effect of Position ($\chi^2(1) = 27.78, p < 0.001$), indicating that the ratings for the items generally decreased over the course of the experiment. The main effect of Completeness was marginally significant ($\chi^2(1) = 3.49, p < 0.09$), indicating a tendency for the full forms to be preferred over the utterances with topic drop. The main effect of Topicality was not significant ($\chi^2(1) = 1.04, p > 0.3$), nor were its interaction with Completeness ($\chi^2(1) = 0.0008, p > 0.9$) or any of the other effects.

9.2.2.4 Discussion

Experiment 6 was designed to investigate again the impact of topicality on the acceptability of topic drop by employing the concept of a discourse topic. By means of two questions, a referent was unambiguously set as the discourse topic, thereby circumventing the possible issue of overwriting the topic of experiment 5, discussed above. Similar to experiment 5, the results of experiment 6 provided no evidence that the omission of a constituent that is coreferential with the current discourse topic is more acceptable. Therefore, there continues to be no support for topicality being a favoring factor for topic drop. In the next two sections,

[7] Models with random slopes for Position did not converge. Since this variable is of least theoretical interest, it seems to be relatively unproblematic for the analysis not to include the corresponding random slopes. The formula of the full model was as follows: `Ratings ~ (Completeness + Topicality + Position)^2 + (1 + Completeness * Topicality | Subjects) + (1 + Completeness * Topicality | Items)`.

I describe two further experiments that employed a different method of setting the topic.

9.2.3 Experiment 7: topicality (subject method, lexical verbs)

Experiment 7 was similar to experiment 5 in that it investigated whether topicality and grammatical person influence the acceptability of topic drop in the form of a 2 × 2 × 2 acceptability rating study crossing Completeness (full form vs. topic drop), grammatical Person (1SG vs. 3SG), and Topicality (topic vs. no topic).[8] However, while the previous two studies used questions to set the (discourse) topic in the context preceding the critical utterance, in this experiment and experiment 8, the topic was set via the subject function, as described below. The prediction is identical to those of experiments 5 and 6. If topicality is a favoring factor for topic drop, topic drop should be more acceptable when the omitted constituent is the topic. For grammatical person, this experiment sought to replicate the effect found in experiment 5, according to which topic drop of the 1st person singular is more acceptable than topic drop of the 3rd person singular (see Section 10.5.2 for details).

9.2.3.1 Background

In Section 3.1.1, I already mentioned that subjects can be considered unmarked topics (Reinhart 1981: 62) because there is a strong tendency across languages for the topic and subject of a sentence to coincide (Lambrecht 1994: 132). Both Reinhart (1981) and Lambrecht (1994) stress the fact that the concurrence of topicality and subjecthood is only a tendency, i.e., that there are exceptions where subject and topic are distinct. Nevertheless, it seems promising to exploit this tendency to set the topic experimentally.

A further tendency for topics is to build so-called *topic chains* in discourse, i.e., having the same topic over larger parts of a discourse (Givón 1983). In combination, both tendencies predict (tentatively of course) that if a constituent is the subject of a sentence, such as Sabrina in (3), it is also the unmarked topic and it should be more likely that a constituent referring to the same topic is also the topic expression in the following sentence, in particular, if this constituent is also the subject of this sentence, as in (4a) vs. (4b).

[8] This experiment and experiment 8 were part of my bachelor thesis (Schäfer 2019) and were published in Schäfer (2021). In this book, I revised their analyses and newly included a joint analysis of both experiments. All materials and the analysis script can be found online: https://osf.io/zh7tr.

(3) *Vor Kurzem ist Sabrina wieder bei mir eingezogen.*
recently is Sabrina again at me in.moved
'Recently Sabrina moved back in with me.'

(4) a. *Sie gibt mir noch eine Chance.*
she gives me yet a chance
'She gives me another chance.'
b. *Ich gebe ihr noch eine Chance.*
I give her yet a chance
'I give her another chance.'

This prediction can be partially described through the so-called *centering theory* (Grosz et al. 1995, Walker et al. 1998), a framework originally proposed to determine the reference of anaphora. The basic idea of centering theory is that each utterance in a discourse segment except for the first one exhibits a so-called *backward-looking center* C_b. According to Walker et al. (1998: 3), this backward-looking center C_b corresponds to the topic of that utterance. The C_b of an utterance U_n gets selected from the set of all referring expressions of the previous utterance U_{n-1}, which are called *forward-looking centers* C_f. It is assumed that these C_f have a certain order, which for English and German is usually argued to be determined by their syntactic function (see Walker et al. 1998 for English, Speyer 2007 for German, and Walker et al. 1998 for a different assumed hierarchy in Japanese), i.e., subjects > objects > adverbials. It is the highest ranked element of the *forward-looking centers* C_f of the preceding utterance U_{n-1} that occurs in the current utterance U_n that is selected as *backward-looking center* C_b, i.e., as topic, of this utterance U_n.

According to this mechanism, the C_b of (4a) is the expression *sie*, whereas in (4b) it is *ihr*, both referring to Sabrina. As the subject of the preceding utterance U_{n-1} (3), *Sabrina* is the highest-ranked element of the forward-looking centers of this utterance $C_f = \{Sabrina > bei\ mir\}$ that occurs in the target utterances (4a) and (4b). From this, it follows that in (4a), the C_b occurs in the prefield, whereas in (4b) it occurs in the middle field. Put differently, in (4a), the prefield constituent is identical to the topic of the utterance and coreferential with the highest-ranked C_f-element of the previous utterance, whereas in (4b), the prefield constituent is distinct from both.

Note that, under a different analysis than the one proposed by centering theory, one might assume that in (4b) the prefield constituent *ich* is still the topic expression of the utterance, but it refers to a topic that is distinct from the topic of the previous utterance. That means that one has to assume a shift in topics between both sentences, which ends a topic chain. Under this analysis, topic drop

both of (4a) and of (4b) would indeed target a topic but in one case a "better" topic, which is part of a topic chain.

While I adopt the interpretation derived from centering theory, according to which the prefield constituent in (4b) is not a topic expression, both interpretations predict that topic drop of (4a) is more acceptable than topic drop of (4b), provided that one assumes that topicality affects the acceptability of topic drop.

It is important to reiterate at this point that topic continuity and subject continuity coincide in this manipulation of topicality. It has been repeatedly argued that hearers have a preference for interpreting pronouns as coreferential with the subject of the previous utterance (e.g., Crawley et al. 1990, Järvikivi et al. 2005, Kaiser 2009), in particular in German (e.g., Hemforth et al. 2010, Colonna et al. 2012). Therefore, hearers may also prefer elliptical utterances if the omitted constituent is coreferential with the subject of the preceding utterance, as proposed, for example, by Auer (1993: 200). With this experiment and experiment 8, I cannot separate a potential effect of topic continuity from one of subject continuity. From the perspective of the information-theoretic approach, however, such a distinction is not essential because both types of continuity should make the prefield constituent more likely and its omission thus more acceptable.

9.2.3.2 Materials

Items

I constructed 24 items such as (5) as short dialogues between two persons.

(5) a. (i) *A: Hallo, wie sieht's aus?*
hello how looks.it VPART
A: 'Hello, what's the situation?'

(ii) *B: Vor Kurzem bin ich wieder bei Thomas eingezogen.*
recently am I again at Thomas in.moved
B: 'Recently I moved back in with Thomas.'

(iii) *B: (Ich) gebe ihm noch eine Chance.*
I give him yet a chance
B: '(I) give him another chance.'
(topic drop / full form, 1SG, topic)

b. (i) *A: Hallo, wie sieht's aus?*
hello how looks.it VPART
A: 'Hello, what's the situation?'

(ii) *B: Vor Kurzem ist Claudia wieder bei mir eingezogen.*
recently is Claudia again at me in.moved
B: 'Recently Claudia moved back in with me.'

(iii) *B: (Ich) gebe ihr noch eine Chance.*
I give her yet a chance
B: '(I) give her another chance.'
(topic drop / full form, 1SG, no topic)

c. (i) *A: Hallo, wie sieht's aus?*
hello how looks.it VPART
A: 'Hello, what's the situation?'
(ii) *B: Vor Kurzem ist Sabrina wieder bei mir eingezogen.*
recently is Sabrina again at me in.moved
B: 'Recently Sabrina moved back in with me.'
(iii) *B: (Sie) gibt mir noch eine Chance.*
she gives me yet a chance
B: '(She) gives me another chance.'
(topic drop / full form, 3SG, topic)

d. (i) *A: Hallo, wie sieht's aus?*
hello how looks.it VPART
A: 'Hello, what's the situation?'
(ii) *B: Vor Kurzem bin ich wieder bei Patrick eingezogen.*
recently am I again at Patrick in.moved
B: 'Recently I moved back in with Patrick.'
(iii) *B: (Er) gibt mir noch eine Chance.*
he gives me yet a chance
B: '(He) gives me another chance.'
(topic drop / full form, 3SG, no topic)

The conversations had the following pattern: Person A asked a non-specific question, e.g., (5a-i), which, unlike in experiments 5 and 6, did not introduce any discourse referents. Person B answered with two utterances. In the first utterance, the speaker, person B, mentioned themselves and a 3rd person in the form of a proper name such as *Thomas*.[9] More specifically, person B uttered that they and this 3rd person participate in an action such as moving in. It was varied whether the speaker, as in (5a-ii) and (5d-ii), or the 3rd person, as in (5b-ii) and (5c-ii), is the subject of the utterance. The other person was mentioned in a syntactically subordinate role as a prepositional object or adverbial and should therefore be less accessible as a topic (or as a backward-looking center) in the following sentence. In this first utterance of person B, the subject always appeared in the middle field, while the prefield was filled with a temporal adverbial. This way,

[9]Note that in this experiment and in experiment 8, the proper names of this 3rd person were varied within a token set. Of course, there was always a match between the gender of the proper name in the context sentence and the gender of the personal pronoun in the target sentence.

structural parallelism effects similar to those found to impact pronoun resolution, i.e., that pronouns tend to be interpreted as referring to an element in the same syntactic position (Smyth 1994, Chambers & Smyth 1998), should be alleviated.[10] The last utterance picked up both referents, i.e., the speaker and the 3rd person. It was varied which of them appeared as the (realized or omitted) subject in the prefield and which, consequently, as the object in the middle field. If the prefield constituent of the target utterance was identical to the subject of the previous utterance, as in (5a) and (5c), it is a *topic* condition. If both were distinct, as in (5b) and (5d), it is a *no topic* condition. In this experiment, the verb in the left bracket was always a lexical verb in the indicative present with a distinct inflectional marking.

Fillers

The 24 items described above were mixed with 80 fillers, which were basically identical to those used in experiment 5 including seven ungrammatical attention checks. The only difference was that I reduced the number of fillers containing (potential) instances of gapping and right node raising from 28 to 24 for simplification.

9.2.3.3 Procedure

I recruited 48 self-reported native speakers of German between the ages 18 and 50 from Clickworker (Clickworker 2022), who had not participated in any of my other experiments on topic drop. They received €2.50 to participate in the experiment, which I implemented again with LimeSurvey (Limesurvey GmbH 2023). The task of the participants was to rate the naturalness of the last italicized utterance on a 7-point Likert scale (7 = completely natural). I used a Latin square design to distribute the materials among eight lists and showed them to the participants along with the fillers in an individual pseudo-randomized order, ensuring that no two items or fillers of the same type immediately followed each other. Each participant saw each of the 24 critical items once and in only one condition. The materials were again presented in lowercase and as instant messaging dialogues (see Section 3.1.6.2).

9.2.3.4 Results

I excluded the data from five participants who had rated four or more of the seven ungrammatical attention checks with 6 or 7 points on the 7-point scale, i.e., as

[10] Note that this strategy mainly prevents effects of the surface structure, i.e., between two constituents in the prefield. It cannot rule out the possibility that there is an effect of subject continuity, as discussed above.

completely natural or almost completely natural, which suggests that they had been inattentive during the task. The descriptive statistics for the rating data, i.e., the mean ratings and standard deviations per condition, are shown in Table 9.5.

Table 9.5: Mean ratings and standard deviations per condition for experiment 7

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.64	1.44
Topic drop	1SG	Topic	5.38	1.40
Full form	3SG	Topic	5.58	1.43
Topic drop	3SG	Topic	4.65	1.60
Full form	1SG	No topic	5.58	1.37
Topic drop	1SG	No topic	4.91	1.46
Full form	3SG	No topic	5.67	1.45
Topic drop	3SG	No topic	4.21	1.67

Figure 9.3 illustrates the mean ratings also graphically, including 95% confidence intervals. They suggest that topic drop of a non-topical constituent was degraded compared to topic drop of a topical constituent.

I analyzed the data using CLMMs from the package ordinal (Christensen 2019) in R, following the procedure described in Section 3.1.6.3 and similar to experiment 5. The participants' responses were modeled as a function of the three binary predictors Completeness, grammatical Person, and Topicality, coded using deviation coding (full form, 1SG and topic coded as 0.5, topic drop, 3SG and no topic coded as −0.5), as well as the numeric scaled and centered Position at which the trial appeared in the experiment. Additionally, I included the three-way interaction between Completeness, Person, and Topicality, as well as all two-way interactions between the four independent variables. The random effects structure consisted of random intercepts for participants and items and by-item and by-subject random slopes for all independent variables, and the two-way interactions between Completeness, Person, and Topicality.[11] The fixed effects in the final CLMM are shown in Table 9.6.

[11]The formula of the full model was as follows: `Ratings ~ Completeness : Person :Topicality + (Completeness + Person + Topicality + Position)^2 + (1 + (Completeness + Person + Topicality)^2 + Position | Subjects) + (1 + (Completeness + Person + Topicality)^2 + Position | Items)`.

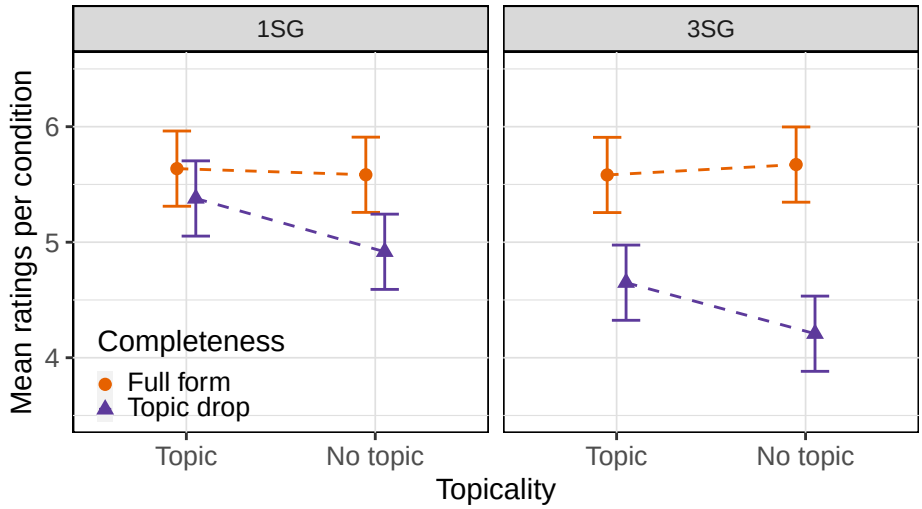


Figure 9.3: Mean ratings and 95% confidence intervals per condition for experiment 7

Table 9.6: Fixed effects in the final CLMM of experiment 7

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.72	0.35	20.75	< 0.001	***
Person	0.52	0.20	6.10	< 0.05	*
Topicality	0.38	0.15	6.69	< 0.01	**
Completeness × Person	−1.33	0.35	12.85	< 0.001	***
Completeness × Topicality	−0.81	0.33	5.95	< 0.05	*

The final model contained a significant interaction between Completeness and Topicality ($\chi^2(1) = 5.95, p < 0.05$), indicating that utterances with topic drop were rated better if the omitted constituent was the topic expression. The significant interaction between Completeness and Person ($\chi^2(1) = 12.85, p < 0.001$) suggests that topic drop of the 1st person singular was more acceptable than topic drop of the 3rd person singular. Additionally, there were significant main effects of all three binary predictors Completeness ($\chi^2(1) = 20.75, p < 0.001$), Person ($\chi^2 = 6.1, p < 0.05$), and Topicality ($\chi^2(1) = 6.69, p < 0.01$). Utterances were more acceptable if they were syntactically complete, if the preverbal constituent was an overt or covert 1st person singular pronoun, and if it was the topic.

9.2.3.5 Discussion

Experiment 7 was designed to investigate the impact of topicality and grammatical person on the acceptability of topic drop in German. In this respect, it was similar to experiment 5, and, concerning topicality, also to experiment 6. However, the topic was set not using a question method as in these two experiments but through the subject function, exploiting the frequent concurrence of the topic and the subject.

Unlike the previous experiments, the results of experiment 7 support an impact of topicality on topic drop. While utterances with a topical prefield constituent were generally preferred, this effect was particularly strong for topic drop. It seems that omitting a topic expression is more acceptable than omitting a non-topic expression if the topic was set using the subject function. As pointed out in Section 9.1, the information-theoretic account can explain this pattern. Given the general tendency to build topic chains in discourse, the topic of an utterance can become predictable from the previous utterance and predictable constituents can better be omitted than unpredictable ones.

However, the conclusion that topicality is a favoring factor for topic drop has to be drawn cautiously. On the one hand, the results of the following study question a general influence (see Section 9.2.4). On the other hand, the topicality manipulation coincides with subject constancy in this experiment and the next one. That is, in the conditions in which the (realized or omitted) subject was the topic of the target utterance, it was at the same time identical to the subject of the previous utterance. As mentioned above, I tried to prevent a positional parallelism effect by placing the subject of the context utterance not in the prefield but in the middle field. But possible parallelism effects due to the same syntactic function remained unaffected by this solution and could simply not be prevented here due to the kind of topic manipulation. As mentioned above, my information-theoretic approach would also be able to explain the effect of subject continuity since subject continuity is also expected to increase the likelihood of the prefield constituent. I revisit this discussion in Section 9.2.4.4.

9.2.4 Experiment 8: topicality (subject method, modal verbs)

Experiment 8 was a replication of experiment 7 with modal verbs instead of lexical verbs.[12] Since the forms of the modal verbs for the 1st and the 3rd person singular are syncretic (e.g., *ich muss* ('I must') and *Tino muss* ('Tino must')), in this experiment, unlike in experiment 7, there was no distinct inflectional ending on the verb that could serve as a cue to facilitate recovering the omitted preverbal

[12] All materials and the analysis scripts are available online: https://osf.io/zh7tr.

subject. This omitted subject could, however, clearly be recovered based on the object pronoun referring to the competing referent that was also present in the utterances. In other words, the corresponding utterances were only locally ambiguous. While this experiment together with experiment 7 with the distinctly marked lexical verbs examined whether the factor of verbal inflection impacts the acceptability of topic drop (see Section 10.5.3), I also again investigated Topicality. The intention was simply to replicate the result of experiment 7 according to which topic drop of a topical prefield constituent is preferred over topic drop of a non-topical one if the topic was set via the subject function. This resulted again in a 2 × 2 × 2 design, which crossed Completeness, grammatical Person, and Topicality.

In this section, in addition to analyzing experiment 8 according to the established scheme, I also present a joint post hoc analysis of experiments 7 and 8, in which I considered Verb Type (full verb vs. modal verb) as a variable manipulated between-subjects. This had the purpose of investigating whether the distinct inflectional marking has an impact on the acceptability of topic drop, in combination with grammatical person or as a factor on its own (see Section 10.5.3).

9.2.4.1 Materials

Items

The items were identical to the 24 items of experiment 7, except that the left bracket was filled with a modal verb instead of a lexical verb. The infinitive lexical verb was consequently moved to the right bracket. The modal verbs were varied between items so that the target utterance was still as natural as possible. Eight utterances contained *wollen* ('want'), five utterances *dürfen* ('may'), four utterances *sollen* ('shall'), three utterances *können* ('can'), three utterances *mögen* ('would like'), and one utterance *müssen* ('must'). The comparison between (6a) and (6b) shows that both verb forms were identical. When encountering the verb, the participants could not use the inflectional ending as a cue to recover the omitted subject. However, the items were only locally ambiguous since the object pronoun referring to one referent always clarified that the other referent had to be the omitted subject.[13]

[13]The acceptability rating study as an offline task is not suitable to observe the disambiguation process itself. To accomplish this, an on-line method like self-paced reading or eye tracking would be necessary. The effort associated with disambiguating could be quantified by measuring reading times on the object pronoun and a spillover region. If speakers have an a priori preference for a 1st person singular pronoun, they should slow down when encountering a 1st person singular object pronoun that rules out this reading compared to a 3rd person singular object pronoun.

(6) a. *(Ich) will ihr noch eine Chance geben.*
I want her yet a chance give
'(I) want to give her another chance.' (topic drop / full form, 1SG)
b. *(Sie) will mir noch eine Chance geben.*
she wants me yet a chance give
'(She) wants to give me another chance.' (topic drop / full form, 3SG)

Fillers

The items were mixed with the same 80 fillers as in experiment 7.

9.2.4.2 Procedure

48 self-reported native speakers of German between 18 and 40 years participated in the study for a reward of €2.50. They had not taken part in any previous experiment on topic drop and were recruited from Clickworker (Clickworker 2022). The study was implemented with LimeSurvey (Limesurvey GmbH 2023) and conducted online. Just like in the previous experiments, the participants rated the naturalness of the target utterances on a 7-point Likert scale (7 = completely natural). The 24 adapted items were distributed across eight lists according to a Latin square design and presented as instant messaging dialogues together with the fillers in individually pseudo-randomized order.

9.2.4.3 Results

9.2.4.3.1 Analysis of experiment 8

Since all participants passed the attention checks, the complete data was used for the analysis. Table 9.7 shows the mean ratings and standard deviations per condition.

In Figure 9.4, the mean ratings are plotted including 95% confidence intervals. While omitting a non-topic seems to be degraded, also the full forms with a non-topical prefield constituent were worse than the conditions with a topical prefield constituent.

Again, I used CLMMs from the package ordinal (Christensen 2019) for the inferential statistical analysis (see Section 3.1.6.3 for details on the procedure). The full model was identical to the one described for experiment 7. All predictors were coded as described in Section 9.2.3.4.[14] The fixed effects in the final model

Table 9.7: Mean ratings and standard deviations per condition for experiment 8

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.41	1.55
Topic drop	1SG	Topic	5.24	1.67
Full form	3SG	Topic	5.52	1.43
Topic drop	3SG	Topic	4.43	1.93
Full form	1SG	No topic	5.35	1.67
Topic drop	1SG	No topic	4.86	1.77
Full form	3SG	No topic	5.25	1.74
Topic drop	3SG	No topic	4.27	1.77

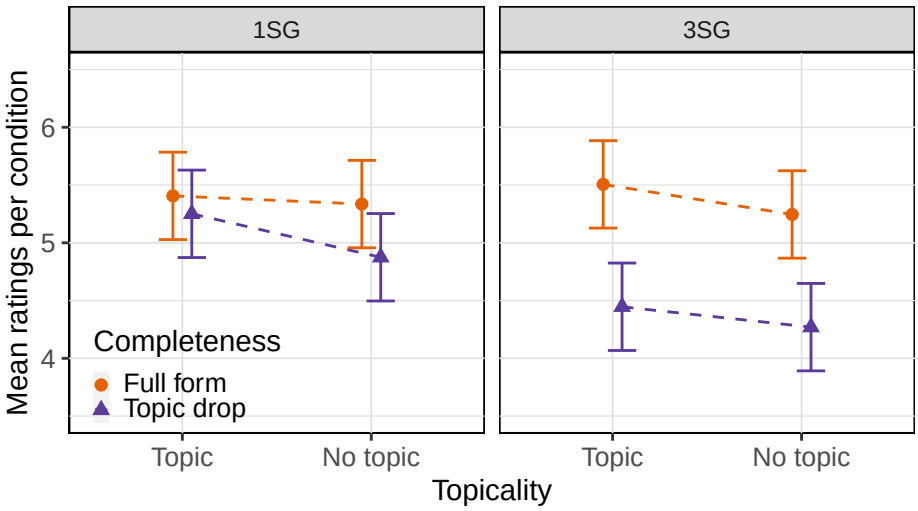


Figure 9.4: Mean ratings and 95% confidence intervals per condition for experiment 8

Table 9.8: Fixed effects in the final CLMM of experiment 8

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.08	0.23	18.07	< 0.001	***
Person	0.61	0.17	10.67	< 0.01	**
Completeness × Person	−0.98	0.29	9.43	< 0.01	**

with flexible thresholds are shown in Table 9.8.
There were significant main effects of Completeness ($\chi^2(1) = 18.07$, $p < 0.001$) and Person ($\chi^2(1) = 10.67$, $p < 0.01$) and a significant interaction between both predictors ($\chi^2(1) = 9.43$, $p < 0.01$). Utterances with topic drop were degraded compared to complete utterances, utterances with the 3rd person singular were degraded compared to utterances with the 1st person singular, and utterances with topic drop of a 3rd person singular prefield constituent were particularly degraded. Neither the main effect of Topicality was significant ($\chi^2(1) = 1.18$, $p > 0.2$), nor its interaction with Completeness ($\chi^2(1) = 1.64$, $p > 0.1$), nor any other effect.

9.2.4.3.2 Analysis of experiments 7 and 8

In a post hoc analysis, I combined the data from experiments 7 and 8 to look for a potential influence of the verb in the left bracket and its inflectional ending on topic drop. I added Verb Type as a new predictor that distinguishes between the lexical verbs with a distinct inflectional ending used in experiment 7 and the modal verbs with syncretic forms used in experiment 8. I coded this new predictor using deviation coding: 0.5 for the lexical verbs and −0.5 for the modal verbs. I analyzed the combined data again with CLMMs (Christensen 2019). Since the Position of the trial in the experiment was not involved in any significant effect in the analyses of experiments 7 and 8, I did not include it as a predictor in this post hoc analysis, which allowed me to use a simpler model. I modeled the ordinal ratings as a function of the deviation-coded binary predictors Completeness, Person, Topicality, and Verb Type, including all three- and two-way interactions between them. The random effects structure consisted of random intercepts for subjects and items and by-subject random slopes for Completeness, Person,

[14]The formula of the full model was as follows: `Ratings ~ Completeness : Person : Topicality + (Completeness + Person + Topicality + Position)^2 + (1 + (Completeness + Person + Topicality)^2 + Position | Subjects) + (1 + (Completeness + Person + Topicality)^2 + Position | Items)`.

Topicality, and their two-way interactions and by-item random slopes for Completeness, Person, Topicality, Verb Type, and their two-way interactions.[15]

Table 9.9 shows the fixed effects in the final model, which had flexible thresholds and was obtained with a backward model selection.

Table 9.9: Fixed effects in the final CLMM of the joint analysis of experiments 7 and 8

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.10	0.18	30.41	< 0.001	***
Person	0.38	0.12	8.52	< 0.01	**
Topicality	0.24	0.10	5.73	< 0.05	*
Completeness × Person	−1.02	0.19	22.24	< 0.001	***
Completeness × Topicality	−0.53	0.20	6.98	< 0.01	**

The predictor Verb Type was not involved in any significant effect. In particular, neither the interaction with Completeness ($\chi^2(1) = 0.3, p > 0.5$), nor the interaction with Completeness and Person ($\chi^2(1) = 0.005$, $p > 0.9$) were significant. Topic drop before a modal verb was not significantly more or less acceptable than topic drop before a lexical verb, nor did topic drop of the 1st person singular have an advantage before a distinctly marked lexical verb in comparison to a syncretic form of a modal verb. The significant effects in the final model replicated those found in the analysis of experiment 7 rather than those in the analysis of experiment 8. There were significant main effects of Completeness ($\chi^2(1) = 30.41, p < 0.001$), Person ($\chi^2(1) = 8.52, p < 0.01$), and Topicality ($\chi^2(1) = 5.73, p < 0.05$). Furthermore, Completeness interacted significantly with Person ($\chi^2(1) = 22.24, p < 0.001$) and with Topicality ($\chi^2(1) = 6.98, p < 0.01$). Full forms were preferred over topic drop, utterances with the 1st person singular were preferred over utterances with the 3rd person singular, and utterances with a topical prefield constituent were preferred over utterances with a non-topical prefield constituent. Topic drop of the 3rd person singular and topic drop of a non-topical prefield constituent were particularly degraded.

[15]The formula of the full model was as follows: `Ratings ~ (Completeness + Person + Topicality + Verb Type)^3 + (1 + (Completeness + Person + Topicality)^2 | Subjects) + (1 + (Completeness + Person + Topicality + Verb Type)^2 | Items)`.

9.2.4.4 Discussion

Concerning topicality, experiment 8 had the purpose of replicating the effect of topicality found in experiment 7. With a joint post hoc analysis of experiments 7 and 8, I furthermore investigated whether there is an impact of verbal inflection on topic drop and a potential interaction with topicality.

For topicality, a complex picture emerges from the analysis of experiment 8 and the joint post hoc analysis of experiments 7 and 8 (see Section 10.5.3.4 for a discussion with respect to grammatical person). While the interaction between Completeness and Topicality was still significant in the joint analysis of both experiments, experiment 8 alone failed to replicate this topicality pattern. Neither were utterances with a topical prefield constituent generally rated as more acceptable, nor was topic drop of a topical constituent particularly preferred. This suggests that the effect in the joint analysis was mainly driven by the ratings for the utterances with lexical verbs, but it also suggests that the ratings for the items with the modal verbs are not diametrically opposed to this. Given that the only factor that was modified between experiment 7 and experiment 8 was the type of the verb in the left bracket, any significant difference between the two experiments must be attributed to this modification. That is, the lack of an interaction between Topicality and Completeness must in fact be related to the syncretic forms of the modal verbs. The result can be interpreted as showing that topic drop of a topic constituent is more acceptable only if, at the same time, the reconstruction of this constituent is facilitated by a distinct inflectional marking on the following verb. That is, topicality and inflectional marking seem to impact the acceptability of topic drop only in combination but not as factors on their own.

9.3 Summary: topicality

In the first part of this book, I argued both theoretically and empirically that topicality is neither a (strictly) sufficient nor a necessary condition for topic drop. In this chapter, I investigated whether topicality favors topic drop, i.e., whether topic drop is more acceptable if it targets a topic. Such an effect of topicality could be explained by the information-theoretic *avoid troughs* principle (see Section 9.1). Given the fact that the topic is what the sentence is about and given the known tendency to have topic chains in discourse, i.e., to keep the topic constant across multiple utterances, a topic should generally be more predictable than a non-topic. Therefore, topic drop should be more acceptable if it targets a more predictable topic than a less predictable non-topic.

In experiments 5 and 6, I attempted to use questions to set the discourse topic and did not find any effect of topicality on topic drop. In experiments 7 and 8, I exploited the fact that subjects are often unmarked topics and used the subject function to set the (sentence) topic. This had the consequence that topicality and subjecthood were intermingled so that I cannot finally decide which of them is the crucial factor. In a future study, it may thus be beneficial to try to disentangle topicality and subjecthood without resorting to using questions to set the topic.[16] From the *UID* perspective, however, what makes the antecedent constituent more probable is ultimately irrelevant, be it its topic or its subject status. The only decisive factor is that it is more predictable thus has a lower surprisal, and should be more likely omitted.

By using the subject method, I found an effect of topicality in the expected direction but only in experiment 7 and in the joint analysis of experiments 7 and 8 but not in experiment 8 in isolation. This suggests that either the effect of topicality (or potentially of subjecthood) is not very robust, and/or that distinct verbal inflection (which was present in experiment 7 but not in experiment 8) and topicality work together. A subject can be better omitted if it is predictable as the topic and if the following verb provides a cue to recover it. From an information-theoretic perspective, if several factors come together to facilitate the resolution of the ellipsis, this could lower the processing effort for the verb in the left bracket to the point where it falls below the channel capacity, whereas the influence of just one factor alone may not suffice. Taken together, there is only sparse evidence of an impact of topicality on the usage of topic drop. This impact could nevertheless be well captured by the *avoid troughs* principle of my information-theoretic account of topic drop usage.

In the next chapter, I turn to the influencing factor that I investigate in most depth in this book, grammatical person, and to verbal inflection and ambiguity avoidance, which are closely related to grammatical person.

[16] Possibly, syntactic means of topic marking such as left dislocation and cleft clauses (see, e.g., Musan 2017: 33–34), which can also be used to mark syntactic objects as topics, would be suitable for this purpose.

10 Grammatical person, verbal inflection, and ambiguity avoidance

As I already mentioned in Chapter 8 on syntactic function, the grammatical person of the omitted constituent is discussed as a further important factor for the usage of topic drop. Since object topic drop is mainly restricted to the 3rd person singular according to the theoretical literature, I focus on the impact of grammatical person and number on the omission of subjects in this chapter and only briefly mention its role for objects. First, I provide a theoretical overview of the literature on grammatical person, and, since the three factors are argued to be closely related, also on the inflectional ending on the verb after topic drop and ambiguity avoidance.[1] After presenting previous empirical results with respect to these three factors and the information-theoretic predictions, I turn to my investigations. Grammatical person is the factor that I investigated in most detail, both corpus-linguistically and experimentally. I first present the corpus results and then turn to the results of five acceptability rating studies. In the corpus study and two of the experiments, I also looked at verbal inflection and ambiguity avoidance.

10.1 Theoretical overview

10.1.1 Theoretical overview of grammatical person and verbal inflection

Volodina (2011: 272) states that subject pronouns of all persons and numbers can be targeted by topic drop. However, according to the previous literature, the 1st person singular subject pronoun *ich* ('I') is especially often omitted (Tesak & Dittmann 1991, Auer 1993, Volodina & Onea 2012, Imo 2013, 2014). This prevalence of 1st person subject topic drop is explained by two different types of hypotheses, which I term *inflectional hypothesis* and *extralinguistic hypotheses*.

[1] Section 10.1.1 was published in a similar form in Schäfer (2021).

The *inflectional hypothesis* was first proposed by Auer (1993: 198)[2] and connects the grammatical person of the omitted constituent and the inflection of the following verb. His claim is that topic drop of the 1st person singular is easily possible because the verbal morphology in the present tense singular in German is still so differentiated that the grammatical person can be expressed only by inflection.[3] So if a (1st person) singular subject is omitted, the inflectional marking on the verb should allow a hearer to reconstruct at least its grammatical person. However, this reasoning applies only to the present tense and not to preterite present verbs.[4] In the preterite and for preterite present verbs even in present tense, there is a syncretism between the forms of the 1st and the 3rd person singular (e.g., *ich war* – 'I was' vs. *sie war* – 'she was' and *ich kann* – 'I can' vs. *er kann* – 'he can'). Consequently, not all 1st person singular verb forms are distinctly marked for inflection, so recoverability through inflectional marking alone cannot explain the frequent omission of subject pronouns of this person. Furthermore, Sigurðsson (2011: 279) notes that in Swedish, which does not mark verb forms for grammatical person at all, ambiguous utterances with topic drop and syncretic verb forms are nevertheless interpreted as 1st person (singular) by default. Therefore, Auer's (1993) *inflectional hypothesis* falls short at least for this other Germanic language. For German, I tested the *inflectional hypothesis* in experiments 7 and 8, where I compared topic drop before lexical verbs with a distinct inflectional marking to topic drop before syncretic forms of modal verbs. I found that topic drop of the 1st person singular was preferred both before distinct and before syncretic verb forms (see Sections 10.5.2 and 10.5.3).

Besides the *inflectional hypothesis*, the prevalence of 1st person (singular) topic drop is also explained by a group of hypotheses that I term *extralinguistic hy-*

[2]It was later adapted by Imo (2013, 2014), who additionally suggests an *extralingustic hypothesis*. See below.

[3]Note that this argumentation has similarities to what is nowadays called *Taraldsen's generalization*, i.e., the hypothesis that null subjects in *pro*-drop languages are licensed by a rich inflectional marking on the verb (Taraldsen 1980, see, e.g., Roberts 2014 for an overview). See Section 2.2.1 for a distinction of topic drop from *pro*-drop and Section 2.2.2 for discourse (*pro*-) drop languages, which allow for null subjects even though the verb is not marked for inflection at all.

[4]According to Zifonun et al. (1997: 1258), in German the preterite present verbs are *wissen* ('to know') and the modal verbs except for *wollen* ('to want'). They are characterized by the fact that their present forms were originally forms of the preterite, which has led to a syncretism of the 1st and the 3rd person singular forms. Such a syncretism is also present in the remaining modal verb *wollen* ('to want'), whose present forms originated from optative forms. Therefore, when I speak of preterite present verbs in the following, I also always mean *wollen*, which is a sensible simplification since we are concerned here exclusively with syncretism as the central property of this verb group.

potheses because they postulate that the omitted constituent can be recovered through some form of extralinguistic context or knowledge (see Section 4.2 for the general possibility to have topic drop with extralinguistic referents). Two groups can be further distinguished within the *extralinguistic hypotheses*: The first group argues that only the 1st person (singular) can be more easily omitted and recovered (*extralinguistic 1SG*). The second group states that both the 1st and the 2nd person (singular) have this advantage (*extralinguistic 1SG+2SG*).

Imo (2013, 2014) is not only an advocate of Auer's *inflectional hypothesis*, but he also proposes a further explanation for the prevalence of 1st person singular subject topic drop that belongs to the group of *extralinguistic 1SG hypotheses*. According to him, hearers can easily process utterances where a 1st person singular subject pronoun is omitted because "the default 'origo' of speaking, i.e. 'I-here-now', can be activated in most cases so that the recipients can assume that the 'missing' element is the unmarked 'I' " (Imo 2014: 153–154). Thus, the recovery of the omitted element also depends on a form of extralinguistic context, namely the hearer's knowledge about the deictic center of the utterance in question. A further representative of the *extralinguistic 1SG hypotheses* is Volodina (2011: 272). She points out that topic drop of the 1st person singular or plural can be used without anaphoric reference in certain text types. Similarly, Schalowski (2015: 4, footnote 1) states that 1st person pronouns can be omitted without a licensing context. According to Volodina & Onea (2012: 218), this is because the reference of 1st person pronouns is usually evident through text type knowledge. 1st person subject topic drop is particularly frequent in certain conceptually spoken text types such as diaries and personal letters, which led some authors to treat it as a phenomenon in its own right, using terms such as *diary drop* (e.g., Haegeman 1990 for English, see Volodina 2011: 272).

I discuss the second group of *extralinguistic hypotheses* under the name *extralinguistic 1SG+2SG*. A central representative of these variants is the IDS grammar (Zifonun et al. 1997: 413). They discuss 1st and 2nd person subject topic drop under their own name as *Person-Ellipse* ('person ellipsis'), a subtype of situative ellipsis. According to this concept, pronouns referring to the speaker and the hearer – and in parallel to the author and the addressee in written discourse (Zifonun et al. 1997: 416) – can be omitted in a shared speech situation, where the roles of speaker and hearer are clearly determined (Zifonun et al. 1997: 414). Reference to a group of speakers or hearers is also possible but rarely occurs (Zifonun et al. 1997: 415). According to this *extralinguistic hypothesis*, topic drop of the 1st and the 2nd person should be more frequent because both speaker and hearer are recoverable from the speech situation and through knowledge about certain text types.

A similar argumentation can be found in Ariel 1990 (see Section 4.4 for her accessibility theory). She claims that 1st and 2nd person pronouns "correspond to assigned roles in conversations, while third-person pronouns refer to any person, excluding the above two" (Ariel 1990: 47).[5] This difference has consequences not only for topic drop but for null pronouns in several languages where covert 1st and the 2nd person pronouns are more common than covert 3rd person pronouns. Ariel (1990: 48) stresses that speaker and hearer are such salient referents that the overt 1st and 2nd person pronouns are only used under special circumstances in languages with null pronouns, i.e., if they are less accessible in Ariel's terms.

A different variant of the *extralinguistic 1SG+2SG hypotheses* is put forward by Volodina & Onea (2012: 218). They state that 1st and 2nd person pronouns can be better omitted because they are not marked for grammatical gender,[6] which they believe reduces potential ambiguity. While it may well be that ambiguity plays a role in the usage of topic drop (see also Section 10.1.2), it seems to me that the lack of gender marking is only a side effect of the uniqueness of speaker and hearer roles, which is the real factor that reduces ambiguity. In a communication situation, there is usually one speaker at a time and one or several hearers to which the 1st and 2nd person pronouns unambiguously refer. 3rd person pronouns, in turn, can theoretically refer to any person, object, or proposition that is somehow present in the linguistic or extralinguistic context, provided that the grammatical gender matches in the singular cases. Therefore, gender marking indeed reduces ambiguity in German but for the 3rd person singular.

In sum, two or rather three explanations for the prevalence of 1st person (singular) topic drop can be distinguished: First, there is Auer's (1993) *inflectional hypothesis* according to which the 1st person singular can be easily recovered because the following verb (often) has a distinct inflectional ending expressing the grammatical person of the omitted subject. If one takes the hypothesis seriously, one would have to assume that, in principle, all subjects followed by a distinct verb form can be recovered well and, thus, be omitted, i.e., the *inflectional hypothesis* would then not only concern the 1st person singular. Second, there are two variants of *extralinguistic hypotheses*. The first variant, which I termed *extralinguistic 1SG*, argues that topic drop of the 1st person should be more felicitous

[5] Zdorenko (2010: 128) makes a similar statement, arguing that "[f]irst- and second-person subjects [...] are always topics, i.e., they always have a discourse antecedent, because the identities of the speaker and the hearer are taken as a background assumption in any conversation."

[6] Note that there are languages that also or only mark the 1st and 2nd person pronouns for grammatical gender, such as Ngala (Ndu, Sepik; north-western Papua New Guinea), in which all three persons in the singular are marked for gender, and Korana (Central Khoisan; South Africa), which distinguishes gender for all three persons and numbers (Siewierska 2013).

because of the special status of the speaker in certain text types and as the deictic center. The second variant, *extralinguistic 1SG+2SG*, explains an advantage of both 1st and 2nd person topic drop by the uniqueness of speaker and hearer roles. The results of experiments 9 and 10 in Sections 10.5.4 and 10.5.5 suggest that there is no difference in acceptability between topic drop of the 1st and topic drop of the 2nd person singular, supporting the second type of *extralinguistic hypotheses*.

A more radical approach to explain the special status of 1st and 2nd person topic drop in German is taken by Trutkowski (2011, 2016). She assumes that in German, the omission of a prefield constituent can be caused by two distinct mechanisms and distinguishes topic drop, i.e., the omission of objects and 3rd person subjects, from "out of the blue-drop" (OBD), i.e., the omission of 1st and 2nd person subjects. Consequently, she assumes two different ellipsis types depending on the grammatical person of the omitted constituent. In the following, I reject this distinction for two reasons. On the one hand, the justification behind it is problematic. On the other hand, assuming two phenomena instead of one unnecessarily complicates the grammatical system. According to Occam's razor, the simpler system is preferable if the more complicated one offers no advantages.

By OBD, Trutkowski refers to cases where "[t]he gap is identified/licensed independently of a salient discourse antecedent" (Trutkowski 2016: 187). According to her, OBD is limited to "out of the blue-contexts", i.e., real null contexts without salient discourse antecedents, which she assumes to exist in the form of book and song titles, headlines, and novel and conversation beginnings (Trutkowski 2016: 188). She argues that OBD is distinct from English diary drop (Haegeman 1990, see Section 2.4.1) because the latter would require "a (default) referent or an antecedent/addressee" (Trutkowski 2016: 192), while OBD does not. In this context, she discusses, among others, the examples (1b), to which I added the corresponding context in (1a) from the corpus,[7] and (2).

(1) a. „Du hättest ruhig diplomatischer sein können", sage ich zu meiner Schwester. Sie schaut mich verwundert an und fragt, was ich denn Diplomatisches zu unserer Unterhaltung beigetragen hätte. „Hätte ich sagen sollen, tolle Frisur, steht dir bestimmt gut?", fragt meine Schwester.

[7] Context added from corpus search results for "Dann: Die lässt sich die Haare trotzdem so schneiden. Kennst sie doch.", from the aggregated reference and newspaper corpus of the digital dictionary of the German language (Digitales Wörterbuch der deutschen Sprache), https://www.dwds.de/r/?q=Dann%3A++Die+l%C3%A4sst+sich+die+Haare+trotzdem+so+schneiden.+Kennst+sie+doch.+&corpus=public&date-start=1465&date-end=2018&genre=Belletristik&genre=Wissenschaft&genre=Gebrauchsliteratur&genre=Zeitung&format=max&sort=date_desc&limit=10 (visited on 01/02/2025).

' 'You could have been more diplomatic,' I say to my sister. She looks at me in surprise and asks what I would have contributed diplomatically to our conversation. 'Should I have said, great hairstyle, looks good on you?' my sister asks.' [DWDS: Corpus, Berliner Zeitung, 01/22/2005]

b. *Dann: „Die lässt sich die Haare trotzdem so*
then she.DEM gets REFL the hair nevertheless in.this.way
schneiden. Δ *Kennst sie doch"*
cut you.2SG know her PART
'Then: 'Anyway, she will get her hair cut like that. (You) know her.' '
(Trutkowski 2016: 190)

(2) *Bitte einmal abstimmen.* Δ *Würdet mir sehr mit eurer*
please PART vote you.2PL would me very with your.2PL
Meinung helfen!
opinion help
'Please vote. (You) would help me a lot with your opinion!' (Trutkowski 2016: 191)

In Trutkowski's view, (1) resembles the English diary drop. Therefore, she does not consider it to be OBD because "the hearer (that is addressed by the speaker) is provided by the discourse situation" (Trutkowski 2016: 191). According to her, however, (2) is an undoubted case of OBD. This distinction is questionable. (2) is a post in a forum that is clearly addressed to the other users of this forum. The addressee can just as well be traced back to the discourse situation as in example (1), even if this situation is digitally mediated. It is also questionable whether the context in (2) is really an "out of the blue-context" or whether the imperative infinitive does in fact evoke the addressee as discourse antecedent.

The concept of OBD seems to be disputable in general. Even in the null contexts that Trutkowski (2016) proposes, such as the beginning of a song or a headline, it is clear to the hearers or readers that they are in some kind of communication situation involving the roles of speaker and hearer. Therefore, they can easily recover any omitted 1st or 2nd person subject pronouns. Further evidence against OBD as a concept is given in the following. I discuss a variant of the *inflectional hypothesis* that Trutkowski (2016) proposes to explain the licensing of OBD:

(a) $1^{st}/2^{nd}$ person null subjects are well-formed *out of the blue* because they can be identified/licensed by discrete verbal inflectional endings.
(b) 3^{rd} person null subjects are identified and licensed in dependence of [sic!] an external antecedent (cf. topic drop) – as a consequence it does not

matter whether 3rd person verbal inflectional endings are syncretic with 1st/2nd person verbal inflectional endings. (Trutkowski 2016: 218)

Trutkowski (2016) claims that syncretic verb forms that are ambiguous between the 1st and the 3rd person do not pose a problem for interpreting null subjects. If there is a suitable antecedent that licenses 3rd person topic drop, the utterance will be interpreted as 3rd person topic drop. If there is no such antecedent, the null subject will be interpreted as OBD of the 1st person. She backs up her claim with examples such as (3) with null subjects before syncretic verb forms.

(3) *Hans und ich haben den Film schon gesehen.*
Hans and I have the movie already seen
'Hans and I have already seen the movie.'

a. *Δ *Will deshalb lieber zuhause bleiben.*
he want.3SG(/1SG) therefore rather at.home stay
'(He) therefore prefers to stay at home.' (3rd person singular)

b. Δ *Will deshalb lieber zuhause bleiben.*
I want.1SG(/3SG) therefore rather at.home stay
'(I) therefore prefer to stay at home.' (1st person singular)
(Trutkowski 2016: 208, her judgments)

According to her, the coordinated subject *Hans und ich* in the first sentence cannot serve as the antecedent for 3rd person topic drop in (3a), whereas OBD of the 1st person singular is possible in (3b). The similar example (4) questions this intuition.[8]

(4) *Hans und ich kommen heute später zum Training.*
Hans and I come today later to.the training
'Hans and I will come later today for training.'

a. Δ *Musste sich noch rasieren.*
he must.3SG(/1SG).PRET himself still shave
'(He) still had to shave.' (Lit. '(He) still had to shave himself.')
(3rd person singular)

b. *Δ *Musste sich noch rasieren.*
I must.1SG(/3SG).PRET himself still shave
'(I) still had to shave.' (Lit. '(I) still had to shave himself.')
(1st person singular)

[8] I thank Robin Lemke for suggesting the example and for the helpful discussion.

The difference is that despite the syncretic verb form, the reflexive *sich* makes it clear that the omitted subject has to be a 3rd person singular subject (4a). An interpretation as 1st person singular is not possible (4b). Therefore, the topic drop in (4a) is possible and, importantly, *Hans* in the coordinated subject of the first sentence can now serve as an antecedent, whereas (4b), which Trutkowski (2016) refers to as OBD of the 1st person, is no longer an option.[9]

This suggests that the contrast observed by Trutkowski in (3) is not categorical in nature.[10] Rather, there seems to be a preference for interpreting the covert constituent as a 1st person singular pronoun, which is based on probabilities and frequencies and which can change in an appropriate context, as example (4) shows. This line of reasoning, therefore, raises again the question of whether it is legitimate to assume two different ellipsis types, i.e., to distinguish between OBD and topic drop. I refrain from making such a distinction because it does not provide additional explanatory power but would complicate the grammatical system.

The theoretical overview has shown that many authors agree in assuming that the 1st person can be omitted particularly well under topic drop but disagree on how to explain this assumption. Auer (1993) proposes an *inflectional hypothesis*, which was at least originally put forward to explain only a part of the data, i.e., only 1st person singular subject topic drop before distinctly marked verb forms. Zifonun et al. (1997), Volodina (2011), Volodina & Onea (2012), Imo (2013, 2014), and Schalowski (2015) are advocates of what I termed *extralinguistic hypotheses*, which come in two varieties. Some authors assume that only the 1st person has an advantage for being the deictic center and the speaker/writer of certain text types, whereas others refer to the default roles of speaker and hearer in every communicative situation and assume that both the 1st and the 2nd person can be omitted particularly well. The differentiation between 1st and 2nd person on the one hand and 3rd person on the other culminates in Trutkowski (2016) even assuming two different ellipsis types, topic drop and OBD, a view I rejected as unnecessary.

10.1.2 Theoretical overview of ambiguity avoidance

Ambiguity avoidance could be a further factor that impacts the usage of topic drop. It predicts that speakers avoid ambiguous linguistic structures to facilitate

[9] Of course, if *sich* is replaced by *mich*, i.e., the 1st person singular reflexive, the pattern is inverse and only the reading with a covert 1st person singular subject is possible.

[10] However, an experimental investigation of both Trutkowski's (2016) argument and my counterexample is still pending.

the processing for the hearer. Like *UID*, ambiguity avoidance is, thus, based on the concept of audience design, i.e., the adaptation of linguistic structures by the speaker for the benefit of the hearer (see also Section 6.2).

A linguistic expression is ambiguous if it can be interpreted in more than one way, i.e., if it has more than one possible meaning.[11] Sentences can be either globally (fully) ambiguous if it cannot be decided from the utterance alone which reading is the intended one, or locally (temporarily) ambiguous if the intended meaning becomes evident when the processing of the utterance is complete (Pritchett 1988, Ferreira 2007). While there are several forms of ambiguity, e.g., lexical, structural, and scope ambiguity (Wasow 2015),[12] for topic drop, a special type of lexical ambiguity is important: the ambiguity of inflectional morphemes called syncretism (Wasow et al. 2005). It is common to distinguish systematic syncretisms from accidental syncretisms. Systematic syncretisms can be captured by underspecification since they are simplifications of the feature space of a natural class. This is not possible for the latter because the features involved

[11]Kennedy (2019: 238–239) delimits ambiguity from vagueness by stating that ambiguity is an uncertainty that results in variation with respect to truth conditions, whereas terms are vague if there remains "uncertainty [...] about precisely what properties these terms ascribe to the objects to which they are applied". The expression *funny* in (i) is ambiguous because it has two meanings. Depending on which meaning is the intended one, the same utterance can be judged true or false in the exact same context. For instance, (i) can be successfully uttered to say that Sterling's cousin is good at making people laugh but that she is not strange.

(i) Sterling's cousin is funny, but she is not funny. (following Kennedy 2019: 237)

This is different for (ii), where the adjective *tall* is vague. *Tall* does not have two meanings but only one, something like "having a height above the average", therefore (ii) is not possible. It is this fuzzy meaning that causes the vagueness because the average is not fixed.

(ii) #Sterling's cousin is tall, but she is not tall.

[12]For instance, the sentence (i) is lexically ambiguous because the German word *Bank* can refer to a financial institution ('bank') or a seat ('bench'), among other things. The classic sentence (ii) is structurally ambiguous because the PP *with a telescope* can be attached either as an attribute to *a man* or function as an instrumental adverbial specifying the action of seeing. The sentence (iii) exhibits scope ambiguity because it can either mean that every person in the room speaks any two languages or that two specific languages are spoken by all the people in the room.

(i) *Treffen wir uns bei der Bank!*
meet we us at the bench/bank
'Let's meet at the bench/bank!'

(ii) I saw a man with a telescope.

(iii) Everyone in the room speaks two languages. (Wasow 2015: 35)

are not part of a common natural class (Korth 2017). I already discussed a type of systematic syncretism in Section 10.1.1. In the present tense, the inflected forms of modal verbs are identical for the 1st and the 3rd person singular (*ich kann* ('I can') vs. *sie kann* ('she can')), just like they are for the 1st and the 3rd person plural (*wir können* ('we can') vs. *sie können* ('they can')). Following Korth (2017), the paradigm is underspecified as it lacks a feature expressing the opposition between the 1st and the 3rd person. Consequently, an utterance with topic drop such as (5) with a corresponding syncretic form is globally ambiguous between the 1st and the 3rd person singular (see also Trutkowski 2016: 10 for an overview of this syncretism listed in tabular form).

(5) Δ *Kann heute arbeiten.*
I/she/he/it can.1SG/3SG today work
'(I/She/He/It) can work today.'

Poitou (1993), as well as Zifonun et al. (1997), discuss another ambiguity arising from further syncretic forms. They note that several forms of the 1st person singular present tense indicative are formally identical to the forms of the imperative singular (6). Since the indicative forms and the imperative do not form a natural class, this syncretism is best described as an accidental syncretism.

(6) *Arbeite heute nicht so viel.*
work.1SG.IND.PRS/IMP.SG today not so much
'(I) will not work so much today.' / 'Do not work so much today!'

Zifonun et al. (1997: 415) add that the two forms are also prosodically identical as both exhibit a falling boundary tone. Poitou (1993) argues that this ambiguity is mostly negligible because topic drop would occur preferably before modal verbs, which do not have an imperative, and auxiliaries, where the forms would not be syncretic. The latter point, however, is only correct for the auxiliary *sein* ('to be') where the imperative *sei* is distinct from the 1st person singular present tense indicative form *bin*, but the forms of *haben* ('to have') and *werden* ('will') are identical: *hab(e)* and *werd(e)*. More importantly, the corpus studies by Androutsopoulos & Schmidt (2002) and Frick (2017), discussed in Section 11.2, show that although topic drop may be more frequent at least before copular and modal verbs, it is by no means rare before lexical verbs so that the ambiguity may well be relevant.

The syncretism between the imperative and the 1st person singular in present tense indicative occurs for verbs such as *öffnen* ('to open'), for which both forms end in *e* (schwa) (7a) (Leibniz-Institut für Deutsche Sprache 2018b). But also verbs

such as *anrufen* ('to call'), for which the imperative is usually formed without a final *e*, in particular in spoken language, exhibit syncretic forms. This is because the final *e* of the 1st person singular is frequently omitted in colloquial speech, turning the actual standard form (*ich*) *rufe an* into the syncretic form (*ich*) *ruf an* (7b) (Leibniz-Institut für Deutsche Sprache 2018b). In Standard High German, the forms of verbs with an *e/i-alternation* such as *essen* ('to eat') are not syncretic (8) (Leibniz-Institut für Deutsche Sprache 2018b). However, in certain dialects and occasionally also in colloquial speech, the present tense form is used as the imperative for these verbs as well, as the example (9) from the FOLK corpus (Schmidt 2014) shows.

(7) a. *Öffne die Tür!*
open.1SG.IND.PRS/IMP.SG the door!
'(I) open the door!' / 'Open the door!'

b. *Ruf(e) gleich an!*
call.1SG.IND.PRS/IMP.SG right.away VPART
'(I) will call right away!' / 'Call right away!'

(8) a. *Iss den Kuchen!*
eat.IMP.SG the cake
'Eat the cake!' (Not: '(I) eat the cake!')

b. *(Ich) esse den Kuchen.*
I eat.1SG.IND.PRS the cake
'(I) am eating the cake.' (Not: 'Eat the cake!')

(9) *ess mal dein zimt[hörnchen]*
eat.1SG.IND.PRS/eat.IMP.SG PART your cinnamon.croissant
'Eat your cinnamon croissant!' [FOLK_E_00309_SE_01_T_02]

In summary, in colloquial speech, there is potentially an ambiguity between topic drop of the 1st person singular in present tense and the singular imperative, which can arise in appropriate contexts, and if no other words in the utterance such as reflexive or object pronouns disambiguate.

The concept of ambiguity avoidance predicts that the existence of ambiguities such as those just discussed for topic drop constrains language usage. Already Grice famously formulated "Avoid ambiguity!" as one of the submaxims of his maxim of manner (Grice 1975: 46). Accordingly, one interlocutor should assume that if the other is cooperative, they will not use ambiguous linguistic structures, or if they do, it will be to trigger an implicature. The concept of ambiguity avoidance is, thus, similar to *UID* (see Section 6.2) in that it is based on the idea of

audience design, i.e., that speakers adapt their utterances to the hearer. Speakers should not use ambiguous structures in their language production for two reasons: first, to avoid an increased processing effort on the part of the hearers, who need to disambiguate the utterance, and, second, to avoid misunderstandings if the hearers disambiguate in an unintended way. This line of reasoning has been extensively discussed for so-called *garden path sentences* such as the famous (10) (see, e.g., Sanz et al. 2013). Here, a temporarily ambiguous utterance can cause processing difficulties but only when the hearer realizes at some point that the structure they initially considered to be the more likely one is not accurate. That is, increased processing effort occurs not at the ambiguous but at the disambiguating point.

(10) The horse raced past the barn fell. (Bever 1970: 316)

Psycholinguistic research has found little evidence that speakers systematically avoid linguistic ambiguities, in particular, structural ones (see Ferreira 2008, Wasow 2015 for more detailed overviews). For example, the experimental results of Ferreira & Dell (2000) and the corpus results of Rohdenburg (2021) do not suggest that speakers use optional function words such as *that* to disambiguate (temporarily) ambiguous clauses. In experiments by Arnold et al. (2004), speakers did not use constituent ordering to disambiguate local PP-attachment ambiguities. Snedeker & Trueswell (2003) found in their experiments that speakers avoided ambiguity by providing prosodic cues depending on the current situation. However, the results by Kraljic & Brennan (2005) and Schafer et al. (2005) show that while speakers also disambiguated prosodically in spontaneous speech, they did so regardless of whether the current situation demanded it or the hearer benefited from it, i.e., it could not be interpreted as a strategy of ambiguity avoidance. In a study of lexical ambiguity by Ferreira et al. (2005), speakers took linguistic ambiguity into account to some extent when describing objects presented in ambiguous displays, but they considered non-linguistic ambiguity more strongly and more consistently. This pattern was confirmed in an eye-tracking study by Rabagliati & Robertson (2017), who furthermore showed that speakers even proactively monitored for non-linguistic ambiguities, as well as self-monitored their produced utterances subsequently for how informative they were.[13] Concerning the processing rather than the production of ambiguous structures, Levy (2008) points to the results of Traxler et al. (1998), van Gompel et al. (2001), and van Gompel et al. (2005). They found that several types of global attachment

[13]Rabagliati & Robertson (2017) argue that for younger children, these two processes are still limited.

ambiguities did not result in slower but faster reading times, i.e., they were processed more easily. This suggests that hearers do not necessarily experience difficulties processing ambiguous structures (potentially even the opposite), thus, speakers may not be required to avoid them. Levy (2008: 1152–1153) argues that this pattern follows naturally from the surprisal theory. He assumes a "fully parallel, incremental probabilistic parser capable of online inference (that is, inference before input is complete)" which assigns "a probability distribution over the complete structures to which the already-seen input may possibly extend" (Levy 2008: 1132). In this framework, ambiguity is of relevance only for conditional word probabilities. If at a given point in an incrementally parsed sentence, a local structural ambiguity at a word occurs so that the word would, up to this point, be consistent with several structural analyses of the complete sentence, the conditional probability of that word is fed by all these consistent structures. This makes the word overall more probable and, therefore, easier to process.

While the research just discussed seems to call into question the role of ambiguity avoidance for speech production (and perception), it is debatable whether it can be generalized to syncretisms relevant to topic drop given that primarily structural ambiguity was tested (see also the study by Soares et al. 2019 in Section 10.2.2). For instance, Volodina & Onea (2012) argue that ambiguity avoidance does play a role in how topic drop is used. They state that the existing number of syncretisms in the German verbal paradigm could lead to the situation that the hearer cannot reliably identify the referent of the omitted constituent. Therefore, they postulate that topic drop is restricted to those cases where the omission does not lead to ambiguity (Volodina & Onea 2012: 214–215).

Given the conflicting predictions of whether or not ambiguity avoidance affects speech production and, in particular, topic drop, it seems appropriate to test the impact of ambiguity avoidance empirically. The first steps in this direction are considering verbal inflection as a factor in the corpus analysis in Section 10.4.4 and comparing topic drop before unambiguously marked lexical verbs with topic drop before ambiguous verb forms in locally ambiguous utterances.[14] To this end, I conducted a post hoc analysis comparing data from experiment 7 (see Section 10.5.2) with distinctive verb forms and experiment 8 (see Section 10.5.3) with syncretic forms.

[14] In another experiment, I also tested utterances that were globally ambiguous between a reading as an utterance with 1st person singular topic drop and a singular imperative reading in ambiguity-promoting and non-ambiguity-promoting contexts. However, the results here were not conclusive. Moreover, the experiment suffered from the fact that the baseline, which was actually intended to be non-ambiguous, was itself unintentionally a syncretic form. For this reason, I do not report the results of this experiment here.

10.2 Previous empirical evidence

10.2.1 Previous empirical evidence regarding grammatical person

In the following, I discuss four corpus studies from the literature (Auer 1993, Androutsopoulos & Schmidt 2002, Döring 2002, Frick 2017) that evidence differences in the frequency of topic drop with respect to grammatical person. According to them, the 1st person singular is indeed frequently omitted. In particular, it is more often omitted than realized in text messages. Topic drop of the 2nd person singular and the 1st person plural are also relatively frequent. The 3rd person singular seems to be omitted particularly often when the prefield could be filled with the neuter demonstrative pronoun *das* ('that'), which often refers to propositions. The remaining plural persons are hardly ever targeted by topic drop but are also rare in the prefield of the corresponding full forms.

10.2.1.1 Auer (1993)

Auer (1993) presents data from a corpus of spoken conversations that he does not specify further in terms of size, source, or date of creation. He states that he extracted a bit more than 100 instances of topic drop and V1 declaratives (Auer 1993: 195). Topic drop accounts for about 75% of the cases (Auer 1993: 198). The 3rd person subject and object pronoun *das* ('that'), which refers to propositions, groups of propositions, verbs, or predicatives, is omitted most often (Auer 1993: 200). Auer adds that the 1st person singular subject pronoun *ich* ('I') is frequently omitted (Auer 1993: 198), as well as the 2nd person singular, whereas the 1st and 2nd person plural are rarely unrealized (Auer 1993: 199). He does not provide information about the frequency of omitted 3rd person singular pronouns other than *das*, nor about 3rd person plural pronouns but only presents an example of the latter (Auer 1993: 199, example 10). Since he does not report absolute or relative numbers, frequency tables, or a statistical analysis, his results can only be interpreted as anecdotal.

10.2.1.2 Androutsopoulos & Schmidt (2002)

In one of the earliest empirical studies of text messages in German, Androutsopoulos & Schmidt (2002) analyzed a corpus of 934 text messages. The messages had been produced during a period of 8 weeks in the year 2000 by a small group of mainly 5 people in their late twenties (Androutsopoulos & Schmidt 2002: 55–56). In their corpus, Androutsopoulos & Schmidt (2002: 68–69) found 229 instances of subject topic drop and compared them to 197 complete utterances where the

subject pronoun in the prefield could have been omitted. They did not look at object topic drop.[15] Table 10.1 shows the omission rates by pronoun that they determined. For the ambiguous pronoun *sie*, which can either be the 3rd person singular feminine ('she') or the 3rd person plural ('they'), they do not provide a subdivision.

Table 10.1: Omission rates by pronoun in the text message corpus of Androutsopoulos & Schmidt (2002), taken from Androutsopoulos & Schmidt (2002: 69), adapted

Pronoun	Person, number, gender	Full form	Topic drop	Total	Omission rate
ich ('I')	1SG	124	187	311	60.13%
du ('you')	2SG	20	7	27	25.93%
er ('he')	3SG M	1	3	4	75.00%
sie ('she'/'they')	3SG F/3PL	4	0	4	0.00%
es/das ('it'/'that')	3SG N	19	20	39	51.28%
wir ('we')	1PL	28	12	40	30.00%
ihr ('you')	2PL	1	0	1	0.00%

The 1st person singular subject pronoun is omitted most often in absolute terms,[16] followed by *es* ('it') and *das* ('that'), i.e., 3rd person singular neuter pronouns. The difference between *ich* and *es*/*das* is not significant according to a Pearson's chi-squared test with Yates's continuity correction, which I calculated in R (R Core Team 2021) ($\chi^2(1) = 0.79, p > 0.37$).[17] The 1st person plural *wir* ('we') is omitted only half as often as the 1st person singular (significant, $\chi^2(1) = 11.91, p < 0.001$) but only slightly more often than the 2nd person singular *du* ('you') (not significant, $\chi^2(1) = 0.007, p > 0.9$).[18] These quantitative results are similar to the tentative results of Auer (1993): *Das* (*es*) and *ich* are most often omitted.

[15] They also state that they excluded "elliptische Formeln" ('elliptical formulae', my translation) such as *geht so* ('Not too bad', e.g., as a response to the question *How are you?*, lit. '(it) goes so') without further discussing what they consider to be a formula and how many of them were excluded (Androutsopoulos & Schmidt 2002: 68, footnote 24).

[16] Actually, *er* ('he') has the highest omission rate of 75%, but this rate is unreliable because it is based on only four occurrences of *er* in the entire corpus. The same is true for the omission rates of *sie* ('she') and *ihr* ('you.PL'), which are also based on only four or even only one data point, respectively.

[17] The other two tests reported in the following were calculated in the same way.

[18] Note, however, that there are relatively few instances of *wir* in the corpus and even fewer of *du*. Thus, the result of their comparison may not be reliable, i.e., for a larger data set there might be a significant difference between their omission rates.

10.2.1.3 Döring (2002)

A further text message corpus of comparable size was analyzed by Döring (2002). The corpus consists of 1 000 text messages with about 13 000 tokens and was created in April and September 2001 (Döring 2002: 102). Döring investigated several types of what she terms syntactic reduction (Döring 2002: 105), among which is the omission of subject pronouns, i.e., topic drop of subjects. She does not provide absolute or relative numbers, but her results are in line with the results by Androutsopoulos & Schmidt (2002) according to which the subject pronoun *ich* ('I') is most often omitted (Döring 2002: 107). By way of qualification, she notes that *ich* ('I') and *du* ('you') were at the same time the most frequent words in the text message corpus. However, this information is of limited value because she does not consider the syntactic position of the pronouns. Consequently, it remains unclear how many of them occur in the prefield and could theoretically be targeted by topic drop.

10.2.1.4 Frick (2017)

In her study of Swiss German text messages (see Section 8.2 for details), Frick (2017) also assessed the grammatical person of her 2 326 instances of subject topic drop and the 2 059 realized subject pronouns in the prefield.[19] The distribution is shown in Table 10.2.

Table 10.2: Omission rates by grammatical person in the text message corpus of Frick (2017), taken from Frick (2017: 88)

Grammatical person	Full form	Topic drop	Total	Omission rate
1SG	1 314	1 916	3 230	59.32%
2SG	195	173	368	47.01%
3SG	229	158	387	40.83%
1PL	279	70	349	20.06%
2PL	25	4	29	13.79%
3PL	17	5	22	22.73%

[19] Given that the number of instances is about ten times higher in Frick's (2017) corpus than in the corpus of Androutsopoulos & Schmidt (2002), Frick's results should be more reliable. While her results are based on Swiss German data, the results of the other corpus studies discussed here and of my corpus study presented in Section 10.4 on German Standard German show a similar basic tendency.

Again, the omission rate of the 1st person singular subject pronoun is the highest. More specifically, it is significantly higher than the omission rates of the other grammatical persons, as Frick (2017: 90, footnote 102) reports. *Ich* is omitted more often than it is realized. Consequently, Frick (2017: 89) argues that in (Swiss German) text messages, topic drop of the 1st person singular has become the unmarked default case. The 2nd person singular exhibits the second highest omission rate with about 47.01% (according to Frick 2017: 90, footnote 102, it is significantly higher than those of the remaining persons). It is slightly more often realized in the prefield than omitted. 3rd person singular pronouns are omitted in 40.83% of the cases and, thus, still significantly more frequently than all the plural persons, which have omission rates between about 14% and 23% (Frick 2017: 90, footnote 102).

At first glance, it is surprising that the omission rate of the 3rd person singular is not much lower than that of the 2nd person singular. Androutsopoulos & Schmidt's (2002) results suggest that the omitted 3rd person singular pronouns are rather *es* ('it') and *das* ('that'), which often refer to propositions, instead of *sie* ('she') or *er* ('he'), which refer to persons and objects. While there are many persons and objects in the world, to which such a pronoun may refer, there is often only one very salient proposition in discourse, namely the last one mentioned. This proposition is therefore highly predictable and the pronoun referring to it can be easily omitted.[20] This hypothesis is supported by the results of Helmer's (2016) corpus study, where she compared the type of referent for utterances with topic drop and utterances with the anaphoric subject and object pronoun *das* in the prefield (see Section 11.2 for details on this reference data). Helmer (2016: 210–211) states that in her data of spoken dialogues, most instances of topic drop (41.6%) refer to a proposition, followed by 20.2% that refer to a noun phrase, and 20% that constitute cases of indirect topic drop (see Section 4.1). The reference of the full forms with *das* is mostly a noun phrase (33.5%), followed by cases with indirect reference (26.5%), and the reference to propositions in the third place (23.5%). Helmer (2016) states that topic drop, thus, refers significantly more often to propositions than the anaphor *das*. She adds that it occurs particularly often in evaluations, consents, objections, and responses that refer to propositions (this is similar to the socio-pragmatic functions of topic drop discussed in Section 5.4).

[20] It is not clear from Frick's (2017) statements whether in her corpus data, the omitted 3rd person singular subjects also most frequently refer to propositions. While in her examples the covert 3rd person singular constituents are coreferential with concrete NPs in the previous discourse (Frick 2017: 106), she adds that the reference can also be a more abstract entity (Frick 2017: 107, footnote 120).

10.2.2 Previous empirical evidence regarding ambiguity avoidance

While there is, to the best of my knowledge, no study of topic drop and ambiguity avoidance in German, Soares et al. (2019) looked at the interplay of 1st person null subjects and syncretic verb forms in Brazilian Portuguese. Considering ambiguity avoidance in a different language and with a different form of null subjects takes me further away from topic drop in German than the studies mentioned in the context of the other factors do. It still seems reasonable to me because the principle of ambiguity avoidance should be at work across languages if it plays any role at all. This means that the conclusions drawn here should be transferable.

Soares et al. (2019) conducted a corpus study and two acceptability rating experiments. For the corpus study, they used a corpus that was previously analyzed by Duarte (1995) and that contains 18 oral interviews with a total of 8 032 of what they term "inflected clauses" (Soares et al. 2019: 3585). Soares et al. (2019: 3586–3587) found that for the 1st person singular, the subject is more often realized overtly (80% overt subjects) before a syncretic verb form than before a verb form that is distinctly marked (70% overt subjects).

To verify this result in a controlled setting, they conducted two acceptability rating experiments in the form of 2 × 2 designs crossing Inflection (syncretic vs. distinct)[21] and Subject (null vs. overt). The items consisted of a context sentence, a question, and a target sentence with the 1st person singular as the overt subject of the main clause and as the overt vs. covert subject of a subordinate clause, as shown in example (11). In experiment 1, the target sentence contained an animate 3rd person singular object,(11a) and (11b), whereas in experiment 2 this object was inanimate, (11c) and (11d).

(11) A: *Maria estava muito nervosa. Você sabe quando ela ficou mais calma?*
A: 'Mary was very nervous. Do you know when she's got calmer?'

a. *B: Eu tranquilizei a Maria quando (eu) divulguei os*
I calm.down.1sg.pst the Maria when I publish.1sg.pst the
resultados do exame.
results of.the exam
B: 'I calmed Mary down when I published the results of the exam.'
(Soares et al. 2019: 3588)

b. *B: Eu tranquilizei a Maria quando (eu) ia*
I calm.down.1sg.pst the Maria when I was.going.to

[21] They kept the verb and the grammatical person constant and achieved variation in inflection by using verb forms in different tenses that are either syncretic or distinctive.

divulgar os resultados do exame.
publish.INF the results of.the exam
B: 'I calmed Mary down when I was going to publish the results of the exam.' (Soares et al. 2019: 3588)

c. *B: Eu resolvi o problema quando (eu) divulguei os resultados do exame.*
I solve.1SG.PST the problem when I publish.1SG.PST the results of.the exam
B: 'I solved the problem when I published the results of the exam.' (Soares et al. 2019: 3592)

d. *B: Eu resolvi o problema quando (eu) ia divulgar os resultados do exame.*
I solve.1SG.PST the problem when I was.going.to publish.INF the results of.the exam
B: 'I solved the problem when I was going to publish the results of the exam.' (Soares et al. 2019: 3592)

In experiment 1, they found that utterances with null subjects were rated significantly worse before syncretic verb forms than before distinct ones compared to the overt subjects (Soares et al. 2019: 3591). In experiment 2, no such interaction was present, but null subjects were generally preferred over overt subjects (Soares et al. 2019: 3593). Soares et al. (2019: 3593) interpret these results as evidence of ambiguity avoidance, which, however, is only relevant if there is a syncretic verb form *and* a "competition between potential antecedents for the null subject in contexts where the verb is ambiguous between first and third persons" (Soares et al. 2019: 3593). Such a competition was present in experiment 1, where the animate object represented a competing referent for the null subject but not in experiment 2, where the object was inanimate. In sum, these results suggest that ambiguity avoidance does not generally play a role in null subjects and syncretic verb forms but only if there is competition between multiple potential antecedents.

In fact, there was such competition in my experiments 7 and 8, which I used indirectly to study ambiguity avoidance. The items also contained two potential referents for the covert constituent, like in Soares et al.'s (2019) experiment 1. A major difference to Soares et al.'s (2019) study was, however, that the target utterances in experiment 8 with syncretic verb forms were not globally but only locally ambiguous. It is unclear whether speakers also avoid this local ambiguity. I discuss this point again in Section 10.5.3.4.

10.3 Information-theoretic predictions for grammatical person and verbal inflection

In the following, I outline the information-theoretic predictions for grammatical person and distinct inflectional ending and argue that their impact can be captured by the *avoid troughs* and the *facilitate recovery* principles. On the one hand, those grammatical persons and numbers that are highly predictable in context should be better omitted. On the other hand, a distinct inflectional ending on the verb following subject topic drop should facilitate recovering the omitted subject as it provides information about the person and number of that subject.

The overview of previous corpus studies that investigated grammatical person in Section 10.2.1 has shown that topic drop of the 1st person singular subject pronoun *ich* ('I') is particularly frequent. Also the 2nd person singular subject pronoun *du* ('you') and the 3rd person singular neuter subject pronouns *das* ('that') and *es* ('it') still exhibit relatively high omission rates. It is reasonable to also infer for certain grammatical persons how well they can be omitted from their general frequency in the prefield, similar to the argument for syntactic function in Section 8.3. However, such a strategy falls short. For instance, the data from the DeReKo study, discussed in Section 8.3, partly repeated here in Table 10.3, show that *du* occurs strikingly less frequently in the corpus than the 1st person and the 3rd person singular masculine subject pronouns, both generally and in the prefield. However, the existing studies of topic drop suggest that omitting *du* is more frequent than omitting *er*.

Table 10.3: Frequency of the 1st and 2nd person singular, as well as of the 3rd person singular masculine in the DeReKo TAGGED-T archive (repeated from page 216)

Pronoun	Person, number	Gender	Case	Total occurrences	Occurrences in prefield	Proportion in prefield
ich	1SG	–	NOM	2 068 332	570 809	28.60%
du	2SG	–	NOM	130 166	10 230	7.86%
er	3SG	M	NOM	3 808 561	721 514	18.94%

While one could attribute the difference between frequency and omission rate to the different text types used in the studies, news articles in the DeReKo study and text messages in the studies by Androutsopoulos & Schmidt (2002) and Frick (2017), this does not explain the whole pattern. Also in Frick's (2017) text message

data, the subject pronoun of the 1st person plural occurs more often in the prefield than that of the 2nd person singular but is nevertheless omitted less than half as often. Thus, it stands to reason that there are further factors that influence the predictability of the prefield constituent and how well it can be omitted than its frequency and the occurrence in certain text types.

For the 1st and the 2nd person, such factors are considered in what I termed *extralinguistic 1SG+2SG hypotheses*. They assume that the recovery of an omitted prefield 1st or 2nd person subject pronoun is possible through some form of extralinguistic context, e.g., knowledge about the speaker and hearer being uniquely determinable and integral parts of any speech situation. Given that the speaker and the hearer share this knowledge, the speaker knows that pronouns referring to themselves or the hearer are highly predictable to both of them. So if the speaker performs audience design, as *UID* assumes, they should be more likely to omit a 1st or 2nd person pronoun referring to the speaker(s) or hearer(s) when using topic drop than a 3rd person pronoun referring to a person. The case of the 3rd person singular subject and object pronouns *das* ('that') and *es* ('it') may be special because they frequently refer to propositions that are highly salient in the current discourse, more specifically, to the last proposition expressed by the previous utterance. Consequently, this proposition and the pronouns referring to it are very predictable so the latter should be omitted by the speaker to avoid surprisal minima. The results of Helmer's (2016) corpus study mentioned in Section 10.2.1 and of my corpus study discussed in Section 10.4 show that covert 3rd person singular constituents do indeed frequently refer to propositions or VPs but also that not all cases can be explained in this way. About 40% of the instances of topic drop refer to propositions in Helmer's (2016) data. In my data, between 40% and 50% of the subjects have a propositional reference, for the objects the figure is higher, at almost 70%.

An information-theoretic explanation of the potential impact of grammatical person does not only capture the rationale behind the *extralinguistic hypotheses*, but it can also account for Auer's (1993) *inflectional hypothesis* and a potential role of verbal inflection in general. In Section 6.3, I discussed how processing an utterance with topic drop requires additional effort on the verb because the hearer needs not only to process the verb but also to resolve the ellipsis. If this effort is too high, i.e., if there is an information peak exceeding the channel capacity, the processing capacities of the hearer are overburdened resulting in processing difficulties. I argued that the effort associated with recovering an omitted subject could be reduced if the following congruent verb is distinctly marked for inflection. If it is marked for the 1st or the 2nd person, the referent of topic drop becomes evident because the roles of the speaker(s) and hearer(s) are uniquely

defined in the utterance situation. In contrast, the set of possible referents is only narrowed for the 3rd person. This difference can explain why topic drop of the 1st person (and of the 2nd person, see experiments 9 and 10 in Sections 10.5.4 and 10.5.5) is more felicitous than topic drop of a 3rd person pronoun that does not refer to the most salient proposition in the current discourse.

10.4 Corpus study of grammatical person and verbal inflection

As mentioned above, I investigated the influence of the grammatical person on topic drop with a corpus study and a series of experiments. In this section, I present the corpus results, more specifically a descriptive overview of grammatical person in the FraC-TD-Comp and the FraC-TD-SMS data sets, which were derived from the fragment corpus FraC (see Section 7.2.1) and the results of the logistic regression analysis that I performed on the FraC-TD-SMS-Part data set.[22] In the regression analysis, I furthermore considered information about the verb following topic drop including its inflectional ending, and interpreted the results also with respect to ambiguity avoidance. I investigated whether topic drop of the 1st person singular subject pronoun *ich* is indeed particularly frequent and whether such a preference could then be explained by the *inflectional hypothesis*, i.e., whether the rate of topic drop increases if the verb in the left bracket has a distinct inflectional marking.

10.4.1 Grammatical person in FraC-TD-Comp

Table 10.4 shows the grammatical person, number, and gender of the subjects in the prefield in the FraC-TD-Comp data set, which consists of all instances of topic drop and the corresponding full forms in the FraC (see Section 7.2.3.1).[23] The

[22] The corpus data and the analysis scripts can be found online: https://osf.io/zh7tr. For copyright reasons, I can only provide the IDs and the annotations of each instance but not the actual linguistic material.

[23] I follow Zifonun (1995: 43) in considering the anaphor *es* to be of neuter gender. This means that all occurrences of the expletive subject *es* in this table and the following tables in this chapter are contained within the total number of 3rd person singular instances. With Eisenberg (2002: 68), however, it should be noted that already in the case of the pronominal *es* the neuter gender is unmarked compared to the masculine and the feminine. It can even be used when the antecedent is, for example, a sentence and, thus, cannot be a carrier of gender congruence (Eisenberg 2002: 68). For expletive *es* occurrences, this reasoning should apply even more strongly since there is no antecedent and no reference at all. The category indeterminate consists of overt and covert indefinite pronouns *man* ('one') and *alle* ('everyone') and of omitted proper names referring to products or brands whose gender is uncertain.

data suggest a correlation between the frequency of the full forms with a certain grammatical person and number combination and the corresponding instances of topic drop,[24] which I argue hints at a causal relationship. The information-theoretic account that I advocate in this book predicts that the likelihood of omitting a certain grammatical person increases with its frequency in the prefield.

Table 10.4: Full forms, instances of topic drop, and omission rates as a function of the grammatical person, number, and gender for the subjects in the FRAC-TD-COMP data set

Grammatical person, number	Full form	Topic drop	Total	Omission rate
1SG	1 077	465	1 542	30.16%
2SG	83	11	94	11.70%
3SG	1 186	278	1 464	18.99%
– 3SG masculine	137	21	158	13.29%
– 3SG feminine	131	23	154	14.94%
– 3SG neuter	825	206	1 031	19.98%
– indeterminate	93	28	121	23.14%
1PL	387	44	431	10.21%
2PL	26	0	26	0.00%
3PL	292	9	301	2.99%

Across all text types, 1st person singular subjects occur very frequently in the prefield both overtly and covertly, as well as 3rd person singular subjects, in particular those with neuter gender. The fact that most instances of subject topic drop are omissions of the 1st person singular is in line with the statements in the literature and with the results of the previous corpus studies, discussed in Section 10.2.1. Also, the observation that the number of utterances with topic drop of subject pronouns of the 3rd person singular neuter is about ten times higher than the number of utterances where a 3rd person singular masculine or feminine subject is omitted is in accord with the previous results. Of these 200 cases, 28 are non-referential expletives, about 100 are cases where the antecedent of topic

[24] I correlated the number of full forms and the number of instances of topic drop for each combination of grammatical person and number. For the 3rd person singular, I only included the subgroups obtained by considering gender and not the summed total. I obtained a Pearson's r of 0.93 ($p < 0.001$), which hints at a strong positive correlation. This value needs to be handled with care, though, since there are not many data points and in particular the high frequency of the 1st person singular seems to be the driving force for the observed effect.

Table 10.5: Full forms, instances of topic drop, and omission rates as a function of the grammatical person, number, and gender for the objects in the FraC-TD-Comp data set

Grammatical person, number	Full form	Topic drop	Total	Omission rate
1SG	0	0	0	0.00%
2SG	0	0	0	0.00%
3SG	150	61	211	28.91%
– 3SG masculine	7	6	13	46.15%
– 3SG feminine	5	0	5	0.00%
– 3SG neuter	138	55	193	28.50%
1PL	1	0	0	0.00%
2PL	0	0	0	0.00%
3PL	9	0	0	0.00%
undefined	0	5	5	100.00%

drop is a proposition or a VP that is salient in the current discourse, and in the remaining cases, the antecedent is mainly a DP denoting concrete or abstract things. This suggests that the explanation discussed in Section 10.3 that topic drop of the 3rd person singular neuter is so frequent because it targets mainly pronouns referring to salient propositions is only tentatively correct. Pronouns referring to physical or abstract entities are also frequently omitted.[25] A look at the full forms suggests that there is a similar difference in frequency per grammatical gender for the utterances with overt 3rd person singular pronouns as well, suggesting that the high proportion of the neuter gender is not a peculiarity of topic drop. I return to this issue below.

Table 10.5 shows the grammatical person, number, and gender of the object data. Of the 160 overt objects in the prefield, 150 are 3rd person singular, nine are 3rd person plural and one is 1st person plural. Of the 66 covert objects, 61 are 3rd person singular and for five, the person and number cannot be determined due to the missing precontext. As for the 3rd person singular subjects, also the vast majority of the 3rd person singular cases are of neuter gender, both in the instances of topic drop and the full forms. Most of the topic drop cases, around 50, have a proposition or a VP as their antecedent. Potentially, the explanation that

[25] Since in German neuter DPs are often, though not always, inanimate, perhaps the animacy could play a role here too (similar to object topic drop in Dutch, see Section 2.3).

topic drop of the 3rd person singular neuter is frequent because it often targets pronouns referring to salient propositions is more accurate for objects than for subjects. Further research is needed to investigate whether there is a systematic difference. While the numbers for the objects are overall relatively low, the result is in line with the observation made in Section 8.1 that topic drop of 1st and 2nd person objects is at least very infrequent. Similarly, the corresponding full forms appear to be rare, which is compatible with a frequency-based approach to explain their omission. They cannot be omitted (well) because they are not very predictable in the prefield (see Section 8.3). This argumentation has to be qualified by the fact that objects of the 3rd person singular masculine were omitted almost as often as they were realized, even though they were clearly rarer than the 3rd person singular neuter objects with only 13 occurrences in total. Thus, it seems that the frequency hypothesis (simplified: if the full form is frequent, topic drop is frequent) cannot fully explain the occurrence of object topic drop or that at least larger amounts of data are needed to make conclusive statements.

10.4.2 Grammatical person in FraC-TD-SMS

The FraC-TD-SMS data set primarily contains overt or covert subjects in the prefield (see Section 7.2.3.2). Since there are only seven objects,[26] a number too low for meaningful conclusions, I focus only on the subjects in this section. Table 10.6 shows their distribution by grammatical person.[27]

The majority of omitted or realized subjects in the prefield position are 1st person singular pronouns. Their omission rate of about 67% is the highest. This is generally in line with the previous research presented in Section 10.2.1, according to which the 1st person singular is especially often targeted by topic drop (Auer 1993), in particular in text messages (Androutsopoulos & Schmidt 2002, Döring 2002, Frick 2017), although the 1st person plural and the 3rd person singular neuter have compatible high omission rates.[28] It seems that, on the one

[26] There are two realized objects in the form of the 3rd person singular demonstrative pronoun *das* ('that'). For the five instances of topic drop, the grammatical person and number cannot be unambiguously determined due to the missing precontext.

[27] The two indeterminate cases are omissions of the indefinite pronoun *man* ('one') in the prefield.

[28] The omission rate of the 1st person singular even is about 7% higher in absolute terms in the FraC-TD-SMS data set than in the corpora of Androutsopoulos & Schmidt (2002) and Frick (2017). The difference to Androutsopoulos and Schmidt is marginally significant according to a Pearson's chi-squared test with Yates's continuity correction calculated in R (R Core Team 2021) ($\chi^2(1) = 3.11$, $p < 0.1$); the difference to Frick is significant ($\chi^2(1) = 7.98$, $p < 0.01$). (Recall that the amount of data investigated in Frick 2017 is considerably higher than that in Androutsopoulos & Schmidt 2002, yielding higher statistical power.) Also the omission rate of the

hand, pronouns referring to the speaker(s) are particularly often omitted in the FraC-TD-SMS data set.

Table 10.6: Full forms, instances of topic drop, and omission rates as a function of grammatical person, number, and gender for the subjects in the FraC-TD-SMS data set

Grammatical person, number	Full form	Topic drop	Total	Omission rate
1SG	131	264	395	66.84%
2SG	17	8	25	32.00%
3SG	29	41	70	58.57%
– 3SG masculine	5	1	6	16.67%
– 3SG feminine	5	3	8	37.50%
– 3SG neuter	17	37	54	68.52%
– indeterminate	2	0	2	0.00%
1PL	21	35	56	62.50%
2PL	1	0	1	0.00%
3PL	0	0	0	0.00%

On the other hand, pronouns of the 3rd person singular are also more often omitted than realized. Of the 41 instances of 3rd person singular subject topic drop in the data set, 37 are neuter. Their omission rate of 68.52% is even higher than the rate of the 1st person singular. Of these neuter pronouns, 11 are expletives and 17 refer to propositions or VPs in the previous discourse, while the remaining 9 refer to entities. This indicates, similar to the results from the FraC-TD-Comp data set, that a 3rd person singular pronoun is indeed often omitted when it refers to a salient proposition, as discussed in Sections 10.2.1 and 10.3, but it is by no means restricted to only these cases. Additionally, for the 17 neuter full forms a

3rd person singular with about 59% is slightly higher than the corresponding rates by Androutsopoulos and Schmidt (51% for *es* ('it') and *das* ('that'); not significant ($\chi^2(1) = 0.28, p > 0.5$)), and Frick (41%, significant ($\chi^2(1) = 6.69, p < 0.01$)). The rate of about 63% for the 1st person plural is significantly higher than the 30% in Androutsopoulos & Schmidt (2002) ($\chi^2(1) = 8.6, p < 0.01$) and the 20% in Frick (2017) ($\chi^2(1) = 43.08, p < 0.001$). With 32% the omission rate of the 2nd person singular lies between the 26% that Androutsopoulos and Schmidt report and Frick's 47%, the differences are not significant however ($\chi^2(1) = 0.103, p > 0.8$; $\chi^2(1) = 1.56, p > 0.2$). These discrepancies with the data from the previous studies cannot be readily explained. Here, a more detailed examination of the text messages contained in each data set might be required.

similar picture emerges. There are 5 expletives and only 2 neuter pronouns referring back to an entity, while the remaining neuter pronouns are either clearly (3 instances) or presumably (7 instances) referring back to a proposition or a VP. To detect a potential difference between topic drop and full forms in terms of how frequently the overt or covert prefield element refers to propositions or VPs, a larger data set is needed.

10.4.3 Inflectional ending in FraC-TD-SMS-Part

In the FraC-TD-SMS-Part data set, which contains all utterances from the text message subcorpus with an overt or covert 1st and 3rd person singular subject pronoun in preverbal position (see Section 7.2.3.3 and the repeated Table 10.7, for an overview), I investigated not only the role of the grammatical person of the subjects but also the role of the inflectional ending on the verb in the left bracket.

Table 10.7: Full forms, instances of topic drop, and omission rates as a function of grammatical person in the FraC-TD-SMS-Part data set (repeated from page 191)

Grammatical person	Full form	Topic drop	Total	Omission rate
1SG	131	264	395	66.84%
3SG	29	41	70	58.57%
Total	160	305	465	65.59%

For each instance in FraC-TD-SMS-Part, I manually annotated whether the verb in the left bracket following a realized or omitted subject is distinctly marked for inflection or whether the verb form is syncretic. I pursued a strict approach so that any verb form that could have more than one meaning/function was annotated as syncretic, regardless of how frequent/common the forms are respectively. The results of this annotation process are shown in Table 10.8.

In absolute terms, there are 376 syncretic verb forms and 89 distinct ones. The omission rate is higher before distinctly marked verb forms (71.91%) than before syncretic forms (64.1%). Nevertheless, the subject is more often omitted than realized before syncretic verb forms. This observation speaks against the literal, i.e., categorical version of Auer's (1993) *inflectional hypothesis* because topic drop also frequently occurs before verb forms that do not indicate the grammatical person and number of the omitted subject through their inflection. However, it is in line

Table 10.8: Full forms, instances of topic drop, and omission rates as a function of inflectional ending in the FRAC-TD-SMS data set according to a strict annotation

Verb form	Full form	Topic drop	Total	Omission rate
Syncretic	135	241	376	64.10%
– 1SG and 3SG	44	71	115	61.74%
– 1SG and IMP.SG	84	157	241	65.15%
– 3SG and 2PL	7	13	20	65.00%
Distinct	25	64	89	71.91%

with a gradual variant and with the information-theoretic *facilitate recovery* principle because the preference for topic drop seems to be higher before distinctly marked verb forms.

In FRAC-TD-SMS-PART, there are three relevant types of syncretisms, which I annotated manually: (i) The first one is the form equivalence between the 1st and 3rd person singular in the preterite and for preterite-present verbs also in the present tense, which I discussed in Section 10.1.1. This concerns 115 verb forms. In the majority of the cases (35 or 79.55% for the full forms, 57 or 80.28% for the utterances with topic drop), the intended form is that of the 1st person singular. (ii) The second type is the syncretism between the 1st person singular present tense and the imperative singular discussed in Section 10.1.2. It contributes the largest proportion of syncretic forms with 241 instances (84 full forms, 157 utterances with topic drop). Their omission rate of about 65% is the highest (together with that for the syncretism between the 3rd person singular and the 2nd person plural), which suggests that this syncretism is especially unproblematic for topic drop. This hypothesis is supported by the fact that besides the 241 reported instances of this type of syncretism that are indicative forms, there are additionally 80 imperative occurrences, of which 77 are syncretic with the indicative forms.[29] This sheer amount of 234 sentence-initial verb forms (157 instances of topic drop and 77 syncretic imperative forms) that are ambiguous between the imperative singular and the 1st person singular indicative present tense suggests indeed that writers seem to consider this ambiguity relatively unproblematic. However, this has to be qualified by the fact that many of the 1st person singular verb forms

[29] I manually searched the complete text message subcorpus for imperative singular forms at the left edge of the utterances. The search revealed 7 ambiguous forms and 3 clear imperative forms (*vergiss* ('forget'), *nimm* ('take'), and *gib* ('give')).

that are formally identical to imperative forms are not very common as an imperative, such as *komme* ('come'), which is usually formed without the final schwa as *komm*, or are not very plausible as a demand, such as *brauch*(*e*) ('need'). Therefore, it seems useful, in particular for this group of syncretic verb forms, to distinguish between the theoretical form equality and the ambiguity present in practice, as suggested below. (iii) As a third type, I considered a previously unmentioned form equivalence for verbs without stem vowel alternation between the 3rd person singular present tense and the 2nd person plural present tense or the imperative plural, e.g., *er geht* ('he goes') vs. *ihr geht* ('you go') vs. *Geht!* ('Go!'). With only 20 instances (7 full forms, 13 utterances of topic drop), this group is the rarest syncretism in FraC-TD-SMS-Part but together with type (ii) the one with the highest omission rate (however, this rate is less expressive for only 20 instances). Given the fact that the 2nd person plural generally occurs very rarely in the text message subcorpus of the FraC, as does the imperative plural, this syncretism may not be problematic. It may be that this type of potentially ambiguous verb form behaves more like a distinct form. That is, the hearer simply interprets it as the a priori more likely 3rd person singular form, and only realizes that there is an ambiguity when they are forced to reanalyze the structure because the subsequent context does no longer fit with their initial interpretation. The data of these three types of syncretisms as a result of the strict annotation are used in the regression analysis, which is presented in Section 10.4.4.

The inflection predictor also includes an "informed" annotation of syncretisms to indicate which of the theoretical existing syncretisms matter in practice, i.e., which verb forms are ambiguous for a hearer or reader. To this effect, three student assistants,[30] undergraduate students of German studies from Saarland University, manually annotated all verb forms in the FraC-TD-SMS-Part data set. They read the immediately preceding context utterance, if available, and then only the verb form of the target utterance, i.e., the prefield constituent was removed if originally present, just like the rest of the utterance. This was intended to simulate incremental parsing at the verb, i.e., to mimic a hearer who processes the verb of an utterance with topic drop in context. The student assistants' task was then to annotate this verb form for grammatical person, number, verb mode, and tense. They were told that it is possible to indicate as a comment that there was more than one possible reading for the verb form, but it was not spelled out explicitly that they should pay special attention to potential ambiguities. I considered every verb form as syncretic in the sense of this informed annotation for

[30] I thank Annika Schäfer, Ricarda Scherer, and the third person who wished to remain anonymous.

which an alternative analysis was proposed by at least one of the three student assistants. Table 10.9 shows the distribution of syncretic and distinct forms across full forms and utterances with topic drop for both the strict and the informed annotation in comparison.

Table 10.9: Full forms, instances of topic drop, and omission rates as a function of inflectional ending in the FRAC-TD-SMS data set according to a strict vs. an informed annotation

Verb form	Full form	Topic drop	Total	Omission rate
Distinct (strict)	25	64	89	71.91%
Syncretic (strict)	135	241	376	64.10%
Total	160	305	465	65.59%

Verb form	Full form	Topic drop	Total	Omission rate
Distinct (informed)	87	187	274	68.25%
Syncretic (informed)	73	118	191	61.78%
Total	160	305	465	65.59%

The total number of informed syncretic forms (191) is significantly lower than the total number of strictly syncretic forms (376) ($\chi^2(1) = 152.98$, $p < 0.001$). A closer inspection reveals that the former are a proper subset of the latter. Additionally, the omission rates before the syncretic verb forms are lower than before the distinct verb forms for both the strict and the informed annotation. This suggests that distinct inflection may indeed play a role in how topic drop is used. In the informed annotation, the majority of the recognized syncretisms are cases of ambiguity between the 1st person singular and the imperative singular (115), which is also the largest group in the strict annotation, followed by the syncretism between the 1st and the 3rd person singular (73),[31] whereas there are only 3 cases of the syncretism between 3rd person singular and 2nd person plural. It seems that in most cases, a reader notices whether a verb form is syncretic, especially for

[31]Interestingly, one annotator consistently provided as alternatives for these syncretisms a declarative utterance with topic drop of the 1st person singular and an interrogative utterance with an overt 3rd person singular subject in the middle field. This can be interpreted as anecdotal evidence that topic drop is considered to be a more likely alternative for the 1st than for the 3rd person singular.

the 1st person singular and the imperative. In these cases, as Table 10.9 suggests, the syncretism and the resulting ambiguity lead writers to use topic drop less frequently, but with over 60% omission rate, topic drop is still clearly preferred over the full form. In the next section, I investigate whether these descriptively observed tendencies are confirmed in a statistical analysis.

10.4.4 Grammatical person and inflectional ending in FraC-TD-SMS-Part – logistic regression analysis

As the centerpiece of my corpus study, I conducted a logistic regression analysis on the FraC-TD-SMS-Part data set. I investigated whether the likelihood of subject topic drop of the 1st and the 3rd person singular varies as a function of grammatical person, inflectional ending, verb surprisal, and verb type. In this section, I outline the analysis but focus only on grammatical person and inflection in the interpretation. The other factors are discussed in the corresponding Chapter 11, namely in Section 11.4.4.

Based on 305 utterances with topic drop and 160 full forms, I predicted the likelihood of the binary dependent variable Completeness (full form vs. topic drop) from the independent variables grammatical Person, Inflection, verb Surprisal, and Verb Type in a logistic regression analysis. Person was a binary predictor distinguishing between the 1st and the 3rd person singular. Inflection had three levels: *distinct*, *strictly syncretic*, and *informed syncretic*. *Strictly syncretic* accounted for every theoretically possible form of syncretism, while *informed syncretic* considered only those cases where the ambiguity resulting from the syncretism mattered for at least one of three annotators, as described in Section 10.4.3. I used forward coding to create two contrasts for this predictor: Inflection Distinct compared *distinct* to both *strictly* and *informed syncretic* to see whether any form of syncretism is of relevance, while Inflection Informed compared *informed syncretic* to *distinct* and *strictly syncretic* to look at a potential difference between theoretically and practically relevant ambiguity. Surprisal was a numeric predictor for the unigram surprisal per verb lemma of the verb following topic drop. The binary predictor Verb Type encoded whether this verb belonged either to the classes copular, auxiliary, or modal verb or to the classes lexical or reflexive verbs (see Section 11.4.4 for details).

10.4.4.1 Predictions

From the isolated hypotheses developed in the theoretical literature discussed in Section 10.1.1 and from the predictions of my unifying information-theoretic

account presented in Section 10.3, I derive the following predictions concerning the likelihood of topic drop as a function of grammatical person and verbal inflection (see Table 10.10 for an overview and Section 11.4.4.1 for the other two factors).

Table 10.10: Predictions for grammatical person and inflection to be tested on FraC-TD-SMS-Part: a checkmark indicates an expected effect on the likelihood of topic drop.

	Information-theoretic account	Theoretical literature		Ambiguity avoidance
		Inflectional hypothesis	*Extralinguistic hypotheses*	
(i–Person)	✓	(✓)	✓	–
(ii–Inflection)	✓	✓	–	✓

(i–Person) My information-theoretic account predicts a significant main effect of Person, according to which topic drop should more frequently target the 1st person singular than the 3rd person singular. This effect is the result of the increased predictability of the 1st person singular that feeds on two sources. First, the reference of an omitted 1st person singular pronoun can be uniquely determined in the text type text messages as the writer of the message, whereas for an omitted 3rd person singular constituent, there are usually several potential referents. This means that the probability mass is usually split among several potential referents in the 3rd person singular case, whereas it is completely concentrated on the speaker/writer in the 1st person singular case. Second, the predicted effect is a consequence of general frequency considerations. The writer is not only uniquely determinable, but they are also very prominent in text messages and often text about themselves. This, in turn, results in the 1st person singular pronoun *ich* ('I') occurring more frequently in text messages than 3rd person singular pronouns, as discussed in Sections 10.2.1 and 10.4. Therefore, *ich* ('I') becomes highly predictable and is very likely to be omitted.

In this reasoning, the information-theoretic account draws on what I termed *extralinguistic hypotheses* as one source of the predictability of the prefield constituent. These *extralinguistic hypotheses* likewise predict a significant main effect of Person. The idea is that the information from the extralinguistic context such as world knowledge, knowledge about the components of a communication situation, as well as text type knowledge facilitate the reconstruction and, thus, the omission of the 1st and, possibly, also the 2nd person. In the case of text messages, it is, as mentioned above, the knowledge that in a text messaging setting

there is usually exactly one writer of a text message and that this writer often texts about themselves that serves as an explanation for the higher frequency of 1st person singular subject topic drop.

Partly, also Auer's (1993) *inflectional hypothesis* predicts an effect of PERSON but only in interaction with INFLECTION. Topic drop of the 1st person singular should be more frequent before verbs that are explicitly marked for inflection but not (necessarily) before syncretic verb forms.

(ii–INFLECTION) However, if we take Auer's (1993) *inflectional hypothesis* literally,[32] it can be generalized to mean that topic drop should generally be possible only (or at least be more frequent) if the verb form in the left bracket is not syncretic but uniquely marked for inflection. This predicts not an interaction between PERSON and INFLECTION, as stated in the last paragraph, but a main effect of INFLECTION. If all that matters is being able to derive the grammatical person from the verb form, the relevant annotation should be the strict one, i.e., the main effect should be visible at the new contrast INFLECTION DISTINCT that compares distinct to strictly and informed syncretic verb forms.

While the *extralinguistic hypotheses* do not make specific predictions for verbal inflection, my information-theoretic account does. It likewise predicts a main effect of INFLECTION according to which topic drop should be more frequent if the following verb is distinctly marked. Following from the *facilitate recovery* principle, discussed in Section 6.3, such a distinct inflectional ending should reduce the overall processing effort on the verb because it provides information about the omitted subject. Thus, it should increase the likelihood of topic drop. For the information-theoretic account, it may matter how likely the two or more alternative meanings of a syncretic form are. That is, a syncretic form that has one very likely and one (or more) very unlikely meaning(s) may cause less processing effort if the intended meaning is the likely one than a syncretic form where the competing meanings are approximately equally likely. Therefore, it may not be the contrast INFLECTION DISTINCT for which the significant main effect is expected but rather INFLECTION INFORMED, where the distinct verb forms and the forms that are only theoretically syncretic are pooled and jointly compared to the informed syncretic verb forms.

A main effect of inflection is also predicted by the principle of ambiguity avoidance. Topic drop should be less frequent if the following verb is not distinctly

[32] "Die Morphologie des Deutschen ist im Singular des Präsensparadigmas noch differenziert genug, um auch ohne pronominale Markierung die Person flektivisch ausdrücken zu können" (My translation: 'The morphology of German is still differentiated enough in the singular of the present tense paradigm to be able to express the person inflectively even without pronominal marking') (Auer 1993: 198).

marked for inflection because syncretic verb forms lead to ambiguous utterances and potentially increased processing effort for the hearer (as described in Section 10.1.2). As for the information-theoretic account, it might also matter from the perspective of ambiguity avoidance how likely the two or more different meanings of a syncretic verb form are. That is, a speaker might be more likely to avoid ambiguity if the meanings are approximately likely than if the meaning they intend to communicate is the most likely one anyway. Consequently, it is rather Inflection Informed than Inflection Distinct that is predicted to have a significant effect on the likelihood of topic drop according to ambiguity avoidance.

10.4.4.2 Results

I conducted the logistic regression analysis using general linear models with family binomial in R (R Core Team 2021). I predicted the likelihood of topic drop from the binary independent variables grammatical Person (1SG vs. 3SG) and Verb Type (copular, auxiliary, or modal verbs vs. lexical or reflexive verbs), which were coded using deviation coding (1SG and lexical or reflexive verb were coded as 0.5, 3SG and copular, auxiliary, or modal verb as −0.5 respectively), from the numeric unigram Surprisal of the verb following topic drop and from Inflection, as well as from all the two-way interactions between the predictors. Inflection had three levels (distinct vs. strictly syncretic vs. informed syncretic) and was forward coded: For Inflection Distinct *distinct* was coded as $^2/_3$, *strictly syncretic* and *informed syncretic* as $-^1/_3$ respectively. For Inflection Informed *informed syncretic* was coded as $-^2/_3$, *distinct* and *strictly syncretic* as $^1/_3$ respectively.[33] The fixed effects in the final model, which I obtained with a backward model selection, as described in Section 3.1.6.3, are presented in Table 10.11.

There was a significant interaction between grammatical Person and Inflection Informed ($\chi^2(1) = 5.85$, $p < 0.05$). The likelihood of topic drop is increased for the 1st person singular before practically unambiguous verb forms (see Figure 10.1).

The significant interaction between Person and Verb Type ($\chi^2(1) = 4.72$, $p < 0.05$) suggests furthermore that topic drop is more likely if the covert constituent is a 1st person singular pronoun and if, at the same time, the following verb is a copular, auxiliary, or modal verb. There is a significant main effect of Verb Type ($\chi^2(1) = 6.3$, $p < 0.05$), which runs contrary to the interaction: Accordingly, topic

[33]The formula of the full model was as follows: `Completeness ~ (Person + Surprisal + Verb Type + Inflection Distinct)^2 + (Person + Surprisal + Verb Type + Inflection Informed)^2.`

Table 10.11: Fixed effects in the final GLM analyzing FRAC-TD-SMS-PART

Fixed effect	Est.	SE	χ^2	p-value	
INTERCEPT	2.81	0.57	27.20	< 0.001	***
PERSON	0.35	0.29	1.46	> 0.2	
SURPRISAL	−0.27	0.07	17.85	< 0.001	***
VERB TYPE	0.95	0.39	6.30	< 0.05	*
INFLECTION INFORMED	−0.33	0.31	1.15	> 0.2	
PERSON × VERB TYPE	−1.30	0.61	4.72	< 0.05	*
PERSON × INFLECTION INFORMED	1.47	0.63	5.85	< 0.05	*

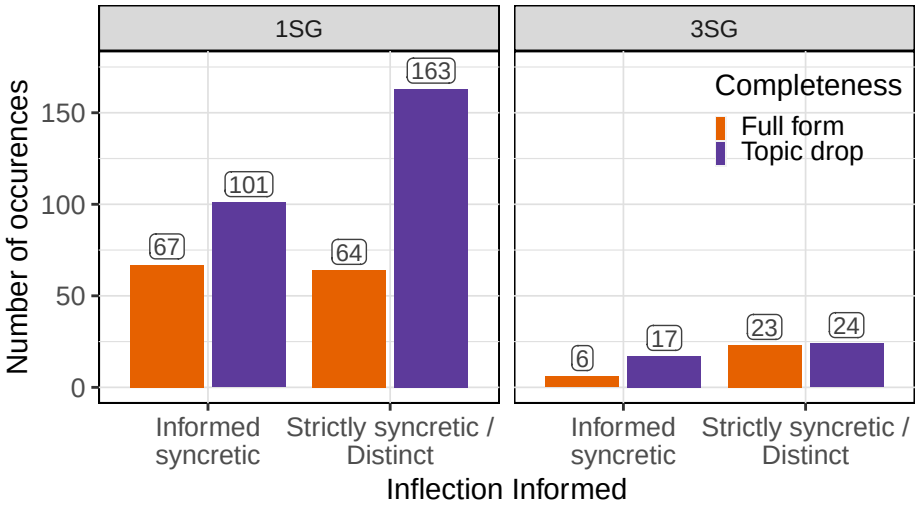


Figure 10.1: Frequency of the full forms and the instances of topic drop as a function of grammatical PERSON and INFLECTION INFORMED in the FRAC-TD-SMS-PART data set

drop is less likely before copular, auxiliary, or modal verbs. Finally, the significant main effect of verb SURPRISAL ($\chi^2(1) = 17.85$, $p < 0.001$) in the negative direction suggests that the likelihood of topic drop decreases with a higher verb surprisal. All other main effects and interactions were not significant (all $p > 0.2$).

10.4.4.3 Discussion

The logistic regression analysis revealed two significant interactions in which grammatical person participated. The results suggest that topic drop of the 1st person singular is more likely than topic drop of the 3rd person singular if (i) the following verb is not ambiguous to the hearer and (ii) if this verb is a copular, auxiliary, or modal verb. While I turn to result (ii) in Section 11.4.4.3, result (i) is in line with the narrower interpretation of Auer's (1993) *inflectional hypothesis*: A 1st person singular subject is more likely to be omitted if the following verb relatively unambiguously indicates the grammatical person of the congruent subject. When discussing prediction (ii–INFLECTION) above, I argued that if we take Auer's (1993) *inflectional hypothesis* literally, we would actually expect not an interaction but a main effect of INFLECTION. Topic drop should be more likely whenever the following verb clearly indicates the grammatical person of the omitted subject. Such a main effect was also predicted by my information-theoretic account and by ambiguity avoidance, but it was not significant in my analysis.

An explanation for why there was an interaction with grammatical person instead of the predicted main effect may be found if a semantic perspective is included on how the referent of the omitted prefield constituent is ultimately determined. While the *inflectional hypothesis* comes from a purely formal perspective, i.e., a distinct inflectional ending results in being able to recover the omitted subject, it disregards an important difference between the 1st and 2nd person on the one hand and the 3rd person on the other. For the 1st and the 2nd person (singular), the referent of the omitted constituent can be identified as the speaker or hearer, whereas there are usually several candidates for the referent of an omitted 3rd person singular subject, i.e., any gender-matching entity present in the current discourse or situation. Consequently, using (relatively) unambiguous verb forms may be particularly beneficial for the 1st person singular. In information-theoretic terms: It reduces the processing effort associated with ellipsis resolution more strongly as it allows the hearer to directly identify the intended referent, as discussed in Section 6.3.

The results also suggest a preference for topic drop of the 1st person as predicted by (i–PERSON) and the *extralinguistic hypotheses*, as well as by my informa-

tion-theoretic account. The 1st person singular as the pronoun referring to the speaker is generally more frequent in the text message subcorpus and, thus, more predictable and more easily recoverable. The fact that this preference for the 1st person is only reflected in the two interactions and not in an additional main effect is most likely due to the imbalanced data set, which contains generally a lot fewer instances of the 3rd person singular than the 1st person singular, in combination with the deviation coding, where "the effect of each factor is coded to reflect how it influences the DV while collapsing across any other factors" (Brehm & Alday 2022: 3).

10.5 Experimental investigations of grammatical person and verbal inflection

The results of the corpus study presented in Section 10.4 suggest that grammatical person and verbal inflection influence why and when speakers use topic drop. In the following, I present the results of five studies that aimed to evidence the relevance of these two factors also experimentally.

I start this section with a further discussion of the three acceptability rating studies that I outlined already in Chapter 9 on topicality (see Section 9.2). Experiments 5, 7, and 8 investigated the impact of grammatical person and topicality on the acceptability of topic drop. In this section, I focus on the results related to grammatical person and verbal inflection. More specifically, by comparing the 1st and the 3rd person singular in these three experiments, I investigated whether the difference in frequency evidenced in the corpus study is reflected in acceptability. Additionally, since experiments 7 and 8 were identical except for the use of lexical verbs with a distinct inflectional marking (experiment 7) vs. syncretic forms of modal verbs (experiment 8), a comparison between both studies can also provide insights into the role of verbal inflection.

In the second part of this section, I turn to experiments 9 and 10, with which I extended the investigation of grammatical person. I compared the 1st and 2nd person singular to distinguish between the two types of *extralinguistic hypotheses* that aim to explain a preference for topic drop of only the 1st or of both the 1st and the 2nd person singular. While experiment 9 tested this with the established instant messages design, in experiment 10 the items were presented as if they were spoken. This also allows for a first comparison of topic drop between different communication forms.

10.5.1 Experiment 5: grammatical person (1SG vs. 3SG)

In this section, I return to experiment 5 to discuss the results for grammatical person. In Section 9.2.1, I described in detail the design of this acceptability rating study, as well as the stimuli and the analysis. Here, I only repeat the most central points.[34]

Experiment 5 investigated not only the role of topicality using a question method to set the discourse topic but also the influence of grammatical person on topic drop. It compared utterances with the overt or covert 1st person singular pronoun *ich* ('I') in the prefield to corresponding utterances with an overt or covert 3rd person singular subject pronoun referring to a person. As discussed in Sections 10.1.1 and 10.2.1, the previous literature found a preference for omitting the 1st person singular subject pronoun, which was confirmed in several corpus studies, in particular of text messages. I was able to add further evidence to this by showing that also in the text message subcorpus of the FraC, the 1st person singular was omitted particularly often (see Sections 10.4.2 and 10.4.4). In the existing literature, this preference was explained by an *inflectional hypothesis* (Auer 1993) or two types of *extralinguistic hypotheses* (e.g., Zifonun et al. 1997, Volodina & Onea 2012), which trace back the fact that the 1st person or the 1st and the 2nd person can be better omitted to text type knowledge and fixed speaker and hearer roles (see Section 10.1.1). In Section 10.3, I discussed that the information-theoretic account can capture all three types of hypotheses and explains the potential effect of grammatical person in terms of predictability and reducing processing effort on the verb following topic drop.

Experiment 5 had the form of a $2 \times 2 \times 2$ within-subjects design with the three binary predictors Topicality, grammatical Person, and Completeness. For Person, I expected to find an interaction with Completeness to the effect that topic drop of the 1st person singular subject pronoun should be more acceptable than topic drop of a 3rd person singular subject pronoun. This result would show that the frequency differences found in the corpus, i.e., speaker-sided production preferences, are reflected in acceptability differences, i.e., hearer-sided preferences.

10.5.1.1 Materials

The structure of the 24 items and the fillers has been presented in Section 9.2.1.1.

[34] All items, fillers, and the analysis script can be found online: https://osf.io/zh7tr.

10.5.1.2 Procedure

The procedure of the experiment with 48 participants is described in Section 9.2.1.2.

10.5.1.3 Results

I repeat the descriptive statistics in Table 10.12 and Figure 10.2.

Table 10.12: Mean ratings and standard deviations per condition for experiment 5 (repeated from page 230)

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.48	1.61
Topic drop	1SG	Topic	4.76	1.87
Full form	3SG	Topic	5.71	1.45
Topic drop	3SG	Topic	4.13	1.88
Full form	1SG	No topic	5.45	1.69
Topic drop	1SG	No topic	4.90	1.69
Full form	3SG	No topic	5.60	1.58
Topic drop	3SG	No topic	3.71	1.88

They show the mean ratings based on the data from 42 participants. From visual inspection, there seems to be a difference for grammatical person. Topic drop of the 3rd person singular seems to be less acceptable than topic drop of the 1st person singular.

As described in Section 9.2.1.3, the final CLMM (repeated as Table 10.13) contained significant main effects of the two predictors Completeness ($\chi^2(1) = 26.67, p < 0.001$) and Person ($\chi^2(1) = 5.30, p < 0.05$) and a significant interaction between Completeness and Person ($\chi^2(1) = 15.07, p < 0.001$). Full forms were significantly preferred over utterances with topic drop. Utterances with an overt or covert 1st person singular pronoun in the prefield received better ratings than corresponding 3rd person utterances. Utterances in which a 3rd person singular constituent was targeted by topic drop were especially degraded.

10.5.1.4 Discussion

Experiment 5 was designed to examine the effects of topicality and grammatical person on the acceptability of topic drop. For grammatical person, it provided

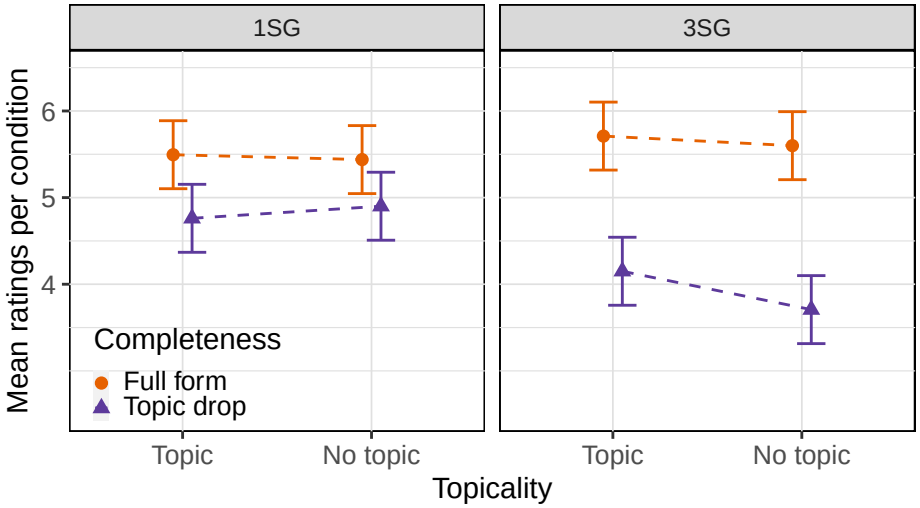


Figure 10.2: Mean ratings and 95% confidence intervals per condition for experiment 5 (repeated from page 231)

Table 10.13: Fixed effects in the final CLMM of experiment 5 (repeated from page 231)

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	2.03	0.34	26.67	< 0.001	***
Person	0.55	0.23	5.30	< 0.05	*
Completeness × Person	−1.57	0.34	15.07	< 0.001	***

evidence that, although participants generally rated utterances with a 1st person subject as more natural than utterances with a 3rd person subject, this preference was particularly strong for topic drop. That is, topic drop of the 1st person singular subject pronoun *ich* ('I') was perceived as more natural than topic drop of a 3rd person singular subject constituent referring to another human referent in the discourse. This result is in line with previous claims from the theoretical literature, with the corpus results by Auer (1993), Androutsopoulos & Schmidt (2002), Frick (2017), and with my own corpus results (see Section 10.4). The production preference for topic drop with respect to grammatical person is also reflected in the addressees' preferences during perception. This can be captured by the information-theoretic account. The 1st person singular subject pronoun is highly predictable in instant messages through text type knowledge and gen-

eral frequency. The writers of instant messages are just like the writers of text messages clearly identifiable and often text about themselves, which results in *ich* ('I') being very frequent and very predictable. According to the *avoid troughs* principle, this enhanced predictability increases the preference to omit the *ich*.

Since topic drop was always followed by a verb whose inflectional marking allowed readers to distinguish between the 1st and 3rd person singular, the result of this experiment is generally also in line with Auer's (1993) *inflectional hypothesis*. Topic drop of the 1st person singular could be more acceptable because the verb's inflectional ending indicated the grammatical person. However, the 3rd person is just as clearly recognizable purely from the verb ending, but it can nevertheless not be omitted just as well. It seems that the *inflectional hypothesis* must be supplemented at least by *extralinguistic hypotheses* or a processing account to explain the data (as I did similarly in the corpus study, see Section 10.4.4.3). There and in Section 6.3, I argued that due to the uniqueness of the hearer and speaker roles in a conversational situation and certain text types such as text messages, it is quite clear for the 1st person singular what the (covert or overt) pronoun refers to. In contrast, the 3rd person pronoun can theoretically refer to a variety of things and people in the extralinguistic context and the wider linguistic context, even if there is only one matching linguistic antecedent present in the immediate precontext, as was the case in experiment 5. This way, an inflectional ending that distinctly encodes the 1st person singular can help to recover the omitted subject to a larger extent than a 3rd person singular ending and, therefore, reduce the processing effort associated with ellipsis resolution more strongly (see the *facilitate recovery* principle in Section 6.3).

10.5.2 Experiment 7: grammatical person (1SG vs. 3SG, lexical verbs)

In Section 9.2.3, I presented the details about experiment 7, which was again a 2 × 2 × 2 acceptability rating study crossing Completeness (full form vs. topic drop), grammatical Person (1SG vs. 3SG), and Topicality (topic vs. no topic).[35]

There, I discussed its results for topicality, which was manipulated through the subject function. In this section, I return to the study, but this time, I focus on grammatical person. The goal was to replicate the result of experiment 5 that evidenced a preference for topic drop of the 1st person singular.

[35]The items and fillers, as well as the analysis script is accessible online: https://osf.io/zh7tr.

10.5.2.1 Materials

I described the 24 items and the fillers in detail in Section 9.2.3.2. Recall that the verb following the overt or covert subject in the prefield was a lexical verb in present tense and, therefore, distinctly marked for grammatical person.

10.5.2.2 Procedure

Information on the procedure and the 48 participants can be found in Section 9.2.3.3.

10.5.2.3 Results

For convenience, I repeat Table 9.5 as Table 10.14 and Figure 9.3 as Figure 10.3 from Section 9.2.3.4 to show again the descriptive statistics based on the rating data from 43 participants. From visual inspection, it seems that topic drop of the 1st person singular was rated better than topic drop of the 3rd person singular.

Table 10.14: Mean ratings and standard deviations per condition for experiment 7 (repeated from page 242)

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.64	1.44
Topic drop	1SG	Topic	5.38	1.40
Full form	3SG	Topic	5.58	1.43
Topic drop	3SG	Topic	4.65	1.60
Full form	1SG	No topic	5.58	1.37
Topic drop	1SG	No topic	4.91	1.46
Full form	3SG	No topic	5.67	1.45
Topic drop	3SG	No topic	4.21	1.67

A detailed description of my analysis of the rating data with CLMMs can be found in Section 9.2.3.4. The final model (repeated as Table 10.15) contained significant main effects of all three predictors Completeness ($\chi^2(1) = 20.75, p < 0.001$), Person ($\chi^2 = 6.1, p < 0.05$), and Topicality ($\chi^2(1) = 6.69, p < 0.01$), a significant interaction between Completeness and Topicality ($\chi^2(1) = 5.95, p < 0.05$), and a significant interaction between Completeness and Person ($\chi^2(1) = 12.85, p < 0.001$). For grammatical person, it seems that utterances with the 3rd person singular in preverbal position were generally degraded but particularly strongly in the case of topic drop.

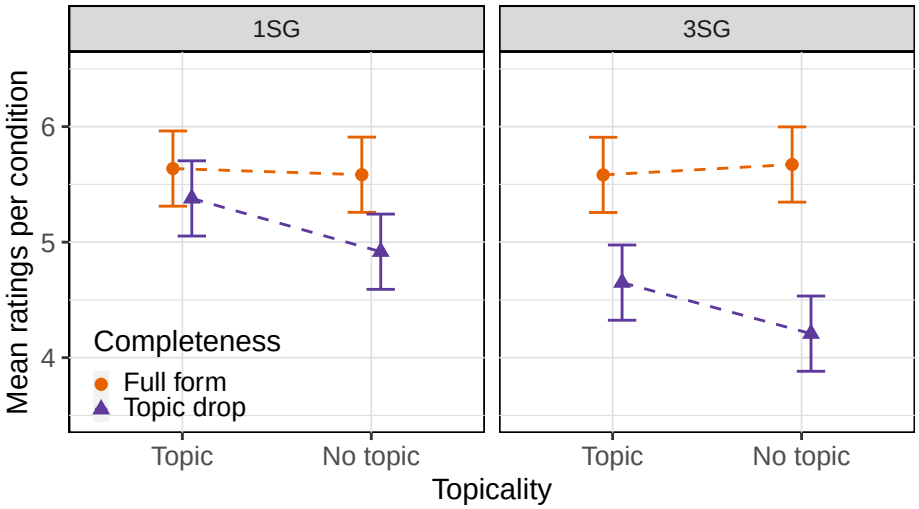


Figure 10.3: Mean ratings and 95% confidence intervals per condition for experiment 7 (repeated from page 243)

Table 10.15: Fixed effects in the final CLMM of experiment 7 (repeated from page 243)

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.72	0.35	20.75	< 0.001	***
Person	0.52	0.20	6.10	< 0.05	*
Topicality	0.38	0.15	6.69	< 0.01	**
Completeness × Person	−1.33	0.35	12.85	< 0.001	***
Completeness × Topicality	−0.81	0.33	5.95	< 0.05	*

10.5.2.4 Discussion

With experiment 7, I investigated again whether grammatical person impacts the acceptability of topic drop, focusing on the contrast between 1st and 3rd person singular subjects. Its results replicate the results for grammatical person of experiment 5: The omission of a 1st person singular pronoun was preferred over the omission of a 3rd person singular constituent. Since the verb following topic drop was likewise a lexical verb with distinct inflectional marking for grammatical person, this result is still compatible with Auer's (1993) *inflectional hypothesis*. It can also be captured by the *extralinguistic hypotheses* and by the information-theoretic account, as outlined in Section 10.5.1.4. I discuss the role of grammatical person in more detail in Section 10.5.3.4, in the context of the next experiment.

10.5.3 Experiment 8: grammatical person (1SG vs. 3SG, modal verbs)

As mentioned already in Section 9.2.4, experiment 8 replicates experiment 7 but uses modal verbs instead of lexical verbs. It was again a 2 × 2 × 2 design, which crossed Completeness, grammatical Person, and Topicality.[36]

In this experiment, I used the syncretic verb forms of modal verbs (e.g., *ich will* ('I want') and *Sabrina will* ('Sabrina wants')). Since they do not encode the grammatical person of the congruent subject, they do not help to recover this subject if it is omitted. Nevertheless, the hearers could still recover the omitted subject in this experiment because the target utterance always contained an object pronoun in the middle field referring to one of the two possible referents. This way, it was indicated that the covert subject refers to the other referent.

Considered together, experiment 7 with the distinctly marked forms of lexical verbs and experiment 8 with the syncretic form of modal verbs investigated whether the inflectional ending of the verb in the left bracket impacts the acceptability of topic drop. Such an impact is implicitly predicted by Auer's (1993) *inflectional hypothesis*. According to a narrow interpretation, there should be an interaction between Completeness, Person, and Verb Type. Topic drop of the 1st person singular should be more acceptable only if it precedes a lexical verb that is distinctly marked for grammatical person. According to a wider interpretation, any distinct inflectional marking should improve topic drop, i.e., there could be an interaction between Completeness and Verb Type. The same effect is predicted by the information-theoretic account because a distinct inflectional marking should generally reduce the processing effort on the verb following topic drop.

[36] All materials and the analysis scripts can be found online: https://osf.io/zh7tr/.

An impact of distinct vs. syncretic verb forms is also predicted by ambiguity avoidance. Utterances with ambiguous syncretic verb forms should be avoided by speakers to facilitate the hearer's processing. However, the utterances used in this experiment are not globally ambiguous but only locally. Since the disambiguation later in the sentence by the object pronoun is sufficient to avoid ambiguity, it is questionable whether the syncretic verb form needs to be avoided at all.

10.5.3.1 Materials

The adjustment of the materials is described in Section 9.2.4.1.

10.5.3.2 Procedure

Details on the procedure can be found in Section 9.2.4.2.

10.5.3.3 Results

10.5.3.3.1 Analysis of experiment 8

The mean ratings of all 48 participants are repeated in Table 10.16 and Figure 10.4. They indicate a preference for the 1st over the 3rd person singular for topic drop.

Table 10.16: Mean ratings and standard deviations per condition for experiment 8 (repeated from page 247)

Completeness	Person	Topicality	Mean rating	Standard deviation
Full form	1SG	Topic	5.41	1.55
Topic drop	1SG	Topic	5.24	1.67
Full form	3SG	Topic	5.52	1.43
Topic drop	3SG	Topic	4.43	1.93
Full form	1SG	No topic	5.35	1.67
Topic drop	1SG	No topic	4.86	1.77
Full form	3SG	No topic	5.25	1.74
Topic drop	3SG	No topic	4.27	1.77

In Section 9.2.4.3, I already outlined the statistical analysis of this experiment with CLMMs. The final model (repeated as Table 10.17) contained significant

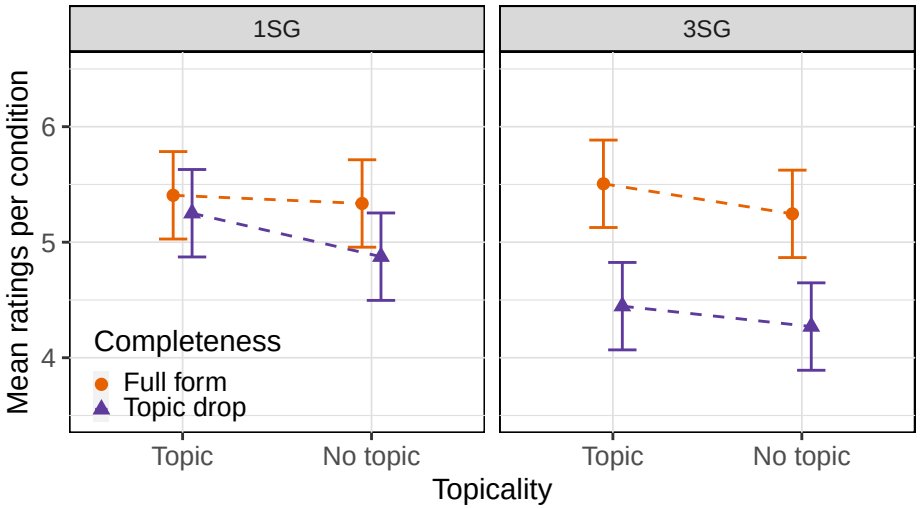


Figure 10.4: Mean ratings and 95% confidence intervals per condition for experiment 8 (repeated from page 247)

main effects of Completeness ($\chi^2(1) = 18.07, p < 0.001$) and Person ($\chi^2(1) = 10.67, p < 0.01$) and a significant Completeness × Person interaction ($\chi^2(1) = 9.43, p < 0.01$). Full forms were preferred over topic drop and utterances with the 1st person singular were preferred over utterances with the 3rd person singular. Those instances of topic drop that targeted a 3rd person singular constituent were especially degraded.

Table 10.17: Fixed effects in the final CLMM of experiment 8 (repeated from page 248)

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.08	0.23	18.07	< 0.001	***
Person	0.61	0.17	10.67	< 0.01	**
Completeness × Person	−0.98	0.29	9.43	< 0.01	**

10.5.3.3.2 Analysis of experiments 7 and 8

In Section 9.2.4.3, I presented a post hoc analysis of the combined data from experiments 7 and 8. The newly added predictor Verb Type was not involved in

any significant effect in the final model (repeated as Table 10.18). The two-way interaction between Verb Type and Completeness was not significant ($\chi^2(1) = 0.3, p > 0.5$), nor was the three-way interaction between Verb Type, Completeness, and Person ($\chi^2(1) = \chi^2(1) = 0.005, p > 0.9$).[37] The acceptability of topic drop before a syncretic form of a modal verb did not differ significantly from that of topic drop before a distinctly marked form of a lexical verb, neither in general nor for only the 1st person singular.

As mentioned above, the fixed effects in the final model were similar to those of experiment 7. What is relevant to grammatical person is the significant interaction between Completeness and Person ($\chi^2(1) = 22.24, p < 0.001$) and the significant main effect of Person ($\chi^2(1) = 8.52, p < 0.01$). There was a preference for utterances with a 1st person singular subject, in particular, in the case of topic drop.

Table 10.18: Fixed effects in the final CLMM of the joint analysis of experiments 7 and 8 (repeated from page 249)

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	1.10	0.18	30.41	< 0.001	***
Person	0.38	0.12	8.52	< 0.01	**
Topicality	0.24	0.10	5.73	< 0.05	*
Completeness × Person	−1.02	0.19	22.24	< 0.001	***
Completeness × Topicality	−0.53	0.20	6.98	< 0.01	**

10.5.3.4 Discussion

With experiment 8, I investigated whether the preference for topic drop of the 1st person singular persists if the verb is not distinctly marked for inflection. The joint post hoc analysis of experiments 7 and 8 investigated whether a distinct verbal inflection on the verb in the left bracket impacts the acceptability of topic drop.

For grammatical person, experiment 8 replicated the result of experiments 5 and 7. Topic drop of the 1st person singular was preferred over topic drop of the 3rd person singular. However, experiment 8 showed that this preference is independent of a distinct inflectional ending on the verb following topic drop. The

[37] It might be that even when collapsing the two experiments, the power was too low to find such a three-way interaction. Recall that the studies were not originally designed to find this effect.

preference was also present for topic drop before syncretic verb forms and, thus, does not hinge on the encoding of the grammatical person through a distinct inflectional verb ending. This argues against Auer's (1993) *inflectional hypothesis*, according to which the 1st person singular is particularly frequently targeted by topic drop because the following verb identifies the omitted subject through its inflectional ending.

Instead, the result supports those *extralinguistic hypotheses* that trace back the preference for topic drop of the 1st person singular to the prominence of the speaker per se and in certain text types. These hypotheses can be covered by the information-theoretic account according to which differences in likelihood affect how well the prefield constituent can be omitted, as predicted by the *avoid troughs* principle. In Sections 10.5.4 and 10.5.5, I look more closely at the two different types of *extralinguistic hypotheses* to refine the overall picture.

Concerning verbal inflection, the joint post hoc analysis of experiments 7 and 8 does not provide any evidence for a general influence of a distinct verbal inflection on topic drop. This is unexpected under the information-theoretic *facilitate recovery* principle, discussed in Section 6.3. This principle predicts that subject topic drop should be easier to process and, therefore, be more acceptable if it is followed by a distinctly marked verb form. The inflectional ending is argued to facilitate recovering the omitted subject and this way to reduce the overall processing effort on the verb. The fact that there was no difference in acceptability between topic drop before syncretic and before distinct verb forms suggests that the distinct inflectional marking does not provide a (substantial) advantage in processing that would be reflected in acceptability.

The absence of an effect of the verb type, i.e., the verbal inflection, is at first glance also unexpected from the perspective of ambiguity avoidance. The study by Soares et al. (2019), discussed in Section 10.2.2, suggested a tendency to avoid ambiguity by using an overt pronoun instead of a covert constituent before a syncretic verb form but only if there is a competition between several potential antecedents. Although such a competition was present in the items of my experiment, the ambiguous utterances with a syncretic form of a modal verb were not degraded. Unlike in Soares et al.'s (2019) study, however, the utterances were only locally ambiguous. As hypothesized already above, it is conceivable that such a local ambiguity does not need to be avoided, since the object pronoun later in the sentence allows the hearer to determine the intended meaning of the utterance with topic drop.

10.5.4 Experiment 9: grammatical person (1SG vs. 2SG)

The previously discussed experiments 5, 7, and 8 evidenced a preference for topic drop of the 1st person singular subject pronoun over topic drop of a 3rd person singular subject pronoun in line with the production preferences observed in the corpus study in Section 10.4.4. In this experiment, which was part of the same study as experiment 1, I likewise focused on grammatical person, but this time, I investigated whether there is a difference in acceptability between topic drop of 1st and 2nd person singular subjects.[38] The study had the form of a 2 × 2 design crossing grammatical Person of the preverbal subject (1st person singular vs. 2nd person singular) and Completeness (full form vs. topic drop).

10.5.4.1 Background

In Section 10.1.1, I summarized several approaches under the heading *extralinguistic hypotheses* that attempt to explain why the 1st and partly also the 2nd person singular are particularly frequently omitted. I distinguished between two types of these hypotheses: *extralinguistic 1SG* and *extralinguistic 1SG+2SG*. The central representatives of the *extralinguistic 1SG hypotheses* are Imo (2013, 2014) and Volodina & Onea (2012). Imo (2014: 153–154) claims that the 1st person singular is especially easy to recover because the speaker it refers to is part of the default origo. Volodina & Onea (2012: 218) attribute the more frequent omission of the 1st person singular, in particular in certain text types, to knowledge about these text types and their typical speakers or writers. While Imo (2013, 2014) and Volodina & Onea (2012) focus on the 1st person, the *extralinguistic 1SG+2SG hypotheses* also take the 2nd person into account. For instance, the IDS grammar (Zifonun et al. 1997: 414) argues that both the 1st and the 2nd person can be omitted well because in a shared speech situation, the roles of speaker and hearer, to which the corresponding pronouns refer, are clearly established. A similar claim is made by Ariel (1990: 47), who points out that cross-linguistically null pronouns for the 1st and the 2nd person are usually more common than null pronouns for the 3rd person. Thus, it turns out that one half of the *extralinguistic hypotheses*, *extralinguistic 1SG*, (implicitly) predicts that the omission of the 1st person should have an advantage over the omission of the 2nd person because the 1st person refers to the origo or to the speaker/writer, who is particularly prominent in some text types. The other half, *extralinguistic 1SG+2SG*, (explicitly) predicts that both persons should behave comparably since they refer to the constitutive components of any communicative situation, speaker and hearer.

[38] All items, fillers, and the analysis script are available online: https://osf.io/zh7tr.

From the quantitative corpus studies by Androutsopoulos & Schmidt (2002), Frick (2017), and myself (Section 10.4) a mixed picture emerges. While in all three studies, the 1st person singular was most often targeted by topic drop (from an omission rate of 60% in Androutsopoulos & Schmidt (2002) and Frick (2017) to a rate of 67% in my corpus study of FraC-TD-SMS), there were greater differences for the 2nd person singular. In Frick's (2017) corpus, the omission rate of the 2nd person singular was the second highest with about 47%, but the 2nd person singular was omitted only in about 35% of the cases in my data set FraC-TD-SMS (see Section 10.4.2) and in about 26% of the cases in the corpus of Androutsopoulos & Schmidt (2002). Consequently, the corpus frequencies suggest that topic drop of the 1st person singular and topic drop of the 2nd person singular do not occur with equal absolute and relative frequency, supporting the first half of the *extralinguistic hypotheses*.

Experiment 9 addressed the question of whether these production preferences are also reflected in acceptability. If this were the case, there should be a significant interaction between the predictors Completeness and grammatical Person, according to which topic drop of the 1st person singular is more acceptable than topic drop of the 2nd person singular. This would also be predicted by an information-theoretic approach that relies mainly on frequencies to predict differences in likelihood. In the other case, and if the *extralinguistic 1SG+2SG hypotheses* are correct, there should be no such interaction, i.e., topic drop of the 1st person singular and topic drop of the 2nd person singular should be comparably acceptable. This could be captured by the information-theoretic account if not only plain corpus frequencies are considered but also the predictability through the extralinguistic context and its impact on recoverability. For instance, the knowledge about the roles of the speaker and hearer allows for easily identifying and recovering the reference of an omitted 2nd person subject pronoun although *du* as a token may be less frequent in a corpus than *ich*.

10.5.4.2 Materials

Items

I constructed 24 items such as (12), which consisted of a context utterance and a target utterance, both texted by the same person. The target utterance always contained a modal verb in the left bracket, which was equally balanced between *können* ('can'), *sollen* ('shall'), and *müssen* ('must') and which had an inflectional ending that indicated the grammatical person. The utterance was either a statement about the speaker with a preverbal (realized or omitted) 1st person singular

subject (12a) or a request to the hearer with a preverbal (realized or omitted) 2nd person singular subject (12b). The context utterance was designed to make both options, i.e., the statement about the speaker and the request or advice to the hearer, as equally natural as possible.[39]

(12) *Papa kriegt es scheinbar nicht gebacken, seinen neuen Laptop*
dad gets it apparently not baked his new laptop
einzurichten🤦
set.up
'Dad can't seem to get it together to set up his new laptop🤦'

a. *(Ich) sollte ihn besser mal anrufen*
I should him better PART call
'(I) better give him a call' (1SG)

b. *(Du) solltest ihn besser mal anrufen*
you should him better PART call
'(You) better give him a call' (2SG)

Fillers

As described for experiment 1 in Section 3.1.6.1, I included 80 fillers: 24 items with an over or covert expletive subject in the prefield, 24 gapping or right node raising structures and corresponding full forms, 24 sluicing or sprouting structures and corresponding full forms, and 8 ungrammatical catch trials.

10.5.4.3 Procedure

The procedure is described in Section 3.1.6.2 for experiment 1, which was part of the same study.

10.5.4.4 Results

As described in Section 3.1.6.3, the data from 42 participants could be used for the analysis. Table 10.19 shows the mean ratings and standard deviations per condition, which are plotted with 95% confidence intervals in Figure 10.5. The visual inspection shows no differences between conditions.

I analyzed the responses with CLMMs from the package ordinal (Christensen 2019), following the procedure described in Section 3.1.6.3. I modeled the ratings

[39] A possible difference in naturalness between the PERSON conditions would also show up in the full forms that function as a baseline and, thus, would be captured by a main effect of the predictor PERSON.

Table 10.19: Mean ratings and standard deviations per condition for experiment 9

Completeness	Person	Mean rating	Standard deviation
Full form	1SG	5.90	1.21
Topic drop	1SG	5.85	1.29
Full form	2SG	5.95	1.23
Topic drop	2SG	5.94	1.26

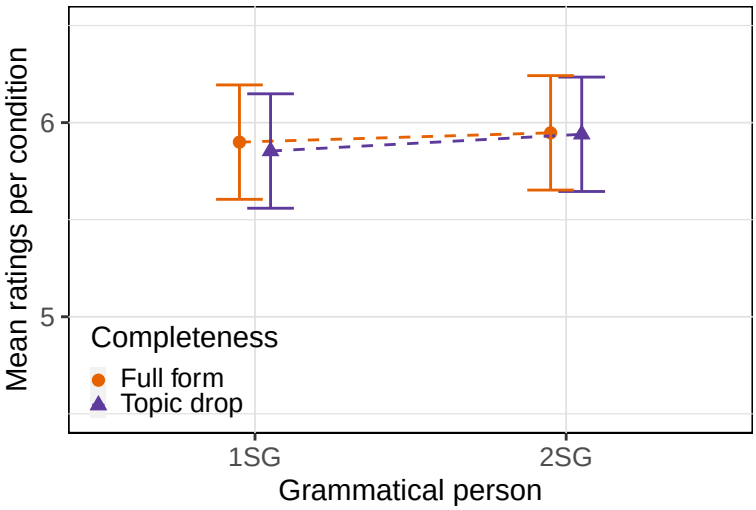


Figure 10.5: Mean ratings and 95% confidence intervals per condition for experiment 9

as a function of the binary predictors Completeness and Person, coded using deviation coding as 0.5 (full form, 1SG) and −0.5 (topic drop, 2SG), as well as of the numeric scaled and centered Position of the trial in the experiment and included all two-way interactions between the independent variables. The maximal random effects structure was included.[40]

The final model obtained with a backward model selection did not contain any significant fixed effects. Neither the interaction between Completeness and Person was significant ($\chi^2(1) = 0.57, p > 0.4$), nor the main effect of Completeness ($\chi^2(1) = 0.01, p > 0.9$), nor any of the Position effects. There was a marginally significant main effect of Person ($\chi^2(1) = 3.65, p > 0.05$) indicating that there was a general tendency for utterances with the 2nd person singular as the (realized or omitted) preverbal subject to be preferred over those utterances with the 1st person singular.

10.5.4.5 Discussion

Experiment 9 had the purpose of comparing 1st and 2nd person singular topic drop in appropriate contexts. Its results indicate that they are both equally acceptable in such contexts. This is somewhat unexpected given the corpus results by Androutsopoulos & Schmidt (2002), Frick (2017), and myself (Section 10.4). Apparently, the production preferences are not fully reflected in the perception preferences. Instead, the results are in line with the *extralinguistic 1SG+2SG hypotheses* as proposed by Zifonun et al. (1997) and Ariel (1990), which rely on the fact that the speaker and hearer are constitutive parts of any communicative situation. Therefore, the pronouns of the 1st and 2nd person singular referring to these roles are easily recoverable, and to the same degree. This result also suggests that the information-theoretic account should not rely on frequencies alone but also consider further influences on the likelihood such as the presence of the speaker and hearer in virtually any communicative situation. In the next section, I present a replication of this experiment in a different text type.

10.5.5 Experiment 10: grammatical person (1SG vs. 2SG, dialogues)

Experiment 10 replicated experiment 9 by also investigating whether there is a difference in acceptability between topic drop of the 1st and the 2nd person singular but in another text type or rather using a different form of presentation.[41]

[40]The formula of the full model was as follows: `Ratings ~ (Completeness + Person + Position)^2 + (1 + (Completeness + Person + Position)^2 | Subjects) + (1 + (Completeness + Person + Position)^2 | Items).`

[41]The materials and the analysis script can be accessed online: https://osf.io/zh7tr.

Instead of showing the critical utterance as an instant message, it was presented in quotation marks as direct speech. In this way, I wanted to validate the reliability of the result of the previous experiment. At the same time, this experiment allows for an exploratory look at possible differences between text types with regard to topic drop, as a continuation of the considerations made in Section 2.1.5. This mainly concerns the question of whether topic drop is generally perceived as less natural in spoken language than in instant messages. I do not expect any difference between text types for grammatical person because, although the communicative situation is different in talking face-to-face vs. texting, the roles of speaker or writer and hearer or reader are equally fixed and unique. The pronouns *ich* ('I') and *du* ('you') should therefore be equally predictable and recoverable and topic drop should be equally acceptable.

Like experiment 9, experiment 10 had the form of a 2 × 2 design crossing grammatical Person of the preverbal constituent (1st person singular vs. 2nd person singular) and Completeness (full form vs. topic drop). I expected to replicate the result of experiment 9 that topic drop of the 1st and the 2nd person singular is equally acceptable.

10.5.5.1 Materials

Items

The 24 items used in this experiment were adaptions of the 24 items of experiment 9, as shown in (13). Instead of two instant messages, they consisted of a context story and the target utterance, which was presented as direct speech. This target utterance was slightly adapted compared to experiment 9 to better fit with the context. Each context story was three sentences long and introduced two characters. One of these characters produced the target utterance, which was introduced by a reporting clause such as *Jessica sagt* ('Jessica says'). The target utterance itself was enclosed in quotation marks to mark it as direct speech. Like in experiment 9, this target utterance always contained a modal verb in the left bracket and was either a statement about the speaker (1SG conditions) or a request to the hearer (2SG conditions).

(13) Marcel und Jessica kochen gemeinsam Ratatouille in Marcels Küche. Während sie das Gemüse klein schneiden, tauschen sie Neuigkeiten über ihren Freund Sven aus. Sie unterhalten sich darüber, dass Sven zögert, sich auf eine interessante Stelle zu bewerben. Jessica meint:
'Marcel and Jessica cook ratatouille together in Marcel's kitchen. While they chop the vegetables, they exchange news about their friend Sven.

They talk about how Sven is hesitant to apply for an interesting job. Jessica says:'

a. *„Ich muss ihn überreden."*
   I must.1SG him persuade
   'I must persuade him.' (full form, 1SG)
b. *„Muss ihn überreden."*
   must.1SG him persuade
   'Must persuade him.' (topic drop, 1SG)
c. *„Du musst ihn überreden."*
   you.2SG must.2SG him persuade
   'You must persuade him.' (full form, 2SG)
d. *„Musst ihn überreden."*
   must.2SG him persuade
   'Must persuade him.' (topic drop, 2SG)

Fillers

I tested 72 fillers together with the 24 items. Those fillers contained 24 items of an experiment on the verb *brauchen* ('need'), 24 items of an experiment on preposition omission in short answer fragments, and 24 items on predictable vs. unpredictable fragment answers vs. full answers. I included 8 ungrammatical fillers with two finite verbs in the left bracket as attention checks to be able to exclude inattentive participants.

10.5.5.2 Procedure

48 self-reported native speakers of German between the ages of 18 and 40 who had not taken part in any previous experiment on topic drop participated in the study. They were recruited from *Clickworker* and received a compensation of €4.00. The experiment was presented online with LimeSurvey (Limesurvey GmbH 2023). The task of the participants was to rate the naturalness of the final utterance on a 7-point Likert scale (7 = completely natural), thereby considering the context story. The 24 items were distributed across 4 lists according to a Latin square design so that each participant saw each token set only once and in one condition. I mixed the items with the fillers and presented them to the participants in individual pseudo-randomized order.

10.5.5.3 Results

The data from 6 participants who had rated more than 4 of the 8 ungrammatical fillers with 6 or 7 points on the 7-point Likert scale, i.e., as completely natural or as almost completely natural, were excluded from the data set. Table 10.20 shows the mean ratings and standard deviations per condition. The mean ratings are also plotted with 95% confidence intervals in Figure 10.6. Unlike in experiment 9, there seems to be a difference between the two Completeness conditions. Full forms received higher ratings than utterances with topic drop. Additionally, there seems to be a tendency that the two Completeness conditions differed more strongly for the 1st than for the 2nd person singular.

Table 10.20: Mean ratings and standard deviations per condition for experiment 10

Completeness	Person	Mean rating	Standard deviation
Full form	1SG	5.92	1.35
Topic drop	1SG	4.46	1.87
Full form	2SG	5.85	1.53
Topic drop	2SG	4.68	1.85

For the inferential statistics, I analyzed the responses of the remaining 42 participants with CLMMs from the package ordinal (Christensen 2019) in R (R Core Team 2021), as described in Section 3.1.6.3. I modeled the ratings as a function of the binary predictors Completeness and Person, as well as of the numeric scaled and centered Position of the trial in the experiment, and included all two-way interactions between the independent variables. I coded Completeness and Person using deviation coding. Full forms and the 1st person singular were coded as 0.5, topic drop and the 2nd person singular as −0.5 respectively. The random effects consisted of random intercepts for participants and items and of by-participant and by-item random slopes for all predictors and their two-way interactions.[42] Table 10.21 summarizes the fixed effects in the final model with symmetric thresholds.

There was only a significant main effect of Completeness ($\chi^2(1) = 43.71, p < 0.001$), according to which participants preferred full forms over topic drop. The

[42]The formula of the full model was as follows: `Ratings ~ (Completeness + Person + Position)^2 + (1 + (Completeness + Person + Position)^2 | Subjects) + (1 + (Completeness + Person + Position)^2 | Items)`.

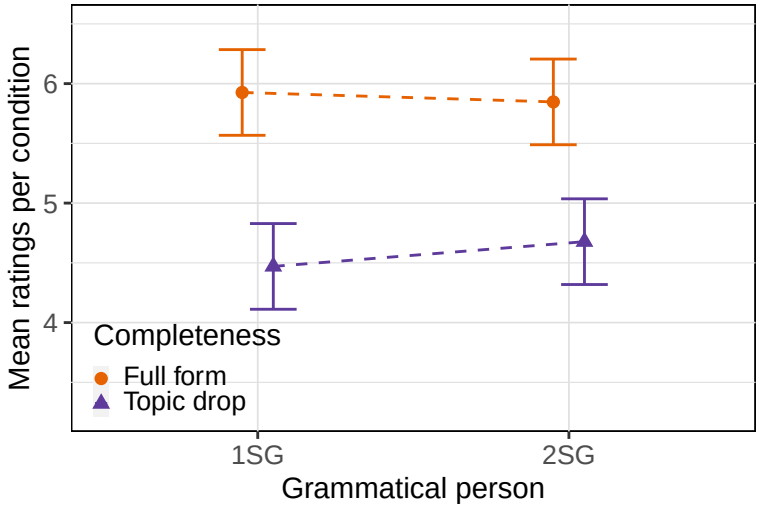


Figure 10.6: Mean ratings and 95% confidence intervals per condition for experiment 10

Table 10.21: Fixed effects in the final CLMM of experiment 10

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	2.09	0.25	43.71	< 0.001	***

interaction between Completeness and Person was not significant ($\chi^2(1) = 0.42, p > 0.5$), which indicates that topic drop of the 2nd person singular was not perceived as degraded compared to topic drop of the 1st person singular (or vice versa). The main effect of Person was not significant either ($\chi^2(1) = 1.55, p > 0.2$). Utterances with the 1st and 2nd person singular were rated equally acceptable, which indicates that the contexts made both readings of the target utterance equally possible.

10.5.5.4 Discussion

Experiment 10 compared 1st and 2nd person singular topic drop in appropriate contexts and replicated the results of experiment 9. Topic drop of the 1st and the 2nd person singular were equally acceptable. This result is in line with the *extralinguistic 1SG+2SG hypotheses*: If the communicative situation makes it obvious how the roles of speaker and hearer are distributed, the pronouns referring

to each of them can be omitted because the recovery is easy. This seems to hold for communication via instant messaging but also for spoken utterances in face-to-face conversations (at least if presented in written form).

Unlike in experiment 9, topic drop received generally worse ratings than the full forms in this experiment. This may be due to the presentation as direct speech. As I showed in Section 2.1.5, at least in the FraC, topic drop is less frequent in dialogues (omission rate of 16%) than in text messages (omission rate of almost 64%), the direct ancestors of instant messages. Furthermore, while in text messages the 1st person singular subject pronoun *ich* is most frequently targeted by topic drop (264 instances, 66.84% of all omissions), this is less so in dialogues (11 instances, 12.94% of all omissions). In dialogues, it is mainly the subject and object demonstrative pronoun *das* that is omitted (50 instances, 58.82%), which in turn is less often covert in text messages (19 instances, 5.44%). Consequently, the degraded ratings for topic drop in this experiment may be caused by the fact that (i) topic drop is generally less frequent in dialogues and (ii) that topic drop of the 1st (and probably also the 2nd) person singular is even less common. In summary, the two experiments 9 and 10 provide first experimental evidence that topic drop occurs preferably in certain text types, which needs to be validated and extended in future research.

10.6 Summary: grammatical person, verbal inflection, and ambiguity avoidance

In Section 10.1.1, I discussed in detail that according to the previous literature (e.g., Tesak & Dittmann 1991, Volodina & Onea 2012, Imo 2013, 2014), 1st person (singular) subjects are particularly often targeted by topic drop. This observation was confirmed through corpus studies, anecdotally by Auer (1993), Döring (2002) and quantitatively by Androutsopoulos & Schmidt (2002) and Frick (2017) (see Section 10.2.1), as well as by my own corpus study in Section 10.4. Both in the complete FraC and the text message subcorpus, the 1st person singular has the highest omission rate.

The acceptability rating studies, experiments 5, 7, and 8, presented in this chapter consistently confirmed the pattern also experimentally. The observed production preference was reflected in an acceptability difference, i.e., in perception. In all three experiments, topic drop of the 1st person singular was significantly preferred over topic drop of the 3rd person singular. It has to be noted, though, that in these studies, the 3rd person singular always referred to a human and was, thus, of either feminine or masculine gender, while my corpus study suggests

that topic drop of the 3rd person singular is most frequent for the neuter gender. It could be that for non-human 3rd person singular subjects of neuter gender, the effect would be less pronounced. This should be tested in a future study.

Also in Section 10.1.1, I presented potential explanations for the preference to omit the 1st person (singular) from the theoretical literature. On the one hand, there is Auer's (1993) *inflectional hypothesis*, which traces back the preference to a distinct inflectional ending on the verb following topic drop. On the other hand, there are *extralinguistic hypotheses* of two variants. First, authors such as Imo (2014) and Volodina & Onea (2012) argue that the 1st person singular can be omitted well because it refers to the speaker, who is easily recoverable as being part of the origo of speaking and through text type knowledge (*extralinguistic 1SG*). Second, Zifonun et al. (1997) and Ariel (1990) propose that both speaker and hearer are easily recoverable as they are essential parts of every communicative situation. Therefore, the omission of both 1st and 2nd person (singular) should work well (*extralinguistic 1SG+2SG*). The results of experiment 8 speak against Auer's (1993) *inflectional hypothesis* because they evidenced that the preference for topic drop of the 1st person singular persists even before syncretic verb forms. The two experiments 9 and 10 provided evidence for the *extralinguistic 1SG+2SG hypotheses*. Topic drop of the 1st person singular and the 2nd person singular are comparable in their acceptability, which suggests that both speaker and hearer can be equally well recovered.

As discussed in Section 10.3, my information-theoretic account of topic drop usage explains the effect of grammatical person by the *avoid troughs* principle. According to this principle (see Section 6.2.1), speakers should omit predictable expressions to distribute information more uniformly. In the case of the 1st (and the 2nd) person singular, this predictability seems to be impacted by at least the following two factors: (i) the frequency of the respective overt pronoun in certain text types and (ii) the likelihood of the referent in a communicative situation. (i) Corpus studies have evidenced that, in particular in text messages, the 1st person singular subject pronoun *ich* ('I') is very frequent in sentence-initial position, which makes it predictable there and likely to be omitted. (ii) Speaker and hearer are necessary parts of every communicative situation. This makes the 1st and 2nd person (singular) pronouns, which refer to the speaker and hearer, predictable and easier to omit. Given the results of experiments 9 and 10, which did not indicate an acceptability difference between 1st and 2nd person, it is reasonable to assume that factor (ii) has priority over factor (i). Accordingly, the likelihood of a referring expressing may be more strongly impacted by the presence of its referent in a situation than by its pure frequency in a text type. In summary,

the results for grammatical person are consistent with the information-theoretic account and can be explained by the *avoid troughs* principle.

As discussed above, Auer (1993) and, following him also Imo (2013, 2014), explain the high frequency of topic drop of the 1st person singular subject pronoun *ich* with the easy recoverability through the distinct inflectional marking on the verb following topic drop. They argue that a subject can be easily omitted if the grammatical person is also expressed through the inflectional ending at the congruent verb. While they only use this argument to explain the high frequency of the 1st person singular, this reasoning should apply in principle to every grammatical person for which the following verb is distinctly marked. In Section 6.3, I proposed to explain the potential impact of verbal inflection as the consequence of trying to reduce the processing effort on this verb, as predicted by the *facilitate recovery* principle. Since a distinct inflectional ending indeed indicates the grammatical person of the omitted congruent subject, this ending can reduce the overall effort associated with ellipsis resolution, i.e., determining the reference of the omitted constituent. I furthermore argued that this reduction of processing effort could be stronger for the 1st person singular than for the 3rd person singular. If the hearer identifies the omitted subject as 1st person singular, they can automatically conclude that the speaker is the intended referent, while for the 3rd person singular there are still, at least theoretically, always several potential referents present in the context.

First support for this line of reasoning came from the corpus study, where I found that topic drop of the 1st person singular was indeed more frequent before distinctly marked verb forms. Experiments 7 and 8, however, did not provide evidence in this direction. Topic drop of the 1st person singular was preferred to an equal degree over topic drop of the 3rd person singular, regardless of whether the following verb was distinctly marked for inflection or not. There was no effect of a distinct vs. a syncretic verb form at all in the joint analysis of both studies. Thus, the result remains inconclusive concerning verbal inflection.

Also from the perspective of ambiguity avoidance, it is predicted that syncretisms should be relevant to topic drop. Utterances with subject topic drop followed by a syncretic verb form are ambiguous and should be avoided in order not to cause processing difficulties for the hearer. I only found partial evidence for this hypothesis in the corpus study but not in the experiments. The corpus study suggests that topic drop of the 1st person singular but not of the 3rd person singular is rarer before verbs that are perceived as ambiguous. In experiments 7 and 8, there was no difference in acceptability between the utterances in which topic drop was followed by a distinct verb form and those in which the verb form was syncretic and, thus, ambiguous. This cannot yet be taken as evidence against

ambiguity avoidance because the object pronoun referring to the competing referent made the utterances with topic drop preceded by syncretic verb forms only locally ambiguous, not globally. The investigation of globally ambiguous topic drop structures needs to be left to future research (but see the attempt discussed in Footnote 14 on page 265). In the next chapter, I discuss two further verb-related influencing factors: verb type and verb surprisal.

11 Verb type and verb surprisal

After having reviewed, tested, and discarded Auer's (1993) *inflectional hypothesis*, according to which a distinct inflectional marking on the verb following topic drop can facilitate the recovery of omitted 1st person subject pronouns, in Chapter 10, I turn to two further verb-related factors in this chapter: verb type and verb surprisal. While verb type has been discussed at least briefly in the theoretical literature as a factor relevant to topic drop and has been examined in corpus studies, the same is not true for verb surprisal. An influence of verb surprisal is not predicted by any hypotheses from the theoretical literature but exclusively by the information-theoretic approach presented in this book (see Sections 6.2.2 and 6.3).

The structure of this chapter is as follows: For verb type, I give a brief theoretical overview and discuss previous studies. Then, I turn to the information-theoretic predictions for both factors. Finally, I present, first, my corpus results and, second, my experimental results, which partially evidence an impact of verb type and verb surprisal on topic drop.

11.1 Theoretical overview of verb type

In the theoretical literature on topic drop, some authors claim that it occurs preferably before certain verb types. However, they do not provide any explanation or theoretical account of why there should be an effect of the type of verb following topic drop at all. According to the IDS grammar (Zifonun et al. 1997: 415), subject topic drop of the 1st and the 2nd person, which Zifonun et al. term "person ellipsis", is particularly frequent before copular, auxiliary, and modal verbs. They argue that topic drop of these grammatical persons before finite lexical verbs is more marked in terms of style (Zifonun et al. 1997: 415). However, they do not justify these claims nor do they substantiate them empirically or even with introspective data. Also without empirical validation, Imo (2013: 297) claims that topic drop occurs particularly frequently before verba dicendi and verba sentiendi, such as *wissen* ('to know'), *finden* ('to find'), or *glauben* ('to believe').

11.2 Previous empirical evidence regarding verb type

Although a sophisticated account of verb type effects is pending, some corpus studies have investigated them. I summarize the results of four studies (Poitou 1993, Androutsopoulos & Schmidt 2002, Helmer 2016, Frick 2017), which generally suggest that the omission rates for modal, auxiliary, and copular verbs are higher than for lexical verbs, while there is only sparse evidence for a higher frequency of topic drop before verba dicendi and verba sentiendi. The most detailed study is again that by Frick (2017), who not only has the most data but also distinguishes the largest number of verb categories and even briefly discusses the interplay between verb type and grammatical person.

11.2.1 Poitou (1993)

In his corpus study (for details see Section 8.2), Poitou (1993) observes that topic drop is almost always followed by modal verbs and auxiliaries, which form a verbal complex with the element(s) in the right bracket, but he does not support this claim with count data.

11.2.2 Androutsopoulos & Schmidt (2002)

Androutsopoulos & Schmidt (2002) report the frequency and the omission rates of verbs that follow topic drop in their text message corpus (see Section 10.2.1 for details on their data set). Table 11.1 shows both figures for verb lemmas with more than 10 occurrences in the corpus.[1]

The verbs with the highest omission rates in the corpus of Androutsopoulos & Schmidt (2002) are the modal verbs *müssen* ('must'), *können* ('can'), and *wollen* ('to want') with over 70%. The verbs *sein* ('to be') and *haben* ('to have'), which are auxiliary verbs but can also function as a copular verb (*sein*) or a lexical verb (*haben*), have significantly lower omission rates of about 55% to 60% ($\chi^2(1) = 4.25, p < 0.05$).[2] The lexical verb *sitzen* ('to sit') has a similar omission rate of just under 55%, while the omission rates of the remaining two lexical verbs *gehen* ('to go') and *kommen* ('to come') are even lower with about 40% to 45%. The rates are on average significantly lower than those of the modal verbs

[1] It is not clear from their paper why they drew the line at 10 occurrences and whether they considered only subjects or also objects for this table.

[2] I compared the summed frequencies of the modal verbs (*müssen*, *können*, and *wollen*) and the summed frequencies of the two potential auxiliary verbs with a Pearson's chi-squared test with Yates's continuity correction in R (R Core Team 2021).

Table 11.1: Full forms, instances of topic drop, and omission rates as a function of the following verb lemma for lemmas with more than 10 occurrences in the text message corpus of Androutsopoulos & Schmidt (2002), taken from Androutsopoulos & Schmidt (2002: 69)

Verb	Full form	Topic drop	Total	Omission rate
sein ('to be')	51	76	127	59.84%
haben ('to have')	34	44	78	56.41%
wollen ('to want')	5	12	17	70.59%
gehen ('to go')	10	7	17	41.18%
müssen ('must')	3	11	14	78.57%
können ('can')	3	10	13	76.92%
sitzen ('to sit')	5	6	11	54.55%
kommen ('to come')	6	5	11	45.45%

($\chi^2(1) = 6.89, p < 0.01$), but there is no significant difference from the potential auxiliaries ($\chi^2(1) = 1.57, p > 0.2$).[3] Androutsopoulos & Schmidt (2002) do not provide a breakdown by grammatical person, but most of the instances in their corpus are 1st person singular pronouns (see Section 10.2.1). Therefore, it is possible to interpret the data at least tentatively in terms of the claim by the IDS grammar (Zifonun et al. 1997) that copular, auxiliary, and modal verbs most often follow topic drop of the 1st and the 2nd person. For the modal verbs, it turns out that, at least based on the data in Table 11.1,[4] they do indeed follow topic drop (especially of the 1st person singular) particularly frequently, to be more precise, significantly more frequently than the listed lexical verbs. This is consistent with the claim by the IDS grammar. At the same time, however, their omission rates are also significantly higher than those of the potential auxiliaries, which, in turn, contrary to the prediction by the IDS grammar, are not significantly dif-

[3] Again, I used Pearson's chi-squared tests with Yates's continuity correction in R (R Core Team 2021) to compare the summed frequencies between the verb types. But consider that the numbers of occurrences for all verbs except *sein* and *haben* in Androutsopoulos & Schmidt's (2002) data set are very small, resulting in potentially unreliable statistical results even after summing up.

[4] It should be noted that Androutsopoulos & Schmidt (2002) do not present the whole data. It might be possible that there are verb lemmas with fewer than 10 occurrences in their corpus that behave differently. For example, there could be lexical verbs with a higher omission rate than the lexical verbs presented in the table, or, in turn, it could be that the remaining modal verbs have lower omission rates. Thus, the implications discussed in the following should be taken with caution.

ferent from the lexical verbs in the table. Thus, overall, we find at most partial evidence for the claim by the IDS grammar.

In the data by Androutsopoulos & Schmidt (2002), there is no evidence that topic drop occurs preferably before verba dicendi or verba sentiendi, as predicted by Imo (2013). There is no corresponding verb of saying or thinking in Table 11.1.

11.2.3 Helmer (2016)

As sketched in Section 3.2.1, Helmer (2016) investigated 541 instances of topic drop in the corpora of spoken conversations FOLK and GIF. She compared them to a random sample of 200 syntactically complete utterances from the FOLK, each of which contains the pronoun *das* ('that') in subject or object function.[5] She argues that utterances with what she refers to as the "unspecific" anaphor *das* are similar to utterances with topic drop in their semantic vagueness. In her view, this allows for a comparison between overt and covert variants independently of the semantic content that a more specific anaphor would have (Helmer 2016: 197).

Helmer (2016) presents the frequency of topic drop and the full forms with *das* with the four biggest[6] semantic verb classes in her data set, as shown in Table 11.2: copular *sein* ('to be'), modal verbs, what she terms mental verbs, and verba dicendi. Note that what I term "proportion" in Table 11.2 is different from the omission rate that I report in many tables in this book. It indicates the proportion of instances of topic drop with a particular verb type out of the total number of 541 instances of topic drop, or the proportion of full forms with *das* and a particular verb class out of the total number of all 200 full forms with *das*. Since Helmer (2016) does not provide numbers for the verb classes that are not listed in Table 11.2, these proportions do not sum to 100%.

From the table, it can be seen that topic drop occurs most frequently before *sein* as a copula in Helmer's (2016) corpus of spoken dialogues. However, this verb class is also the most frequent for the variants with overt *das*, even by a larger margin. The second and third most frequent verb classes after topic drop are modal and mental verbs, but they differ by only about 2%. Verba dicendi follow in fourth place. The finding that topic drop occurs most frequently before the copular verb *sein* is partly consistent with the claim by Zifonun et al. (1997). However, the result should be qualified in that the full forms with *das* occur even

[5] Recall that Helmer (2016) considers topic drop to be possible in the middle field as well. For this reason, her reference data also contain 29 full forms with *das* in the middle field, i.e., after the finite verb in the left bracket, and 10 for which the topological position could either not be determined or that occurred in the postfield (Helmer 2016: 214).

[6] She does not provide a complete list of all verb classes that she assumes.

Table 11.2: Frequency and proportions of the full forms with *das* and of topic drop for the four most frequent semantic verb classes in Helmer's (2016) data set, taken from Helmer (2016: 212)

	Full form		Topic drop	
Verb class	Frequency	Proportion	Frequency	Proportion
Copular *sein*	109	54.5%	169	31.2%
Modal verbs	19	9.5%	78	14.4%
Mental verbs	6	3.0%	91	16.8%
Verba dicendi	5	2.5%	41	7.6%

more frequently with this verb type and Zifonun et al. (1997) made their claim only for 1st and 2nd person subject topic drop.

Helmer (2016) interprets her result as support for Imo's (2013) claim that topic drop is particularly frequent before verba dicendi and sentiendi. While Helmer's (2016) instances of topic drop do seem to occur more often with these verb classes than her full forms with *das*, they are at the same time even more frequent with *sein* and modal verbs.

It must be noted, however, that her approach prevents true comparability and the computation of reasonable omission rates for two reasons. First, the ellipses and the full forms come from different corpora, the ellipses from the FOLK and the GIF, the full forms only from the FOLK. The Forschungs- und Lehrkorpus Gesprochenes Deutsch FOLK ('The research and teaching corpus of spoken German') (Schmidt 2014) contains recordings of informal everyday conversations (Helmer 2016: 65). The corpus Gespräche im Fernsehen GIF ('conversations on TV') consists of recordings from talk shows, discussions, and programs concerning the German federal election campaign of 2002 (Institut für deutsche Sprache, Mannheim 2014). Second, the ellipses are not restricted in terms of their grammatical person, number, and gender while the full forms are.[7] Additionally, the data set seems to contain both subjects and objects. Helmer (2016) does not provide information about their respective proportions.

[7]With the exclusive fixation on full forms with *das*, all of Helmer's (2016) full forms are 3rd person singular neuter. In contrast, her topic drop data also contain omission of constituents of other grammatical persons, numbers, and genders. She does not provide a corresponding breakdown and in most of her examples the omitted prefield constituent can be reconstructed with a *das*, but she also discusses examples where the 1st person singular pronoun *ich* is omitted (e.g., Helmer 2016: 98, example 19).

In sum, there seems to be some tentative support for the two claims from the literature concerning verb type in Helmer's (2016) data set, but this support must be qualified by the fact that her restricted full forms with *das* do not allow for a real comparison between overt and covert variants.

11.2.4 Frick (2017)

For the subjects in her Swiss German text message corpus, Frick (2017) observes a similar tendency to Androutsopoulos & Schmidt (2002) concerning verb type, as shown in Table 11.3.

Table 11.3: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb for the subjects in Frick's (2017) text message corpus, taken from Frick (2017: 93)

Verb type	Full form	Topic drop	Total	Omission rate
Lexical verbs	1 381	1 353	2 734	49.49%
Copular verbs	377	482	859	56.11%
Modal verbs	261	340	601	56.57%
Reflexive verbs	34	148	182	81.32%
Auxiliary verbs	7	3	10	30.00%

Subject topic drop is about as common before copular verbs as it is before modal verbs ($\chi^2(1) = 0.01$, $p > 0.9$). In comparison to lexical verbs, it is significantly more frequent before copular verbs ($\chi^2(1) = 11.21$, $p < 0.001$) and before modal verbs ($\chi^2(1) = 9.61$, $p < 0.01$).[8] What is striking are the results for auxiliaries and reflexive verbs. Auxiliaries seem to be surprisingly rare in Frick's data. Their omission rate is lower than the rate determined by Androutsopoulos & Schmidt (2002) for the potential auxiliaries *sein* and *haben*,[9] but it is not very meaningful

[8] I compared the numbers for the verb types with Pearson's chi-squared tests with Yates's continuity correction in R (R Core Team 2021).

[9] Recall that Androutsopoulos & Schmidt (2002) did not distinguish between copular and auxiliary usage of *sein* nor between auxiliary and lexical verb usage of *haben*. It is questionable whether this distinction is even relevant to topic drop. As discussed in Section 6.3, I assume that topic drop is usually resolved on the verb in the left bracket, i.e., at a point in time when it is not yet clear to the reader whether the form of *sein* is a copula or an auxiliary or whether *haben* is used as an auxiliary or a full verb. This would be different under parallel parsing theories, such as the one proposed by Levy (2008) and briefly discussed in Section 10.1.2. Under his account, the ambiguity would not be resolved directly on the verb but only when the intended structure becomes apparent, in the right bracket or at the end of the clause. In this framework, the difference between, e.g., a usage of *haben* as auxiliary and a usage as lexical verb might be of relevance.

with only 10 data points. The highest omission rate of all verb types is exhibited by reflexive verbs with over 80%.[10] However, without providing a proportion, Frick states that these are mainly instances of the verb *sich freuen* ('to be glad') (Frick 2017: 94). In other words, topic drop may not generally be frequent before reflexive verbs but only or mainly before *sich freuen*. Frick (2017: 95) speculates that the Swiss German utterance Δ *freu mi* ('(I) am glad'), i.e., 1st person singular subject topic drop before *sich freuen* in the present tense, is currently developing into a fixed formula.

Frick (2017: 96) notes that the verb type following topic drop seems to differ depending on the grammatical person of the omitted subject. This may be relevant to the claim by Zifonun et al. (1997) that subject topic drop of the 1st and the 2nd person is particularly frequent before copular, auxiliary, and modal verbs.[11] Frick states that the 2nd person is preferably omitted before modal verbs (Frick 2017: 96). The omission rate of the 2nd person singular is 64.41% before modals vs. only 37.79% before lexical verbs. For the 2nd person plural, there are only 4 instances of topic drop, which admittedly all occur before modals, but which are too few to make a quantitative statement. In contrast to the 2nd person, Frick (2017: 96) notes that the 3rd person singular is rarely omitted before modal verbs (13.33% omissions vs. 30.77% for lexical verbs). She attributes this to the syncretism with the 1st person singular discussed above and to ambiguity avoidance: "Due to the danger of confusion with the first person singular, the omission before modal verbs is therefore blocked or only possible as an exception"[12] (Frick 2017: 108). However, in her corpus, only 13 overt 3rd person singular pronouns and 2 corresponding instances of topic drop occur before modal verbs.[13] These figures are too small to base conclusions on them. The results of my experiments 7 and 8 (Sections 10.5.2 and 10.5.3), which tested topic drop of the 1st and 3rd person singular before lexical verbs and before modal verbs, do not suggest that topic drop of the 3rd person singular is more degraded before modal than before lexical verbs. For the 1st person, Frick (2017) does not discuss the effect of verb type, but her figure 11 (Frick 2017: 96) shows that both the 1st person singular and the 1st

[10] It is significantly higher than that of the modal verbs with the second highest rate ($\chi^2(1) = 35.39$, $p < 0.001$).

[11] Note that the IDS grammar does not explicitly claim that topic drop of the 3rd person is less frequent before these verb types. The 1st and 2nd person topic drop is discussed separately as "person ellipsis" and the verb type claim is only made there (see Zifonun et al. 1997: 414–417).

[12] My translation, the original: "Durch die Verwechslungsgefahr mit der ersten Person Singular ist die Auslassung vor Modalverben deshalb geblockt bzw. nur ausnahmsweise möglich" (Frick 2017: 108).

[13] This may be because in text messages, it is more likely to give the addressee advice of the type *you should* / *you can*, etc. than a third person.

person plural are omitted most often before lexical verbs. Despite her remarks, Frick (2017) also does not provide an explanation for the different omission rates per verb type.

In short, the corpus studies are tentatively in line with the claim by the IDS grammar (Zifonun et al. 1997) that topic drop of the 1st and the 2nd person occurs particularly often before copular and modal verbs. For the auxiliaries, the results are less conclusive, which hinges partly on the extremely low number of auxiliaries that Frick (2017) claims to have found in her corpus. Concerning Imo's (2013) claim that topic drop occurs preferably before verba sentiendi and dicendi, there is only tentative support from Helmer's (2016) study. In her corpus, verbs of saying and thinking were in the top four verb classes that occurred most often with topic drop. It is important to note that even the authors who empirically studied verb type do not explain why topic drop should appear particularly frequently with certain verb types. In the following, I propose such an explanation from the perspective of information theory.

11.3 Information-theoretic predictions for verb type and verb surprisal

In Section 6.2.2, I outlined how verb surprisal should impact the usage of topic drop according to my information-theoretic account. I briefly repeat the corresponding reasoning here. According to the *avoid peaks* principle, the likelihood of topic drop should decrease with the surprisal of the following verb: the higher

Table 11.4: Verbs in the frequency classes 3 to 7 in the DeReWo: the lower the frequency class, the more frequent the corresponding verb

Verb lemma	Frequency class	Verb lemma	Frequency class
sein ('to be')	3	*stehen* ('to stand')	6
werden ('will')	3	*sehen* ('to see')	7
haben ('to have')	3	*finden* ('to find')	7
können ('can')	5	*lassen* ('to let')	7
sollen ('shall')	6	*bleiben* ('to remain')	7
müssen ('must')	6	*liegen* ('to lie')	7
geben ('to give')	6	*stellen* ('to put')	7
sagen ('to say')	6	*nehmen* ('to take')	7
kommen ('to come')	6	*zeigen* ('to show')	7
wollen ('want')	6	*dürfen* ('may')	7
gehen ('to go')	6	*halten* ('to hold')	7
machen ('to make')	6		

the surprisal of the verb in the left bracket, the lower the likelihood of topic drop. This is because the overt prefield constituent may be needed to reduce the processing effort on the verb for two reasons. First, it can make the following verb more predictable, and, second, it avoids the processing effort required to resolve the ellipsis.

My information-theoretic account also predicts effects of verb type as an indirect consequence of the *avoid peaks* principle. Copular, auxiliary, and modal verbs are generally very frequent in German. This is illustrated in Table 11.4 based on the Korpusbasierte Wortgrundformenliste DeReWo (Leibniz-Institut für Deutsche Sprache 2013b),[14] which shows all verbs in the frequency classes 3 to 7 from the list.[15]

Sein ('to be'), *werden* ('will'), *haben* ('to have'), and all modal verbs except for *dürfen* ('may') are more frequent than or at least equally frequent as the most frequent lexical verbs, such as *geben* ('to give') and *sagen* ('to say'). Given that copular, auxiliary, and modal verbs are so frequent, it is generally unlikely for them to cause a peak in the information density profile. Thus, the speaker does not need to insert the preverbal constituent to smooth the profile but can omit it whenever possible.

There are two qualifications to this conclusion. First, the table does not consider the syntactic position of the verb, so the occurrences in the left and right brackets are combined, whereas only the left bracket is relevant to the information-theoretic frequency-based explanation of topic drop. Second, the data are based on the DeReKo, i.e., the German reference corpus, which consists largely of news articles. It could be that the distribution would be different in a correspondingly large text message corpus or another corpus of conceptually spoken text types. For the text message subcorpus of the fragment corpus FraC (see Section 7.2.1 for details), however, a similar distribution can be observed if we look at the 10 most frequent verb lemmas, again independent of the topological position in Table 11.5.

[14] The Korpusbasierte Wortgrundformenliste ('corpus-based list of word base forms') DeReWo v-ww-bll-320000g-2012-12-31-1.0 (Leibniz-Institut für Deutsche Sprache 2013b) is based on the Deutsches Referenzkorpus DeReKo ('German reference corpus') (Leibniz-Institut für Deutsche Sprache 2012). The frequency of words is indicated by their grouping into frequency classes. Frequency class 0 contains the most frequent word, i.e., the definite determiner with the forms *der*, *die*, *das*. Each subsequent frequency class n contains words that are approximately 2^n times less frequent than the word in frequency class 1 (Leibniz-Institut für Deutsche Sprache 2013a: 7). This means that *sein* ('to be') in frequency class 3 is about $2^n = 2^3 = 8$ times less frequent than the definite determiner *der*, *die*, *das*. The higher the frequency class, the rarer the words contained in this class.

[15] I limited myself to frequency classes 1 through 7 because that way all copular, auxiliary, and modal verbs are included in my selection. Of these, *dürfen* ('may') is the rarest in class 7.

Table 11.5: The ten most frequent verb lemmas and their frequency in the text message subcorpus of the FraC

Verb lemma	Frequency
sein ('to be')	363
haben ('to have')	221
gehen ('to go')	76
kommen ('to come')	75
können ('can')	67
machen ('to make')	47
müssen ('must')	47
werden ('will')	44
wollen ('to want')	42
sagen ('to say')	35

Again, the auxiliary and copular verb *sein* ('to be') and the auxiliary and lexical verb *haben* ('to have') are the most frequent verbs by a large margin. Therefore, the predictions regarding auxiliary and copular verbs should translate to text message data. Although the modal verbs seem to be a bit less frequent relative to the most frequent lexical verbs, still three of them are in the top ten of verb lemmas in the text message corpus. Therefore, at least for these modal verbs, it can also be expected that they tend to have a rather low surprisal. Note that the information-theoretic prediction according to which topic drop before modal verbs should be more felicitous because they are more frequent clashes with the predictions of ambiguity avoidance discussed in Section 10.1.2. According to ambiguity avoidance, speakers should not use topic drop before modal verbs, at least of the 1st and the 3rd person singular, because they exhibit syncretic forms resulting in ambiguity.

11.4 Corpus study of verb type and verb surprisal

In addition to grammatical person, verbal inflection and ambiguity avoidance, I also investigated the remaining verb-related factors both in my corpus study and experimentally. In this section, I first present and discuss the corpus results.[16] I

[16]The corpus data and the analysis scripts are accessible online: https://osf.io/zh7tr. For copyright reasons, I only provide the IDs and the annotations of each instance but not the linguistic material.

start with the descriptive results concerning verb type in the FraC-TD-Comp data set. With this data set, I furthermore refute Imo's (2013) claim concerning verb types. As discussed in Section 7.2.3, given the typical occurrence of topic drop in certain text types, it is not reasonable to determine relative numbers and omission rates on the whole FraC. Therefore, I focused on the text message data set FraC-TD-SMS to determine relative numbers for the type of the verb in the left bracket and on FraC-TD-SMS-Part to calculate the surprisal of this verb. For all three data sets, I discuss the claim of the IDS grammar (Zifonun et al. 1997) that topic drop of the 1st and the 2nd person occur preferably before copular, auxiliary, and modal verbs. However, it was tested systematically in the logistic regression analysis of FraC-TD-SMS-Part, which I discussed already in Section 10.4.4 and which I revisit in Section 11.4.4.

11.4.1 Verb type in FraC-TD-Comp

In the first step, I determined the verb type in the FraC-TD-Comp data set, which consists of all instances of topic drop and all full forms in the FraC. Each instance is annotated for the type of the verb in the left bracket distinguishing between copular verbs, auxiliaries, modal verbs, reflexive verbs, and lexical verbs (see Section 7.2.2). As Table 11.6[17] shows for the subjects and Table 11.7 for the objects, the instances of topic drop and the full forms principally behave in parallel concerning the verb type. For subjects, lexical verbs are the most common, followed by copular verbs, auxiliaries, and modal verbs, while reflexive verbs are the least common. Overt and covert objects occur most frequently before lexical verbs, followed by modal verbs and auxiliaries but rarely before copular and reflexive verbs. The omission rates for both subjects and objects are the highest before copular verbs, followed by lexical verbs.

This only partly matches the claim of the IDS grammar (Zifonun et al. 1997: 415) according to which subject topic drop, at least of the 1st and the 2nd person, is more frequent before copular, auxiliary, and modal verbs than before lexical verbs. Below in Sections 11.4.2 and 11.4.3, I come back to this issue in more detail and to verb type in general by looking at the FraC-TD-SMS and FraC-TD-SMS-Part data sets.

As discussed in Section 7.2.3, I only conducted statistical analyses of the text message data sets. This is because we may not only expect differences between the text types in terms of the rate of utterances with topic drop but also in the

[17] The verb type is unascertainable for three full forms. There, the sentence is interrupted right after the finite verb so that for verbs such as *sein* ('to be') or *haben* ('to have'), I could not decide whether it was used as a copular, auxiliary, or lexical verb.

Table 11.6: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb for the subjects in the FraC-TD-Comp data set

Verb type	Full form	Topic drop	Total	Omission rate
Lexical verb	1 450	398	1 848	21.54%
Copular verb	602	205	807	25.40%
Auxiliary verb	450	94	544	17.28%
Modal verb	421	73	494	14.78%
Reflexive verb	125	31	156	19.87%
Unascertainable	3	0	3	0.00%

Table 11.7: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb for the objects in the FraC-TD-Comp data set

Verb type	Full form	Topic drop	Total	Omission rate
Lexical verb	78	35	113	30.97%
Copular verb	1	1	2	50.00%
Auxiliary verb	29	16	45	35.56%
Modal verb	51	20	71	28.17%
Reflexive verb	1	0	1	0.00%

frequency of the verb types. For instance, the verb type is also influenced by which tenses are frequently used, e.g., using the perfect leads to a higher rate of auxiliary verbs, etc. Theoretically, text types with little or no topic drop could distort the results if certain verb types are particularly frequent or rare there.

The claim by Imo (2013: 297), according to which topic drop is said to occur most frequently before verba dicendi and verba sentiendi, such as *wissen* ('to know'), *finden* ('to find'), or *glauben* ('to believe'),[18] is challenged by the data in Table 11.8.

[18]The original claim in German was: "Es ist offensichtlich, dass uneigentliche Verbspitzenstellungen mit bestimmten, hoch frequenten verba dicendi et sentiendi wie *wissen*, *finden* oder *glauben* besonders häufig auftreten" (Imo 2013: 297). My translation: 'It is obvious that the improper verb top positionings occur particularly frequently with certain high-frequency verba dicendi and sentiendi such as to know, to find, or to believe.'

Table 11.8: Absolute frequencies and omission rates of selected verb lemmas in the left bracket of the full forms and the utterances with topic drop in the FraC-TD-Comp data set

Rank	Verb lemma	Full form	Topic drop	Total	Omission rate
1	*sein* ('to be')	701	216	917	23.56%
2	*haben* ('to have')	419	117	536	21.83%
3	*können* ('can')	195	41	236	17.37%
4	*sich freuen* ('to be glad')	34	36	70	51.43%
5	*werden* ('to become')	167	34	201	16.92%
6	*müssen* ('must')	117	28	145	19.31%
7	*gehen* ('to go')	56	19	75	25.33%
8	*suchen* ('to search')	23	19	42	45.24%
9	*machen* ('to make')	38	18	56	32.14%
10	*kommen* ('to come')	30	17	47	36.17%
…					
13	*wissen* ('to know')	69	14	83	16.87%
14	*finden* ('to find')	35	13	48	27.08%
…					
17	*denken* ('to think')	53	9	62	14.52%
…					
39	*glauben* ('to believe')	40	3	43	6.98%

In absolute terms, the verbs that follow topic drop most often are forms of *sein* ('to be') with more than 200 occurrences and *haben* ('to have') with more than 100. The most frequent reflexive verb is *sich freuen* ('to be glad') with about 36 occurrences.[19] The lexical verbs *wissen* (11 occurrences), *finden* (13 occurrences), and *glauben* (3 occurrences) are at most mid-table[20] and their omission rates, except for *finden*, are even below the mean of 22.08% for the lexical verbs. Taken together, there is nothing that suggests that these verbs are particularly frequent after topic drop or that topic drop occurs most often before them. With this result, I consider Imo's (2013) strong claim that topic drop occurs very frequently with certain verba dicendi and sentiendi to be sufficiently invalidated and do not discuss it further.

[19] As in Frick's (2017) corpus, this makes it by far the most frequent reflexive verb following topic drop (see Section 11.2).

[20] Recall that *wissen* is also a preterite present verb with syncretic forms for the 1st and 3rd person singular in present tense (*ich weiß* and *sie weiß*).

11.4.2 Verb type in FraC-TD-SMS

Table 11.9 shows the omission rates in the FraC-TD-SMS data set as a function of the type of the following verb. Because of the higher frequency of subject topic drop, I restrict myself to the verb type following the subjects, disregarding the 8 objects.[21]

Table 11.9: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb for the subjects in the FraC-TD-SMS data set

Verb type	Full form	Topic drop	Total	Omission rate
Auxiliary verb	22	57	79	72.15%
Copular verb	43	85	128	66.41%
Lexical verb	95	151	246	61.38%
Modal verb	29	42	71	59.15%
Reflexive verb	10	13	23	56.52%

As discussed in Section 11.1, Zifonun et al. (1997: 415) claim in the IDS grammar that subject topic drop of the 1st and the 2nd person is more frequent before copular, auxiliary, and modal verbs than before other verb types. The observed frequency differences in the FraC-TD-SMS data set are only partially in line with this claim.[22] While descriptively auxiliaries and copulas follow topic drop

[21] Again, the numbers for the objects are too small anyway to base conclusions on them.

[22] While the overall tendency in FraC-TD-SMS and in Frick's (2017) study on Swiss German text messages using a comparable method (see Section 11.2) is similar, the individual omission rates differ, in two cases considerably. Topic drop is the rarest before auxiliaries in her data (30%), but in FraC-TD-SMS data set it was the most frequent (72%). Since Frick (2017) reports to have found only 7 full forms and 3 instances of topic drop before auxiliaries, this discrepancy is most likely caused by different criteria for assigning a verb to a category. Conversely, the verb type that most frequently follows topic drop in Frick's (2017) data, reflexives (80%, based on 182 occurrences), has the lowest omission rate (57%) in FraC-TD-SMS. As in Frick's (2017) case, the data for reflexive verbs mostly stem from the verb *sich freuen* ('to be glad'). In German Standard German, the topic drop variant, *freu(e) mich* ('am glad'), does not yet seem to have become equally fixed as a formula as in Swiss German (at least at the time of the data collection), so the full form with *ich* is still frequent. The omission rates of copular verbs and lexical verbs are around 10% higher in FraC-TD-SMS than in Frick's (2017) corpus (significant according to a Pearson's chi-squared tests with Yates's continuity correction calculated in R (R Core Team 2021), $\chi^2_{cop}(1) = 4.42, p < 0.05$; $\chi^2_{lex}(1) = 12.3, p < 0.001$), while the rate for modal verbs is only slightly higher (57% in Frick (2017) and 59% in FraC-TD-SMS, not significant, $\chi^2(1) = 0.08, p > 0.7$). There is no obvious explanation for the differences in modal verbs and lexical verbs other than that of general differences between the corpora used.

more often than lexical verbs, the differences are not significant according to a simple logistic regression predicting the probability of topic drop from the verb type and setting lexical verbs as the baseline ($\chi^2_{aux}(1) = 3.09, p > 0.07$; $\chi^2_{cop}(1) = 0.92, p > 0.3$), although there is a tendency in this direction for the auxiliaries. What is unexpected is that the omission rate before the modal verbs in the data set is relatively low, even a bit lower than for the lexical verbs in absolute terms (the difference is not significant, $\chi^2(1) = 0.11, p > 0.7$). Both the IDS grammar and my information-theoretic approach would have predicted a higher omission rate for modal verbs. While the IDS grammar does not justify this expectation, the information-theoretic approach argues that verbs in the closed class of modal verbs should have lower surprisal on average than members of the open class of lexical verbs. Therefore, it should not be necessary to use the full form to facilitate their processing. I discuss the role of verb type more closely below.

11.4.3 Verb type and surprisal in FraC-TD-SMS-Part

I also analyzed the verb type and additionally the verb surprisal of the instances in the FraC-TD-SMS-Part data set, which contains only the 1st and 3rd person subjects from FraC-TD-SMS (see Section 7.2.3.3). Table 11.10 shows the omission rates per verb type. The total figures are very similar to the slightly larger FraC-TD-SMS and also the order of verb types as a function of their omission rate is identical. The omission rates seem to provide partial support for the claim from the IDS grammar (Zifonun et al. 1997: 415) that subject topic drop of the 1st and the 2nd person is particularly frequent before copular, auxiliary, and modal verbs.

Table 11.10: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb in the FraC-TD-SMS-Part data set

Verb type	Full form	Topic drop	Total	Omission rate
Auxiliary verb	14	45	59	76.27%
Copular verb	30	72	102	70.59%
Lexical verb	86	142	228	62.28%
Modal verb	20	33	53	62.26%
Reflexive verb	10	13	23	56.52%

Table 11.11 shows the omission rates per verb type subdivided by the two grammatical persons.

Table 11.11: Full forms, instances of topic drop, and omission rates as a function of the type of the following verb in the FraC-TD-SMS-Part data set divided by grammatical person

	1SG				3SG			
Verb type	Full form	Topic drop	Total	Omission rate	Full form	Topic drop	Total	Omission rate
Auxiliary	13	41	54	75.93%	1	4	5	80.00%
Copular	14	52	66	78.79%	16	20	36	55.56%
Lexical	75	126	201	62.69%	11	16	27	59.26%
Modal	19	32	51	62.75%	1	1	2	50.00%
Reflexive	10	13	23	56.52%	0	0	0	0.00%

It can be seen that both the 1st and the 3rd person singular are particularly frequently omitted before auxiliaries, whereas only the 1st person singular is also especially often omitted before copulas. However, the omission rates of neither the 1st nor the 3rd person singular are higher before modal verbs than before lexical verbs. It is worth noting that the rates of the 1st person singular are generally higher than those of the 3rd person singular for all verb types except auxiliaries and that overall the figures for the 3rd person singular are relatively low. This suggests that there is a tendency that partly matches Zifonun et al.'s (1997) claim. To also test this claim statistically, I included a binary predictor Verb Type in my regression analysis, which contrasted copular, auxiliary, and modal verbs on the one hand with lexical and reflexive verbs on the other. I revisit this analysis in Section 11.4.4.

With the data set FraC-TD-SMS-Part, I also looked at the surprisal of the verb following topic drop to test my information-theoretic approach. Following from the *uniform information density hypothesis* and the *avoid peaks* principle (see Section 6.2.2), topic drop should be less frequent if the surprisal of the following verb is higher because the insertion of an overt prefield constituent may be necessary to lower the surprisal and, thus, the processing effort on this verb. This is supported by Figure 11.1,[23] which shows the proportion of topic drop as a function of the unigram surprisal of the verb in the left bracket and indicates that the rate of topic drop decreases when the verb surprisal increases.

I relied on the lemma-based unigram surprisal of the verb in the left bracket as a measure of surprisal. It was calculated on the lemmatized text message

[23] Figures 11.1, 11.2, and 11.2 were created in R with the package ggplot2 (Wickham 2016).

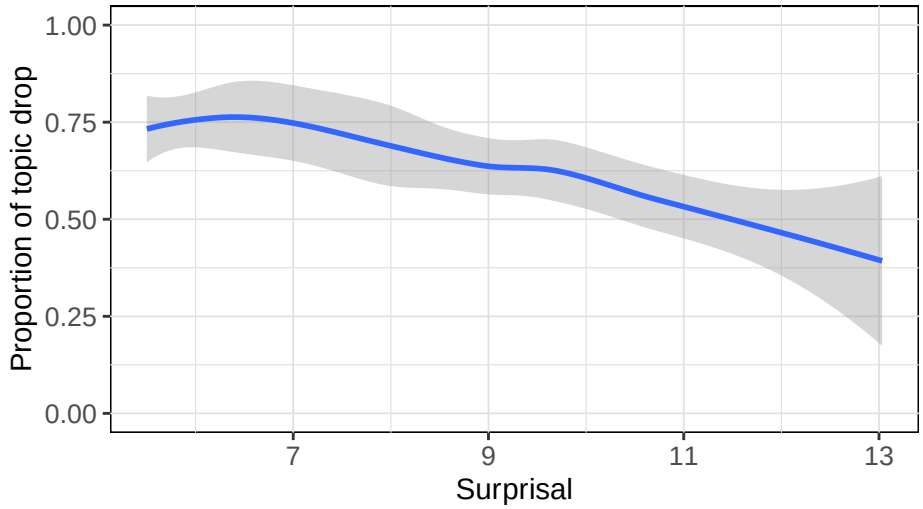


Figure 11.1: Proportion of topic drop as a function of the unigram surprisal of the verb in the left bracket in the FraC-TD-SMS-Part data set

subcorpus of the FraC using the SRILM toolkit (Stolcke 2002) and corresponds to the frequency of the lemma in this subcorpus regardless of its position in the clause. That is, the more frequently a verb occurs in the text message corpus in any topological position, the lower its surprisal (measured in bits). With this measure, I approximated the likelihood of the verbs in text messages, i.e., the context considered in the surprisal measure is the text type text message: $S(\textit{verb lemma}) \approx -\log_2\ p(\textit{verb lemma} \mid \textit{text message})$. The lemma frequency is only a rough approximation to what would be a "psychologically realistic" surprisal, i.e., one that considers all relevant aspects of the linguistic and extralinguistic context. Given, however, the data sparsity of FraC-TD-SMS-Part, the unigram surprisal conditioned on the text type text message seems to be a reasonable approximation. This holds also because, following from the V2 word order in German, the verbs in my data set are predominantly either in the first or second position, i.e., higher n-gram models, which would require more data, may not be a lot more informative.

The violin plot in Figure 11.2 shows the unigram surprisal per verb type. The mean surprisal of all verbs in the corpus is at 8.18 (SD = 2.23). The small and closed classes of copulas and auxiliaries are very frequent and, thus, have a low surprisal. The mean surprisal of the likewise closed class of modal verbs lies between that of the auxiliaries and the reflexive and lexical verbs. While the mean surprisal of

the reflexive verbs is close to that of the lexical verbs, the latter is additionally distributed over almost the entire surprisal range from 5.5 to 13 with the mean at 9.6 (SD = 1.94).

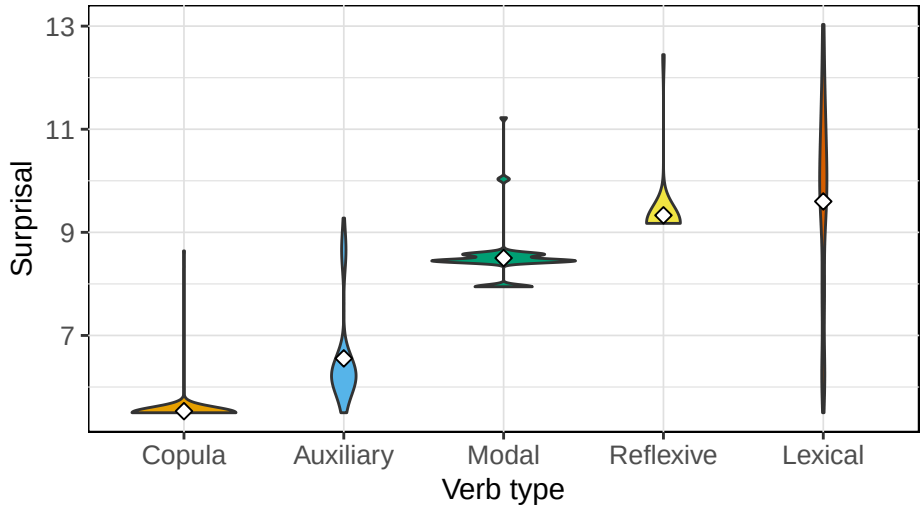


Figure 11.2: Relation between verb surprisal and verb type in the FraC-TD-SMS-Part data set (diamonds are mean values)

The order of verb types based on their omission rates is very similar to that based on the mean unigram surprisal but not identical, as shown in Figure 11.3. The auxiliaries have a higher total omission rate than the copulas (but not after the 1st person singular), but the copulas have a lower mean surprisal. This suggests that the verb surprisal is generally a good predictor for the omission rate before certain verb types, but it might not be the only relevant factor.

11.4.4 Verb type and surprisal in FraC-TD-SMS-Part – logistic regression analysis

In Section 10.4.4, I presented the logistic regression analysis conducted on the FraC-TD-SMS-Part data set and discussed its results for grammatical person and verbal inflection. In this section, I return to this analysis and focus on the two remaining factors in my discussion: verb type and verb surprisal. Recall that I predicted the likelihood of topic drop vs. full form from the independent variables grammatical Person, verb Surprisal, Verb Type, and Inflection. The binary Person predictor compared the 1st to the 3rd person singular and the three-level

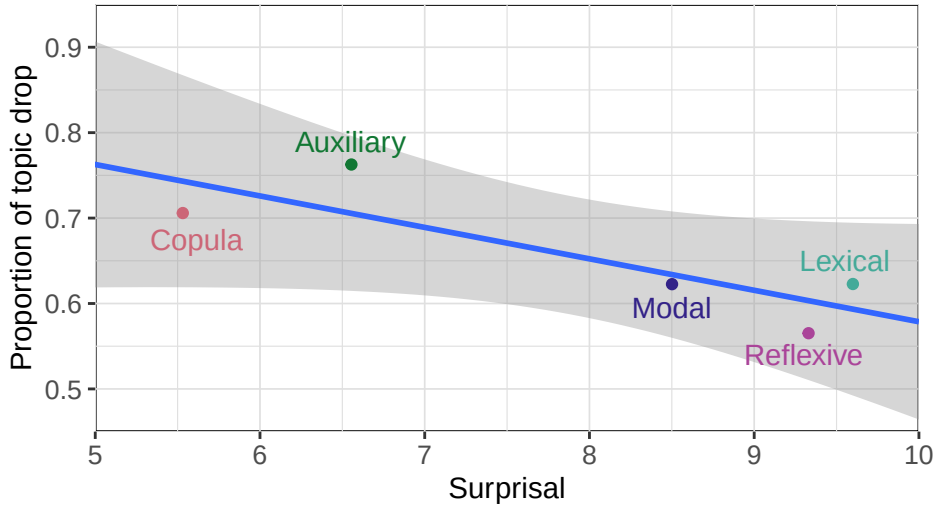


Figure 11.3: Relation between verb surprisal, omission rate, and verb type in the FraC-TD-SMS-Part data set

predictor Inflection compared *distinct*, *strictly syncretic*, and *informed syncretic* verb forms with each other (for details see Section 10.4.4). The independent variable Verb Type had two levels. It encoded whether the verb in the left bracket was either part of the group of copular, auxiliary, and modal verbs or part of the group of lexical and reflexive verbs. This way, I tested the claim from the IDS grammar that topic drop of the 1st and 2nd person is particularly frequent before copular, auxiliary, and modal verbs. The numeric Surprisal predictor encoded the unigram surprisal per verb lemma of the verb in the left bracket.

In this section, I additionally present an analysis of only the instances in FraC-TD-SMS-Part with a lexical verb in the left bracket to evidence an impact of verb surprisal for only them as well. Such an effect would provide stronger support for my information-theoretic account. It would then be excluded that the surprisal effect, already described in Section 10.4.4 for all verb types, is only an artifact of systematic surprisal differences between verb types. Up to now, this possibility has not been ruled out because the predictor Verb Type only captures the difference between copular, auxiliary, and modal verbs on the one hand and lexical and reflexive verbs on the other hand in a binary way but not differences within the two groups. This means that if I find a surprisal effect when focusing only on the lexical verbs, this effect is independent of verb type and, thus, provides genuine evidence for the information-theoretic approach, or more precisely for the *avoid peaks* principle.

11.4.4.1 Predictions

In the following, I complement the predictions (i–Person) and (ii–Inflection) presented in Section 10.4.4, according to which topic drop of the 1st person singular should be more frequent, in particular, before verbs with a distinct inflectional ending, with two further ones regarding verb type and verb surprisal.

(iii–Surprisal) It is exclusively the information-theoretic account that predicts an influence of verb surprisal on topic drop, namely a main effect. Following from the *avoid peaks* principle (Section 6.2.2), topic drop should be less frequent if the following verb has a high surprisal. In this case, the processing effort of this verb is more likely to exceed the hearer's cognitive capacity. Therefore, to prevent processing difficulties for the hearer, the speaker could make the verb more predictable and avoid the effort associated with ellipsis resolution by using the full form with an overt pronoun in the prefield.

(iv–Verb Type) From the theoretical claim in the IDS grammar (Zifonun et al. 1997: 415), we can derive the prediction of an interaction between the predictors Verb Type and grammatical Person. Topic drop of the 1st person should be particularly frequent before copular, auxiliary, and modal verbs. From the information-theoretic perspective, it may be that some verb types are generally less predictable in the left bracket and this may impact whether the prefield constituent is realized or omitted. Thus, the information-theoretic account could explain an effect of Verb Type in addition to the more fine-grained surprisal effect.

11.4.4.2 Results

Details on the logistic regression analysis of all verb types can be found in Section 10.4.4.2. The final model (repeated here as Table 11.12) contained a significant interaction between grammatical Person and Inflection Informed ($\chi^2(1) = 5.85, p < 0.05$), a significant interaction between Person and Verb Type ($\chi^2(1) = 4.72, p < 0.05$), and significant main effects of Verb Type ($\chi^2(1) = 6.3, p < 0.05$) and Surprisal ($\chi^2(1) = 17.85, p < 0.001$). Topic drop of a 1st person singular subject pronoun is more likely if the following verb is practically unambiguous and if it is a copular, auxiliary, or modal verb. Topic drop is generally, i.e., independently of Person, less likely before copular, auxiliary, and modal verbs (see Figure 11.4). Additionally, the likelihood of topic drop decreases with the surprisal of the verb in the left bracket.

In the second step, I conducted a similar logistic regression analysis of only the lexical verbs. The full model predicted the likelihood of topic drop from grammatical Person, Surprisal, and Inflection Informed, as well as from their

Table 11.12: Fixed effects in the final model analyzing FRAC-TD-SMS-PART (repeated from page 287)

Fixed effect	Est.	SE	χ^2	p-value	
INTERCEPT	2.81	0.57	27.20	< 0.001	***
PERSON	0.35	0.29	1.46	> 0.2	
SURPRISAL	−0.27	0.07	17.85	< 0.001	***
VERB TYPE	0.95	0.39	6.30	< 0.05	*
INFLECTION INFORMED	−0.33	0.31	1.15	> 0.2	
PERSON × VERB TYPE	−1.30	0.61	4.72	< 0.05	*
PERSON × INFLECTION INFORMED	1.47	0.63	5.85	< 0.05	*

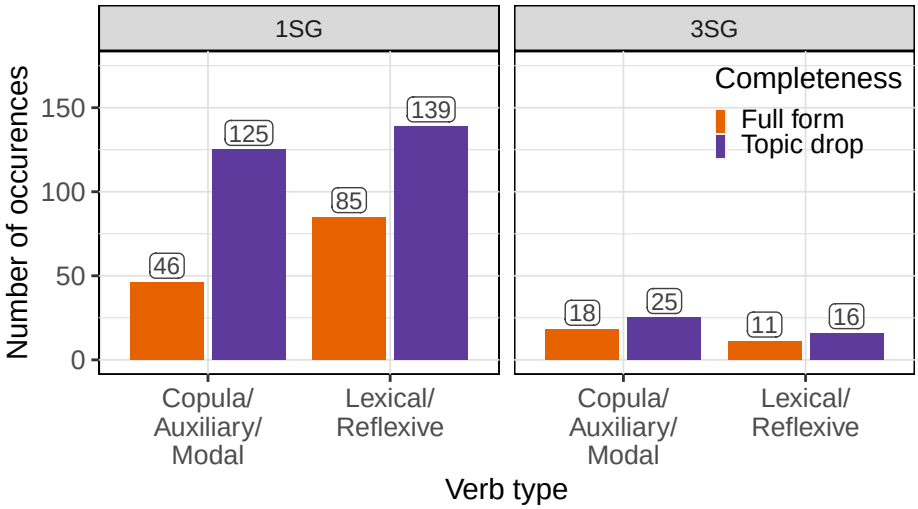


Figure 11.4: Frequency of the full forms and the instances of topic drop as a function of grammatical PERSON and VERB TYPE in the FRAC-TD-SMS-PART data set

two-way interactions. For reasons of data sparsity, I was unable to include the contrast Inflection Distinct. The final model contained a significant main effect of verb surprisal ($\chi^2(1) = 13.85, p < 0.001$), as shown in Table 11.13. Even when only the lexical verbs are considered, the likelihood of topic drop decreases if the verb surprisal increases. All other effects were not significant. That is, for only the instances with lexical verbs, neither grammatical Person nor Inflection informed impacted the likelihood of topic drop.

Table 11.13: Fixed effects in the final model analyzing only the instances with a lexical verb in FraC-TD-SMS-Part

Fixed effect	Est.	SE	χ^2	p-value	
Intercept	3.21	0.79	19.06	< 0.001	***
Surprisal	−0.28	0.08	13.85	< 0.001	***

11.4.4.3 Discussion

In the main analysis, I found a significant interaction between verb type and grammatical person, which supports the prediction (iv–Verb Type), motivated by the claim of the IDS grammar (Zifonun et al. 1997). Topic drop of the 1st person singular indeed seems to be more likely before copular, auxiliary, and modal verbs. The frequency table presented in Section 11.4.3 suggests that this effect is mainly caused by the copular verbs.[24] Topic drop of the 1st person singular is only particularly frequent before this verb type compared to the 3rd person singular. This interaction is qualified by the unexpected main effect of Verb Type in the opposite direction. When averaging over the other predictors, topic drop before copular, auxiliary, and modal verbs is generally less likely than before lexical and reflexive verbs. Moreover, since there are relatively few cases with the 3rd person singular in the data for both groups of verb types overall (see again Figure 11.4), the interaction should be interpreted cautiously. The claim from the IDS grammar should be investigated in future research on a larger data set.

In line with prediction (iii–Surprisal), both analyses evidenced a significant main effect of the surprisal of the following verb on the frequency of topic drop. Topic drop is less likely before a verb with a high surprisal than before a verb with

[24]The numbers for the individual verb types are too small for statistical comparison, especially for the 3rd person singular.

a low surprisal. I found this effect when looking at all verb types and when looking at only the lexical verbs. It provides support for the information-theoretic account of the usage of topic drop, in particular for the *avoid peaks* principle. The writers of text messages are indeed more likely to overtly realize the prefield constituent if the following verb has a high surprisal, i.e., if it causes higher processing effort for the receiver. On the one hand, the overt prefield constituent saves the receiver the cognitive effort required to resolve the ellipsis on the verb. On the other hand, it can also make the following verb more predictable and, thus, reduce its surprisal. In this way, the surprisal peak on the verb can be lowered and the receiver's processing can be facilitated. This result of a consistent surprisal effect in both corpus analyses provides the first genuine empirical support for my information-theoretic account of topic drop usage.

In Section 10.4.4.3, I already discussed the results of the regression analysis with respect to grammatical person and verbal inflection. There, I argued that the likelihood of topic drop is higher for the 1st person singular than for the 3rd person singular if it precedes a verb form perceived as unambiguous. In the analysis of only the lexical verbs, which I presented in this section, I did not replicate this result. One reason could be that the observed effects are primarily due to the other verb types. Another reason could be the small size of the data set. More in-depth studies with larger data sets are needed here.

In summary, the results of the corpus analysis are only partially consistent with the predictions from the IDS grammar for verb type. Even if larger data sets in future studies were to provide clearer evidence, it would nevertheless remain an open question why topic drop of the 1st and 2nd person should be more frequent before copular, auxiliary, and modal verbs. Zifonun et al. (1997) do not provide a theoretical explanation for their prediction. In contrast, my information-theoretic approach does explain the attested influence of verb surprisal on the frequency of topic drop. Two analyses showed that the rate of topic drop decreases as the surprisal of the verb in the left sentence bracket increases and, thus, the effort required to process it. This result is consistent with the predictions of the *avoid peaks* principle and provides the first genuine evidence for the information-theoretic approach since the effect cannot be explained by any theoretical approach to the usage of topic drop discussed in the literature so far.

11.5 Experimental investigations of verb surprisal

The corpus study on the text message subcorpus of the FraC evidenced an effect of verb surprisal on the frequency of topic drop (see Section 11.4.4). Topic drop was rarer if the following verb had a high surprisal. In two experiments, I attempted to show the effect of verb surprisal that I found in the production of topic drop also in its perception. To this end, I tried to explicitly manipulate the surprisal of the following verb in a controlled setting.

11.5.1 Experiment 11: verb surprisal (context)

In the first experiment, I used the likelihood of the verb given the preceding linguistic context to manipulate its surprisal.[25]

To this effect, I exploited the stereotypical association between certain professions and typical actions or tasks that I had beforehand confirmed with a pretest. The assumption is that the knowledge about the profession of a person makes it more or less likely for this person to perform a certain action or task, impacting the surprisal of the verb denoting this action or task. For example, *to program* should be more likely for a computer scientist than for a bricklayer, and, thus, have a lower surprisal. If we consider again the *avoid peaks* principle discussed in Section 6.2.2, this means that topic drop should be more likely and, thus, more acceptable if the omitted subject before *to program* is a pronoun referring to a computer scientist than if it refers to a bricklayer. I argue that this is because, in the case of the bricklayer, the overt subject pronoun in the prefield may be needed to lower the processing effort on the verb by, first, indicating structurally that a congruent finite verb will follow and, second, by saving the effort required to resolve the ellipsis. The main experiment had the form of a 2×2 design crossing the factors Completeness (topic drop vs. full form) and Profession (predictive vs. non-predictive for the main verb). With a pretest, I assessed whether the relations between professions and verbs do indeed hold.

11.5.1.1 Pretest

The pretest was intended to ensure that for each token set, the verb in the target utterance was indeed more likely after a predictive profession than after a non-predictive profession. Using the example from above, *programmieren* (‘to

[25] The items and fillers, as well as the analysis scripts of both the pretest and the main experiment can be found online: https://osf.io/zh7tr.

program') should be more likely after *Informatiker* ('computer scientist') than after *Maurer* ('bricklayer'). I presented the participants the materials in either the predictive or the non-predictive condition (predictor PROFESSION) and assessed the likelihood of (the beginning of) the answer, given the question and the introduction.

11.5.1.1.1 Items

I constructed 38 token sets such as (1), which were always built around a pair consisting of a profession such as *Informatiker* ('computer scientist') and a verb denoting an action that is typically performed by a person carrying on this profession such as *programmieren* ('to program'). This pair was then complemented by a second profession that typically does not perform this action as part of their job such as *Maurer* ('bricklayer').

(1) a. (i) *Anna schreibt mit Jan, der Informatiker ist:*
Anna writes with Jan who computer.scientist is
'Anna is texting with Jan, who is a computer scientist:' (predictive)
(ii) *Anna schreibt mit Jan, der Maurer ist:*
Anna writes with Jan who bricklayer is
'Anna is texting with Jan, who is a bricklayer:' (non-predictive)
b. *Anna: Was gibt's Neues bei dir?*
what gives.it new at you.DAT.2SG
Anna: 'What's new with you?'
c. *Jan: Ich programmiere #####.*
I program
Jan: 'I am programming #####'

The items consisted of an introductory sentence and a question-answer pair. The first sentence (1a) introduced the two interlocutors, e.g., Anna and Jan, and specified the profession of the second person, e.g., Jan was either a computer scientist in the predictive condition (1a-i) or a bricklayer in the non-predictive condition (1a-ii). This sentence was then followed by a general question asked by the first person, e.g., Anna (1b). The second person, e.g., Jan, answered this question with a declarative V2 clause with a finite lexical main verb in the left bracket, e.g., *programmiere* (1c). This verb was the same for the two PROFESSION conditions, but in the predictive condition, it matched the profession, while it did not in the non-predictive condition. In the pretest, I presented the answer as the full form, i.e.,

with the 1st person singular subject pronoun *ich* in the prefield, and presented the placeholder ##### after the finite verb, as shown in Figure 11.5.

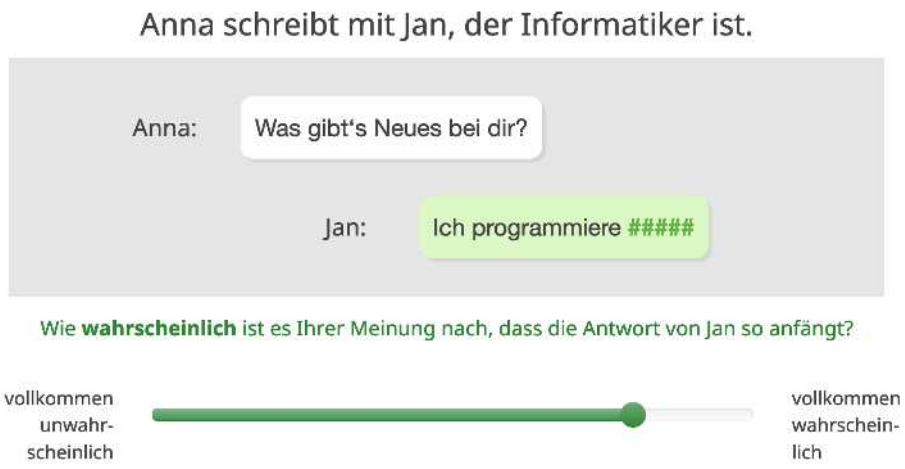


Figure 11.5: Presentation of the pretest of experiment 11 using the item shown in (1)

11.5.1.1.2 Procedure

The experiment was conducted online using LimeSurvey (Limesurvey GmbH 2023). I recruited 48 participants, native speakers of German between the ages of 18 and 40, from Clickworker (Clickworker 2022), who were paid €1.30. They read the introductory sentence with the profession, the question by person A, and the answer by person B cut off after the verb in the left bracket. The participants' task was to rate how probable it is that the answer of person B would begin like this, given person A's question and the introductory sentence. They rated this likelihood using a slider scale with labeled endpoints (from "vollkommen unwahrscheinlich" ('completely unlikely') to "vollkommen wahrscheinlich" ('completely likely')). The slider ranged from 0 to 100, but the values were not visible to participants, i.e., they had to orient themselves only visually, as shown in Figure 11.5.

For each participant, the experiment started with the same two filler trials which were not analyzed. They mimicked the critical items and had the purpose of familiarizing the participants with the task. The items were mixed with 9 catch trials, where the main verb in the answer was a verb that usually has no human subject, such as *laichen* ('to spawn') or *glimmen* ('to smolder'). Consequently, it

should be unlikely for an utterance to start with *ich* plus the congruent form of such a verb.

The items were distributed across two lists so that each participant rated 19 predictive and 19 non-predictive conditions. Due to an error, there were actually 20 predictive and 18 non-predictive items in one list, because a token set appeared in both lists in the predictive condition. This token set was excluded from further analysis. The materials were presented in individually randomized order.

11.5.1.1.3 Results

I excluded the data from two participants whose mean score for the catch trials deviated by more than 2.5 standard deviations from the overall mean score for the catch trials. Table 11.14 shows the mean ratings and standard deviations for the two conditions of the items and the catch trials. The items received higher scores than the catch trials in both conditions and the items received higher scores in the predictive condition than in the non-predictive condition.

Table 11.14: Descriptive overview of the likelihood scores for the items and catch trials in the pretest of experiment 11

	Mean score	Standard deviation
Items, predictive	82.82	20.45
Items, non-predictive	29.77	25.99
Catch trials	8.31	13.45

This visual impression was supported by the statistical analysis. I fitted a linear mixed effects regression model using the lme4 package (Bates et al. 2015) in R with random intercepts for participants[26] to compare the ratings for the critical items to the ratings for the catch trials. I modeled the numeric likelihood scores as a function of the dummy coded predictor Stimulus Type with catch trial as the reference level, which was compared to the critical items. The full and likewise final model (see Table 11.15) revealed a significant main effect of Stimuli Type indicating that the items were generally rated higher than the fillers ($\chi^2(1) = 601.49$, $p < 0.001$).

With a second linear mixed effects regression model, I compared the ratings for the two Profession conditions with each other. I focused on only the items and included the likelihood scores as the dependent variable and Profession as the

[26]More complex random effects structures resulted in singular fits.

Table 11.15: Fixed effect in the final LMER of the pretest of experiment 11 considering Stimulus Type

Fixed effect	Est.	SE	χ^2	p-value	
Stimulus Type.Item	47.99	1.81	601.49	< 0.001	***

independent variable using dummy coding with non-predictive as the reference level. In the full and likewise final model (see Table 11.16), the condition Profession had a significant main effect on the probability scores ($\chi^2(1) = 1602.8$, $p < 0.001$). The beginning of the answer received higher probability scores in the predictive condition than in the non-predictive condition. These results indicate that overall the manipulation worked in the intended way.

Table 11.16: Fixed effect in the final LMER of the pretest of experiment 11 considering Profession

Fixed effect	Est.	SE	χ^2	p-value	
Profession.predictive	53.03	1.02	1602.8	< 0.001	***

On the level of individual token sets, there were clear differences though, as can be seen from Figure 11.6.[27] It shows the mean likelihood scores per token set and Profession condition. For convenience, I added the verbs and the professions in the predictive and non-predictive conditions to the plot.

For example, token set 4 compared the likelihood of the verb *komponieren* ('to compose') for the predictive profession *Musiker* ('musician') to the non-predictive profession *Surflehrer* ('surf instructor'). The mean rating for the predictive condition was 89, while it was 11 for the non-predictive one, resulting in a difference of 78 points. In contrast, the predictive condition of token set 26 received a mean rating of 66 (*Kosmetikerin* ('cosmetologist.fem') and *zupfen* ('to pluck')), the non-predictive condition one of 47 (*Winzerin* ('winemaker.fem') and *zupfen* ('to pluck')), which made a difference of only 19 points. To account for these differences between token sets and to make the contrast between the two item conditions as obvious as possible, I decided to reduce the token sets to those 32 where the difference in terms of z-scores between the predictive and the non-predictive

[27] It should be noted that some of the German equivalents of the verbs have a clearer meaning and therefore match the occupations in the condition better than is suggested by the English translation.

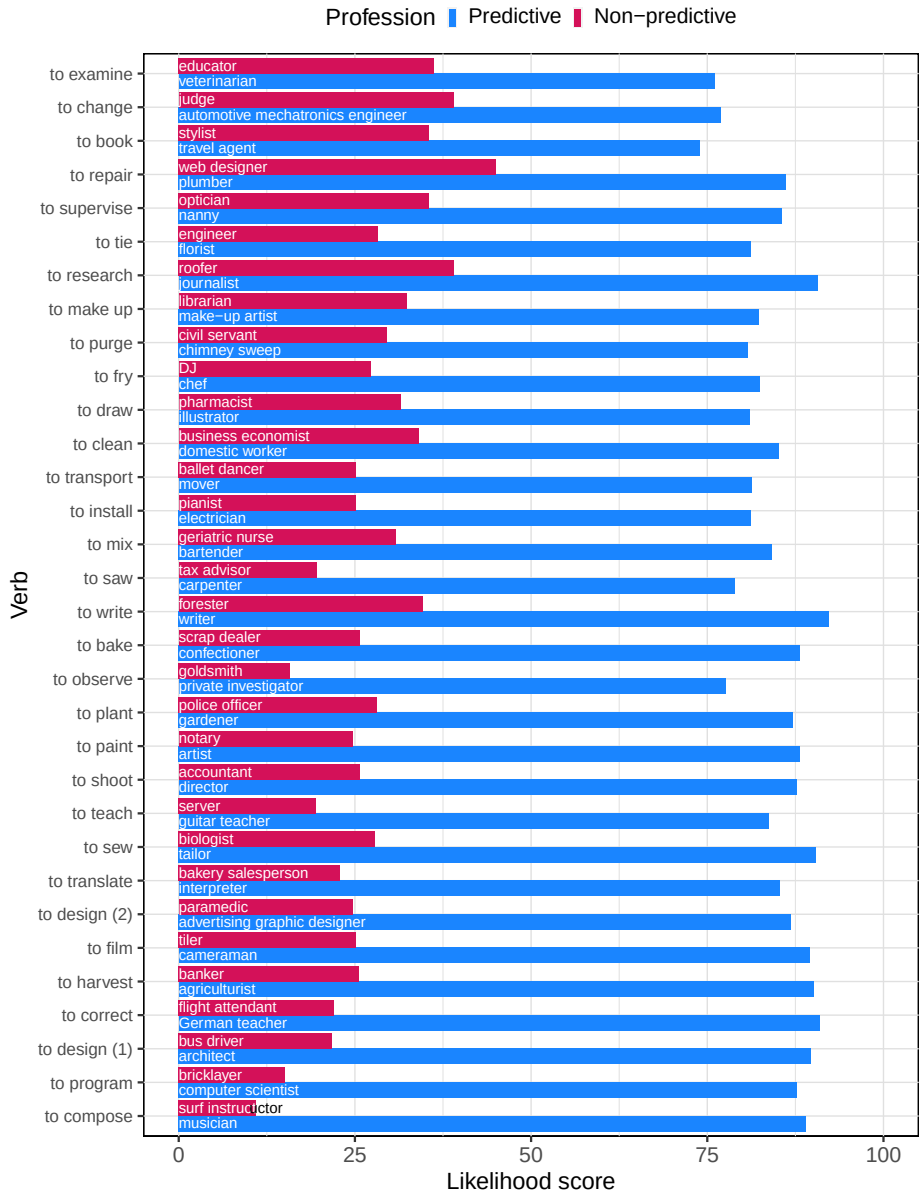


Figure 11.6: Likelihood scores of the 32 token sets selected based on the pretest as a function of the PROFESSION predictor, including verbs and professions

condition was the highest. For these 32 token sets, the difference in the absolute scores between the predictive and the non-predictive condition ranged between a minimum of 37.91 and a maximum of 78.04. There is a clear dichotomy between predictive and non-predictive conditions, which justifies using a categorical predictor for PROFESSION in the analysis of the rating study in Section 11.5.1.4 below.

11.5.1.2 Materials

Items

The 32 token sets selected by the pretest were completed by replacing the placeholder ##### with a meaningful continuation, as shown in example (2). For the target utterance, I varied whether the prefield constituent, the 1st person singular pronoun *ich* ('I'), was realized overtly or covertly, i.e., whether the utterance contained topic drop or not.

(2) a. (i) *Anna schreibt mit Jan, der Informatiker ist:*
Anna writes with Jan who computer.scientist is
'Anna texts with Jan, who is a computer scientist:' (predictive)

(ii) *Anna schreibt mit Jan, der Maurer ist:*
Anna writes with Jan who bricklayer is
'Anna texts with Jan, who is a bricklayer:' (non-predictive)

b. *Anna: Was gibt's Neues bei dir?*
what gives.it new at you.DAT.2SG
Anna: 'What's new with you?'

c. *Jan: (Ich) programmiere gerade eine Software.*
I program just a software
Jan: '(I) am currently programming a piece of software.'
(topic drop / full form)

Fillers

Alongside the 32 critical items for this experiment, I tested a total of 64 fillers in the actual rating study, which had the same overall structure as the items: an introductory sentence followed by a question-answer pair. 16 fillers were the items of another experiment on preposition omission. The introductory sentence specified that one person is texting with a group of people. The question-answer pair contained a fragmentary answer for which the definiteness and the presence or absence of a preposition were varied. 24 fillers were question-answer pairs with

V2 or verb-final subordinate clauses with (potential) gapping structures in the answers, where the introductory sentence explained where the two interlocutors know each other from. The final group of fillers were 24 further question-answer pairs. The introductory sentence explained again where the two interlocutors know each other from, while the answer was a declarative sentence with an adverbial in the prefield.

11.5.1.3 Procedure

The main rating study was conducted as an online study implemented with LimeSurvey (Limesurvey GmbH 2023). 48 participants, native German speakers (18 to 40 years old) who had not participated in any other of my topic drop experiments, were recruited from Clickworker (Clickworker 2022) and received €4.00. Their task was to rate the naturalness of the answer in the context of the question and the introductory utterance on a 7-point Likert scale (7 = completely natural). The 32 items selected based on the pretest were distributed across four lists according to a Latin square design and mixed with 64 fillers. Each participant saw the materials in an individually pseudo-randomized order, ensuring that no two items immediately followed each other. The question-answer pairs of all materials were presented as instant messaging dialogues, as shown for the pretest in Figure 11.5.

11.5.1.4 Results

Table 11.17 shows the mean ratings and standard deviations per condition. In Figure 11.7, the mean ratings and 95% confidence intervals are plotted.

Table 11.17: Mean ratings and standard deviations per condition for experiment 11

Completeness	Profession	Mean rating	Standard deviation
Full form	Predictive	5.41	1.55
Topic drop	Predictive	5.49	1.53
Full form	Non-predictive	3.40	1.89
Topic drop	Non-predictive	3.34	1.85

Their visual inspection indicates an enormous difference between the two Profession conditions, which seems to be independent of the Completeness

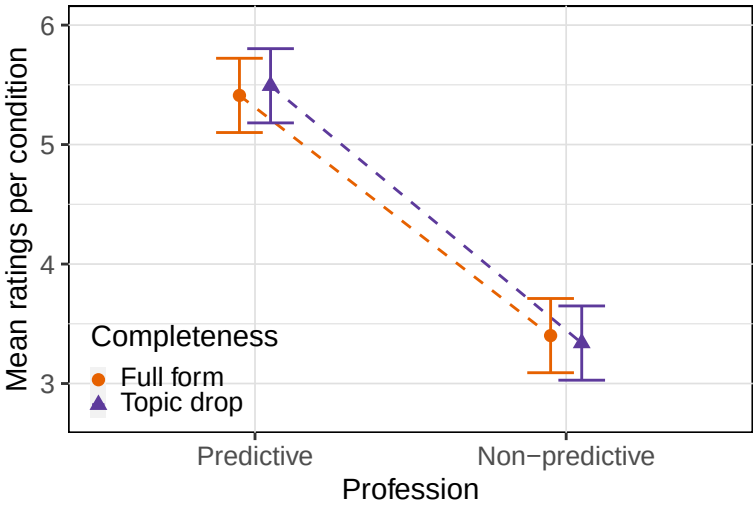


Figure 11.7: Mean ratings and 95% confidence intervals per condition for experiment 11

condition. Utterances in the non-predictive conditions were clearly degraded, regardless of whether they were syntactically complete or contained topic drop.

I analyzed the data with CLMMs in R (Christensen 2019), as described in Section 3.1.6.3. The full model contained the ratings as the dependent variable and as the independent variables Completeness and Profession,[28] coded with deviation coding (full form and predictive coded as 0.5, topic drop and non-predictive coded as −0.5), as well as the numeric scaled and centered Position of the trial in the experiment and all two-way interactions between the predictors. The random effects structure consisted of random intercepts for participants and items and of by-participant and by-item random slopes for Completeness, Profession, and their interaction.[29]

The final model obtained with a backward model selection had symmetric thresholds and contained the following significant fixed effects (see Table 11.18): a significant main effect of Profession ($\chi^2(1) = 44.51$, $p < 0.001$), indicating that

[28] As discussed above in Section 11.5.1.1, I used a categorical predictor for Profession because there was a clear split between predictive and non-predictive conditions in the results of the pretest.

[29] The formula of the full model was as follows: `Response ~ (Completeness + Profession + Position)^2 + (1 + Completeness * Profession | Subjects) + (1 + Completeness * Profession| Items)`. Models including random slopes for Position did not converge.

the predictive conditions were highly preferred over the non-predictive conditions and an interaction between Profession and Position ($\chi^2(1) = 15.3, p < 0.001$), according to which the ratings for the predictive conditions improved in the course of the experiment, just as the ratings for the non-predictive conditions became worse. The latter result suggests that the participants adapted their ratings more and more to the Profession manipulation. The interaction between Completeness and Profession ($\chi^2(1) = 1.55, p > 0.2$) was not significant nor were any other effects.

Table 11.18: Fixed effects in the final CLMM of experiment 11

Fixed effect	Est.	SE	χ^2	p-value	
Profession	3.08	0.38	44.51	< 0.001	***
Position	0.02	0.05	0.12	> 0.7	
Profession × Position	0.40	0.10	15.30	< 0.001	***

11.5.1.5 Discussion

Experiment 11 was intended to test the effect of verb surprisal on the acceptability of topic drop by exploiting the assumed strong association between a certain profession and a typical activity performed by this profession. The idea was that a verb denoting the typical activity has a lower surprisal in the context of the predictive profession than in the context of another unrelated, non-predictive profession. As a consequence, topic drop should be more acceptable before the verb in the context of the predictive profession than in the context of the non-predictive profession. This is because, after the non-predictive profession, it may be necessary to reduce the processing effort on the verb by inserting an overt preverbal constituent. This constituent is argued to reduce the peak on the verb because it may increase the likelihood of that verb and it saves the effort required to resolve the ellipsis.

The results of the experiment do not support this hypothesis. Topic drop and the full forms were rated comparably. The ratings for both Completeness conditions were equally impacted by the manipulation of Profession but this in a massive way. Participants more strongly preferred the target utterances in the context of a predictive profession than in the context of the non-predictive profession, i.e., when the action denoted by the verb was typical for this profession. This preference even increased in the course of the experiment. It could be the

case that this massive effect of the Profession predictor masked any potential influence that the verb surprisal might have on the acceptability of topic drop, i.e., that participants focused on the different professions, neglecting the variation between full forms and topic drop. More abstractly, the manipulation relying on world knowledge and pragmatics might have overridden the more subtle variation in syntactic form. For this reason, I chose a different way to manipulate verb surprisal in the following experiment.

11.5.2 Experiment 12: verb surprisal (verb type)

Experiment 11 indicated no effect of verb surprisal on the usage of topic drop, but this might be caused by properties of the experimental design. It might be that the context manipulation had such a massive effect on the acceptability of the target utterances that any possible difference between full forms and topic drop was masked by it. Therefore, in experiment 12, I manipulated the verb surprisal not via the utterance context but via the verb type used, which I varied between a lexical and an auxiliary verb. This resulted in a 2×2 acceptability rating study, crossing Completeness (full form vs. topic drop) and Verb Type (lexical verb vs. auxiliary).[30]

In Section 11.4.3, I showed that there is a correlation between verb type and verb surprisal. Since auxiliary verbs such as *haben* ('have'), which are used to build the perfect for most verbs, are very frequent, they generally have a low unigram surprisal. Lexical verbs, in turn, vary more strongly in terms of their surprisal, but they are generally expected to be less frequent and, thus, have a higher surprisal. It follows from the *avoid peaks* principle (Section 6.2.2) that topic drop should on the whole be more acceptable before unsurprising auxiliary verbs than before more surprising lexical verbs. Recall that also Zifonun et al. (1997: 415) claim that topic drop of the 1st and the 2nd person is more frequent or more acceptable before copular, auxiliary, and modal verbs, however without substantiating this claim (see Section 11.1). Since in this experiment I only tested utterances with topic drop where the 1st person singular pronoun *ich* ('I') was omitted, this claim would likewise predict that topic drop before auxiliaries should be rated as more acceptable than topic drop before lexical verbs.

[30] All items, fillers, and the analysis script can be accessed online: https://osf.io/zh7tr.

11.5.2.1 Materials

Items

I adapted the 32 items of experiment 11 by varying the tense in the target utterance between the original present tense (3b-i) and the present perfect (3b-ii). As a result, the verb in the left bracket was once a lexical verb with a higher surprisal and once an auxiliary verb with a lower surprisal, namely *haben* ('have').[31] Additionally, I removed the introductory sentence and changed the temporal adverbial in the prefield for some items to make the perfect conditions more natural. For example, I changed *gerade* ('just') to *diese Woche* ('this week') in (3).

(3) a. *Anna: Was gibt's Neues bei dir?*
what gives.it new at you.DAT.2SG
Anna: 'What's new with you?'

b. (i) *Jan: (Ich) programmiere diese Woche eine Software.*
I program this week a software
Jan: '(I) am programming a piece of software this week.'
(lexical verb)

(ii) *Jan: (Ich) habe diese Woche eine Software programmiert.*
I have this week a software programmed
Jan: '(I) have programmed a piece of software this week.'
(auxiliary verb)

Fillers

The same 64 fillers as in experiment 11 were used, complemented by 8 ungrammatical catch trials.

11.5.2.2 Procedure

Another 48 native speakers of German were recruited from Clickworker (Clickworker 2022) and received €4.00 to participate in the study, which was again implemented with LimeSurvey (Limesurvey GmbH 2023). They were between the ages of 18 and 40 and had not taken part in any of my previous studies of topic drop. Their task was again to rate the last utterance of the instant messaging dialogues in terms of its naturalness on a 7-point Likert scale (7 = completely natural). The 32 items were distributed across 4 lists according to a Latin square

[31] While in German some verbs form the perfect with *sein* ('be'), all verbs in this experiment use *haben*.

design to ensure that each participant rated each token set only once and in one condition. The items were mixed with the 64 fillers and presented in individual pseudo-randomized order as instant messaging dialogues.

11.5.2.3 Results

I excluded 5 participants who had rated 4 or more of the 8 ungrammatical catch trials as (almost) completely natural, i.e., with 6 or 7 on the 7-point Likert scale. Table 11.19 and Figure 11.8 provide an overview of the aggregated data for the remaining 43 participants, broken down by condition. It looks as if there is no interaction, just a difference for VerbType and one for Completeness. Full forms seem to be more acceptable than topic drop and, unexpectedly, utterances with lexical verbs seem to be preferred over utterances with auxiliaries.

Table 11.19: Mean ratings and standard deviations per condition for experiment 12

Completeness	Verb Type	Mean rating	Standard deviation
Full form	Lexical verb	5.16	1.79
Topic drop	Lexical verb	5.36	1.65
Full form	Auxiliary verb	4.76	1.79
Topic drop	Auxiliary verb	5.02	1.75

I used CLMMs (package ordinal (Christensen 2019)) to analyze the data, as presented in Section 3.1.6.3. I modeled the ordinal ratings as a function of the numeric Position of the trial in the experiment and of the two binary predictors Completeness and VerbType, as well as of all two-way interactions between them. Position was scaled and centered. Completeness and VerbType were coded using deviation coding (full form and lexical verb as 0.5, topic drop and auxiliary as −0.5). The random effects structure consisted of random intercepts for subjects and items and of by-subject and by-item random slopes for all three predictors and their two-way interactions.[32]

The final model obtained with a backward model selection had flexible thresholds and contained the fixed effects shown in Table 11.20. There were significant

[32]The formula of the full model was as follows: `Ratings ~(Completeness + VerbType + Position)^2 + (1 + (Completeness + VerbType + Position)^2 | Subjects) + (1 + (Completeness + VerbType + Position)^2 | Items)`.

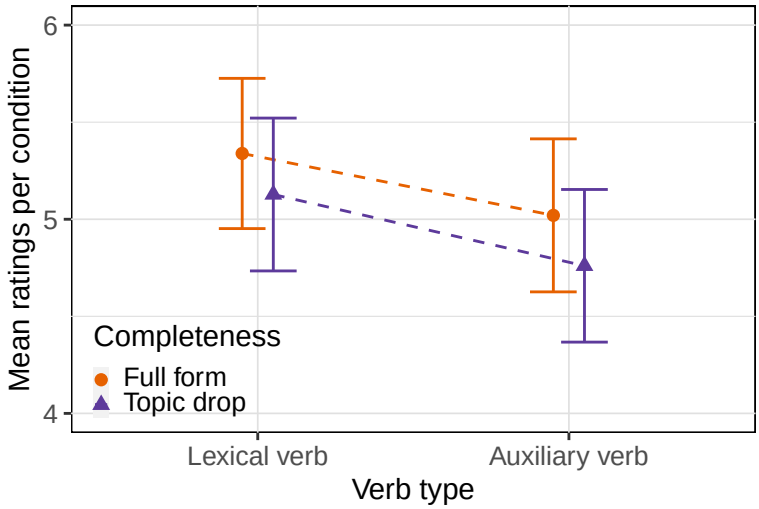


Figure 11.8: Mean ratings and 95% confidence intervals per condition for experiment 12

main effects of both binary predictors Completeness ($\chi^2(1) = 4.34, p < 0.05$) and Verb Type ($\chi^2(1) = 10.96, p < 0.001$), but their interaction was not significant ($\chi^2(1) = 0.38, p > 0.5$). Full forms were preferred over topic drop and utterances with lexical verbs were rated as more acceptable than utterances with auxiliaries.

Table 11.20: Fixed effects in the final CLMM of experiment 12

Fixed effect	Est.	SE	χ^2	p-value	
Completeness	−0.40	0.19	4.34	< 0.05	*
VerbType	0.55	0.16	10.96	< 0.001	***

11.5.2.4 Discussion

Experiment 12 had the purpose of testing again for an effect of verb surprisal on the acceptability of topic drop. It relied on the observation that auxiliary verbs are typically more frequent and, therefore, have a lower surprisal than lexical verbs. Based on the *avoid peaks* principle, this led to the prediction that topic drop should be more acceptable before auxiliary than before lexical verbs.

This prediction was not confirmed by the experimental results, since I found no significant interaction in my data. Somewhat unexpectedly, participants overall preferred utterances with lexical verbs over utterances with auxiliaries, but there was no difference between full forms and topic drop in this respect.

An explanation for why topic drop was not, as predicted, more acceptable before auxiliary verbs, and for the main effect of VerbType is to assume that the participants considered the form *habe* of the auxiliary, which was used in the experimental items, as less natural than the forms of the lexical verbs because they would have expected the more colloquial form *hab* without final *schwa*. This may be relevant for two reasons. First, if the inflected form of the auxiliary verb is unexpected in the given form of communication / text type, it might not have a lower surprisal than the corresponding lexical verbs. This means that the acceptability of topic drop might not vary between conditions because there is an equally strong need to insert an overt prefield constituent to facilitate the processing of the following verb. Second, it could be that the form *habe* is marked in terms of style in the conceptually spoken text type instant messages. If this were the case, this would account for the main effect, i.e., that the full forms and the instances of topic drop with *habe* are degraded.

To check the assumption that *habe* is unnatural in instant messages, I searched the Mobile Communication Database 2 (MoCoDa2) (Beißwenger et al. 2020) for both verb forms *habe* and *hab*. MoCoDa2 is a still growing corpus of WhatsApp instant messages, which at the time of my search in August 2023 consisted of about 980 chats with about 38 000 messages and 309 000 tokens.

A search for *habe* resulted in 683 matches in 287 chats while searching for *hab* yielded 1 300 results in 436 chats. This shows that while *hab* was indeed the preferred form with twice as many matches, *habe* was by no means infrequent as it still occurred in more than a quarter of all chats in the MoCoDa2. *Habe* was also more frequent in this corpus than the forms of all lexical verbs used in the experiment. Of these, the most frequent was *schreibe* ('write.1sg') with 40 occurrences, while several others such as *komponiere* ('compose.1sg'), *entwerfe* ('design.1sg') and *observiere* ('keep.1sg under surveillance') did not occur at all in the corpus. This indicates that although *habe* may be less frequent than the more colloquial *hab* in WhatsApp messages, it was still a lot more frequent than all of the inflected forms of the lexical main verbs. Consequently, it should also have a lower unigram surprisal. Still, though, topic drop was not more acceptable before *habe* than before the lexical verbs.

In sum, both experiments 11 and 12 failed to experimentally evidence an impact of verb surprisal. Further research is required to investigate whether such an

effect is indeed absent or whether only better methods or research designs are required to find it.

11.6 Summary: verb type and verb surprisal

In this chapter on the role of verb type and verb surprisal for topic drop, I first summarized the existing sparse literature on the matter. While there is so far no explanatory approach in the theoretical literature for a possible effect of the following verb on topic drop, my information-theoretic approach provides one, as repeated below. For verb type, there are only two unsubstantiated claims in the theoretical literature, one by the IDS grammar (Zifonun et al. 1997) and one by Imo (2013). According to Zifonun et al. (1997), subject topic drop of the 1st and 2nd person is supposed to be more frequent before copular, auxiliary, and modal verbs. Imo (2013) claims that topic drop is generally more frequent before verbs of saying and thinking.

Of the four previous corpus studies that I discussed, only one provides tentative support for Imo's (2013) claim. Only Helmer (2016) found that verbs of saying and thinking are among the four most frequent verb classes to occur with topic drop. In contrast, all corpus studies are at least consistent with the theoretical claim by Zifonun et al. (1997) that subject topic drop of the 1st and the 2nd person is particularly frequent before copular verbs and modal verbs. The results for topic drop before auxiliaries were mixed (not least because of the surprisingly low number of auxiliary verbs in Frick's (2017) corpus) and still need additional investigation.

My empirical investigations aimed to examine this more systematically and to investigate the extent to which my information-theoretic account can explain the data. In the corpus study, I found no evidence for Imo's (2013) hypothesis that topic drop is particularly frequent before verba dicendi and sentiendi. Concerning the hypothesis by the IDS Grammar (Zifonun et al. 1997) that in particular 1st and 2nd person subject topic drop is more frequent before copular verbs, auxiliaries, and modal verbs than before lexical verbs, the evidence from my corpus study was mixed. While 1st person singular topic drop was indeed more frequent before these verb types in my text message data, the visual inspection has suggested that the effect was mainly caused by the copular verbs. Since I did not test copular verbs in my experiments,[33] this might also explain why I did not find comparable effects of verb type there. In experiment 12 presented in this chapter,

[33] In these experiments, I was less interested in the verb type than in the inflection (experiments 7 and 8) or surprisal (experiment 12) of the verb. Testing copular verbs could be done in a future study.

there was no topic drop-specific difference between lexical verbs and auxiliaries. In Section 10.5.3, I discussed that topic drop was found to be equally acceptable before lexical verbs and before modal verbs in experiments 7 and 8. However, those experiments did not focus on the effect of verb type but on whether a distinct inflectional ending on the verb impacts the acceptability of topic drop. There was a one-to-one mapping between verb type and inflection so their effects could not be separated. Every form of a modal verb tested was syncretic and vice versa, just as every form of a lexical verb was distinct and the other way around. Thus, an explicit experimental investigation of different verb types has yet to be conducted.

For verb surprisal, the corpus study revisited in Section 11.4.4 provided robust evidence that topic drop is less likely before verbs with a high surprisal, as predicted by the *avoid peaks* principle. I found this effect in an analysis including all verb types and replicated it in an additional analysis of the lexical verbs only. The latter analysis rules out potential confounds due to the correlation between surprisal and verb type. The result provides the first genuine evidence for my *UID*-based information-theoretic account.

In my experimental studies, experiments 11 and 12, I could not strengthen this evidence. When I tried an exploratory approach of manipulating the surprisal of individual lexical verbs via the context in experiment 11, I did not find a topic drop-specific difference in acceptability. I argued that the pragmatic manipulation of the likelihood of verbs using world knowledge was probably too strong and thus obscured a possible difference concerning the variation between full forms and topic drop. As mentioned above, also the manipulation of surprisal via the verb type, auxiliaries as less surprising than lexical verbs, did not result in an effect on the acceptability of topic drop. Therefore, experimental evidence of a surprisal effect on topic drop is still pending.

All in all, my information-theoretic approach with the *avoid peaks* principle provides an explanation for the influence of the following verb on the usage of topic drop. Such an influence has been discussed sporadically in the theoretical literature for the type of this verb but has not yet been theoretically substantiated. My approach remedies this situation by predicting, on the one hand, an influence of the inflectional ending of this verb on the processing effort of the entire utterance, using the *facilitate recovery* principle discussed in Chapter 10. On the other hand, it also states that this processing effort is influenced or, more precisely, can be lowered by the presence of an overt prefield constituent. I was able to empirically demonstrate such an effect in my corpus study, independent of the influence of verb type. Thus, I provided genuine evidence for my information-theoretic approach, which should be further extended using different experimental methods.

12 General discussion

The goal of this work was to provide a comprehensive account of the ellipsis type topic drop in German that incorporates both aspects of licensing and aspects of usage. To this end, I followed a strongly empirical approach, relying on corpus and experimental data. After having defined topic drop, I focused on two research questions, each of which I will also summarize and discuss in a section below. At the same time, this book also contributes to a systematic discussion and, most importantly, an empirical investigation of isolated claims from the theoretical literature, in some cases for the first time. I will mention and highlight the corresponding results concerning these controversies at the appropriate places.

12.1 Definition and typological perspective

Before turning to the investigation of the two research questions, I first preliminarily defined topic drop in German as an ellipsis type that is restricted to the prefield of declarative V2 clauses. I discussed that topic drop is typically used in certain text types. This property is not part of its definition, but it is nevertheless characteristic of topic drop. By means of my corpus study on the text type-balanced fragment corpus FraC (Horch & Reich 2017), I could show that topic drop is not bound to mainly one specific text type, like null articles, null copulas (headlines), or object omission (recipes and instructions) are. Instead, there are several text types in which topic drop occurs frequently, such as text messages, colloquial dialogues, blogs, and online chats, many of which share several features that Koch & Oesterreicher (1985) describe as characteristic of conceptually spoken text types.

In the second step, I distinguished topic drop from similar phenomena such as *pro*-drop and V1 declaratives. In doing so, I was taking a stand on the first topic drop controversy because in the literature topic drop has occasionally been both analyzed as *pro*-drop and equated with V1 declaratives. Furthermore, I provided a typological overview of topic drop in other Germanic V2 languages and also discussed similarities of topic drop to omissions in English, French, and Russian, i.e., the left-peripheral positioning of topic drop and null subjects in English and

French and the typical occurrence in certain registers or text types of topic drop and the null arguments in English, French, and Russian. I argued that a joint consideration of, in particular, left-peripheral register- or text-type-dependent ellipsis types in several languages could be a promising perspective for future research.

12.2 When is topic drop licensed?

The first research question of this work was the question of when topic drop is syntactically licensed in German.

12.2.1 Prefield restriction as licensing condition

Here, I looked in detail at the prefield restriction of topic drop. I reviewed several hypotheses and observations from the literature and refined the restriction based on my results and on proposals by Rizzi (1994) and Freywald (2020). I argued on theoretical and empirical grounds that topic drop is not restricted to the deletion of topics in the prefield but may target any recoverable constituent in a prefield that is not c-commanded sentence-internally by a potential identifier or that is the highest prefield of a root clause.

In a first step, I rejected the widely held view in the literature that topic drop is restricted exclusively to topics, thereby touching upon another topic drop controversy. I first argued with Frey (2000) against characterizing the prefield position as the topic position in German main clauses. Second, I concluded, referring to the impossibility of omitting contrastive and non-recoverable topics, that topicality is not a (strictly) sufficient condition for topic drop. Building on corpus data and a rating study, I was furthermore able to show that non-referential and, thus, non-topical expletive subjects, especially those of weather verbs, can be targeted by topic drop. Consequently, I concluded that topicality is not a necessary condition for topic drop either. It would presumably be more accurate to term topic drop not *topic drop* but *prefield ellipsis*.

I argued that the prefield restriction of topic drop is purely positional and narrowed down its nature more precisely with the help of three experiments. In experiment 2, I compared topic drop in the prefield to corresponding omissions in the middle field and found that the omission from the prefield is significantly more acceptable. This result is in line with the prefield restriction of topic drop and contributes to clarifying a further topic drop controversy as it calls into question Helmer's (2016) claim that the omission of verb arguments in the middle

field should also be considered as topic drop. Experiment 3 provided evidence that topic drop is not licensed in dependent V2 clauses that follow the matrix clause. However, if the V2 complement precedes the matrix clause, topic drop is acceptable. It is not fully clear, though, what the syntax of preceding V2 complements exactly looks like. They can be analyzed as occupying the prefield (but see Reis 1997), as main clauses with an integrated V1 parenthetical, or as forming a quasi-paratactic structure together with another clause. Thus, since it cannot be conclusively clarified whether they are embedded clauses, it is also not possible to answer the question, controversially discussed in the literature, whether topic drop in embedded clauses is possible or not.

Following the proposals by Rizzi (1994) and Freywald (2020), I took the results concerning the prefield restriction and the dependent V2 clauses as evidence that topic drop cannot occur in any prefield but only in a prefield that is not c-commanded sentence-internally by a potential identifier or that is the highest prefield of a root clause. These proposals can also be reconciled with the results of experiment 4. This study provided evidence that conjunctions such as *und* and *aber* can precede topic drop, i.e., that it is not strictly restricted to the sentence-initial position. Even if the conjunction is taken to belong to the second conjunct, this conjunct would be a root clause and the prefield constituent would still not be c-commanded by a potential identifier but by a non-referential functional category. Referring to previous research, I discussed the three main generative approaches to modeling topic drop, the operator approach, the *pro*-approach, and the PF-deletion approach. After weighing their respective advantages and disadvantages, I argued, following Occam's Razor, for the PF-deletion approach because it can account for expletive omission while requiring the fewest additional assumptions.

In summary, I was able to define the prefield restriction of topic drop as its central syntactic licensing condition more precisely and argued that the PF-deletion approach currently offers the most promising analysis. By simultaneously providing evidence against two claims from the literature, namely that topic drop can occur in the middle field and that it is restricted to topics, I have thus contributed to an empirically adequate account of topic drop licensing that can inform further theoretical, as well as empirical engagement with this ellipsis type.

12.2.2 Recoverability as felicity condition

Also related to my first research question, I examined the role of recoverability, which is not exclusive to topic drop but a prerequisite for any type of omission. I

concluded that it is a felicity or usage condition rather than a licensing condition. This means that an utterance with topic drop in which the omitted constituent cannot be recovered is not ungrammatical but infelicitous. I reviewed the previous literature on the role of recoverability and argued that omitted referential constituents can have antecedents in the linguistic as well as in the extralinguistic context. In the case of the former, the antecedent and topic drop tend to occur close to each other. Moreover, there can be a direct or indirect relationship between the antecedent and the covert constituent. For expletive subjects, I argued that they are trivially recoverable because they convey no semantic content.

I proposed to model the recoverability of referential omitted constituents by a gradual givenness relation. Following Chafe (1994) and Ariel (1990), I consider a constituent to be more given when it is more strongly activated or accessible. A higher degree of givenness also leads to better recoverability and a reduced processing effort on the part of the hearer who has to recover the ellipsis. Thus, while there are likely cases in which recovery always succeeds (e.g., with covert expletives) and cases in which it always fails (e.g., if the omitted constituent is neither mentioned in the preceding discourse nor present in the utterance situation), most cases will occupy an intermediate position. That is, in these cases, how easy it is to recover topic drop depends on the current situation and the cognitive state of the hearer. With this reasoning, I established a link between recoverability as a felicity condition and recoverability as a factor determining the usage of topic drop. Here it became apparent that the boundary between licensing or felicity and usage cannot always be drawn sharply.

12.3 When is topic drop used?

The second part of my book was concerned with the usage of topic drop, i.e., with answering the question of when speakers decide to use topic drop instead of the corresponding full form, given that the ellipsis is licensed and recoverable. To this end, I proposed an information-theoretic approach, according to which the usage of topic drop is guided by the intention to distribute processing effort efficiently across utterances, following the three principles *avoid troughs*, *avoid peaks*, and *facilitate recovery*.

12.3.1 Information-theoretic account of topic drop usage

The information-theoretic account that I advocated in this book is based on a probabilistic notion of information (surprisal) going back to the seminal work

by Shannon (1948). It also draws on the *uniform information density hypothesis* (*UID*) proposed by Levy & Jaeger (2007). *UID* predicts that to communicate efficiently, speakers should prefer utterances with a more uniform distribution of information over utterances with a less uniform one, provided they are grammatical. I argued, first, that conforming to *UID* amounts to an efficient use of the processing resources available to the hearer, and, second, that the two principles derivable from *UID*, *avoid troughs* and *avoid peaks*, guide the usage of topic drop. Accordingly, speakers are assumed to use topic drop to omit predictable expressions and to prevent underutilization of the hearer's processing resources. Speakers are predicted not to use topic drop if the overt prefield constituent is needed to reduce the processing effort of the following constituent. In addition to these two principles derived from *UID*, I proposed a third principle, *facilitate recovery*. It predicts that, for example, a distinct inflectional ending on the verb after topic drop can also reduce the processing effort on that verb that is caused by ellipsis resolution.

The proposed information-theoretic approach offers two advantages. First, by analyzing the usage of topic drop in terms of information theory, I place myself in a tradition of similar approaches that have successfully accounted for the usage of various types of omissions by building on *UID*. Second, this approach explains how general processing principles can account for the usage of a very specific phenomenon such as topic drop, principles that are argued to shape language in general anyway. Certain properties of topic drop, which were previously explained by rules in the grammatical system, can be argued to largely follow from these principles, which unburdens and simplifies the grammatical system.

12.3.2 Evidence for the *avoid troughs* priciple

Evidence for the *avoid troughs* priciple was provided mainly by the attested influence of grammatical person. In my investigations of this factor, I focused on the observation made in the literature and confirmed by several previous corpus studies that the 1st person singular subject pronoun *ich* ('I') is omitted particularly frequently. As possible explanations, I discussed Auer's (1993) *inflectional hypothesis*, as well as two types of *extralinguistic hypotheses*. The *extralinguistic 1SG hypotheses* predict that the 1st person singular subject pronoun referring to the speaker can be easily recovered because they are part of the origo of speaking and known through text type knowledge. The *extralinguistic 1SG+2SG hypotheses* argue that both the 1st and 2nd person can be readily omitted because they refer to the speaker and the hearer as integral components of every communicative situation. I argued that the *inflectional* and both *extralinguistic hypotheses*

are covered by the information-theoretic approach, more specifically by the *facilitate recovery* priciple (for the *inflectional hypothesis)* and the *avoid troughs* priciple (for the *extralinguistic hypotheses*).

In the corpus study, I found that 1st singular subjects indeed had the highest omission rate of all grammatical persons, both in the whole FraC and in the text message data sets. 1st person singular subjects were significantly more frequently omitted than 3rd person singular subjects before copular verbs, auxiliaries, and modal verbs and before verb forms perceived as unambiguous. These production preferences were also reflected in acceptability judgments. In experiments 5, 7, and 8, topic drop of a 1st person singular subject was clearly preferred over topic drop of a 3rd person singular subject referring to a person, in the last experiment even in the absence of a distinct inflectional ending at the following verb (contra the *inflectional hypothesis*). In experiments 9 and 10, I furthermore found that topic drop of 1st and 2nd person singular subjects was rated comparably well. This result provides support for those *extralinguistic hypotheses* that predict that pronouns referring to the speaker *and* the hearer can be omitted equally well.

In sum, I showed in line with the *avoid troughs* priciple that those grammatical persons that are more predictable can also be more readily omitted because pronouns referring to them occur more frequently in general and/or because their referents as the speaker or hearer are predictable from the communicative situation.

Concerning topicality, I concluded in the first part of this book that it is neither a (strictly) sufficient nor a necessary condition for topic drop. In this second part, I argued that its potential facilitating influence on topic drop could be attributable to the *avoid troughs* priciple too. The topic as what is talked about and what is often held constant across multiple utterances should be more predictable than a non-topic. There was no corresponding effect in experiments 5 and 6, in which I manipulated the discourse topic. However, in experiment 7, where topic drop was followed by a lexical verb with a distinct inflectional ending, and in the joint analysis of experiments 7 and 8, in which the sentence topic was set via the subject function, topic drop of a topical constituent was indeed rated better. However, the effect did not show up in experiment 8 alone, where topic drop was followed by syncretic verb forms. This suggests that topicality affects the acceptability of topic drop in particular when not only the *avoid troughs* priciple is at work but also the *facilitate recovery* priciple through the distinct ending of a subsequent full verb.

12.3.3 Evidence for the *avoid peaks* priciple

For the second principle derived from *UID*, the *avoid peaks* priciple, the influence of verb surprisal provided particular evidence. In my corpus study, I found a clear effect of unigram surprisal on the frequency of topic drop, i.e., that the probability/frequency of topic drop decreases when the surprisal of the following verb lemma increases. However, I could not demonstrate an effect of verb surprisal in experiments 11 and 12, possibly in part for methodological reasons. The experimental proof of a surprisal effect is, thus, still pending.

With respect to verb type, it also appeared that not all verb-specific effects can be explained by surprisal alone. In addition to the surprisal effect at the lemma level, the corpus study also revealed an effect of verb type. More specifically, 1st person singular topic drop was more common before copular verbs, auxiliaries, and modal verbs, as predicted by the IDS grammar (Zifonun et al. 1997). Further research is needed to investigate verb type and verb surprisal even more clearly. However, it should be noted that the experimental studies do not support these observations. In experiment 12, which contrasted lexical verbs with auxiliaries, and in experiments 7 and 8, which compared lexical verbs and modal verbs, there were no topic drop specific acceptability differences for verb type, which suggested that it was the copular verbs that were the driving force for the effect in the corpus study.

12.3.4 Evidence for the *facilitate recovery* priciple

An effect of the *facilitate recovery* priciple was shown, at least in part, by the influence of a distinct inflectional ending on the verb after topic drop. The role of verbal inflection is closely related to the influence of grammatical person and to ambiguity avoidance. The relation between inflection and grammatical person becomes evident through Auer's (1993) *inflectional hypothesis*, which, if understood literally, actually predicts that not only topic drop of the 1st person singular but topic drop of any person should be favored if followed by a verb with a distinct inflectional ending. Similarly, the *facilitate recovery* priciple predicts that unambiguous forms help to resolve ellipsis on the verb following topic drop and, thus, facilitate processing this verb.

In my empirical studies, I found mixed evidence for the impact of verbal inflection and ambiguity avoidance. In the corpus study, topic drop of the 1st person singular was more frequent than topic drop of the 3rd person singular before verbs that were perceived as unambiguous. This is consistent with the *facilitate recovery* priciple. Accordingly, a verb form perceived as unambiguous is more

helpful for resolving topic drop of a 1st person singular subject because it allows direct reovery of the speaker as the intended referent. In contrast, the joint analysis of experiment 7 with distinct (unambiguous) verb forms and experiment 8 with syncretic (ambiguous) verb forms showed no effect of verbal inflection or ambiguity on the acceptability of topic drop. It might be that the difference between production (in the corpus study) and perception (in the experiments) plays a role here, i.e., speakers tend to avoid topic drop before ambiguous verb forms, but hearers do not really need this "help" and still find topic drop acceptable ceteris paribus. However, an alternative explanation could be that there was actually no need to avoid the ambiguity because the items were only locally but not globally ambiguous. That is, while hearers could not recover the intended meaning through the verb form, they could when looking at the competing object pronoun.

12.3.5 Summary: information-theoretic account

Overall, I was able to provide initial evidence for an information-theoretic account of the usage of topic drop. According to this approach, the usage of topic drop can be described as being governed by three principles: the *avoid troughs* principle, the *avoid peaks* principle, and the *facilitate recovery* principle. Adherence to these principles leads to an efficient distribution of information and thus facilitates the hearer's processing. It was also shown that the individual factors do not only have an effect in isolation but partly interact and jointly impact the usage of topic drop.

12.4 Open questions and outlook

Although I have taken the first steps to a comprehensive account of topic drop, there are still some unanswered questions, some of which I briefly mention here.

What this book could not accomplish, and where I see a major research desideratum, is to use online methods to investigate the processing of topic drop in real time. Self-paced reading, eye-tracking, and ERPs could be used to explicitly test the implicitly assumed link between processing effort on the one hand and production and perception preferences for topic drop on the other. Specifically, it should be investigated whether the assumed processing difficulties for the troughs and peaks show up in corresponding correlates: reading times, eye movements, and ERP components. According to the *avoid troughs* principle, the prefield constituent of a full form should be read faster in a predictive context,

where it has a lower surprisal, than in a non-predictive context, where it has a higher surprisal. In eye-tracking, participants should have faster first-pass and total reading times on that constituent, shorter fixation times, and fewer regressions to it in the predictive context (see, e.g., Demberg & Keller 2008 and Boston et al. 2008 for eye tracking studies using surprisal as a predictor). In an ERP study, I would expect an attenuated N400 and/or P600 in the predictive context, indicating lower surprisal of the prefield constituent (while, e.g., Frank et al. 2015, Michaelov et al. 2023 argue that surprisal explains N400 effects, according to Brouwer et al. 2021 it is the P600 that indexes at least surprisal effects related to plausibility). According to the *avoid peaks* priciple, it is expected that if a verb following topic drop has a high surprisal, this should be reflected in longer reading times on that verb, more regressions to it, and an increased amplitude of the ERP component related to surprisal, all compared to the same verb in a predictive context.

Furthermore, I consider it worthwhile to study topic drop more cross-linguistically and to work out the similarities and differences between the Germanic V2 languages more strongly, also empirically. Here, it would make sense to use the same or similar experimental designs in several languages or to examine similar or even parallel corpora. This applies to the syntactic licensing of topic drop as well as to its usage. For example, it could be investigated, taking into account possible syntactic differences, whether a refined prefield restriction can also be found in the other Germanic V2 languages and how the influence of topicality is to be evaluated there. Since the information-theoretic approach is based on universal processing principles, the usage of topic drop in other languages should also be determined by the three principles described. In a second step, the experiments and corpus analyses could be expanded to similar ellipsis types in non-V2 languages, such as subject omissions in English and French.

Thirdly, I consider it important to examine the typical occurrence of topic drop in certain text types more closely. While I tried to give first impulses in this direction, it should be investigated in more depth whether the text types in which topic drop occurs particularly frequently exhibit certain properties that predestine them for this ellipsis type. Such properties could be, for example, a more fixed word order with the subject in the first position, a higher proportion of 1st and 2nd person (singular) pronouns, or a lower degree of formality. Here, it seems to me indispensable to create and use a much larger corpus, preferably also containing several different text types.

Such a larger corpus is also needed to investigate the role of object topic drop of different cases in more detail. In particular, dative and genitive objects were not omitted in the FraC. However, since the corresponding full forms were also at

least extremely rare, it could not be decided whether a preference for the overt realization of these objects is indeed evident here or whether it was only data sparsity that prevented finding corresponding instances of topic drop that are only rare but still exist. An alternative or supplement to an extended corpus study is the experimental examination of topic drop of different objects, as a continuation of Trutkowski (2018). Here, however, formality (genitive objects) and contrast (effect of objects in the prefield), among other things, must be considered as potential confounding factors.

References

Ackema, Peter & Ad Neeleman. 2007. Restricted pro drop in Early Modern Dutch. *The Journal of Comparative Germanic Linguistics* 10(2). 81–107. DOI: 10.1007/s10828-007-9010-0.

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. *Natural Language & Linguistic Theory* 17(4). 673–711. DOI: 10.1023/A:1006335629372.

Alday, Phillip M. 2022. *Coding schemes for categorical variables*. https://phillipalday.com/stats/coding.html (19 April, 2023).

Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3). 247–264. DOI: 10.1016/s0010-0277(99)00059-1.

Altmann, Hans. 1987. Zur Problematik der Konstitution von Satzmodi als Formtypen. In Jörg Meibauer (ed.), *Satzmodus zwischen Grammatik und Pragmatik: Referate anläßlich der 8. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* (Linguistische Arbeiten 180), 22–56. Tübingen: Niemeyer. DOI: 10.1515/9783111560588-003.

Anand, Pranav, Daniel Hardt & James McCloskey. 2021. The Santa Cruz sluicing data set. *Language* 97(1). e68–e88. DOI: 10.1353/lan.2021.0009.

Androutsopoulos, Jannis & Gurly Schmidt. 2002. SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe. *Zeitschrift für Angewandte Linguistik* 36. 49–79.

Ariel, Mira. 1990. *Accessing noun-phrase antecedents* (Croom Helm Linguistics Series). London: Routledge. DOI: 10.4324/9781315857473.

Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4(4). 187–203. DOI: 10.1111/j.1749-818X.2010.00193.x.

Arnold, Jennifer E., Thomas Wasow, Ash Asudeh & Peter Alrenga. 2004. Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language* 51(1). 55–70. DOI: 10.1016/j.jml.2004.03.006.

Asr, Fatemeh Torabi & Vera Demberg. 2015. Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. In Matthew Purver, Mehrnoosh Sadrzadeh & Matthew Stone (eds.), *Proceedings of the 11th International Conference on Computational Semantics*, 118–

128. London: Association for Computational Linguistics. https://aclanthology.org/W15-0117/ (23 April, 2025).

Attneave, Fred. 1959. *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. Oxford: Holt.

Auer, Peter. 1993. Zur Verbspitzenstellung im gesprochenen Deutsch. *Deutsche Sprache* 3. 193–222.

Auer, Peter. 1997. Formen und Funktionen der Vor-Vorfeldbesetzung im gesprochenen Deutsch. In Peter Schlobinski (ed.), *Syntax des gesprochenen Deutsch*, 55–91. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-322-88924-9_3.

Axel, Katrin & Helmut Weiß. 2011. *Pro*-drop in the history of German – From Old High German to the modern dialects. In Melani Wratil & Peter Gallmann (eds.), *Null pronouns* (Studies in Generative Grammar 106), 21–51. Berlin: De Gruyter. DOI: 10.1515/9783110238716.21.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56. DOI: 10.1177/00238309040470010201.

Bachmann-Stein, Andrea. 2011. Kontaktanzeigen im World Wide Web. In Patrick Schäfer & Christine Schowalter (eds.), *In mediam linguam: Mediensprache – Redewendungen – Sprachvermittlung. Festschrift für Heinz-Helmut Lüger*, 95–110. Landau: Verlag empirische Pädagogik.

Bader, Markus, Emilia Ellsiepen, Vasiliki Koukoulioti & Yvonne Portele. 2017. Filling the prefield: Findings and challenges. In Constantin Freitag, Oliver Bott & Fabian Schlotterbeck (eds.), *Two perspectives on V2: The invited talks of the DGfS 2016 workshop "V2 in grammar and processing: Its causes and its consequences"*, 27–49. Konstanz: University of Konstanz. http://nbn-resolving.de/urn:nbn:de:bsz:352-0-390234 (23 April, 2025).

Bader, Markus & Jana Häussler. 2010. Word order in German: A corpus study. *Lingua* 120(3). 717–762. DOI: 10.1016/j.lingua.2009.05.007.

Barbosa, Pilar. 2011. *Pro*-drop and theories of *pro* in the minimalist program part 2: Pronoun deletion analyses of null subjects and partial, discourse and semi *pro*-drop. *Language and Linguistics Compass* 5(8). 571–587. DOI: 10.1111/j.1749-818X.2011.00292.x.

Barnickel, Katja. 2017. *Deriving asymmetric coordination in German: A non-monotonic approach*. Leipzig: Universität Leipzig. (Doctoral dissertation).

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-

crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. DOI: 10.1007/s10579-009-9081-4.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.

Barton, Ellen L. 1998. The grammar of telegraphic structures: Sentential and non-sentential derivation. *Journal of English Linguistics* 26(1). 37–67. DOI: 10.1177/007542429802600103.

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

Beißwenger, Michael, Marcel Fladrich, Wolfgang Imo & Evelyn Ziegler. 2020. Die Mobile Communication Database 2 (MoCoDa 2). In Konstanze Marx, Henning Lobin & Axel Schmidt (eds.), *Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig* (Jahrbuch des Instituts für Deutsche Sprache 2019), 349–352. Berlin: De Gruyter. DOI: 10.1515/9783110679885-018.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204. DOI: 10.1017/S004740450001037X.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. New York: Wiley & Sons. DOI: 10.1093/acprof:oso/9780199677139.003.0001.

Bhat, D. N. S. 2004. *Pronouns* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199230242.001.0001.

Biber, Douglas. 1994. An analytical framework for register studies. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic perspectives on register*, 31–56. New York: Oxford University Press.

Bîlbîie, Gabriela & Joanna Nykiel. 2023. Introduction to the special issue experimental and corpus-based approaches to ellipsis. *Journal of Linguistics* 59(2). 229–234. DOI: 10.1017/S0022226722000585.

Blatz, Friedrich. 1896. *Neuhochdeutsche Grammatik mit Berücksichtigung der historischen Entwickelung der deutschen Sprache.* 3rd edn., vol. 2 Satzlehre (Syntax). Karlsruhe: Lang.

Blümel, Andreas, Christopher Götze & Anke Holler. 2017. Revisiting continuative relative clauses: Towards a unified account. *Linguistica Brunensia* 65(1). 81–96.

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

Bosch, Peter, Graham Katz & Carla Umbach. 2007. The non-subject bias of German demonstrative pronouns. In Monika Schwarz-Friesel, Manfred Consten & Mareile Knees (eds.), *Anaphors in text: Cognitive, formal and applied approaches to anaphoric reference* (Studies in Language Companion Series 86), 145–164. Amsterdam: John Benjamins. DOI: 10.1075/slcs.86.13bos.

Boston, Marisa Ferrara, John Hale, Reinhold Kliegl, Umesh Patil & Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2(1). DOI: 10.16910/jemr.2.1.1.

Boswijk, Vincent & Matt Coler. 2020. What is salience? *Open Linguistics* 6(1). 713–722. DOI: 10.1515/opli-2020-0042.

Brehm, Laurel & Phillip M. Alday. 2022. Contrast coding choices in a decade of mixed models. *Journal of Memory and Language* 125. 104334. DOI: 10.1016/j.jml.2022.104334.

Breindl, Eva. 2017. Gemeinsam einsam: Was eint satzverknüpfende Einzelgänger? In Sandra Döring & Jochen Geilfuß-Wolfgang (eds.), *Probleme der syntaktischen Kategorisierung: Einzelgänger, Außenseiter und mehr* (Stauffenburg Linguistik 90), 1–42. Tübingen: Stauffenburg.

Breindl, Eva, Anna Volodina & Ulrich Hermann Waßner. 2014. *Handbuch der deutschen Konnektoren 2: Semantik der deutschen Satzverknüpfer*. Berlin: De Gruyter. DOI: 10.1515/9783110341447.

Brouwer, Harm, Francesca Delogu, Noortje J. Venhuizen & Matthew W. Crocker. 2021. Neurobehavioral correlates of surprisal in language comprehension: A neurocomputational model. *Frontiers in Psychology* 12. 615538. DOI: 10.3389/fpsyg.2021.615538.

Brysbaert, Marc, Denis Drieghe & Françoise Vitu. 2005. Word skipping: Implications for theories of eye movement control in reading. In Geoffry Underwood (ed.), *Cognitive processes in eye guidance*, 53–77. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198566816.003.0003.

Bühler, Karl. 1965. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. 2nd edn. Stuttgart: Gustav Fischer.

Büring, Daniel. 1997. *The meaning of topic and focus: The 59th street bridge accent* (Routledge Study in German Linguistics 3). London: Routledge. DOI: 10.4324/9780203448014.

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

Burger, Susanne, Karl Weilhammer, Florian Schiel & Hans G. Tillmann. 2000. Verbmobil data collection and annotation. In Wolfgang Wahlster (ed.), *Verbmo-*

bil: Foundations of speech-to-speech translation (Artificial Intelligence Series), 537–549. Berlin: Springer. DOI: 10.1007/978-3-662-04230-4_39.

Cardinaletti, Anna. 1986. Topicalization in German: Movement to COMP or base-generation in TOP? *GAGL: Groninger Arbeiten zur germanistischen Linguistik* (28). 202–231.

Cardinaletti, Anna. 1990. Subject/object asymmetries in German null-topic constructions and the status of specCP. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress*, 75–84. Berlin: De Gruyter Mouton. DOI: 10 . 1515 / 9783110867848.75.

Carifio, James & Rocco Perla. 2008. Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 42(12). 1150–1152. DOI: 10.1111/j.1365-2923.2008.03172.x.

Chafe, Wallace. 1974. Language and consciousness. *Language* 50(1). 111–133. DOI: 10.2307/412014.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chambers, Craig G. & Ron Smyth. 1998. Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language* 39(4). 593–608. DOI: 10.1006/jmla.1998.2575.

Chiarcos, Christian, Berry Claus & Michael Grabski. 2011. Introduction: Salience in linguistics and beyond. In Christian Chiarcos, Berry Claus & Michael Grabski (eds.), *Salience. Multidisciplinary perpectives on its function in discourse* (Trends in Linguistics. Studies and Monographs 227), 1–28. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110241020.1.

Chomsky, Noam. 1981. *Lectures on government and binding. The Pisa Lectures* (Studies in Generative Grammar 9). Dordrecht: Foris. DOI: 10 . 1515 / 9783110884166.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding* (Linguistic Inquiry Monographs 6). Cambridge: MIT Press.

Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monographs 13). Cambridge: MIT Press.

Christensen, Rune Haubo Bojesen. 2019. *Ordinal. Regression models for ordinal data. R package version 2019.12-10.* https : / / rdrr . io / cran / ordinal/ (30 April, 2020).

Cinque, Guglielmo. 1990. *Types of Ā-dependencies* (Linguistic Inquiry Monographs 17). Cambridge: MIT Press.

Clark, Herbert H. 1975. Bridging. In *TINLAP '75: Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*, 169–174. Stroudsburg: Association for Computational Linguistics. DOI: 10.3115/980190.980237.

Clickworker. 2022. *Über uns. Crowdsourcing Services & Content Provider.* https://www.clickworker.de/ueber-uns/ (24 March, 2022).

Colonna, Saveria, Sarah Schimke & Barbara Hemforth. 2012. Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. *Linguistics* 50(5). 991–1013. DOI: 10.1515/ling-2012-0031.

Consten, Manfred & Monika Schwarz-Friesel. 2007. Anapher. In Ludger Hoffmann (ed.), *Handbuch der deutschen Wortarten* (De Gruyter Studienbuch), 265–292. Berlin: De Gruyter.

Corver, Norbert & Hans Broekhuis. 2016. *Syntax of Dutch: Verbs and verb phrases.* Vol. 3. Amsterdam: Amsterdam University Press. DOI: 10.26530/OAPEN_614910.

Crawley, Rosalind A., Rosemary J. Stevenson & David Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 19(4). 245–264. DOI: 10.1007/BF01077259.

Crump, Matthew J. C., Danielle J. Navarro & Jeffrey Suzuki. 2019. *Answering questions with data: Introductory statistics for psychology students.* DOI: 10.17605/OSF.IO/JZE52.

D'Alessandro, Roberta. 2015. Null subjects. In Antonio Fábregas, Jaume Mateu & Michael T. Putnam (eds.), *Contemporary linguistic parameters* (Contemporary Studies in Linguistics), 201–226. London: Bloomsbury Academic. DOI: 10.5040/9781474219549.

Daneš, František. 1970. One instance of Prague School methodology: Functional analysis of utterance and text. In Paul L. Garvin (ed.), *Method and theory in linguistics* (Janua Linguarum. Series Maior 40), 132–146. The Hague: Mouton. DOI: 10.1515/9783110872521.132.

Davidson, Brad. 1996. 'Pragmatic weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish. *Journal of Pragmatics* 26(4). 543–565. DOI: 10.1016/0378-2166(95)00063-1.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210. DOI: 10.1016/j.cognition.2008.07.008.

den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In Werner Abraham (ed.), *On the formal syntax of the Westgermania: Papers from the 3rd Groningen Grammar Talks (3e Groninger Grammatikgespräche), Groningen, January 1981* (Linguistik Aktuell / Linguistics Today 3). Amsterdam: John Benjamins. DOI: 10.1075/la.3.03bes.

den Dikken, Marcel (ed.). 2013. *The Cambridge handbook of generative syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511804571.

Deppermann, Arnulf. 2021. Imperative im Deutschen: Konstruktionen, Praktiken oder social action formats? In Beate Weidner, Katharina König, Wolfgang Imo & Lars Wegner (eds.), *Verfestigungen in der Interaktion* (Empirische Linguistik / Empirical linguistics 13), 195–230. Berlin: De Gruyter. DOI: 10.1515/9783110637502-009.

Dittmann, Jürgen, Hedy Siebert & Yvonne Staiger-Anlauf. 2007. *Medium und Kommunikationsform – am Beispiel der SMS* (Networx 50). Hannover: Gottfried Wilhelm Leibniz Universität, Seminar für deutsche Literatur und Sprache. DOI: 10.15488/2941.

Döring, Nicola. 2002. "Kurzm. wird gesendet". Abkürzungen und Akronyme in der SMS-Kommunikation. *Muttersprache. Vierteljahresschrift für Deutsche Sprache* 112(2). 97–114.

Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt/Main: Diesterweg.

Dryer, Matthew S. 2013a. Expression of pronominal subjects. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/101 (5 May, 2023).

Dryer, Matthew S. 2013b. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/81 (15 June, 2023).

Duarte, Maria Eugênia Lamoglia. 1995. *A perda do princípio "evite pronome" no português brasileiro*. Campinas: Universidade Estadual de Campinas. (Doctoral dissertation).

Duden. 2016. *Duden. Die Grammatik*. Angelika Wöllstein & Dudenredaktion (eds.). 9th, completely revised and updated edition. Berlin: Dudenverlag.

Duden. 2022. *Duden. Die Grammatik*. Angelika Wöllstein & Dudenredaktion (eds.). 10th, completely rewritten edition. Berlin: Dudenverlag.

Dudenredaktion. 2021. *„erwehren"*. Duden online. https://www.duden.de/rechtschreibung/erwehren (23 December, 2021).

Dürscheid, Christa. 2016. Nähe, Distanz und neue Medien. In Helmuth Feilke & Mathilde Hennig (eds.), *Zur Karriere von 'Nähe und Distanz'*. Berlin: De Gruyter. DOI: 10.1515/9783110464061-013.

Dürscheid, Christa & Sarah Brommer. 2009. Getippte Dialoge in neuen Medien. Sprachkritische Aspekte und linguistische Analysen. *Linguistik Online* 37(1). 3–20. DOI: 10.13092/lo.37.511.

Dürscheid, Christa & Karina Frick. 2014. Keyboard-to-screen-Kommunikation gestern und heute: SMS und WhatsApp im Vergleich. In Jens Runkehl, Alexa Mathias & Torsten Siever (eds.), *Sprachen? Vielfalt! Sprache und Kommunikation in der Gesellschaft und den Medien. Eine Online-Festschrift zum Jubiläum von Peter Schlobinski* (Networx 64), 149–181. Hannover: Gottfried Wilhelm Leibniz Universität, Seminar für deutsche Literatur und Sprache.

Duguine, Maia. 2014. Argument ellipsis: A unitary approach to pro-drop. *The Linguistic Review* 31(3-4). 515–549. DOI: 10.1515/tlr-2014-0010.

Eckert, Miriam. 1998. *Discourse deixis and null anaphora in German*. Edinburgh: University of Edinburgh. (Doctoral dissertation).

Eide, Kristin Melum & Hilde Sollid. 2011. Norwegian main clause declaratives: Variation within and across grammars. In Peter Siemund (ed.), *Linguistic universals and language variation* (Trends in Linguistics. Studies and Monographs 231), 327–360. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110238068.327.

Eisenberg, Peter. 2002. Es gibt Gerüchte über eine Rettung Österreichs. Zur Rekonstruktion der Grammatik von Karl Kraus. In Ulrike Hass-Zumkehr, Werner Kallmeyer & Gisela Zifonun (eds.), *Ansichten der deutschen Sprache: Festschrift für Gerhard Stickel zum 65. Geburtstag* (Studien zur deutschen Sprache 25), 55–87. Tübingen: Narr.

Eisenberg, Peter. 2020. *Grundriss der deutschen Grammatik: Der Satz*. 5th edn. Stuttgart: Metzler. DOI: 10.1007/978-3-476-05094-6.

Ellipsis. 2023. Wiktionary. https://en.wiktionary.org/w/index.php?title=ellipsis&oldid=72791181 (25 May, 2023).

Engel, Ulrich. 1970. *Regeln zur Wortstellung* (Forschungsberichte des Instituts für deutsche Sprache 5). Mannheim: Institut für Deutsche Sprache.

Engel, Ulrich. 1972. Regeln zur "Satzgliedfolge". Zur Stellung der Elemente im einfachen Verbalsatz. In Hugo Moser (ed.), *Linguistische Studien*, vol. 1 (Sprache der Gegenwart 19), 17–75. Düsseldorf: Schwann. DOI: 10.1515/9783110854442.17.

Erdmann, Oskar. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung dargestellt. Erste Abteilung*. Stuttgart: J. G. Cotta'sche Buchhandlung.

Erteschik-Shir, Nomi. 1997. *The dynamics of focus structure* (Cambridge Studies in Linguistics 84). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511519949.

Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface* (Oxford Surveys in Syntax and Morphology 3). Oxford: Oxford University Press.

Falk, Cecilia. 1993. Non-referential subjects and agreement in the history of Swedish. *Lingua* 89(2-3). 143–180. DOI: 10.1016/0024-3841(93)90051-W.

Fanselow, Gisbert & Stefan Frisch. 2006. Effects of processing difficulty on judgements of acceptability. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*, 291–316. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199274796.003.0015.

Fenk, August & Gertraud Fenk. 1980. Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß? *Zeitschrift für experimentelle und angewandte Psychologie* 27(3). 400–414.

Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27(3). 517–556. DOI: 10.1515/ling.1989.27.3.517.

Fenk-Oczlon, Gertraud. 1990. Ökonomieprinzipien in Kognition und Kommunikation. In Norbert Boretzky, Werner Enninger & Thomas Stolz (eds.), *Spielarten der Natürlichkeit – Spielarten der Ökonomie* (Beiträge zum 5. Essener Kolloquium über Grammatikalisierung, Natürlichkeit und Systemökonomie vom 6. 10.–8.10.1988 an der Universität Essen), 37–50. Bochum: Brockmayer.

Ferreira, Victor S. 2007. How are speakers' linguistic choices affected by ambiguity? In Antje Meyer, Linda Wheeldon & Andrea Krott (eds.), *Automaticity and control in language processing*, 63–92. Hove: Psychology Press.

Ferreira, Victor S. 2008. Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation. Advances in Research and Theory* 49. 209–246. DOI: 10.1016/S0079-7421(08)00006-6.

Ferreira, Victor S. & Gary S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40(4). 296–340. DOI: 10.1006/cogp.1999.0730.

Ferreira, Victor S., L. Robert Slevc & Erin S. Rogers. 2005. How do speakers avoid ambiguous linguistic expressions? *Cognition* 96(3). 263–284. DOI: 10.1016/j.cognition.2004.09.002.

Féry, Caroline. 2006. Laute und leise Prosodie. In Hardarik Blühdorn, Eva Breindl & Ulrich Hermann Waßner (eds.), *Text – Verstehen. Grammatik und darüber hinaus* (Jahrbuch des Instituts für Deutsche Sprache 2005), 164–186. Berlin: De Gruyter. DOI: 10.1515/9783110199963.1.164.

Filippova, Katja & Michael Strube. 2007. The German Vorfeld and local coherence. *Journal of Logic, Language, and Information* 16(4). 465–485. DOI: 10.1007/s10849-007-9044-3.

Firbas, Jan. 1971. On the concept of communicative dynamism in the theory of functional sentence perspective. *Sborník prací Filozofické fakulty brněnské univerzity* (A19). 135–144.

Firbas, Jan. 1982. Has every sentence a theme and a rheme? In John A. E. Anderson (ed.), *Language form and linguistic variation. Papers dedicated to Angus McIntosh* (Current Issues in Linguistic Theory 15), 97–116. Amsterdam: John Benjamins. DOI: 10.1075/cilt.15.07fir.

Firbas, Jan. 1992. *Functional sentence perspective in written and spoken communication* (Studies in English Language). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511597817.

Fischer, Kerstin. 2006. Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume. In Kerstin Fischer (ed.), *Approaches to discourse particles* (Studies in Pragmatics 1), 1–20. Leiden: Brill. DOI: 10.1163/9780080461588_002.

Fortmann, Christian. 2005. Die Lücken im Bild von der Subjektlücken-Koordination. *Linguistische Berichte* 204. 441–476.

Fox, Barbara A. 2007. Principles shaping grammatical practices: An exploration. *Discourse Studies* 9(3). 299–318. DOI: 10.1177/1461445607076.

Frank, Austin F. & T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In Cognitive Science Society (ed.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci 2008)*, 939–944. Austin.

Frank, Nicola. 2000. Probleme lexikalischer Selektion und abhängige Verbzweitsätze. *Linguistische Berichte* 184. 469–483.

Frank, Stefan L., Leun J. Otten, Giulia Galli & Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140. 1–11. DOI: 10.1016/j.bandl.2014.10.006.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax* (Oxford Studies in Comparative Syntax). New York: Oxford University Press. DOI: 10 . 1093 / oso / 9780195089707.001.0001.

Frey, Werner. 2000. Über die syntaktische Position der Satztopiks im Deutschen. *ZAS Papers in Linguistics* 20. 137–172. DOI: 10.21248/zaspil.20.2000.82.

Frey, Werner. 2005. Zur Syntax der linken Peripherie im Deutschen. In Franz-Josef d'Avis (ed.), *Deutsche Syntax: Empirie und Theorie* (Göteborger Germanistische Forschungen 46), 147–171. Göteburg: Acta Universitatis Gothoburgensis.

Freywald, Ulrike. 2009. Kontexte für nicht-kanonische Verbzweitstellung: V2 nach *dass* und Verwandtes. In Veronika Ehrich, Christian Fortmann, Ingo Reich & Marga Reis (eds.), *Koordination und Subordination im Deutschen* (Linguistische Berichte. Sonderheft 16), 113–134. Hamburg: Buske.

Freywald, Ulrike. 2013. Uneingeleiteter V1- und V2-Satz. In Jörg Meibauer, Markus Steinbach & Hans Altmann (eds.), *Satztypen des Deutschen*, 317–337. Berlin: De Gruyter. DOI: 10.1515/9783110224832.317.

Freywald, Ulrike. 2016. "V2-Nebensätze" – ein eigener Satztyp? In Rita Finkbeiner & Jörg Meibauer (eds.), *Satztypen und Konstruktionen* (Linguistik – Impulse & Tendenzen 65), 326–372. Berlin: De Gruyter. DOI: 10.1515/9783110423112-011.

Freywald, Ulrike. 2020. Nach dem Öffnen rasch verbrauchen! Objektauslassung in direktiven Infinitiven – ein Fall von Topikdrop. In Robert Külpmann, Laura Neuhaus & Vilma Symanczyk Joppe (eds.), *Variation in der Argumentstruktur des Deutschen* (Linguistische Berichte. Sonderheft 28), 147–177. Hamburg: Buske.

Frick, Karina. 2017. *Elliptische Strukturen in SMS: Eine korpusbasierte Untersuchung des Schweizerdeutschen* (Empirische Linguistik = Empirical Linguistics 7). Berlin: De Gruyter. DOI: 10.1515/9783110517859.

Fries, Norbert. 1988. Über das Null-Topik im Deutschen. *Sprache & Pragmatik* 3. 19–49. DOI: 10.18452/9478.

Gallmann, Peter. 2015. Das topologische Modell: Basisartikel. In Angelika Wöllstein & Saskia Ripp (eds.), *Das topologische Modell für die Schule* (Thema Sprache – Wissenschaft für den Unterricht 19), 1–36. Baltmannsweiler: Schneider-Verlag Hohengehren.

Gallmann, Peter & Horst Sitta. 2007. *Deutsche Grammatik*. 5th edn. (Lehrmittel der Interkantonalen Lehrmittelzentrale). Zürich: Lehrmittelverlag des Kantons Zürich.

Gansel, Christina & Frank Jürgens. 2009. *Textlinguistik und Textgrammatik. Eine Einführung*. New edition of the 2nd, revised and supplemented edition of 2007 (UTB Sprachwissenschaft 3265). Göttingen: Vandenhoeck & Ruprecht. DOI: 10.36198/9783838532653.

Genzel, Dmitriy & Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics – ACL '02*, 199–206. Philadelphia: Association for Computational Linguistics. DOI: 10.3115/1073083.1073117.

Genzel, Dmitriy & Eugene Charniak. 2003. Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 65–72. Sapporo: Association for Computational Linguistics. DOI: 10.3115/1119355.1119364.

Givón, Talmy (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study* (Typological Studies in Language 3). Amsterdam: John Benjamins. DOI: 10.1075/tsl.3.

Gohl, Christine & Susanne Günthner. 1999. Grammatikalisierung von *weil* als Diskursmarker in der gesprochenen Sprache. *Zeitschrift für Sprachwissenschaft* 18(1). 39–75. DOI: 10.1515/zfsw.1999.18.1.39.

Goldrick, Matthew, Victor S. Ferreira & Michele Miozzo (eds.). 2014. *The Oxford handbook of language production* (Oxford Library of Psychology). Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199735471.001.0001.

Gordishevsky, Galina & Sergey Avrutin. 2004. Optional omissions in an optionally null subject language. In Jacqueline van Kampen & Sergey Baauw (eds.), *Proceedings of GALA (Generative Approaches to Language Acquisition) 2003*, vol. 1, 187–198. Utrecht: LOT.

Grewendorf, Günther. 1988. *Aspekte der deutschen Syntax: eine Rektions-Bindungs-Analyse* (Studien zur deutschen Grammatik 33). Tübingen: Narr.

Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts* (Syntax and Semantics 3), 41–58. New York: Academic Press. DOI: 10.1163/9789004368811_003.

Grice, Herbert Paul. 1989. *Studies in the way of words*. Cambridge: Harvard University Press.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225. DOI: 10.21236/ADA324949.

Günthner, Susanne. 2000. *Vorwurfsaktivitäten in der Alltagsinteraktion. Grammatische, prosodische, rhetorisch-stilistische und interaktive Verfahren bei der Konstitution kommunikativer Muster und Gattungen* (Germanistische Linguistik 221). Tübingen: Niemeyer. DOI: 10.1515/9783110919974.

Günthner, Susanne. 2006. Grammatische Analysen der kommunikativen Praxis – 'Dichte Konstruktionen' in der Interaktion. In Arnulf Deppermann, Reinhard Fiehler & Thomas Spranz-Fogasy (eds.), *Grammatik und Interaktion: Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Interaktionsprozessen*, 95–121. Radolfszell: Verlag für Gesprächsforschung.

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology* (Typological Studies in Language 17), 209–239. Amsterdam: John Benjamins. DOI: 10.1075/tsl.17.16gun.

Gundel, Jeanette K. 2010. Topic. In Louise Cummings (ed.), *The pragmatics encyclopedia*, 471–472. Abingdon: Routledge.

Habscheid, Stephan (ed.). 2011. *Textsorten, Handlungsmuster, Oberflächen: Linguistische Typologien der Kommunikation* (De Gruyter Lexikon). Berlin: De Gruyter. DOI: 10.1515/9783110229301.

Haegeman, Liliane. 1990. Non-overt subjects in diary contexts. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress. GLOW essays for Henk van Riemsdijk* (Studies in Generative Grammar 36), 167–174. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110867848.167.

Haegeman, Liliane. 1996. Verb-second, the split CP and null subjects in early Dutch finite clauses. *GenGenP (Geneva Generative Papers)* 4(2). 133–175.

Haegeman, Liliane. 1997. Register variation, truncation, and subject omission in English and in French. *English Language & Linguistics* 1(2). 233–270. DOI: 10.1017/S1360674300000526.

Haegeman, Liliane. 2000. Adult null subjects in the non *pro*-drop languages. In Marc-Ariel Friedemann & Luigi Rizzi (eds.), *The acquisition of syntax: Studies in comparative developmental linguistics*, 129–169. London: Routledge.

Haegeman, Liliane. 2007. Subject omission in present-day written English. On the theoretical relevance of peripheral data. *Rivista di Grammatica Generativa* 32. 91–124.

Haegeman, Liliane. 2013. The syntax of registers: Diary subject omission and the privilege of the root. *Lingua* 130. 88–110. DOI: 10.1016/j.lingua.2013.01.005.

Haegeman, Liliane. 2017. Unspeakable sentences. Subject omission in written registers: A cartographic analysis. *Linguistic Variation* 17(2). 229–250. DOI: 10.1075/lv.14019.hae.

Haegeman, Liliane & Tabea Ihsane. 1999. Subject ellipsis in embedded clauses in English. *English Language & Linguistics* 3(1). 117–145. DOI: 10.1017/S1360674399000155.

Haider, Hubert. 2010. *The syntax of German* (Cambridge Syntax Guides). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511845314.

Håkansson, Gisela. 1994. Verb-initial sentences in the development of Swedish. *Working papers / Lund University, Department of Linguistics and Phonetics* 42. 49–65.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 159–166. Pittsburgh: Association for Computational Linguistics. DOI: 10.3115/1073336.1073357.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger*, 111–176. Cambridge: MIT Press.

Halliday, Michael Alexander Kirkwood. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244. DOI: 10 . 1017 / S0022226700016613.

Halliday, Michael Alexander Kirkwood. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, Michael Alexander Kirkwood & Ruqaiya Hasan. 1976. *Cohesion in English* (English Language Series 9). London: Longman. DOI: 10 . 4324 / 9781315836010.

Hamann, Cornelia & Kim Plunkett. 1998. Subjectless sentences in child Danish. *Cognition* 69(1). 35–72. DOI: 10.1016/S0010-0277(98)00059-6.

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

Hansen-Schirra, Silvia & Silke Gutermuth. 2017. Multimodale Daten für die empirische Analyse grammatischer Variation in Übersetzungen. In Marek Konopka & Angelika Wöllstein (eds.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung* (Jahrbuch des Instituts für Deutsche Sprache 2016), 309–318. Berlin: De Gruyter. DOI: 10.1515/9783110518214-018.

Hansen-Schirra, Silvia, Stella Neumann & Erich Steiner (eds.). 2012. *Cross-linguistic corpora for the study of translations: Insights from the language pair English-German* (Text, Translation, Computational Processing 11). Berlin: De Gruyter Mouton. DOI: 10.1515/9783110260328.

Hård af Segerstad, Ylva. 2005. Language use in Swedish mobile text messaging. In Rich Ling & Per E. Pedersen (eds.), *Mobile communications: Re-negotiation of the social sphere* (Computer Supported Cooperative Work 31), 313–333. London: Springer. DOI: 10.1007/1-84628-248-9_21.

Hartmann, Katharina. 1994. Zur Koordination von V-2 Sätzen. *Zeitschrift für Sprachwissenschaft* 13(1). 3–19. DOI: 10.1515/zfsw.1994.13.1.3.

Hartmann, Katharina. 2000. *Right node raising and gapping: Interface conditions on prosodic deletion*. Philadelphia: John Benjamins. DOI: 10.1075/z.106.

Haspelmath, Martin. 2009. Terminology of case. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 505–517. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199206476.013.0034.

Heidolph, Karl Erich. 1992. „Satz" als Kategorie der Grammatik. In *Deutsche Syntax. Ansichten und Aussichten* (Jahrbuch des Instituts für Deutsche Sprache 1991), 396–407. Berlin: De Gruyter. DOI: 10.1515/9783110622447-018.

Helmer, Henrike. 2016. *Analepsen in der Interaktion. Semantische und sequenzielle Eigenschaften von Topik-Drop im gesprochenen Deutsch* (Oralingua 13). Heidelberg: Winter.

Helmer, Henrike. 2017a. Analepsen aus konstruktionsgrammatischer Perspektive. In Arnulf Deppermann, Nadine Proske & Arne Zeschel (eds.), *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*, 407–450. Tübingen: Narr.

Helmer, Henrike. 2017b. Analepsen mit Topik-Drop. Zur Notwendigkeit einer diskurssemantischen Perspektive. *Zeitschrift für germanistische Linguistik* 45(1). 1–39. DOI: 10.1515/zgl-2017-0001.

Hemforth, Barbara, Lars Konieczny, Christoph Scheepers, Saveria Colonna, Sarah Schimke, Peter Baumann & Joël Pynte. 2010. Language specific preferences in anaphor resolution: Exposure or Gricean maxims? In Stellan Ohlsson & Richard Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society (CogSci 2010)*, 2218–2223. Austin: Cognitive Science Society.

Herling, Simon Heinrich Adolf. 1821. Ueber die Topik der deutschen Sprache. In *Abhandlungen des frankfurtischen Gelehrtenvereines für deutsche Sprache. Drittes Stück*, 296–362, 394. Frankfurt/Main: Franz Varrentrapp.

Hinrichs, Erhard W., Julia Bartels, Yasuhiro Kawata, Valia Kordoni & Heike Telljohann. 2000. The Tübingen Treebanks for spoken German, English, and Japanese. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of speech-to-speech translation* (Artificial Intelligence Series), 550–574. Berlin: Springer. DOI: 10.1007/978-3-662-04230-4_40.

Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache* (Heutiges Deutsch Reihe 1, Linguistische Grundlagen. Forschungen des Instituts für deusche Sprache 10). Munich: Hueber.

Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.

Höhle, Tilman N. 1983. Subjektlücken in Koordinationen. Unpublished manuscript. Tübingen.

Höhle, Tilman N. 1986. Der Begriff ‚Mittelfeld': Anmerkungen über die Theorie der topologischen Felder. In Albrecht Schöne (ed.), *Kontroversen, alte und neue. Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985.* Vol. 3: *Textlinguistik contra Stilistik? – Wortschatz und Wörterbuch – Grammatische oder pragmatische Organisation von Rede?* Ed. by Walter E. Weiss, Herbert E. Wiegand, and Marga Reis, 329–340. Tübingen: Niemeyer.

Höhle, Tilman N. 1990. Assumptions about asymmetric coordination in German. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress*, 221–236. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110867848.221.

Höhle, Tilman N. 1992. Über Verum-Fokus im Deutschen. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, 112–141. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-663-12176-3_5.

Hoffmann, Ludger. 1999. Ellipse und Analepse. In Angelika Redder & Jochen Rehbein (eds.), *Grammatik und mentale Prozesse* (Stauffenburg Linguistik), 69–90. Tübingen: Stauffenburg.

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2014. Processing effects in linguistic judgment data. (Super-)additivity and reading span scores. *Language and Cognition* 6(1). 111–145. DOI: 10.1017/langcog.2013.7.

Hofmeister, Philip, Peter Culicover & Susanne Winkler. 2015. Effects of processing on the acceptability of "frozen" extraposed constituents. *Syntax* 18(4). 464–483. DOI: 10.1111/synt.12036.

Hofmeister, Philip, T. Florian 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. DOI: 10.1080/01690965.2011.572401.

Hole, Daniel. 2014. *Dativ, Bindung und Diathese* (Studia grammatica 78). Berlin: De Gruyter. DOI: 10.1515/9783110347739.

Horch, Eva. 2014. Zur Natur der Lücke in SLF-Konstruktionen – Zwei experimentelle Studien. *Linguistische Berichte* 238. 141–163.

Horch, Eva & Ingo Reich. 2017. The Fragment Corpus (FraC). In *Proceedings of the 9th International Corpus Linguistics Conference*, 392–393. Birmingham. https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2017/general/paper194.pdf (23 April, 2025).

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15(4). 531–574.

Huang, Yan. 2000. *Anaphora: A cross-linguistic approach* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Imo, Wolfgang. 2012. Wortart Diskursmarker? In Björn Rothstein (ed.), *Nichtflektierende Wortarten* (Linguistik – Impulse & Tendenzen 47), 48–88. Berlin: De Gruyter. DOI: 10.1515/9783110276619.48.

Imo, Wolfgang. 2013. Interaktionale Linguistik: Ellipsen und Verstehen in der Interaktion. In Mathilde Hennig (ed.), *Die Ellipse. Neue Perspektiven auf ein altes Phänomen*, 281–319. Berlin: De Gruyter. DOI: 10.1515/9783110312560.281.

Imo, Wolfgang. 2014. Elliptical structures as dialogical resources for the management of understanding. In Susanne Günthner, Wolfgang Imo & Jörg Bücker (eds.), *Grammar and Dialogism*, 139–178. Berlin: De Gruyter. DOI: 10.1515/9783110358612.139.

Imo, Wolfgang. 2016. *Grammatik. Eine Einführung*. Stuttgart: Metzler.

Institut für deutsche Sprache, Mannheim. 2014. *Korpus Gespräche im Fernsehen (GF)*. https://agd.ids-mannheim.de/download/Korpus_GF_extern.pdf (19 August, 2023).

Jaccard, James & Choi K. Wan. 1996. *LISREL approaches to interaction effects in multiple regression* (Quantitative Applications in the Social Sciences 114). Thousand Oaks: Sage. DOI: 10.4135/9781412984782.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jaensch, Carol. 2005. UG and poverty of the stimulus – Topic deletion evidence from L2 learners of German. *Essex Graduate Student Papers in Language and Linguistics* 7. 97–117.

Järvikivi, Juhani, Roger P. G. van Gompel, Jukka Hyönä & Raymond Bertram. 2005. Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science* 16(4). 260–264. DOI: 10.1111/j.0956-7976.2005.01525.x.

Jamieson, Susan. 2004. Likert scales: How to (ab)use them. *Medical Education* 38(12). 1217–1218. DOI: 10.1111/j.1365-2929.2004.02012.x.

Jansen, Frank. 1981. *Syntaktische konstrukties in gesproken taal*. Amsterdam: Huis aan de Drie Grachten.

Johannessen, Janne Bondi. 1998. *Coordination* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Kaatari, Henrik. 2016. Variation across two dimensions: Testing the complexity principle and the uniform information density principle on adjectival data. *English Language & Linguistics* 20(3). 533–558. DOI: 10.1017/S1360674316000344.

Kaiser, Elsi. 2009. Investigating effects of structural and information-structural factors on pronoun resolution. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 332–353. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199570959.003.0014.

Kaiser, Georg A. 2003. Syntaktische Variation und generative Syntaxtheorie. In Elisabeth Stark & Ulrich Wandruszka (eds.), *Syntaxtheorien: Modelle, Methoden, Motive* (Tübinger Beiträge zur Linguistik 469), 257–272. Tübingen: Narr.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1). 63–99.

Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23(6). 533–575. DOI: 10.1023/A:1005677819813.

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Stanford: CSLI Publications.

Kempen, Gerard & Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Stephan Kepser & Marga Reis (eds.),

Linguistic evidence. Empirical, theoretical, and computational perspectives (Studies in Generative Grammar 85), 329–349. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110197549.329.

Kennedy, Christopher. 2019. Ambiguity and vagueness: An overview. In *Semantics – Lexical structures and adjectives* (Mouton Reader), 236–271. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110626391-008.

Klein, Wolfgang. 1985. Ellipse, Fokusgliederung und thematischer Stand. In Reinhard Meyer-Herrmann & Hannes Rieser (eds.), *Ellipsen und fragmentarische Ausdrücke*, vol. 1 (Linguistische Arbeiten 148/1), 1–24. Tübingen: Niemeyer. DOI: 10.1515/9783111351308.1.

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 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 9/1), 763–799. Berlin: De Gruyter. DOI: 10.1515/9783110095869.1.12.763.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–43. DOI: 10.1515/9783110244922.15.

Konopka, Agnieszka E. & Sarah Brown-Schmidt. 2014. Message encoding. In Matthew Goldrick, Victor S. Ferreira & Michele Miozzo (eds.), *The Oxford handbook of language production* (Oxford Library of Psychology), 3–20. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199735471.013.010.

Korth, Manuela. 2017. *Synkretismus*. In Stefan J. Schierholz (ed.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft online*. Berlin: De Gruyter. https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_artikel_19474/html (9 May, 2023).

Kraljic, Tanya & Susan E. Brennan. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology* 50(2). 194–231. DOI: 10.1016/j.cogpsych.2004.08.002.

Kravtchenko, Ekaterina. 2014. Predictability and syntactic production: Evidence from subject omission in Russian. In Paul Bello, Marcello Guarini, Marjorie McShane & Brian Scassellati (eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci 2014)*, 785–790. Austin: Cognitive Science Society.

Kravtchenko, Ekaterina & Vera Demberg. 2022. Informationally redundant utterances elicit pragmatic inferences. *Cognition* 225. 105159. DOI: 10.1016/j.cognition.2022.105159.

Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *The notions of information structure* (Interdisciplinary Studies on Information Structure 6), 13–55. Potsdam: Universitätsverlag Potsdam.

Kroll, Margaret & Deniz Rudin. 2017. Identity and interpretation: Syntactic and pragmatic constraints on the acceptability of sluicing. In Andrew Lamont & Katerina Tetzloff (eds.), *Proceedings of the Forty-Seventh Annual Meeting of the North East Linguistic Society (NELS 47)*. Amherst: University of Massachusetts.

Külpmann, Robert. 2021. *Der Einfluss der Satztypwahl auf die Möglichkeit zur definiten Komplementweglassung*. Wuppertal: Bergische Universität Wuppertal. (Doctoral dissertation).

Külpmann, Robert & Vilma Symanczyk Joppe. 2016. Argument omission in imperative-declarative conjunctions. In Ingo Reich & Augustin Speyer (eds.), *Co- and subordination in German and other languages* (Linguistische Berichte. Sonderheft 21), 221–234. Hamburg: Buske.

Kupietz, Marc, Cyril Belica, Holger Keibel & Andreas Witt. 2010. The German reference corpus DeReKo: A primordial sample for linguistic research. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 2010)*, 1848–1854. Valletta: European Language Resources Association (ELRA).

Kurumada, Chigusa & T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152–178. DOI: 10.1016/j.jml.2015.03.003.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Leibniz-Institut für Deutsche Sprache. 2012. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2012-I* (Release 02/29/2012). Mannheim: Leibniz-Institut für Deutsche Sprache. www.ids-mannheim.de/dereko (25 April, 2025).

Leibniz-Institut für Deutsche Sprache. 2013a. DeReWo Benutzerdokumentation / DeReWo manual. In *Korpusbasierte Wortgrundformenliste DeReWo, v-ww-bll-320000g-2012-12-31-1.0, mit Benutzerdokumentation*. Mannheim: Institut für Deutsche Sprache, Programmbereich Korpuslinguistik. https://www.ids-mannheim.de/digspra/kl/projekte/methoden/derewo/ (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2013b. *Korpusbasierte Wortgrundformenliste DeReWo, v-ww-bll-320000g-2012-12-31-1.0, mit Benutzerdokumentation*. Mannheim: Institut für Deutsche Sprache, Programmbereich Korpuslinguistik. https://www.ids-mannheim.de/digspra/kl/projekte/methoden/derewo/ (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2018a. *Das generalisierende Personalpronomen* Man. "Propädeutische Grammatik". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/progr@mm/5204 (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2018b. *Imperativ. Formenbestand*. "Systematische Grammatik". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/systematische-grammatik/1275 (9 February, 2022).

Leibniz-Institut für Deutsche Sprache. 2019. *Wortstellung in den Außenfeldern*. "Propädeutische Grammatik". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/progr@mm/4553 (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2022. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2022-I* (Release 03/08/2022). Mannheim: Leibniz-Institut für Deutsche Sprache. www.ids-mannheim.de/dereko (25 April, 2025).

Leibniz-Institut für Deutsche Sprache. 2023a. *Glauben (Lesart 1)*. "Wörterbuch zur Verbvalenz". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/verbs/view/400667/1 (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2023b. *Helfen (Lesart 1)*. "Wörterbuch zur Verbvalenz". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/verbs/view/400561/1 (9 May, 2023).

Leibniz-Institut für Deutsche Sprache. 2023c. *Helfen (Lesart 2)*. "Wörterbuch zur Verbvalenz". Grammatisches Informationssystem grammis. https://grammis.ids-mannheim.de/verbs/view/400561/2 (9 May, 2023).

Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Berlin: Language Science Press. DOI: 10.5281/zenodo.5596236.

Lemke, Robin, Eva Horch & Ingo Reich. 2017. Optimal encoding! – Information theory constrains article omission in newspaper headlines. In Mirella Lapata, Phil Blunsom & Alexander Koller (eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 131–135. Valencia: Association for Computational Linguistics.

Lemke, Robin, Lisa Schäfer & Ingo Reich. 2022. Can identity conditions on ellipsis be explained by processing principles? In Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.), *Proceedings of Linguistic*

Evidence 2020: Linguistic theory enriched by experimental data, 541–561. Tübingen: University of Tübingen.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge: MIT Press. DOI: 10.7551/mitpress/6393.001.0001.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in neural information processing systems*, vol. 19, 849–856. Cambridge: The MIT Press. DOI: 10.7551/mitpress/7503.003.0111.

Liang, Yiming, Pascal Amsili & Heather Burnett. 2021. New ways of analyzing complementizer drop in Montréal French: Exploration of cognitive factors. *Language Variation and Change* 33(3). 359–385. DOI: 10 . 1017 / S0954394521000223.

Limesurvey GmbH. 2023. *LimeSurvey. An open source survey tool.* http://www.limesurvey.org (9 May, 2023).

Mahowald, Kyle, Evelina Fedorenko, Steven T. Piantadosi & Edward Gibson. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* 126(2). 313–318. DOI: 10.1016/j.cognition.2012.09.010.

Mandelbrot, Benoît. 1965. Information theory and psycholinguistics. In Benjamin B. Wolman & Ernest Nagel (eds.), *Scientific psychology: Principles and applications*, 550–562. New York: Basic Books.

Marslen-Wilson, William D. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244. 522–523. DOI: 10.1038/244522a0.

Marslen-Wilson, William D. 1975. Sentence perception as an interactive parallel process. *Science* 189(4198). 226–228. DOI: 10.1126/science.189.4198.226.

Meinunger, André. 2004. Verb position, verbal mood and the anchoring (potential) of sentences. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*, 313–342. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110912111.313.

Meinunger, André. 2006. On the discourse impact of subordinate clauses. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 459–488. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110922011.459.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis* (Oxford Studies in Theoretical Linguistics 1). Oxford: Oxford University Press.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27(6). 661–738. DOI: 10.1007/s10988-005-7378-3.

Merriam-Webster.com Dictionary. 2021. *Information.* https://www.merriam-webster.com/dictionary/information (6 December, 2021).

Michaelov, James A., Megan D. Bardolph, Cyma K. Van Petten, Benjamin K. Bergen & Seana Coulson. 2023. Strong prediction: Language model surprisal explains multiple N400 effects. *Neurobiology of Language* 5(1). 107–135. DOI: 10.1162/nol_a_00105.

Mörnsjö, Maria. 2002. *V1 declaratives in spoken Swedish: Syntax, information structure and prosodic pattern.* Lund: Lund University. (Doctoral dissertation).

Molnár, Valéria. 1991. *Das Topik im Deutschen und im Ungarischen* (Lunder germanistische Forschungen 58). Stockholm: Almqvist & Wiksell International.

Morgan, Jerry. 1973. Sentence fragments and the notion 'sentence'. In Braj B. Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli & Sol Saporta (eds.), *Issues in linguistics. Papers in honor of Henry and Renée Kahane*, 719–751. Urbana: University of Illinois Press.

Müller, Stefan. 2005. Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung. *Linguistische Berichte* 203. 297–330.

Musan, Renate. 2002. Informationsstrukturelle Dimensionen im Deutschen. Zur Variation der Wortstellung im Mittelfeld. *Zeitschrift für Germanistische Linguistik* 30(2). 198–221. DOI: 10.1515/zfgl.2002.015.

Musan, Renate. 2017. *Informationsstruktur.* 2nd edn. (Kurze Einführungen in die germanistische Linguistik 9). Heidelberg: Winter.

Neeleman, Ad & Kriszta Szendrői. 2005. Pro drop and pronouns. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 299–307. Somerville: Cascadilla Proceedings Project.

Neeleman, Ad & Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38(4). 671–714. DOI: 10.1162/ling.2007.38.4.671.

Nygård, Mari. 2014. Norwegian discourse ellipses in the left periphery – interacting structural and semantic restrictions. In Tor A. Åfarli & Brit Mæhlum (eds.), *The sociolinguistics of grammar* (Studies in Language Companion Series 154), 171–189. Amsterdam: John Benjamins. DOI: 10.1075/slcs.154.09nyg.

Nygård, Mari. 2018. *Norwegian discourse ellipsis: Clausal architecture and licensing conditions* (Studies in Germanic Linguistics 2). Amsterdam: John Benjamins. DOI: 10.1075/sigl.2.

Önnerfors, Olaf. 1997. *Verb-erst-Deklarativsätze: Grammatik und Pragmatik* (Lunder germanistische Forschungen 60). Stockholm: Almqvist & Wiksell International.

Oppenrieder, Wilhelm. 1987. Aussagesätze im Deutschen. In Jörg Meibauer & Deutsche Gesellschaft für Sprachwissenschaft (eds.), *Satzmodus zwischen Grammatik und Pragmatik. Referate anläßlich der 8. Jahrestagung der*

Deutschen Gesellschaft für Sprachwissenschaft, Heidelberg 1986 (Linguistische Arbeiten 180), 161–189. Tübingen: Niemeyer. DOI: 10.1515/9783111560588-010.

Oppenrieder, Wilhelm. 1991. Preposition Stranding im Deutschen? – Da will ich nichts von hören! In Gisbert Fanselow & Sascha W. Felix (eds.), *Strukturen und Merkmale syntaktischer Kategorien* (Studien zur deutschen Grammatik 39), 159–173. Tübingen: Narr.

Oppermann, Sophia J. 2021. Subject gap coordination: A diachronic view. *Journal of Historical Syntax* 5(21). 1–28. DOI: 10.18148/HS/2021.V5I16-25.58.

Pafel, Jürgen. 2011. *Einführung in die Syntax: Grundlagen – Strukturen – Theorien.* Stuttgart: Metzler. DOI: 10.1007/978-3-476-00467-3.

Pasch, Renate, Ursula Brauße, Eva Breindl & Ulrich Hermann Waßner (eds.). 2003. *Handbuch der deutschen Konnektoren: Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer (Konjunktionen, Satzadverbien und Partikeln)*, vol. 1 (Schriften des Instituts für Deutsche Sprache 9). Berlin: De Gruyter. DOI: 10.1515/9783110201666.

Paul, Hermann. 1919. *Deutsche Grammatik. Band III. Teil IV: Syntax (Erste Hälfte).* Halle/Saale: Niemeyer.

Pauly, Dennis. 2013. *Grenzfälle der Subordination : Merkmale, Empirie und Theorie abhängiger Nebensätze.* Potsdam: Universität Potsdam. (Doctoral dissertation).

Perlmutter, David M. 1971. *Deep and surface structure constraints in syntax* (The Transatlantic Series in Linguistics). New York: Holt, Rinehart & Winston.

Pešková, Andrea. 2013. Experimenting with *pro-drop* in Spanish. *SKY Journal of Linguistics* 26. 117–149.

Pittner, Karin & Judith Berman. 2021. *Deutsche Syntax: ein Arbeitsbuch.* 7th, revised and expanded edition (Narr Studienbücher). Tübingen: Narr Francke Attempto.

Platzack, Christer. 1996. Germanic verb second languages. Attract vs. repel: On optionality, A-bar movement and the symmetrical/asymmetrical verb second hypothesis. In Ewald Lang & Gisela Zifonun (eds.), *Deutsch – Typologisch* (Jahrbuch des Instituts für Deutsche Sprache 1995), 92–120. Berlin: De Gruyter. DOI: 10.1515/9783110622522-006.

Platzack, Christer. 2013. Spurious topic drop in Swedish. In Terje Lohndal (ed.), *In search of universal grammar. From Old Norse to Zoque* (Linguistik Aktuell / Linguistics Today 202), 27–50. Amsterdam: John Benjamins. DOI: 10.1075/la.202.04pla.

Poitou, Jacques. 1993. Zu assertiven V1-Sätzen. In Jean-François Marillier (ed.), *Satzanfang – Satzende. Syntaktische, semantische und pragmatische Untersuchungen zur Satzabgrenzung und Extraposition im Deutschen* (Eurogermanistik 3), 111–131. Tübingen: Narr.

Portele, Yvonne & Markus Bader. 2016. Accessibility and referential choice: Personal pronouns and d-pronouns in written German. *Discours. Revue de linguistique, psycholinguistique et informatique* 18. 1–41. DOI: 10.4000/discours.9188.

Postal, Paul M. 1974. *On raising: One rule of English grammar and its theoretical implications* (Current Studies in Linguistics 5). Cambridge: MIT Press.

Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell. DOI: 10.1002/9780470759400.

Prince, Ellen F. 1981. Toward a taxonomy of given-new formation. In Peter Cole (ed.), *Radical pragmatics*, 223–254. New York: Academic Press.

Prince, Ellen F. 1999. Subject-Prodrop in Yiddish. In Peter Bosch & Rob van der Sandt (eds.), *Focus: Linguistic, cognitive, and computational perspectives.* (Studies in Natural Language Processing), 82–104. Cambridge: Cambridge University Press.

Pritchett, Bradley L. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64(3). 539–576. DOI: 10.2307/414532.

Proske, Nadine. 2014. *ˀh ach KOMM; hör AUF mit dem klEInkram.* Die Partikel *komm* zwischen Interjektion und Diskursmarker. *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion* 15. 121–160.

R Core Team. 2021. *R: A language and environment for statistical computing. Version 4.1.0 (05/18/2021)*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/ (23 April, 2025).

Rabagliati, Hugh & Alexander Robertson. 2017. How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language* 94. 15–27. DOI: 10.1016/j.jml.2016.09.007.

Rapp, Irene. 2015. Zur Distribution von infinitiven Komplementsätzen im Deutschen. Fragen, Fakten und Faktoren. In Stefan Engelberg, Meike Meliss, Kristel Proost & Edeltraut Winkler (eds.), *Argumentstruktur zwischen Valenz und Konstruktion* (Studien zur deutschen Sprache 68), 177–200. Tübingen: Narr.

Rasinger, Sebastian M. 2013. *Quantitative research in linguistics: An introduction.* 2nd edn. (Research Methods in Linguistics). London: Bloomsbury Academic.

Rauth, Philipp. 2020. *Ditransitive Konstruktionen im Deutschen: Geschichte und Steuerung der Objektabfolge im Mittelfeld* (Studien zur deutschen Grammatik 100). Tübingen: Stauffenburg.

Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3). 372–422. DOI: 10.1037/0033-2909.124.3.372.

Rayner, Keith & George W. McConkie. 1976. What guides a reader's eye movements? *Vision Research* 16(8). 829–837. DOI: 10.1016/0042-6989(76)90143-7.

Rayner, Keith, Timothy J. Slattery, Denis Drieghe & Simon P. Liversedge. 2011. Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology. Human Perception and Performance* 37(2). 514–528. DOI: 10.1037/a0020990.

Reich, Ingo. 2003. *Frage, Antwort und Fokus* (Studia Grammatica 55). Berlin: Akademie Verlag. DOI: 10.1515/9783050082097.

Reich, Ingo. 2007. Toward a uniform analysis of short answers and gapping. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form. Generalizations across languages* (Linguistik Aktuell / Linguistics Today 100), 467–484. Amsterdam: John Benjamins. DOI: 10.1075/la.100.25rei.

Reich, Ingo. 2008. From discourse to "odd coordinations": On asymmetric coordination and subject gaps in German. In Cathrine Fabricius-Hansen & Wiebke Ramm (eds.), *'Subordination' versus 'coordination' in sentence and text. A cross-linguistic perspective* (Studies in Language Companion Series 98), 281–303. Amsterdam: John Benjamins. DOI: 10.1075/slcs.98.15rei.

Reich, Ingo. 2009a. *"Asymmetrische Koordination" im Deutschen* (Studien zur deutschen Grammatik 75). Tübingen: Stauffenburg.

Reich, Ingo. 2009b. Ist *Asymmetrische Koordination* wirklich ein Fall asymmetrischer Koordination? In Veronika Ehrich, Christian Fortmann, Ingo Reich & Marga Reis (eds.), *Koordination und Subordination im Deutschen* (Linguistische Berichte. Sonderheft 16), 203–222. Hamburg: Buske.

Reich, Ingo. 2011. Ellipsis. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 33), 1849–1874. Berlin: De Gruyter. DOI: 10.1515/9783110255072.1849.

Reich, Ingo. 2013. Asymmetrische Koordination. In Jörg Meibauer, Markus Steinbach & Hans Altmann (eds.), *Satztypen des Deutschen*, 356–371. Berlin: De Gruyter. DOI: 10.1515/9783110224832.356.

Reich, Ingo. 2017. On the omission of articles and copulae in German newspaper headlines. *Linguistic Variation* 17(2). 186–204. DOI: 10.1075/lv.14017.rei.

Reich, Ingo. 2018. Ellipsen. In Frank Liedtke & Astrid Tuchen (eds.), *Handbuch Pragmatik*, 240–251. Stuttgart: Metzler. DOI: 10.1007/978-3-476-04624-6_24.

Reich, Ingo. Forthcoming. Coordination. In Anton Benz, Werner Frey, Hand-Martin Gärtner, Manfred Krifka, Mathias Schenner & Marzena Żygis (eds.), *Handbook of clausal embedding*. Berlin: Language Science Press.

Reich, Ingo & Marga Reis. 2013. Koordination und Subordination. In Jörg Meibauer, Markus Steinbach & Hans Altmann (eds.), *Satztypen des Deutschen*, 536–569. Berlin: De Gruyter. DOI: 10.1515/9783110224832.536.

Reinhart, Tanya. 1976. *The syntactic domain of anaphora.* Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–94. DOI: 10.21825/philosophica.82606.

Reinhart, Tanya & Tal Siloni. 2005. The lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry* 36(3). 389–436. DOI: 10.1162/0024389054396881.

Reis, Marga. 1982. Zum Subjektbegriff im Deutschen. In Werner Abraham (ed.), *Satzglieder im Deutschen: Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung* (Studien zur deutschen Grammatik 15), 171–211. Tübingen: Narr.

Reis, Marga. 1995. *Wer glaubst du hat recht?* On so-called extractions from verb-second clauses and verb-first parenthetical constructions in German. *Sprache und Pragmatik* 36. 27–83.

Reis, Marga. 1996. Extractions from verb-second clauses in German? In Uli Lutz & Jürgen Pafel (eds.), *On extraction and extraposition in German* (Linguistik Aktuell / Linguistics Today 11), 45–88. Amsterdam: John Benjamins. DOI: 10.1075/la.11.03rei.

Reis, Marga. 1997. Zum syntaktischen Status unselbständiger Verbzweit-Sätze. In Christa Dürscheid, Karl Heinz Ramers & Monika Schwarz (eds.), *Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag*, 121–144. Tübingen: Niemeyer.

Reis, Marga. 2000. Anmerkungen zu Verb-erst-Satz-Typen im Deutschen. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (eds.), *Deutsche Grammatik in Theorie und Praxis*, 215–227. Berlin: De Gruyter. DOI: 10.1515/9783110933932.215.

Rizzi, Luigi. 1982. *Issues in Italian syntax* (Studies in Generative Grammar 11). Dordrecht: Foris. DOI: 10.1515/9783110883718.

Rizzi, Luigi. 1990a. *Relativized minimality* (Linguistic Inquiry Monographs 16). Cambridge: MIT Press.

Rizzi, Luigi. 1990b. Speculations on verb second. In Joan Mascaró & Marina Nespor (eds.), *Grammar in Progress. GLOW essays for Henk van Riemsdijk* (Studies in Generative Grammar 36), 375–386. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110867848.375.

Rizzi, Luigi. 1994. Early null subjects and root null subjects. In Teun Hoekstra & Bonnie D. Schwartz (eds.), *Language acquisition studies in generative grammar* (Language Acquisition and Language Disorders 8), 151–176. Amsterdam: John Benjamins. DOI: 10.1075/lald.8.09riz.

Rizzi, Luigi. 1996. Residual verb second and the *wh*-criterion. In Adriana Beletti & Luigi Rizzi (eds.), *Parameters and functional heads. Essays in comparative syntax* (Oxford Studies in Comparative Syntax), 63–90. New York: Oxford University Press. DOI: 10.1093/oso/9780195087932.003.0003.

Rizzi, Luigi. 2000. Remarks on early null subjects. In Marc-Ariel Friedemann & Luigi Rizzi (eds.), *The acquisition of syntax: Studies in comparative developmental linguistics*, 269–292. Harlow: Longman. DOI: 10.4324/9781315839899-10.

Rizzi, Luigi. 2001. Relativized minimality effects. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory* (Blackwell Handbooks in Linguistics), 89–110. Malden: Blackwell. DOI: 10.1002/9780470756416.ch4.

Roberts, Ian. 2014. Taraldsen's generalization and language change: Two ways to lose null subjects. In Peter Svenonius (ed.), *Functional structure from top to toe: The cartography of syntactic structures, vol. 9*, 115–147. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199740390.003.0005.

Rocha, Zeno. 2022. *WhatsApp in Pure CSS and JS*. CodePen. https://codepen.io/zenorocha/details/eZxYOK (24 March, 2022).

Rohdenburg, Günter. 2021. Ambiguity avoidance by means of function words in English? Providing additional corpus-based counterevidence. *Zeitschrift für Anglistik und Amerikanistik* 69(3). 207–236. DOI: 10.1515/zaa-2021-2022.

Rohrbacher, Bernhard Wolfgang. 1999. *Morphology-driven syntax: A theory of V to I raising and pro-drop* (Linguistik Aktuell / Linguistics Today 15). Amsterdam: John Benjamins. DOI: 10.1075/la.15.

Rosenkvist, Henrik. 2009. Referential null subjects in Germanic languages – an overview. *Working papers in Scandinavian syntax* 84. 151–180.

Rosenkvist, Henrik. 2012. Null subjects in Yiddish. Some preliminary observations. In Johan Brandtler, David Håkansson, Stefan Huber & Eva Klingvall (eds.), *Discourse & grammar. A Festschrift in honor of Valéria Molnár*, 433–444. Lund: Lund University. Centre for Languages & Literature.

Ross, John Robert. 1969. Guess who? In Robert I. Binnick, Alice Davison, Georgia M. Green & Jerry L. Morgan (eds.), *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*, 252–286. Chicago: Chicago Linguistic Society.

Ross, John Robert. 1982. *Pronoun deleting processes in German*. Paper presented at the annual meeting of the Linguistic Society of America, San Diego, California, 27–30 December 1982.

RStudio Team. 2021. *RStudio: Integrated development environment for R*. Boston. http://www.rstudio.com/.

Ruppenhofer, Josef. 2018. Argument omissions in multiple German corpora. In Hans C. Boas & Alexander Ziem (eds.), *Constructional approaches to syntac-*

tic structures in German (Trends in Linguistics. Studies and Monographs 322), 204–242. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110457155-006.

Samson, Edward W. 1953. Fundamental natural concepts of information theory. *ETC: A Review of General Semantics* 10(4). 283–297.

Sandig, Barbara. 2000. Zu einer Gesprächs-Grammatik: Prototypische elliptische Strukturen und ihre Funktionen in mündlichem Erzählen. *Zeitschrift für Germanistische Linguistik* 28(3). DOI: 10.1515/zfgl.2000.28.3.291.

Sanz, Montserrat, Itziar Laka & Michael K. Tanenhaus (eds.). 2013. *Language down the garden path: The cognitive and biological basis for linguistic structures.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199677139.001.0001.

Schafer, Amy J., Shari R. Speer & Paul Warren. 2005. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In John C. Trueswell & Michael K. Tanenhaus (eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 209–228. Cambridge: MIT Press.

Schäfer, Lisa. 2019. *Topic drop in German. Experimental investigations on the impact of topicality and grammatical person.* Saarbrücken: Saarland University. (BA thesis).

Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft* 40(2). 161–197. DOI: 10.1515/zfs-2021-2024.

Schäfer, Lisa. 2022. *Preverbal positioning as a licensing condition for topic drop: experimental evidence.* Poster presented at the 35th Annual Conference on Human Sentence Processing, UC Santa Cruz/online, 03/24/2022.

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. DOI: 10.3389/fpsyg.2021.661087.

Schalowski, Sören. 2009. *Über Topik-Drop im Deutschen. Untersuchung zum Einfluss der grammatischen Funktion und des Merkmals ‚Person'.* Berlin: Humboldt-Universität zu Berlin. (Magisterarbeit).

Schalowski, Sören. 2015. *Wortstellungsvariation aus informationsstruktureller Perspektive: Eine Untersuchung der linken Satzperipherie im gesprochenen Deutsch* (Interdisciplinary Studies on Information Structure 18). Potsdam: Universitätsverlag Potsdam.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, 44–49. Manchester.

Schmid, Helmut. 1995. Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-workshop*, 13–25. Dublin. DOI: 10.1007/978-94-017-2390-9_2.

Schmidt, Thomas. 2014. The research and teaching corpus of spoken German – FOLK. In *Proceedings of the Ninth Conference on International Language Resources and Evaluation (LREC'14)*, 383–387. Reykjavik: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/290_Paper.pdf (23 April, 2025).

Schnitzer, Caroline-Victoria. 2012. *Linguistische Aspekte der Kommunikation in den neueren elektronischen Medien; SMS – E-Mail – Facebook*. Munich: Ludwig-Maximilians-Universität München. (Doctoral dissertation). DOI: 10.5282/edoc.14779.

Schütze, Carson T. & Jon Sprouse. 2014. Judgment data. In Devyani Sharma & Rob Podesva (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139013734.004.

Schulz, Barbara. 2006. Tune in, drop out: Harmonic alignment as a prerequisite for German topic drop. Unpublished manuscript. University of Hawai'i at Manoa.

Schwitalla, Johannes. 2012. *Gesprochenes Deutsch: Eine Einführung*. 4th, newly revised and expanded edition (Grundlagen der Germanistik 33). Berlin: Erich Schmidt.

Selting, Margret. 1999. Kontinuität und Wandel der Verbstellung von ahd. *wanta* bis gwd. *weil*. Zur historischen und vergleichenden Syntax der *weil*-Konstruktionen. *Zeitschrift für Germanistische Linguistik* 27(2). DOI: 10.1515/zfgl.1999.27.2.167.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/44 (5 May, 2023).

Sigurðsson, Halldór Ármann. 1989. *Verbal syntax and case in Icelandic in a comparative GB approach*. Lund: University of Lund. (Doctoral dissertation).

Sigurðsson, Halldór Ármann. 2004a. Meaningful silence, meaningless sounds. *Linguistic Variation Yearbook* 4. 235–259. DOI: 10.1075/livy.4.07sig.

Sigurðsson, Halldór Ármann. 2004b. The syntax of person, tense, and speech features. *Rivista di Linguistica – Italian Journal of Linguistics* 16(1). 219–251.

Sigurðsson, Halldór Ármann. 2011. Conditions on argument drop. *Linguistic Inquiry* 42(2). 267–304. DOI: 10.1162/LING_a_00042.

Sigurðsson, Halldór Ármann & Joan Maling. 2010. The empty left edge condition. In Michael T. Putnam (ed.), *Exploring crash-proof grammars* (Language Faculty and Beyond 3), 59–86. Amsterdam: John Benjamins. DOI: 10.1075/lfab.3.04sig.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies. DOI: 10.5281/zenodo.4688088.

Smyth, Ron. 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23(3). 197–229. DOI: 10.1007/BF02139085.

Snedeker, Jesse & John Trueswell. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48(1). 103–130. DOI: 10.1016/S0749-596X(02)00519-3.

Soares, Eduardo Correa, Philip H. Miller & Barbara Hemforth. 2019. The effect of verbal agreement marking on the use of null and overt subjects: A quantitative study of first person singular in Brazilian Portuguese. *Fórum Linguístico* 16(1). 3579–3600. DOI: 10.5007/1984-8412.2019v16n1p3579.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.

Speyer, Augustin. 2007. Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26(1). 83–115. DOI: 10.1515/ZFS.2007.011.

Speyer, Augustin. 2010. Filling the German Vorfeld in written and spoken discourse. In Sanna-Kaisa Tanskanen, Marja-Liisa Helasvuo, Marjut Johansson & Mia Raitaniemi (eds.), *Discourse in interaction*, 263–290. Amsterdam: John Benjamins. DOI: 10.1075/pbns.203.19spe.

Speyer, Augustin & Ingo Reich. 2020. *Deutsche Sprachwissenschaft: Eine Einführung* (Reclams Studienbuch Germanistik). Ditzingen: Reclam.

Spycher, Samuel. 2004. *"I schribdr de no..." Schweizerdeutsche Umgangsformen in der SMS-Kommunikation* (Networx 36). Hannover: Gottfried Wilhelm Leibniz Universität, Seminar für deutsche Literatur und Sprache. DOI: 10.15488/2927.

Stainton, Robert. 2006. *Words and thoughts: Subsentences, ellipsis, and the philosophy of language*. Oxford: Clarendon Press. DOI: 10 . 1093 / acprof : oso / 9780199250387.001.0001.

Stalnaker, Robert. 1974. Pragmatic presuppositions. In Milton K. Munitz & Peter K. Unger (eds.), *Semantics and philosophy*, 197–213. New York: New York University Press.

Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Syntax and Semantics 9: Pragmatics*, 315–332. New York: Academic Press. DOI: 10.1163/9789004368873_013.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5/6). 701–721. DOI: 10.1023/A:1020867916902.

Stark, Elisabeth & Petra Meier. 2017. Argument drop in Swiss WhatsApp messages: A pilot study on French and (Swiss) German. *Zeitschrift für französische Sprache und Literatur* 127(3). 224–252. DOI: 10.25162/zfsl-2017-0006.

Statista. 2022. *Most popular global mobile messenger apps as of January 2022, based on number of monthly active users*. Statista. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ (24 March, 2022).

Steinbach, Markus. 2007. Integrated parentheticals and assertional complements. In Nicole Dehé & Yordanka Kavalova (eds.), *Parentheticals* (Linguistik Aktuell / Linguistics Today 106), 53–87. Amsterdam: John Benjamins. DOI: 10.1075/la.106.06ste.

Sternefeld, Wolfgang. 1985. Deutsch ohne grammatische Funktionen. Ein Beitrag zur Rektions- und Bindungstheorie. *Linguistische Berichte* 99. 394–439.

Sternefeld, Wolfgang. 2009. *Syntax*. 3rd edn., vol. 2 (Stauffenburg-Linguistik 31.2). Tübingen: Stauffenburg.

Sternefeld, Wolfgang. 2015. *Syntax*. 4th edn., vol. 1 (Stauffenburg-Linguistik 31.1). Tübingen: Stauffenburg.

Stolcke, Andreas. 2002. SRILM – An extensible language modeling toolkit. In *Proceedings of the Seventh International Conference on Spoken Language Processing*, 901–904. Denver. DOI: 10.21437/ICSLP.2002-303.

Stowell, Tim. 1991. Empty heads in abbreviated English. *GLOW Newsletter* 26. 56–57.

Stowell, Tim. 1996. Empty heads in abbreviated English. Revised GLOW abstract. Oakland: UCLA.

Taraldsen, Knut Tarald. 1980. On the nominative island condition, vacuous application and the that-trace filter. Paper distributed by the Indiana University Linguistics Club. Bloomington.

Tesak, Jürgen, Elisabeth Ahlsén, Gábor Györi, Päivi Koivuselkä-Sallinen, Jussi Niemi & Livia Tonelli. 1995. Patterns of ellipsis in telegraphese: A study of six languages. *Folia Linguistica* 29(3-4). 297–316. DOI: 10.1515/flin.1995.29.3-4.297.

Tesak, Jürgen & Jürgen Dittmann. 1991. Telegraphic style in normals and aphasics. *Linguistics* 29(6). 1111–1137. DOI: 10.1515/ling.1991.29.6.1111.

Tesak, Jürgen & Jussi Niemi. 1997. Telegraphese and agrammatism: A cross-linguistic study. *Aphasiology* 11(2). 145–155. DOI: 10.1080/02687039708248461.

Thielmann, Winfried. 2021. *Wortarten: Eine Einführung aus funktionaler Perspektive* (Germanistische Arbeitshefte 49). Berlin: De Gruyter. DOI: 10.1515/9783110667967.

Thiersch, Craig Lee. 1978. *Topics in German syntax*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation).

Thompson, Sandra A. & Elizabeth Couper-Kuhlen. 2005. The clause as a locus of grammar and interaction. *Discourse Studies* 7(4-5). 481–505. DOI: 10.1177/1461445605054403.

Thráinson, Höskuldur & Thóra Björk Hjartardóttir. 1986. Pro-drop, topic-drop...: Where do old and modern Icelandic fit in? In Östen Dahl & Anders Holmberg (eds.), *Scandinavian Syntax. Workshop at the Ninth Scandinavian Conference of Linguistics, Stockholm, January 1986*, 150–161. Stockholm: Institute of Linguistics, University of Stockholm.

Thrift, Erica. 2001. Topicalization and topic drop in Dutch. *Calgary Working Papers in Linguistics* 23. 49–78. DOI: 10.11575/PRISM/28966.

Thurlow, Crispin & Michele Poff. 2013. Text messaging. In Susan C. Herring, Dieter Stein, Tuija Virtanen & Wolfram Bublitz (eds.), *Pragmatics of computer-mediated communication* (Handbooks of Pragmatics 9), 163–190. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110214468.163.

Tomioka, Satoshi. 2003. The semantics of Japanese null pronouns and its cross-linguistic implications. In Kerstin Schwabe & Susanne Winkler (eds.), *The Interfaces. Deriving and interpreting omitted structures* (Linguistik Aktuell / Linguistics Today 61), 321–339. Amsterdam: John Benjamins. DOI: 10.1075/la.61.16tom.

Traxler, Matthew J., Martin J. Pickering & Charles Clifton. 1998. Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language* 39(4). 558–592. DOI: 10.1006/jmla.1998.2600.

Truckenbrodt, Hubert. 2006. On the semantic motivation of syntactic verb movement to C in German. *Theoretical Linguistics* 32(3). 257–306. DOI: 10.1515/TL.2006.018.

Trutkowski, Ewa. 2011. Referential null subjects in German. In *Proceedings of the Sixth Cambridge Postgraduate Conference in Linguistics (CamLing 2010)*, 206–217. University of Cambridge.

Trutkowski, Ewa. 2016. *Topic drop and null subjects in German* (Linguistics & Philosophy 6). Berlin: De Gruyter. DOI: 10.1515/9783110446173.

Trutkowski, Ewa. 2018. *Theta role identity as a condition on ellipsis*. Paper presented at the 40th Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Stuttgart, 03/08/2018.

Tsao, Feng-fu. 1977. *A functional study of topic in Chinese: The first step toward discourse analysis*. Taipeh: Student Book Co.

Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Laboratory Phonology* 1(2). 227–262. DOI: 10.1515/labphon.2010.012.

UCLA: Statistical Consulting Group. 2022. *R library contrast coding systems for categorical variables*. https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables (27 April, 2022).

Ueberwasser, Simone. 2015–2022. *The Swiss SMS Corpus. Documentation, facts and figures*. www.sms4science.ch (9 May, 2022).

Ulrich, Miorita. 1986. Thetisch vs. kategorisch und Informationsstruktur. *Acta Linguistica Academiae Scientiarum Hungaricae* 36(1-4). 221–234.

van Craenenbroeck, Jeroen & Jason Merchant. 2013. Ellipsis phenomena. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 701–745. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511804571.025.

van Craenenbroeck, Jeroen & Tanja Temmerman (eds.). 2018. *The Oxford handbook of ellipsis* (Oxford Handbooks). Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198712398.001.0001.

van Gompel, Roger P. G., Martin J. Pickering, Jamie Pearson & Simon P. Liversedge. 2005. Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language* 52(2). 284–307. DOI: 10.1016/j.jml.2004.11.003.

van Gompel, Roger P. G., Martin J. Pickering & Matthew J. Traxler. 2001. Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language* 45(2). 225–258. DOI: 10.1006/jmla.2001.2773.

van de Velde, Marc. 1986. Zum 65. gratuliere ich Ihnen und biete ?(ich) Ihnen diesen Beitrag an. In Heinrich Leonhard Cox, V. F. Vanacker, Edward Verhofstadt & Gilbert A. R. De Smet (eds.), *Wortes anst, verbi gratia: Donum natalicium Gilbert A. R. de Smet* (ML 331), 503–512. Leuven: Uitgeverij Acco.

van der Wurff, Wim. 1996. Syntactic reconstruction and reconstructibility: Proto-Indo-European and the typology of null objects. In Jacek Fisiak (ed.), *Linguistic reconstruction and typology* (Trends in Linguistics. Studies and Monographs 96), 337–356. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110816501.337.

Volodina, Anna. 2011. Null ist nicht gleich Null: Zur diachronen Entwicklung von Nullsubjekten im Deutschen. *Jahrbuch für Germanistische Sprachgeschichte* 2(1). 269–283. DOI: 10.1515/9783110236620.269.

Volodina, Anna & Edgar Onea. 2012. Am Anfang war die Lücke. In Jochen Bär & Marcus Müller (eds.), *Geschichte der Sprache – Sprache der Geschichte*, 207–238. Berlin: De Gruyter. DOI: 10.1515/9783050058245-009.

von Heusinger, Klaus, Claudia Maienborn & Paul Portner (eds.). 2011. *Semantics. An international handbook of natural language meaning*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 33). Berlin: De Gruyter Mouton. DOI: 10.1515/9783110255072.

von der Gabelentz, Georg. 1868. Ideen zu einer vergleichenden Syntax. Wort- und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6. 376–384.

Walker, Marilyn, Aravind K. Joshi & Ellen Prince. 1998. Centering in naturally-occurring discourse: An overview. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering theory in discourse*, 1–28. Oxford: Oxford University Press.

Ward, Gregory & Betty Birner. 1994. A unified account of English fronting constructions. *University of Pennsylvania Working Papers in Linguistics* 1. 159–165.

Wasow, Thomas. 2015. Ambiguity avoidance is overrated. In Susanne Winkler (ed.), *Ambiguity: Language and communication*, 29–48. Berlin: De Gruyter. DOI: 10.1515/9783110403589-003.

Wasow, Thomas, Amy Perfors & David Beaver. 2005. The puzzle of ambiguity. In C. Orhan Orgun & Peter Sells (eds.), *Morphology and the web of grammar*, 265–282. Stanford: CSLI Publications.

Weaver, Warren. 1949. Recent contributions to the mathematical theory of communication. In Shannon, Claude E. & Weaver, Warren. *The mathematical theory of communication*. Urbana: The University of Illinois Press. 3–28.

Weerman, Fred. 1989. *The V2 conspiracy. A synchronic and a diachronic analysis of verbal positions in Germanic languages* (Publications in Language Science 31). Dordrecht: Foris. DOI: 10.1515/9783110250442.

Wegener, Heide. 1990. Komplemente in der Dependenzgrammatik und in der Rektions- und Bindungstheorie. Die Verwendung der Kasus im Deutschen. *Zeitschrift für germanistische Linguistik* 18(2). 150–184. DOI: 10.1515/zfgl.1990.18.2.150.

Wegener, Heide. 2000. *Da*, *denn* und *weil* – der Kampf der Konjunktionen. Zur Grammatikalisierung im kausalen Bereich. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (eds.), *Deutsche Grammatik in Theorie und Praxis*, 69–82. Tübingen: Niemeyer. DOI: 10.1515/9783110933932.69.

Weil, Henri. 1844. *De l'ordre des mots dans les langues anciennes comparées aux langues modernes. Question de grammaire générale*. Paris: Joubert.

Weir, Andrew. 2012. Left-edge deletion in English and subject omission in diaries. *English Language & Linguistics* 16(1). 105–129. DOI: 10.1017/S136067431100030X.

Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. New York: Springer. https://ggplot2.tidyverse.org (23 April, 2025).

Wilder, Chris. 1996. V2-Effekte: Wortstellungen und Ellipsen. In Ewald Lang & Gisela Zifonun (eds.), *Deutsch – Typologisch* (Jahrbuch des Instituts für Deutsche Sprache 1995), 142–180. Berlin: De Gruyter. DOI: 10.1515/9783110622522-008.

Wilder, Chris. 1997. Some properties of ellipsis in coordination. In Artemis Alexiadou & Tracy Alan Hall (eds.), *Studies on universal grammar and typological variation* (Linguistik Aktuell / Linguistics Today 13), 59–107. Amsterdam: John Benjamins. DOI: 10.1075/la.13.04wil.

Wöllstein, Angelika. 2014. *Topologisches Satzmodell*. 2nd edn. (Kurze Einführungen in die germanistische Linguistik 8). Heidelberg: Winter.

Wöllstein, Angelika. 2018. Topologisches Satzmodell. In Jörg Hagemann & Sven Staffeldt (eds.), *Syntaxtheorien: Analysen im Vergleich*, 2nd edn. (Stauffenburg Einführungen 28), 143–164. Tübingen: Stauffenburg.

Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37(1). 111–130. DOI: 10.1162/002438906775321175.

Wunderlich, Dieter. 1984. Zur Syntax der Präpositionalphrasen im Deutschen. *Zeitschrift für Sprachwissenschaft* 3(1). 65–99. DOI: 10.1515/zfsw.1984.3.1.65.

Wunderlich, Dieter. 1988. Some problems of coordination in German. In Uwe Reyle & Christian Rohrer (eds.), *Natural language parsing and linguistic theories* (Studies in Linguistics and Philosophy 35), 289–316. Dordrecht: Springer. DOI: 10.1007/978-94-009-1337-0_11.

Yung, Frances, Kevin Duh, Taku Komura & Yuji Matsumoto. 2016. Modelling the usage of discourse connectives as rational speech acts. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, 302–313. Berlin: Association for Computational Linguistics. DOI: 10.18653/v1/K16-1030.

Zarcone, Alessandra & Vera Demberg. 2021. A bathtub by any other name: The reduction of German compounds in predictive contexts. In Cognitive Science Society (ed.), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society (CogSci 2021)*, 749–755. Austin: Cognitive Science Society. https://escholarship.org/uc/item/3w6451rz (30 April, 2025).

Zdorenko, Tatiana. 2010. Subject omission in Russian: A study of the Russian National Corpus. In Stefan Th. Gries, Stefanie Wulff & Mark Davies (eds.), *Corpus-linguistic applications. Current studies, new directions* (Language and Computers 71), 119–133. Amsterdam: Rodopi. DOI: 10.1163/9789042028012_009.

Zifonun, Gisela. 1995. Minimalia grammaticalia: Das nicht-phorische *es* als Prüfstein grammatischer Theoriebildung. *Deutsche Sprache* 23(1). 39–60.

Zifonun, Gisela. 2001. *Grammatik des Deutschen im europäischen Vergleich: das Pronomen. Teil 1: Überblick und Personalpronomen* (amades – Arbeitspapiere und Materialien zur deutschen Sprache 4/01). Mannheim: Institut für Deutsche Sprache.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*, vol. 1 (Schriften des Instituts für Deutsche Sprache 7). Berlin: De Gruyter. DOI: 10.1515/9783110872163.

Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin.

Zybatow, Tatjana. 2014. Das Vorfeld und seine Besetzung in Lehrbuchtexten. In Ursula Bredel & Claudia Schmellentin (eds.), *Welche Grammatik braucht der Grammatikunterricht?* (Thema Sprache – Wissenschaft für den Unterricht 8), 87–105. Baltmannsweiler: Schneider Verlag Hohengehren.

Name index

Ackema, Peter, 13, 28, 34, 50, 52, 85, 97, 142
Aissen, Judith, 201
Alday, Phillip M., 72, 289
Altmann, Gerry T. M., 177
Altmann, Hans, 32
Anand, Pranav, 85, 124
Androutsopoulos, Jannis, 4, 7, 21, 191, 262, 266–269, 272, 277, 278, 292, 302, 305, 310, 316–318, 320
Ariel, Mira, 9, 85, 139, 142, 144, 256, 301, 305, 311, 358
Arnold, Jennifer E., 136, 264
Asr, Fatemeh Torabi, 9, 173
Attneave, Fred, 165
Auer, Peter, 4, 7, 13, 15, 18, 21, 32, 33, 49, 50, 57, 58, 99, 101, 151–157, 159, 161, 162, 192, 200, 207, 239, 253, 254, 256, 260, 266, 267, 273, 277, 279, 285, 288, 290, 292, 293, 296, 300, 310–312, 315, 359, 361
Avrutin, Sergey, 44
Axel, Katrin, 31
Aylett, Matthew, 172

Bachmann-Stein, Andrea, 27
Bader, Markus, 136, 137, 212–214
Barbosa, Pilar, 30
Barnickel, Katja, 101
Baroni, Marco, 136
Barr, Dale J., 72
Barton, Ellen L., 24
Bates, Douglas, 341
Beißwenger, Michael, 2, 352
Bell, Allan, 7, 180, 182
Berman, Judith, 16, 50, 58, 99
Bever, Thomas G., 264
Bhat, D. N. S., 144
Biber, Douglas, 19
Bîlbîie, Gabriela, 1
Birner, Betty, 43
Blatz, Friedrich, 13, 138
Blümel, Andreas, 86
Bonitz, Petra-Kristin, 101, 102
Bosch, Peter, 203, 204, 217
Boston, Marisa Ferrara, 363
Boswijk, Vincent, 139
Brehm, Laurel, 289
Breindl, Eva, 102, 103, 113, 114
Brennan, Susan E., 264
Broekhuis, Hans, 36
Brommer, Sarah, 23
Brouwer, Harm, 363
Brown-Schmidt, Sarah, 165, 166
Brysbaert, Marc, 83
Bühler, Karl, 138
Burger, Susanne, 18, 24, 133
Büring, Daniel, 60, 119, 120

Cardinaletti, Anna, 4, 14, 16, 18, 20, 24, 51, 58, 62, 84, 88, 118–120, 122, 129, 130, 138, 193, 202, 203

Carifio, James, 71
Chafe, Wallace, 9, 53, 133, 140–142, 144, 358
Chambers, Craig G., 241
Charniak, Eugene, 172
Chiarcos, Christian, 139
Chomsky, Noam, 28, 119, 123, 196
Christensen, Rune Haubo Bojesen, 71, 81, 94, 109, 230, 236, 242, 246, 248, 303, 308, 346, 350
Cinque, Guglielmo, 119
Clark, Herbert H., 131
Clickworker, 69, 80, 107, 229, 241, 246, 340, 345, 349
Coler, Matt, 139
Colonna, Saveria, 239
Comrie, Bernard, 201
Consten, Manfred, 131
Corver, Norbert, 36
Couper-Kuhlen, Elizabeth, 158
Crawley, Rosalind A., 239
Crump, Matthew J. C., 82

D'Alessandro, Roberta, 30
Daneš, František, 54
Davidson, Brad, 29
Dell, Gary S., 264
Demberg, Vera, 9, 169, 173, 182, 363
den Besten, Hans, 15
den Dikken, Marcel, 1
Deppermann, Arnulf, 158
Dittmann, Jürgen, 20, 21, 23, 24, 151, 229, 253, 310
Döring, Nicola, 4, 7, 21, 23, 266, 268, 277, 310
Drach, Erich, 15, 49, 50
Dryer, Matthew S., 28, 43
Duarte, Maria Eugênia Lamoglia, 270
Duden, 13, 17, 195, 200, 202–204, 214
Dudenredaktion, 199
Duguine, Maia, 30
Dürscheid, Christa, 20, 23, 24

Eckert, Miriam, 14, 29, 152, 162
Eide, Kristin Melum, 38, 42
Eisenberg, Peter, 58, 62, 99, 102, 188, 274
Engel, Ulrich, 16, 212
Erdmann, Oskar, 50
Erteschik-Shir, Nomi, 13, 50, 54, 57, 66, 200

Falk, Cecilia, 65
Fanselow, Gisbert, 183
Fenk, August, 10, 169, 171, 172
Fenk, Gertraud, 10, 169, 171, 172
Fenk-Oczlon, Gertraud, 171, 172
Ferreira, Victor S., 261, 264
Féry, Caroline, 161
Filippova, Katja, 213
Firbas, Jan, 54, 66
Fischer, Kerstin, 101
Fortmann, Christian, 104
Fox, Barbara A., 158
Frank, Austin F., 173
Frank, Nicola, 88
Frank, Stefan L., 363
Franks, Steven, 44, 45
Frey, Werner, 58, 59, 79, 100, 356
Freywald, Ulrike, 9, 11, 14, 15, 51, 52, 60, 84, 86–88, 90–92, 97, 115–118, 122–124, 129, 139, 140, 356, 357
Frick, Karina, 4, 7, 13, 15, 21, 24, 31, 34, 49, 57, 62, 63, 75, 100, 138, 139, 151, 191, 193, 205–207, 221, 222, 262, 266, 268, 269,

272, 277, 278, 292, 302, 305, 310, 316, 320–322, 327, 328, 353
Fries, Norbert, 2, 3, 13–20, 25, 29, 31, 49, 57, 60, 62, 63, 129–131, 138, 139, 157, 193, 194, 197, 200, 201, 204, 205, 219, 223
Frisch, Stefan, 183

Gallmann, Peter, 99, 188
Gansel, Christina, 20
Genzel, Dmitriy, 172
Givón, Talmy, 225, 237
Gohl, Christine, 101
Goldrick, Matthew, 168
Gordishevsky, Galina, 44
Grewendorf, Günther, 88
Grice, Herbert Paul, 56, 263
Grosz, Barbara J., 194, 238
Gundel, Jeanette K., 54, 55, 58
Günthner, Susanne, 4, 13, 57, 101, 130, 152–155, 159, 161
Gutermuth, Silke, 212, 213

Habscheid, Stephan, 20
Haegeman, Liliane, 4, 13, 34, 35, 41–44, 63, 101, 118, 119, 255, 257
Haider, Hubert, 197
Håkansson, Gisela, 39
Hale, John, 165, 169
Halle, Morris, 125
Halliday, Michael Alexander Kirkwood, 19, 53, 151
Hamann, Cornelia, 35
Hankamer, Jorge, 76
Hansen-Schirra, Silvia, 212, 213
Harbusch, Karin, 213
Hård af Segerstad, Ylva, 23
Hartmann, Katharina, 2, 119, 120
Hasan, Ruqaiya, 151
Haspelmath, Martin, 196
Häussler, Jana, 212–214
Heidolph, Karl Erich, 151
Helmer, Henrike, 4, 7, 9, 11, 14, 15, 29, 38, 49, 50, 52, 57, 76–78, 118, 129–134, 141–143, 145, 146, 149–153, 158–162, 192, 269, 273, 316, 318–320, 322, 353, 356
Hemforth, Barbara, 239
Herling, Simon Heinrich Adolf, 50
Hinrichs, Erhard W., 18, 133
Hjartardóttir, Thóra Björk, 13, 14
Hoberg, Ursula, 213
Hockett, Charles F., 54–56
Hoffmann, Ludger, 13
Hofmeister, Philip, 183
Höhle, Tilman N., 15, 49, 50, 101, 102, 104, 114, 160
Hole, Daniel, 18
Horch, Eva, 7, 17, 64, 104, 184, 355
Huang, C.-T. James, 4, 14, 20, 24, 25, 27, 30, 50, 52, 57, 58, 85, 97, 118–120, 193
Huang, Yan, 144

Ihsane, Tabea, 43, 44
Imo, Wolfgang, 4, 13, 18, 21, 32, 33, 101, 188, 192, 253–255, 260, 301, 310–312, 315, 318, 319, 322, 325–327, 353
Institut für deutsche Sprache, Mannheim, 319

Jaccard, James, 71
Jaeger, T. Florian, 9, 10, 161, 163, 170–173, 359
Jaensch, Carol, 13, 16–18, 20, 24, 25, 50, 52, 57, 58, 84, 195, 200

Jamieson, Susan, 71
Jansen, Frank, 36, 37
Järvikivi, Juhani, 239
Johannessen, Janne Bondi, 100, 115
Jürgens, Frank, 20

Kaatari, Henrik, 173
Kaiser, Elsi, 239
Kaiser, Georg A., 62
Kamide, Yuki, 177
Keenan, Edward L., 201
Kehler, Andrew, 113, 114
Keller, Frank, 169, 363
Kempen, Gerard, 213
Kennedy, Christopher, 261
Klein, Wolfgang, 14, 17, 29, 76, 101, 194
Koch, Peter, 14, 20, 25, 26, 113, 355
Konopka, Agnieszka E., 165, 166
Korth, Manuela, 262
Kraljic, Tanya, 264
Kravtchenko, Ekaterina, 173, 182
Krifka, Manfred, 54, 56, 60, 61, 66, 225, 232
Kroll, Margaret, 85, 124
Külpmann, Robert, 77
Kupietz, Marc, 89
Kurumada, Chigusa, 173

Lambrecht, Knud, 53–56, 58, 164, 237
Leibniz-Institut für Deutsche Sprache, 77, 89, 99, 188, 195, 215, 262, 263, 323
Lemke, Robin, 9, 42, 165, 169, 173, 184
Levelt, Willem J. M., 165, 166
Levy, Roger, 9, 10, 161, 163, 165, 169–173, 264, 265, 320, 359
Liang, Yiming, 173
Limesurvey GmbH, 69, 80, 107, 229, 241, 246, 307, 340, 345, 349

Mahowald, Kyle, 173
Maling, Joan, 118, 125–127, 178
Mandelbrot, Benoît, 164
Marantz, Alec, 125
Marslen-Wilson, William D., 177
McConkie, George W., 83
Meier, Petra, 24
Meinunger, André, 88
Merchant, Jason, 1, 2, 42, 85, 149, 184
Merriam-Webster.com Dictionary, 164
Michaelov, James A., 363
Molnár, Valéria, 53, 54
Morgan, Jerry, 42, 184
Mörnsjö, Maria, 39, 118, 122, 125, 202, 203, 223
Müller, Stefan, 16
Musan, Renate, 53, 54, 251

Neeleman, Ad, 13, 28, 30, 34, 50, 52, 85, 97, 142
Niemi, Jussi, 24
Nygård, Mari, 29, 38, 42, 76, 118, 122, 125
Nykiel, Joanna, 1

Oesterreicher, Wulf, 14, 20, 25, 26, 113, 355
Onea, Edgar, 17, 18, 20, 24, 25, 60, 63, 97, 99, 130, 133, 134, 137, 141, 142, 188, 200, 202, 203, 214, 223, 253, 255, 256, 260, 265, 290, 301, 310, 311
Önnerfors, Olaf, 31–34
Oppenrieder, Wilhelm, 4, 32, 33, 49, 57, 123, 152–155

Oppermann, Sophia J., 101

Pafel, Jürgen, 99
Pasch, Renate, 99, 102, 114
Paul, Hermann, 53, 54
Pauly, Dennis, 89
Perla, Rocco, 71
Perlmutter, David M., 28
Pešková, Andrea, 29
Pittner, Karin, 16, 50, 58, 99
Platzack, Christer, 39, 118, 122–124
Plunkett, Kim, 35
Poff, Michele, 23, 24
Poitou, Jacques, 4, 20, 24, 25, 33, 36, 57, 63, 150, 155–157, 205, 206, 222, 262, 316
Portele, Yvonne, 136, 137
Postal, Paul M., 2
Prince, Alan, 201
Prince, Ellen F., 40, 131, 140, 141, 144
Pritchett, Bradley L., 261
Proske, Nadine, 101

Rabagliati, Hugh, 264
Rapp, Irene, 89
Rasinger, Sebastian M., 71
Rauth, Philipp, 213
Rayner, Keith, 83
Reich, Ingo, 1, 7, 13, 14, 17, 42, 49, 64, 75, 85, 86, 99–102, 114–116, 119–121, 166, 184, 355
Reinhart, Tanya, 51, 54–56, 196, 237
Reis, Marga, 3, 13, 15, 19, 24, 29, 32, 36, 49, 63, 86, 88–90, 102, 114, 116, 125, 193, 357
Rizzi, Luigi, 4, 9, 11, 28, 34, 35, 50, 51, 84–88, 90–92, 97, 100, 115, 117–119, 202, 356, 357
Roberts, Ian, 254
Robertson, Alexander, 264
Rocha, Zeno, 70
Rohdenburg, Günter, 264
Rohrbacher, Bernhard Wolfgang, 34, 39, 40
Rosenkvist, Henrik, 31, 40
Ross, John Robert, 2, 3, 13, 85, 193
Rudin, Deniz, 85, 124
Ruppenhofer, Josef, 13, 14, 20, 27, 63, 64, 75, 193, 195, 205, 207, 208, 222

Sag, Ivan, 76
Samson, Edward W., 165
Sandig, Barbara, 4, 13, 32, 33, 49, 57, 133, 151, 152, 154, 159, 161
Sanz, Montserrat, 264
Schafer, Amy J., 264
Schäfer, Lisa, 66, 163, 173, 189, 226, 237, 253
Schalowski, Sören, 14, 17, 18, 32, 49, 194, 195, 255, 260
Schmid, Helmut, 184, 215
Schmidt, Gurly, 4, 7, 21, 191, 262, 266–269, 272, 277, 278, 292, 302, 305, 310, 316–318, 320
Schmidt, Thomas, 76, 159, 263, 319
Schnitzer, Caroline-Victoria, 229
Schulz, Barbara, 60, 201, 219
Schütze, Carson T., 71
Schwarz-Friesel, Monika, 131
Schwitalla, Johannes, 4, 32, 130, 142, 149, 152, 153
Selting, Margret, 113
Shannon, Claude E., 163–165, 167–169, 359
Siewierska, Anna, 256
Sigurðsson, Halldór Ármann, 13, 14, 17, 27, 28, 30, 37–39, 98, 118,

119, 125–127, 178, 202, 219, 223, 254
Siloni, Tal, 196
Silverstein, Michael, 201
Sitta, Horst, 188
Smolensky, Paul, 201
Smyth, Ron, 241
Snedeker, Jesse, 264
Soares, Eduardo Correa, 265, 270, 271, 300
Sollid, Hilde, 38, 42
Sperber, Dan, 150, 152
Speyer, Augustin, 59, 99, 194, 213, 238
Sprouse, Jon, 71
Spycher, Samuel, 25
Stainton, Robert, 128
Stalnaker, Robert, 56, 164
Stark, Elisabeth, 24
Statista, 70
Steinbach, Markus, 89
Sternefeld, Wolfgang, 4, 14, 50, 57, 122, 123, 196, 197, 199, 204, 223
Stolcke, Andreas, 331
Stowell, Tim, 85
Strube, Michael, 213
Symanczyk Joppe, Vilma, 77
Szendrői, Kriszta, 30

Taraldsen, Knut Tarald, 28, 254
Temmerman, Tanja, 1
Tesak, Jürgen, 24, 151, 253, 310
Thielmann, Winfried, 188
Thiersch, Craig Lee, 15
Thompson, Sandra A., 158
Thráinson, Höskuldur, 13, 14
Thrift, Erica, 36, 37
Thurlow, Crispin, 23, 24
Tomioka, Satoshi, 30
Traxler, Matthew J., 264
Truckenbrodt, Hubert, 88
Trueswell, John, 264
Trutkowski, Ewa, 7, 13, 14, 20, 25, 29, 49, 50, 52, 57, 63, 65, 66, 74, 84, 85, 97, 118, 122–124, 130, 131, 139, 140, 157, 194, 195, 197–200, 205, 208–211, 214, 223, 257–260, 262, 364
Tsao, Feng-fu, 120
Turk, Alice, 172

UCLA: Statistical Consulting Group, 72
Ueberwasser, Simone, 206
Ulrich, Miorita, 66

van Craenenbroeck, Jeroen, 1
van de Velde, Marc, 102, 103
van der Wurff, Wim, 39
van Gompel, Roger P. G., 264
Volodina, Anna, 13, 17, 18, 20, 24, 25, 29, 49–51, 57, 58, 60, 63, 84, 88, 97, 99, 130, 133, 134, 137, 138, 141, 142, 188, 194, 200, 202, 203, 214, 223, 253, 255, 256, 260, 265, 290, 301, 310, 311
von der Gabelentz, Georg, 53, 54
von Heusinger, Klaus, 1

Walker, Marilyn, 194, 238
Wan, Choi K., 71
Ward, Gregory, 43
Wasow, Thomas, 261, 264
Weaver, Warren, 164, 168, 169
Weerman, Fred, 36
Wegener, Heide, 113, 209

Weil, Henri, 53
Weir, Andrew, 41, 42
Weiß, Helmut, 31
Wickham, Hadley, 21, 73, 170, 330
Wilder, Chris, 42, 43, 101
Wilson, Deirdre, 150, 152
Wöllstein, Angelika, 15, 50, 51, 99
Woolford, Ellen, 196
Wunderlich, Dieter, 102, 123

Yung, Frances, 173

Zarcone, Alessandra, 173
Zdorenko, Tatiana, 44, 256
Zifonun, Gisela, 4, 13–15, 32, 49, 57, 99, 102, 130, 138, 177, 188, 192, 195, 204, 207, 254, 255, 260, 262, 274, 290, 301, 305, 311, 315, 317–319, 321, 322, 325, 328–330, 334, 336, 337, 348, 353, 361
Zipf, George Kingsley, 164
Zybatow, Tatjana, 212, 213

Language index

Basque, 28

Chinese, 27, 30, 120

Danish, 34–35, 43
Dutch, 35–38, 85, 142, 202, 276

English, 23, 28, 34, 40–44, 58, 85, 120, 173, 194, 238, 255, 257, 258, 355, 356, 363

French, 28, 40–44, 173, 355, 356, 363

German, 13–16, 31–34, 42, 43, 49–52, 57–59, 84, 85, 121–124, 194, 197–204, 212–223, 238–239, 253–260, 262–263, 285, 323–324

Hebrew, 28

Icelandic, 37–38, 202
Italian, 28

Japanese, 30, 173, 201, 238

Korean, 30

Norwegian, 38, 42, 76, 122, 202

Portuguese, 270–271

Russian, 44, 45, 173, 355, 356

Spanish, 28–29
Swedish, 23, 38–39, 43, 65, 122, 202, 203, 254

Yiddish, 39–40

Subject index

Acceptability rating study, 7, 36–37, 66–75, 78–83, 91–97, 104–116, 182–183, 192, 198–199, 208–211, 226–250, 270–271, 290–310, 338–353
Accusative case, 187–188, 194–199, 212–223
Adjunct, 15, 18–19, 32, 42, 88, 90, 120
Adverbial, 17–18, 34, 58, 59, 78, 79, 122, 123, 205, 212, 213
Ambiguity, 38, 39, 178, 200–201, 234, 245, 254, 256, 259–265, 279–283, 297, 320, 324, 360
Ambiguity avoidance, 29, 260–265, 270–271, 285–286, 288, 297, 300, 312–313, 321, 324, 361–362
Animacy, 36–37, 205, 270–271, 276
Antecedent, 13, 17, 21, 40, 52, 85, 97, 113, 115, 117, 124, 128, 130–146, 150, 179, 186, 195, 197–200, 208–211, 214, 223, 251, 256–260, 271, 274–276, 293, 300, 358
Argument, 15–19, 32, 34, 42, 44, 45, 58, 88, 152, 193, 196–197, 211, 212, 356
Argument omission, 28, 40–46, 77–83, 125–127, 173, 356
Article, 42, 83, 323
Article omission, 19, 85, 119–122, 128, 173, 355
Audience design, 7, 180, 182–183, 261–264, 273
Auxiliary, 34, 42, 186, 234, 262, 283, 315–326, 328–337, 348–354, 360, 361

Binding theory, 88, 89, 100

C-command, 9, 11, 46, 50–51, 85, 85, 85–91, 97, 115–117, 127, 356–357
Channel capacity, 164, 167–177, 251, 273, 334
Common ground, 54, 56, 61, 164
Complement clause, 88–90
Complementizer, 31, 42–44, 87
Complementizer omission, 42, 43, 173
Complementizer phrase, 9, 15, 50, 51, 85, 88, 126, 126, 126
Conjunction, 53, 98–117, 189, 357
Conjunction phrase, 100, 115–117, 357
Contrastive topic, 60–61, 75, 99, 116, 203
Copula, 62, 159, 186, 262, 283, 315–326, 328–337, 353–354, 360, 361
Copula omission, 19, 85, 119–122, 128, 355
Corpus, 2, 7–8, 21–27, 36–38, 44, 63–65, 76–78, 89, 98–100, 131–133, 136, 145, 157, 159, 160,

181–191, 195, 203, 205–208, 212–223, 229, 257, 263, 264, 266–270, 272–290, 302, 311, 312, 316–338, 352–356, 359–363

Dative case, 18, 187–188, 194–199, 209, 211–223, 363
Determiner phrase, 1, 17, 85
Discourse *pro*-drop, 27, 30, 45
Discourse topic, 55, 57, 66, 150, 158, 226–227, 232–236, 251, 360

Embedding, 16, 31, 40–44, 51, 52, 84–97, 102, 357
Expletive, 3, 17, 35, 38, 39, 42–43, 55, 58, 62–76, 116, 119–128, 138–139, 144, 150, 153, 186, 207, 208, 220–222, 274–279, 356–358

Focus, 41, 53, 60–61, 89, 115, 125, 159–161
Fragment, 42, 128, 173, 184, 210, 214

Genitive case, 187, 194, 196–199, 209–211, 220–223, 363–364
Givenness, 43, 54, 76, 139–145, 152, 153, 218, 358

Imperative, 15, 77, 177, 262–263, 265, 280–283
Information, 164–167, 169, 170, 172, 173, 358, 362
Information structure, 53–57
Inherent case, 196, 209

Left dislocation, 43, 97–100, 251
Lexical case, 194–199, 209, 211–213, 221–223
Lexical verb, 186, 228, 241, 248–250, 254, 262, 265, 283, 289, 294, 296, 298–300, 315–337, 348–354, 360, 361

Middle field, 11, 15, 29, 45, 49–51, 58, 59, 75–83, 97, 101, 117, 124, 126, 213, 238, 356–357
Modal verb, 42, 186, 201, 244, 245, 248–250, 254, 262, 283, 296–300, 315–326, 328–337, 353–354, 360, 361

Noisy channel, 168–169
Nominative case, 29, 31, 122, 188, 193, 197, 199, 209, 215–217, 220
Null subject, 4, 27–28, 31, 40–46, 50, 84, 119, 122, 173, 254, 257–260, 270–271, 355

Object omission, 86, 355
Operator analysis, 84, 117–122, 127, 128, 357
Optimality theory, 59, 201

Parser, 161, 176–177, 265
PF-deletion analysis, 117, 118, 125–128, 357
Pre-prefield, 99–100
Predictability, 6, 10, 83, 114, 117, 135, 136, 161, 163, 170, 172–175, 177, 179, 183, 213, 214, 217, 218, 225, 226, 244, 250, 251, 269, 272, 273, 277, 284, 292, 311, 323, 334, 337, 359, 360
Prefield, 9, 13–16, 45–46, 49–128, 150, 186, 188–189, 203–204, 212–217, 220–221, 225–226, 355–357

Prepositional object, 36, 123, 187, 194, 194, 195, 199, 220, 221
Processing effort, 6, 9–11, 129–130, 141–142, 144, 146, 150, 152, 163, 169–180, 182–183, 192, 245, 251, 261, 264–265, 273, 285–288, 290, 293, 296, 297, 300, 312, 323, 329, 330, 334–338, 347, 352, 354, 358–359, 361–362
Production, 7–8, 114, 180–183, 264–265, 362
Prosody, 60, 89, 154, 159–162, 172, 177, 204, 223, 262, 264

Recoverability, 9, 15, 17, 19, 21, 30, 36, 60–61, 116, 117, 128–146, 163, 177–179, 186–189, 191, 192, 201, 218–219, 245, 251, 254–256, 258, 272–273, 280, 285, 288–289, 293, 296, 300–302, 305–306, 310–312, 356–362
Reflexive verb, 186, 283, 320–321, 325–337
Reflexivity, 200–201, 219, 260
Relevance theory, 150, 152

Salience, 43, 127, 129–131, 139, 142–143, 150, 201, 256–257, 269, 273–274, 276–278
Sluicing, 2, 85, 124, 173
Structural case, 196–198, 209, 211, 212, 223
Subject gap, 101–116, 119, 128
Syncretism, 39, 177, 186, 189, 200–201, 215, 219, 223, 244, 248–250, 254, 259–263, 265, 270–271, 279–289, 296, 297, 299–300, 311–313, 321, 324, 354, 360, 362

Text type, 2–7, 10, 14, 19–27, 31, 33, 34, 37, 38, 40–42, 44–46, 64, 121, 122, 128, 161, 184, 189–191, 208, 213–214, 221, 255–257, 272–273, 305, 306, 310, 323, 331, 352, 355–356, 359, 363
Topic, 53–75, 79, 139, 142, 149, 152–153, 201, 219, 225–251, 360
Topological field model, 15, 49–50, 99

Uniform information density, 10, 161, 170–177, 182, 211, 226, 251, 273, 330, 358–361
Unigram surprisal, 330–337

V1 declarative, 8, 19, 31–34, 39, 45, 184, 266, 355
V2 word order, 2, 8, 15, 33–34, 40, 45, 50, 119, 331, 363
Verb phrase, 1, 37, 88, 90, 173, 273, 276–279
Verbal inflection, 28, 31, 144, 178–180, 191, 253–260, 270–274, 279–290, 293, 296–300, 311–312, 359–362

The licensing and usage of topic drop in German

This book is concerned with the licensing and usage of the elliptical construction topic drop in German. The term topic drop refers to the omission of the preverbal constituent in declarative verb-second sentences, for example, the omission of the subject *ich* ('I') in the sentence *Bin gleich zurück* ('Am right back'). Topic drop exists in most of the Germanic verb-second languages and typically occurs in spoken language and text types such as SMS, chats, notes, etc.

While much of the previous research has focused on individual specific properties of topic drop, often adopting a purely theoretical perspective, this book presents a systematic investigation of both the syntactic properties and usage conditions of topic drop based on empirical evidence from a corpus study and 12 acceptability rating studies.

The first part of the book investigates the licensing of topic drop, in particular its restriction to the preverbal "prefield" position. The results of four rating studies on topic drop in different prefield configurations lead to a refined prefield condition based on proposals by Rizzi (1994) and Freywald (2020) that is independent of topicality. Moreover, they inform the discussion on the most suitable syntactic analysis of topic drop, supporting a PF-deletion approach.

The second part of the book presents and tests an information-theoretic account of topic drop usage that builds on the Uniform Information Density hypothesis (Levy & Jaeger 2007). In a corpus study and seven rating studies, several potential usage factors are investigated, including grammatical person and verb predictability. The results provide initial evidence suggesting that topic drop usage can be explained by general processing principles: The prefield constituent is omitted when it is redundant, and realized overtly when it facilitates the processing of the following verb. This information-theoretic explanation is based on independently evidenced processing mechanisms, bundles isolated claims from the theoretical literature, and allows for a unified analysis of topic drop with other types of ellipsis and reduction.