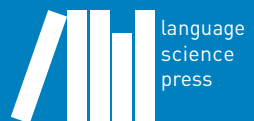


Experimental investigations on the syntax and usage of fragments

Robin Lemke

Open Germanic Linguistics 1



Open Germanic Linguistics

Editors: Michael T. Putnam, B. Richard Page, Laura Catharine Smith

In this series:

1. Lemke, Robin. Experimental investigations on the syntax and usage of fragments.

ISSN (print):

ISSN (electronic):

Experimental investigations on the syntax and usage of fragments

Robin Lemke

Robin Lemke. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/321>

© 2021, Robin Lemke

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: 978-3-96110-331-7 (Digital)

978-3-98554-027-3 (Hardcover)

ISSN (print):

ISSN (electronic):

DOI: 10.5281/zenodo.5596236

Source code available from www.github.com/langsci/321

Collaborative reading: paperhive.org/documents/remote?type=langsci&id=321

Cover and concept of design: Ulrike Harbort

Proofreading: Amir Ghorbanpour, Amy Amoakuh, Eduard S. Lukasiewicz,
Frances Vandervoort, Janina Rado, Jean Nitzke, Jeroen van de Weijer, Marten
Stelling, Sean Stalley, Sebastian Nordhoff

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität  Berlin

Contents

Acknowledgements	v
1 Introduction	1
1.1 What is the syntactic structure of fragments?	2
1.2 Why do speakers use fragments?	3
1.3 Chapter overview	4
1.4 Defining the notion <i>fragment</i>	5
2 Theories of fragments	7
2.1 Fragments as nonsententials	8
2.2 Fragments as elliptical sentences	12
2.2.1 In situ deletion	13
2.2.2 Movement and deletion	16
2.2.3 Discourse-initial fragments under sentential accounts .	21
2.3 Fragments as ungrammatical utterances	26
2.4 Testable predictions of theories of fragments	27
2.4.1 (Anti)connectivity effects: Case marking	27
2.4.2 Constituency	33
2.4.3 Information structure and focus	37
2.4.4 Evidence for movement	41
2.4.5 Summary	46
3 Experiments on the syntax of fragments	49
3.1 Case marking as evidence for sententiality	50
3.1.1 Experiment 1: Default case, acceptability rating study . .	52
3.1.2 Experiment 2: Default case, production study	61
3.1.3 Experiment 3: Mixed accounts?	63
3.1.4 General discussion: Structural case marking	69
3.2 Movement restrictions: Preposition omission	71
3.2.1 Preposition omission as evidence for movement	72
3.2.2 Experiment 4: Preposition omission in German	82
3.2.3 Experiment 5: Preposition omission in English	86

Contents

3.2.4	Experiment 6: Preposition omission and case	91
3.2.5	Experiment 7: Question-answer parallelism	99
3.2.6	General discussion: Preposition omission	111
3.3	Movement restrictions: Complementizer omission	114
3.3.1	Complementizer omission as evidence for movement	114
3.3.2	Experiment 8: CC topicalization in German	118
3.3.3	Experiment 9: CC topicalization in English	129
3.3.4	General discussion: Complementizer omission	133
3.4	Movement restrictions: The German prefield	135
3.4.1	The German prefield and movement in fragments	135
3.4.2	Experiment 10: Multiple prefield constituents	138
3.5	The syntax of fragments: Discussion	147
3.5.1	Fragments are sentential	147
3.5.2	Fragments are not obligatorily moved	148
3.5.3	Conclusion and outlook	150
4	An information-theoretic account of fragment usage	153
4.1	Information theory	154
4.2	Information-theoretic constraints on language	158
4.2.1	Source coding	159
4.2.2	Channel coding	161
4.2.3	UID as efficient distribution of processing effort	169
4.2.4	UID vs. other accounts of predictability-driven reduction	172
5	Evidence for UID effects on omissions in fragments	175
5.1	Scripts as a model of extralinguistic context	176
5.1.1	Script knowledge	178
5.1.2	Estimating event surprisal from script corpora	182
5.2	Experiment 11: Script knowledge, rating	189
5.2.1	Background	189
5.2.2	Materials	189
5.2.3	Procedure	192
5.2.4	Results	193
5.2.5	Discussion	194
5.3	Experiment 12: Script knowledge, production	199
5.3.1	Background	199
5.3.2	Materials	200
5.3.3	Procedure	200
5.3.4	Preprocessing	202

5.3.5	Surprisal estimation	207
5.3.6	Results	215
5.3.7	Discussion	224
5.4	The usage of fragments: Discussion	226
5.4.1	Evidence for UID effects on omissions in fragments . . .	226
5.4.2	UID vs. availability-based production	227
5.4.3	Information theory or information structure?	228
5.4.4	Script knowledge as models of extralinguistic context .	229
5.4.5	Surprisal estimation in elliptical data	230
5.5	Outline of a game-theoretic model of fragment usage	231
5.5.1	Limits of UID effects on the form of utterances	231
5.5.2	Game-theoretic pragmatics	233
5.5.3	Game-theoretic modeling of scalar implicatures	234
5.5.4	Application to fragments	237
5.5.5	Application to natural language data	239
5.5.6	Implications and comparison to UID	239
6	General discussion	243
6.1	Results on the syntax of fragments	244
6.1.1	Fragments are underlyingly sentential	244
6.1.2	Fragments are not obligatorily moved	246
6.1.3	Implications for (generative) syntactic theories	247
6.2	Results on the usage of fragments	248
6.2.1	Results on the form of fragments	248
6.2.2	The usage of fragments is constrained by UID	249
6.3	Implications for predictability effects on language processing .	251
6.3.1	Script-based event chains as a model of context	252
6.3.2	Surprisal estimation in elliptical data	253
6.3.3	UID constrains the omission of content words	253
6.3.4	Effects of extralinguistic context on predictability	254
6.3.5	Psycholinguistic implications of UID	254
6.4	Implications for other reduction phenomena	256
	Appendix: Models	257
	References	265
	Index	287
	Name index	287
	Language index	293
	Subject index	295

Acknowledgements

This book is a significantly revised version of my PhD thesis (Lemke 2020). Writing the thesis, conducting the research reported here and turning it into this book would not have been possible without the help, support and advice of a lot of people whom I'd like to thank in what follows.

First of all, there are my supervisors Ingo Reich, Heiner Drenhaus and Oliver Bott. Without Ingo this book would not exist, not only because he started project B3 in the Sonderforschungsbereich (SFB, Collaborative Research Center) 1102, in whose context the research reported here was conducted, but also because of his way of supervising my PhD thesis. I am grateful to him for always being there for discussing methodological and theoretical issues and for his valuable comments on earlier versions of the manuscript. I would like to thank Heiner, my second supervisor, for the extremely helpful discussions and suggestions on theoretical issues, statistical analysis and the presentation of my results as well as comments on previous versions of the manuscript. I am very grateful to Oliver, my external supervisor, for his comments both in the review and during my PhD defense.

I would also like to thank the other members of my doctoral committee, Julia Knopf, Elke Teich and Stefan Thater, for their time and their questions during my PhD defense. The research reported in this book has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).¹

When I began thinking about the possibility of working on a PhD thesis, I was certain that I wanted to do this in a collaborative environment. Fortunately, in Saarbrücken I found that, both within SFB 1102 and the Department of Modern German Linguistics. I would like to thank everybody with whom I discussed aspects of this work at the joint colloquium of the German and English studies, the SFB's PhD days and of course at the conferences where I could present my research thanks to the DFG's generous funding.

Some colleagues in Saarbrücken helped me out with more specific issues: I would like to thank Philipp Rauth for always sharing his expertise in generative syntax, Julia Stark for drawing the visual stimuli used in experiment 2 and allowing me to publish them in this book and Simon Ostermann for sharing the

¹Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 232722074 – SFB 1102.

Acknowledgements

language model file for pre-processing DeScript. I am particularly indebted to Lisa Schäfer, with whom I shared the office, conference travel and the corresponding leisure, hikes, food and beers during most of the time that I worked on my thesis. She also contributed to preprocessing the DeScript corpus and spent a lot of time proof-reading and commenting on previous versions of this work. I would also like to thank the (former) student assistants Luise Ehrmantraut, Fabian Ehrmantraut, Jonathan Watkins, Natascha Kraushaar and Matt Kuhn for helping with the annotation of production data, the construction of experimental materials and – in the case of Matt – proofreading large parts of the thesis.

At this point, I would like to acknowledge my professors and lecturers at Freie Universität Berlin and Humboldt Universität zu Berlin, in particular Guido Mensching, Anke Lüdeling, Manfred Krifka, Anton Benz and Hubert Truckenbrodt (in order of appearance), who raised my interest in linguistics, convinced me of the importance of doing empirical and quantitative research, and/or allowed me to conduct my first experiments and corpus studies.

The people mentioned so far supported me in one way or another in writing my thesis or even thinking about doing so. For allowing me to publish it in a way that makes the results of my research publicly accessible, I thank the Open Germanic Linguistics series editors, Michael T. Putnam, Laura Catharine Smith and Richard Page and everybody at Language Science Press. In particular, I am grateful to Stefan Müller, Sebastian Nordhoff and Felix Kopecky for their support during the preparation of the final manuscript, the Language Science Press community proof-readers for their time and John T. Hale for providing a very detailed and helpful review on the manuscript.

Schließlich möchte ich meiner Familie und meinen Freund*innen für moralische und logistische Unterstützung in dieser Zeit danken, insbesondere Madaida, Jan, Heike und Julia für zeitaufwändiges Katzenhüten und -taxifahrten. Meinen Eltern Heike, Matthias, Andreas und Carmen und Großeltern Edith und Fritz danke ich dafür, mich während meinem Studium und davor immer unterstützt zu haben.

1 Introduction

The concept of *sentence* occupies a central position in linguistic theory. In (generative) syntax, well-formed expressions are dominated by a node that is related to sententiality, which was originally labeled as S(entence) (see e.g. Chomsky 1965) and which has been more recently redefined as the complementizer phrase (CP) or a CP layer (Rizzi 1997). This layer is taken to host different speech act-related features, such as sentence mood in Rizzi's (1997) ForceP and assertivity (Krifka 1995). From a semantic perspective, it is generally assumed that only sentences can be used to perform speech acts and to communicate propositions.

This theoretically motivated requirement for well-formed utterances to be sentential clashes with linguistic reality. Morgan (1973) observed more than forty years ago that speakers produce nonsentential utterances which still fulfill the same communicative function as their sentential counterparts. For instance, the utterance in (1a) lacks an inflected verb and a subject. At least in this specific context however, (1a) is pragmatically interpreted as being meaning-equivalent to the full sentence in (1b). Morgan (1973) proposed the term *fragment* for these nonsentential utterances, which I adopt in this book.¹

- (1) [Ann and Bill are sharing a pizza. Bill asks Ann:]
- a. Another slice?
 - b. Would you like another slice of pizza?

This book addresses two main research questions which are investigated with experimental methods: First, what is the syntactic structure of these expressions? And, second, why do speakers² use fragments at all?

¹The term *fragment* is roughly meaning-equivalent to that of *nonsentential utterance* used elsewhere (Fernández & Ginzburg 2002, Barton & Progovac 2005, Stainton 2006). None of these notions is theory-neutral, since *fragment* suggests that the utterance is incomplete, whereas the notion of *nonsentential utterance* implies that it is not underlyingly sentential. Besides being more in line with my empirical findings on sententiality I use the notion *fragment* because it has been proposed earlier in the literature for this phenomenon.

²This work focuses on spoken and written language. The term *speaker* refers to the person who produces an utterance and the term *hearer* to the person who processes it. The overarching ideas and results on the question of why people produce a reduced or a syntactically complete utterance might be applied to sign language, what might be investigated in future research.

1.1 What is the syntactic structure of fragments?

The apparent violation of standard assumptions about phrase structure in fragments challenges syntactic theory: For instance, a grammar that requires all well-formed structures to contain an inflected matrix verb is not able to derive a bare DP fragment like (1a). This mismatch between the nonsentential form and the sentential function of fragments has been relatively extensively investigated in theoretical linguistics (see e.g. Morgan 1973, Ginzburg & Sag 2000, Fernández & Ginzburg 2002, Schlangen 2003, Merchant 2004a, Barton & Progovac 2005, Culicover & Jackendoff 2005, Stainton 2006, Reich 2007, Weir 2014a, Ott & Struckmeier 2016), but there is no consensus on the theoretic analysis of fragments. In particular, it is unclear whether fragments result from ellipsis in full sentences, which are derived by the standard syntax rules, or whether fragments require modifications to syntactic theory that allow for the derivation of subsentential output. Furthermore, for now the competing theories rely almost exclusively on partially conflicting introspective data. The first part of the book (Chapters 2 and 3) presents a series of acceptability rating and production studies that investigate the predictions of the competing theories. These experiments provide the first empirical investigation of a set of diverging predictions of the competing theories of fragments.

I focus on three generative accounts of fragments: the *nonsentential* account (e.g. Barton & Progovac 2005, Progovac 2006), the information structure-based *in situ deletion* account (Reich 2007, Ott & Struckmeier 2016) and the *movement and deletion* account (Merchant 2004a, Weir 2014a). These theories make relatively general testable predictions on fragments, such as the requirement that all fragments must be able to appear in the left periphery according to Merchant (2004a). I do not investigate HPSG accounts of fragments (Ginzburg & Sag 2000, Fernández & Ginzburg 2002, Schlangen 2003), which assign different internal structures that are relatively independent from each other to different types of fragments, depending on the context in which the fragment occurs. An empirical study would have to test all of these structures individually in order to determine the appropriateness of such an account.

The accounts that I investigate differ in particular with respect to two issues: First, whether fragments are underlyingly sentential, and second, whether their generation involves obligatory syntactic movement. The first question is a matter of debate between sentential and nonsentential accounts of fragments, whereas the second one is disputed between the different families of sentential accounts. A first series of experiments investigates whether fragments are underlyingly sentential or base-generated nonsentential utterances. These experiments use

structural case marking on DP fragments as a diagnostic for unarticulated syntactic structure. Since the experiments provide evidence for a sentential analysis of fragments, I then use potential parallelisms between fragment and movement restriction as a testing ground for obligatory movement in fragments. Taken together, the experiments support the in situ ellipsis account of fragments, which has been proposed by Reich (2007). This complements the theoretical debate on the syntax of fragments with empirically validated data and settles the ground for the investigation of the usage of fragments in the second part of this book.

1.2 Why do speakers use fragments?

Generative theories of fragments determine which form fragments can take, but they do not explain why speakers use fragments at all, and under which circumstances they prefer a fragment over a complete sentence. Corpus data show that fragments are relatively frequent, and, the frequent usage of fragments³ suggests that speakers have a reason to prefer them over full sentences in particular situations. However, except for a game-theoretic approach with very restricted scope by Bergen & Goodman (2015), the question of what determines this preference is totally unexplored.

The second part of this book (Chapters 4 and 5) is dedicated to the investigation of why and when speakers use fragments. An answer to this question requires establishing (i) why the usage of a fragments or a sentence is sometimes (dis)advantageous, and (ii) why *specific* words are preferably omitted in fragments. For instance, in the case of the pizza example in (1), the speaker might have said *Another slice of pizza* instead, so the choice between competing fragments must be modeled too. At this point, the investigation of the usage of fragments draws on the findings on the syntax of fragments in the first part of the book, since the set of *possible* fragments is necessarily restricted to those which can be derived by syntax.

The account that I propose assumes that the information-theoretic processing principle of Uniform Information Density (UID, Levy & Jaeger 2007) plays a crucial role in the choice of an utterance by the speaker. Two experiments confirm the central predictions of this account: Speakers choose the utterance that makes the most efficient usage of the hearer's processing resources, and they consequently omit words that underutilize these resources but realize words that prevent them from being exceeded. In addition to providing evidence for the

³For instance, Fernández & Ginzburg (2002) find in a corpus study that 11.15% of the utterances in a subcorpus of the British National Corpus (Burnard 2000) are fragments.

information-theoretic account of the usage of fragments, these findings have implications both for the research on ellipsis and for the investigation of the choice between alternative utterances in general. The choice between a reduced (elliptical) form and a complete one is also highly relevant to other ellipsis phenomena like sluicing, gapping and verb phrase ellipsis, and it might be instructive to test whether the conclusions on the usage of fragments apply to these ellipses as well. From a broader psycholinguistic perspective, my results contribute to the growing bulk of evidence for effects of information-theoretic processing constraints, and specifically for UID, on the preferred form of utterances. This supports several implications of UID, such as the close link between predictability and processing effort, the assumption of audience design and the parallel and incremental nature of the human parser.

1.3 Chapter overview

- Chapter 2 outlines three main generative theories of fragments and identifies potential testing grounds for them, that is, phenomena with respect to which their predictions differ.
- Chapter 3 presents a series of experiments that test the predictions of the theories presented in Section 2. The experiments suggest that fragments are derived by ellipsis from regular sentences and that their derivation does not involve obligatory movement to the left periphery.
- Chapter 4 briefly reviews information-theoretic approaches to the omission of linguistic expressions and presents my information-theoretic account of fragment usage.
- Chapter 5 presents two experiments that test and support the predictions of the information-theoretic account: Predictable words are more likely to omitted and words that increase the predictability of following ones are more often realized. The experiments rely on script-based event chains as an approximation to extralinguistic context and a method to estimate word probabilities in elliptical data.
- Chapter 6 summarizes and discusses the main results as well as the empirical and theoretical contributions of this book.

1.4 Defining the notion *fragment*

In the literature on fragments, there is no mutually shared definition of the phenomenon and there is disagreement specifically with respect to which utterances are classified as fragments. This section delimits how the notion *fragment* is used in this book and distinguishes it from other instances of reduced utterances. In order to distinguish fragments from other omission phenomena and instances of apparently incomplete speech, I rely on three criteria: (i) the performance of a speech act (Morgan 1973), (ii) the absence of a finite verb, and (iii) the absence of a linguistic antecedent within the same utterance. First, a fragment must be used to perform a speech act. This excludes labels (Klein 1993) like (2).

- (2) Skim milk / 16 oz. / sugar free ...

Second, the distinction between fragments and sentences is based on the subsentential character of fragments. What counts as “subsentential” depends on the syntactic analysis of the expression in question. For instance, in sentences with null pronouns in argument positions like (3), where a subject has been omitted, the remainder of the sentence is preserved, whereas in the DP fragment in (1a) there is no immediate evidence for any structure above the DP level.

- (3) Will be back soon.

Therefore, the most uncontroversial examples of fragments are XPs that are not of the same category as full sentences, which is most easily evidenced by the absence of an inflected verb. For instance, if English sentences are TPs, the bare DP *another slice* in (1a) must be categorized as a fragment based on this criterion. The same holds for any category below TP, like VP, PP or NP. Consequently, utterances like (3) are not categorized as fragments even though they lack an otherwise obligatory argument, because the auxiliary still evidences that the utterance is a TP. Note that this does not imply that fragments do not ever *contain* TPs but only that they *are* not TPs themselves. The complement clause in (4a) hence counts as a fragment, because in a full sentence it needs to be embedded under a matrix verb (4b) that is missing here.

- (4) a. That he'll be back soon.
b. John said that he'll be back soon.

Third, unlike antecedent-based ellipses (Reich 2011), such as gapping, sluicing, sprouting, and verb phrase ellipsis, fragments do not require an explicit linguistic antecedent. There is some disagreement in the literature about whether this

1 Introduction

condition excludes short answer fragments, too. For instance, Klein (1993) distinguishes discourse-initial fragments from what he calls *adjacency pairs*. According to Klein (1993: 768), ellipses in adjacency pairs “require an explicit linguistic context, [...] on which the elliptical utterance depends [translation from German, R.L.]” This definition of adjacency pairs clearly includes short answers as (5). The distinction between short answers and discourse-initial fragments is explicitly made in Reich (2011). In contrast, all of the researchers whose theories I discuss in Chapter 2 rely on data from short answers in support of their theories of fragments. This suggests that they adopt, at least implicitly, a uniform analysis of short answers and discourse-initial fragments.

- (5) What did John eat?
Pizza.

Even though the status of short answers is theoretically controversial, some of the experiments presented in this book investigate short answers, particularly in the extensions of the experiments by Merchant et al. (2013), who also used short answers in their studies. In experiments 1–3 on default case marking as well as in experiments 11 and 12, which investigate the usage of fragments, I use discourse-initial fragments instead. As for the question of whether there is a categorical distinction between adjacency pairs and genuine fragments, from the probabilistic perspective that my information-theoretic account implies, it seems compelling to attribute potential differences between short answers and discourse-initial fragments to differences in predictability: Material that has been mentioned in an explicit preceding question will be much more predictable than when it must be inferred from extralinguistic context, and the use of fragments will be therefore more strongly preferred. However, testing this experimentally will be complicated due to the necessary correlation between predictability and the type of context. Therefore, I remain agnostic to the question on whether there is a categorical difference between short answers and discourse-initial fragments. Except for the studies that replicate or follow up on previous experiments involving short answers, I rely on discourse-initial fragments, which are the most uncontroversial instances of fragments.

2 Theories of fragments

Since Morgan (1973) introduced the notion of *fragment* and first described the phenomenon, there has been considerable debate and disagreement on the syntax of these expressions. In the first part of this book (Chapters 2 and 3), I therefore discuss and experimentally investigate aspects of the syntax of fragments on which competing theories disagree and which will allow us to test the validity of the theories' predictions. Besides contributing to our theoretical understanding of fragments, the experiments lay the ground for the experiments on their usage in the second part of this work.

This chapter summarizes some representative versions of the most influential generative theories of fragments. Among these, two families of syntactic accounts are to be distinguished. On the one hand, *nonsentential accounts* (Section 2.1) treat fragments as truly nonsentential expressions that lack any sort of unarticulated structure. This requires some modification of syntactic theory in order to allow for well-formed subsentential output (see e.g. Barton & Progovac 2005, Fortin 2007). On the other hand, *sentential accounts* (Section 2.2) claim that fragments are derived by ellipsis from linguistically complete sentences. There are two versions of sentential accounts: the *in situ deletion* account (Reich 2007), which derives fragments from regular sentences, and the *movement and deletion* account (Merchant 2004a), which states that the future fragment has to occupy a left-peripheral position in the full sentence before ellipsis applies. Finally, in Section 2.3 I discuss the claim by Bergen & Goodman (2015) that fragments are actually *ungrammatical*, but that speakers can still use them if they manage to get their message across. The experiments presented in this book do not explicitly address the predictions of theories of fragments in other syntactic frameworks, like HPSG (Ginzburg & Sag 2000, Fernández & Ginzburg 2002, Schlengen 2003). Since these accounts assume relatively complex structures for individual types of fragments, which model connectivity effects and other properties, it is difficult to falsify them empirically and compare their predictions to generative accounts that derive fragments by more abstract and general principles. Nonetheless, in the discussion of the results I address issues that are relevant to the empirical predictions of HPSG accounts.

This chapter is structured as follows: In Sections 2.1–2.3, I present the central ideas of these theories and avoid controversial or conflicting evidence as much as possible. Section 2.4 discusses a series of phenomena that have been argued by the respective authors to constitute evidence for or against specific accounts. As will become clear, most of the theories explain most of the data, but there are some aspects on which they disagree and which will serve as a testing ground for the competing theories in experiments 1–10.

With the exception of Bergen & Goodman (2015), all of the accounts presented here have been developed by authors working in a Chomskyan generative framework. Therefore, they focus on explaining why we observe specific restrictions on the form of fragments, but neglect their processing and psychological reality. This might raise the question of whether modeling the syntactic derivation is relevant at all to the processing and interpretation of fragments by the hearer. For instance, from the hearer's perspective, it might seem irrelevant whether the speaker had a linguistic structure in mind, which is only partially articulated, or there was nothing but a fragment to begin with: The fragment she¹ has to interpret is identical in both cases. However, there are at least two good reasons to take the derivations proposed by the different theories of fragments seriously. First, if fragments are generated by grammatical mechanisms, knowledge about these will guide the hearer in retrieving the intended message. Second, if such grammatical mechanisms restrict the form of possible fragments, they will restrict the set of alternative encodings of a proposition to those fragments which are a well-formed output of grammar.

2.1 Fragments as nonsententials

According to nonsentential accounts, fragments do not contain any sort of unarticulated structure. As Stainton (2006) points out, the assumption that fragments are *genuinely* nonsentential presupposes that there is neither silent material in fragments nor that any parts of the utterance are deleted in course of the derivation. This requires some modification to standard syntax in order to allow for subsentential expressions to be a well-formed output of syntax.

Barton & Progovac (2005) sketch a theory of fragments that is based upon this idea, which is grounded in the minimalist Program (Chomsky 1995). They

¹Throughout this book I use arbitrary gender pronouns in order to refer to abstract hearers and speakers. Sometimes the speaker will be female and the hearer male and vice versa, but I use the same pronoun for the same imaginary person in a situation.

propose two adjustments of the theory in order to allow for syntactically well-formed subsentential objects. First, any maximal projection XP can be a well-formed output of grammar, and second, case checking requirements are relaxed in fragments. Besides bare XP fragments, their theory is designed to explain other omissions found in the ETP corpus (Libben & Tesak 1994), a corpus of elicited ‘telegraphese’ data. As discussed by Barton (1998), this register is characterized by frequent omissions of functional elements like articles, first person subject pronouns and auxiliary verbs.

The first modification to standard syntax that Barton & Progovac (2005) propose is that the derivation can stop at any maximal projection, as long as it is internally well-formed and there are no lexical items left in the numeration. The fragment in (1) is consequently analyzed as a bare VP that does not contain a TP. Barton & Progovac (2005) argue that there is no evidence for a T head in the derivation, because the verb *play* is not inflected for person or tense.

- (1) What does John do all summer? (Barton & Progovac 2005: 81)
Play baseball.

The second modification that Barton & Progovac propose is the Case Feature Corollary (CFC) (2). In minimalism, case-marked DPs are assumed to have uninterpretable case features, which must be checked in a specific syntactic configuration by a head carrying the same feature (3). The CFC loosens this requirement for fragments (2), but not for full sentences.

- (2) Case Feature Corollary (CFC) (Barton & Progovac 2005: 78)
Nonsententials differ from sententials in one property: they are not required to check Case features.

- (3)
-
- ```

graph TD
 VP --> Spec
 VP --> V_prime[V']
 V_prime --> V_Acc[VAcc]
 V_prime --> DP_uAcc[DPuAcc]

```

Barton & Progovac (2005) motivate the CFC with the observation of differing case-marking preferences between fragments and sentences. For instance, English pronominal short answers (4a,b) seem to be more acceptable in accusative case than in nominative. In full sentences (4c,d) the pattern is inverted, even though the pronoun has the same grammatical function in both cases.

## 2 Theories of fragments

- (4) Who can eat another piece of cake? (Barton & Progovac 2005: 77)
- a. ?I/?\*We/?\*He/?\*She
  - b. Me/Us/Him/Her
  - c. I/We/He/She can.
  - d. \*Me/\*Us/\*Him/\*Her can.

Barton & Progovac argue that this follows from the default case status of the English accusative, while nominative is considered structural case. In this line of reasoning, structural case is assigned in specific syntactic configurations and does not contribute to semantics, unlike inherent case, which encodes specific a  $\vartheta$ -role. In the case of English, nominative is checked by a T head. Barton & Progovac (2005) interpret the data in (4) as evidence nominative in fragments is ungrammatical because of the absence of a covert T head, which would check nominative case features. In contrast, accusative is acceptable in fragments according to Barton & Progovac (2005: 78), because it is the default case in English, i.e. the most unmarked form. They argue that the use of accusative in predicative DPs (5) evidences this, because nominative is assigned only to the specifier of TP.

- (5) a. This is me/him/us. (Barton & Progovac 2005: 79)  
 b. ?This is I/he/we.

As the predictions of Barton & Progovac's account crucially rely on the concept of default case, this notion requires some further attention. First of all, it is controversial whether default case exists at all. In minimalism, case is modeled by the assumption of specific features and it has been assumed since Chomsky's (1981) *case filter* that derivations converge only if all DPs are case-marked. However, even Merchant (2004b), who argues against default case, assumes that resumptive pronouns like *who* in his example (6) are base-generated in a left-peripheral position and that they cannot undergo the regular case checking mechanisms.

- (6) Who<sub>i</sub> do you think that if the voters elect him<sub>i</sub>, the country will go to ruin

Schütze (2001) argues that default case *can* be integrated into a minimalist framework if it is defined as a residual category of case-marking that is assigned only to those DPs which are not marked with a more specific case.<sup>2</sup> Default case simply

<sup>2</sup>Schütze adopts concepts from Distributed Morphology (Halle & Marantz 1993), in particular, the idea of *late insertion* of lexical items into the derivation in a postsyntactic spell-out module. Schütze proposes that only arguments receive uninterpretable case features before entering the numeration, which are different from the optional morphological case features that determine which case marking a DP receives when it is selected during late insertion (Schütze 2001: 230–231). Non-arguments do not require syntactic case marking at all.

appears whenever no other case marking is available. According to Schütze, this becomes evident when the DP appears in a position where no syntactic relation to expressions that check case can be established, such as hanging topics (7a) or predicative constructions (7b).

- (7) a. Me/\*I, I like beans. (Schütze 2001: 210)  
b. The real me/\*I is finally emerging. (Schütze 2001: 215)

Since default case has the status of a residual category in Schütze's theory, he predicts considerable crosslinguistic variation both with respect to the contexts where it occurs and to the form that default case takes. First, languages can differ in whether a case checking relation is established in a specific syntactic position, so that the distribution of default case can differ crosslinguistically. With respect to the form, crosslinguistic equivalents of (7) suggest that in some languages, such as English, Irish and Norwegian, accusative is default case, whereas it is nominative in German, Russian or Dutch (Schütze 2001: 229). Progovac (2006: 51) argues that it is also nominative in Serbian (8).

- (8) Ona/\*Nju predsednik kluba?! (Vi se šalite.) (Progovac 2006: 51)  
she.NOM/her president club.GEN you REFL kid  
'Her president of the club?! (You must be kidding!)

Therefore, the nonsentential account makes different predictions on the case marking of DP fragments in examples such as (4) depending on which case is the default case in a language. The German version of the question-answer pair (4), which is given in (9), is in line with this prediction.<sup>3</sup> Note that this example does not contradict sentential accounts, which interpret (9) as evidence *for* an unarticulated T head checking nominative in fragments. I return to this issue in Section 2.4.1.

- (9) Wer kann noch ein Stück Kuchen essen?  
who can more one piece cake eat  
'Who can eat another piece of cake?'  
a. Ich/Wir/Er Nominative  
b. \*Mich/\*Uns/\*Ihn Accusative

<sup>3</sup>It shall be noted that Schütze (2001: 221) also notes that subject DP fragments like (4) receive accusative case marking in English, but nominative in German. However, unlike Barton & Progovac (2005) and Progovac (2006), Schütze does not simply explain this by the different default case in both languages but argues that DP fragments are only a "possible default-case environment" (Schütze 2001: 229): He argues that it is an "actual" one in English, but not in German, which uses the "strategy" of always matching case in question and answer.

The discussion on case checking in fragments in Barton & Progovac (2005) focuses on the distinction between structural and default case, but does not explicitly discuss fragments appearing in inherent case, such as dative or genitive. Progovac et al. (2006: 338–341) argue that in Serbian non-nominative case is inherent case, because it is associated with a specific  $\vartheta$ -role. For instance, “dative objects are typically associated with the theta-role of goal/recipient” (Progovac et al. 2006: 339), so that dative has interpretable case features which do not need to be checked at all. Consequently, the nonsentential account predicts inherent case-marked fragments to be acceptable in an appropriate context and restricts anticonnectivity effects as in (4) to instances where DPs receive structural case marking in complete sentences.

As a syntactic theory, Barton & Progovac’s (2005) nonsentential account is primarily concerned with deriving the fragments that are grammatical in a language. The theory does not explain how fragments are licensed or how they are interpreted. As for licensing, Barton & Progovac (2005: 89) suggest that recoverable expressions can be omitted, be it from linguistic or extralinguistic context. With respect to their interpretation, nonsentential accounts assume that this requires pragmatic enrichment. Stainton (2006) sketches a mechanism for this, which assumes that a salient nonlinguistic conceptual object<sup>4</sup> is used to enrich the fragment to a complete proposition. The crucial difference between sentential and nonsentential theories of fragments is therefore whether the contextually salient objects licensing fragments are linguistic or only conceptual. Reflexes of linguistic structure, like structural case marking or movement restrictions, which are not contained in nonlinguistic representations, will be crucial for differentiating between both families of theories.

## 2.2 Fragments as elliptical sentences

Sentential accounts are motivated by the observation that fragments can be used for the same communicative purposes as full sentences despite their reduced form. For instance, the fragments in (10) appear to be a bare PP (10a) or DP (10b), but in both cases they are used to perform speech acts, just like their fully sentential counterparts in (11). If sentence mood is encoded in the left periphery (see e.g. Rizzi 1997), this possibility of performing speech acts with fragments seems surprising, as there is no direct evidence for a left periphery in these utterances.

---

<sup>4</sup>Stainton (2006: 186–189) terms it “logical form”, but explicitly delimits his use of this term from that referring to the semantic representation of an utterance. Stainton refers to some kind of conceptual nonlinguistic representation instead.

- (10) a. [Passenger to taxi driver:] To the university, please!
- b. [Friends sharing a pizza:] Another slice?
- (11) a. Take me to the university, please!
- b. Would you like another slice?

Since Morgan (1973), one explanation for this apparent mismatch between form and function has been that fragments do not really lack these projections, but that they are actually full sentences, parts of which are deleted by ellipsis.<sup>5</sup> This analysis has the advantage that, beyond mechanisms for licensing ellipsis (which are needed anyway in order to explain other instances of ellipsis), no amendments to syntactic theory are required in order to derive fragments. Their semantics can also be calculated compositionally as in regular sentences. Besides those theoretical advantages of sentential accounts, they are empirically supported by *connectivity effects*, morphosyntactically and semantically identical behavior of a constituent as a fragment and within a full sentence. Such effects concern, for example, case marking (see Section 2.4.1) and binding (Merchant 2004a).

Advocates of a sentential account do not agree on what exactly the underlying structure looks like, specifically on whether fragments involve an obligatory movement step, as suggested by Merchant (2004a) or not, as Reich (2007) argues. In what follows, I present the central ideas of each of these two approaches.

### 2.2.1 In situ deletion

The most straightforward version of a sentential account derives fragments from regular sentences using those ellipsis mechanisms that are needed anyway to account for other types of ellipsis, such as gapping or sluicing. Reich (2007) presents such an account.<sup>6</sup> In a nutshell, he argues that all those parts of the utterance which are not focused are elided, and that the distribution of focus is determined by the relevant Question under Discussion (QuD, Roberts 1996), which can be either implicit or explicit.

The restriction of ellipsis to non-focused expressions follows from *question-answer congruence*, the licensing condition that Reich (2007) imposes on ellipsis. Reich (2007) assumes a question-based discourse structure (following Roberts

---

<sup>5</sup>This “deletion” is assumed to occur only on the phonological form (PF) that determines the acoustic realization of the sentences, but not on the logical form (LF) that determines their meaning in the terminology of Chomsky (1981).

<sup>6</sup>Reich’s theory is specifically motivated by a set of similarities between short answers and gapping. I restrict the presentation of this account to fragments.

1996), so that the information structure of a sentence is determined by the immediately preceding QuD. The QuD can be either explicit or implicit. In Reich's examples of fragments it is explicit, because he discusses short answers and not discourse-initial fragments. However, his theory can account for discourse-initial fragments in the same way, since one of his main goals is a uniform analysis of fragments and gapping, where ellipsis is licensed by an implicit QuD.

Reich (2007) resorts to Rooth's (1992) theory of question-answer congruence in order to formally define the relationship between question and answer. Following Rooth (1992), Reich assumes that the meaning of a question is equivalent to the set of its potential answers, which can be obtained by replacing the *wh*-phrase by an existentially bound variable. For an answer to be well-formed, it must obey the two constraints in (12): First, C-ANSWER (12a) determines that the answer *A* must be included in the denotation of *Q* (Reich 2007: 472). Second, F-ANSWER determines that the answer's focus value, which, following Rooth (1992), is calculated by replacing focused expressions with existentially bound variables, must be a superset of the denotation of the question (Reich 2007: 472).

- (12) a. C-ANSWER:  $[[A]] \in [[Q]]$ .  
 b. F-ANSWER:  $[[Q]] \subseteq [[A]]_F$  (and  $|[[Q]] \cap [[A]]_F| \geq 2$ )

Reich shows how these constraints explain in interaction why (14a), but not (14b) or (14c) are information-structurally well-formed answers to (13).

- (13) a. Which student did John invite *t*? (Reich 2007: 472)  
 b.  $[[ (13a) ]]$  =  $\{p; \exists x[x \text{ a student} \ \& \ p = \text{that John invited } x]\}$   
 (14) a. John invited  $[Sue]_F$ . (Reich 2007: 472)  
 b.  $*[Sue]_F$  invited John.  
 c.  $\#$ John invited  $[Noam \ Chomsky]_F$ .

Reich (2007: 472) defines the focus values for (14a) and (14b) as (15). The focus value of (14a) in (15a) entails the denotation of the question (13b) and thus conforms to F-ANSWER. Since the answer is included in the denotation of the question (provided that Sue is a student), C-ANSWER is also respected. In the case of (14b), its focus value in (15b) does not entail (13b), therefore the answer is not congruent. The focus value of (14c) does entail (13b), but C-ANSWER is violated, because *Noam Chomsky* is not contained in the set of students so that (14c) is not included in the set of possible answers.

- (15) a.  $[[ (14a) ]]$ <sub>F</sub> =  $\{p; \exists x[x \in D_e \ \& \ p = \text{that John invited } x]\}$   
 b.  $[[ (14b) ]]$ <sub>F</sub> =  $\{p; \exists x[x \in D_e \ \& \ p = \text{that } x \text{ invited John}]\}$



Syntactically, Reich (2007: 472) links the question to the answer by assuming a squiggle operator  $\sim$ , which adjoins to the highest node of the syntax tree of the answer, the CP. The operator introduces a variable  $\Gamma$ , which is coindexed with the question (16). The operator presupposes that the answer is congruent to the question with respect to the two constraints in (12a) discussed above. This notion of question-answer congruence is the licensing condition for ellipsis.<sup>7</sup>

- (16) a. [Which professor did John invite  $t$  ?]<sub>1</sub> (Reich 2007: 472)  
 b. [John invited [Noam Chomsky]<sub>F</sub>]  $\sim \Gamma_1$ .

Reich (2007) defines ellipsis as PF-deletion, which can target only non-focused constituents, because the F-mark on focused ones requires them to receive a pitch accent (Selkirk 1984).<sup>8</sup> Defining ellipsis as a post-spellout phenomenon, which applies to PF only, explains why it has no effects on LF. Technically, Reich proposes that PF-deletion proceeds top-down starting at the sister node of  $\Gamma$  (CP, the root node of the answer) according to the rules in (17).

- (17) PF-deletion (Reich 2007: 473)  
 a. F-markers are upper bounds to PF-deletion.  
 b. Maximize PF-deletion. (short answers and gapping)

Taking the sentential answer (16b) as a starting point, the application of these PF-deletion rules yields the fragment in (18a) as the only acceptable outcome of the operation. Preserving larger parts of the structure, e.g. (18b), is ruled out by the need to maximize PF-deletion spelled out in (17b). Reich suggests that this second clause of the rule is specific to short answers and gapping, whereas (17a) applies to all types of ellipses.

<sup>7</sup>See Reich (2007: 474–477) for a comparison to Merchant's (2001) notion of e-givenness.

<sup>8</sup>Ott & Struckmeier (2016) sketch a very similar account but argue that it is the background of the utterance that can be deleted rather than the focus that cannot. They argue that this accounts better for the ability of German modal particles (MPs) to survive ellipsis (i), because MPs do not encode propositional meaning but the attitude of the speaker. According to Ott & Struckmeier, MPs neither belong to the focus nor to the background, so that the PF-deletion rules in Reich (2007) predict them to be omitted, while their own account does not.

- (i) Who did Peter invite? (Ott & Struckmeier 2016: 227–228)  
 a. Er hat wohl seine *Freunde* eingeladen.  
    he.NOM has PRT his friends invited  
    'Presumably he has invited his friends.'  
 b. Wohl seine *Freunde*.  
    PRT his friends  
    'Presumably his friends.'

## 2 Theories of fragments

- (18) a. [Noam Chomsky]<sub>F</sub>  
b. \*Invited [Noam Chomsky]<sub>F</sub>

The theory by Reich (2007) makes a series of testable predictions on the form of fragments. First, just like other sentential accounts, it predicts fragments to exhibit connectivity effects due to the unarticulated syntactic structure which they contain. Second, linguistic context, and specifically the relevant QuD, should have a strong effect on the form of fragments, because ellipsis is licensed only if the answer is congruent to the question. This is specifically expected when the QuD is explicit, as it is in question-answer sequences or other adjacency pairs. Implicit QuDs must be inferred by the hearer, who will try to accommodate a QuD that is congruent to the fragment. If the speaker is cooperative, such a QuD will always be accessible, because otherwise the speaker would prefer to utter a full sentence. Third, the form of fragments will also be constrained by focus projection rules, because only F-marked constituents survive ellipsis and the background is PF-deleted. Language-specific differences with respect to these rules will be reflected in different possible forms of fragments. Finally, Reich (2007) allows for discontinuous non-constituent fragments. This contrasts with most of the other accounts of fragments discussed in this section, which require fragments to be a single constituent. If multiple independent constituents are F-marked in a specific context, e.g. in case of multiple *wh*-questions (19), all of them must survive ellipsis.

- (19) [Waiter serving a couple their food:] Who ordered what?  
Customer: She ~~ordered~~ the pizza.

### 2.2.2 Movement and deletion

While Reich (2007) develops a unified account of fragments and gapping, Merchant (2004a) observes a set of similarities between fragments and sluicing. This motivates the extension of his theory of sluicing (Merchant 2001), which derives sluices by regular *wh*-movement followed by ellipsis of the remnant, to fragments. The central claim of the account is that all fragments undergo movement to a left-peripheral position before ellipsis applies to the remnant.

According to Merchant (2004a), ellipsis is triggered by a specific syntactic item, the E feature. Merchant argues that there are different varieties of E, each of which is related to a specific type of ellipsis, such as sluicing (Merchant 2001), fragments (Merchant 2004a) and VP ellipsis (Merchant 2013). Each variety of E has its own lexicon entry, which encodes its syntactic, phonologic and semantic

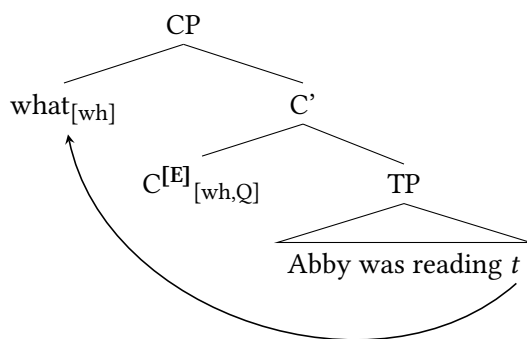


Figure 2.1: Derivation of the sluice in (20) according to Merchant (2004a: 670).

properties. To illustrate the idea, the derivation that Merchant assumes for the sluice in (20) is given in Figure 2.1.<sup>9</sup>

- (20) Abby was reading something, but I don't know what ⟨Abby was reading *t*⟩  
(Merchant 2004a: 670)

E is always located on the head of a functional projection, like CP in Figure 2.1. The syntactic properties of E, which consist of a set of uninterpretable features, determine which head can host the feature. For instance,  $E_S$ , the E feature found in sluicing, has the features  $[uwh^*, uQ^*]$  (Merchant 2004a: 670). This ensures that it can be hosted only by heads that are  $[wh, Q]$  and that therefore can check these features, such as C in interrogatives. The variants of E found in other types of ellipsis may have different feature specifications and are thereby restricted to other functional heads. Merchant (2004a: 671) suggests that the varieties of E are identical with respect to their phonology and semantics and differ only in these syntactic specifications. The phonological effect of the E feature is that the complement of the head it is located on remains unarticulated at PF. In (20), this concerns the complete TP of the second conjunct in (20). Both sentential accounts discussed so far, Merchant (2004a) and Reich (2007), agree that no syntactic structure is deleted during the derivation. Even though parts of it are unarticulated at PF, the unarticulated words are still present on LF. In (20), this results in the

<sup>9</sup>Merchant (2004a: 671) notes that the assumption of independent lexical entries for the specific varieties of E also accounts for crosslinguistic variation. For instance, he argues that German has no VP ellipsis because this language lacks the corresponding variety of E, while it shares with English the varieties found in fragments and sluicing.

wh-phrase being the only articulated word in the sluice, because it leaves the ellipsis site through *wh*-movement to [Spec, CP].

According to Merchant (2004a), the licensing condition on omissions in fragments is *e-givenness*, which is included in the semantics of the E feature (21): E requires the complement of the head hosting E to be e-given. E-givenness is the identity condition licensing ellipsis in Merchant's theory and consists basically in a bidirectional givenness relation in the sense of Schwarzschild (1999). An expression E counts as e-given when it has a salient antecedent A which entails the existential closure of the focus value of A and vice versa.

- (21)  $[[E]] = \lambda p: \text{e-given } (p) [p]$  (Merchant 2004a: 672)

The requirement for the complement of the head hosting E to be e-given ensures that ellipsis is licensed only if there is a structurally parallel antecedent available in context, and that it is blocked if there remains a constituent within the complement that is not e-given. (22) exemplifies the mechanism for the sluicing example in (20): The antecedent has the focus structure in (22a), whose existential closure (22b) is entailed by the sluice (22c). As the existential closure of (22c) is identical to the one of the antecedent in (22b), the opposite relation also holds, so that the ellipsis in (20) is licensed by e-givenness.

- (22) a. Abby was reading [something]<sub>F</sub>  
 b.  $\exists x$ . Abby was reading x  
 c. Abby was reading [what]<sub>F</sub>

Merchant (2004a) extends this analysis to fragments. His theory accounts for discourse-initial fragments (see below for details), but he focuses mostly on short answer fragments like (23), for which he assumes the structure in Figure 2.2. Again, the E feature is hosted by C in the left periphery, while the fragment is moved to the specifier of a functional projection FP immediately above CP. This movement operation proceeds cyclically through [Spec, CP].

- (23) a. Who did she see? (Merchant 2004a: 673)  
 b. John.

The major difference between sluicing and fragments is that E<sub>F</sub>, the variety of E found in fragments, and E<sub>S</sub> have different syntactic features, which are [*uC\**, *uF*] for E<sub>F</sub> and [*uwh\**, *uQ\**] for E<sub>S</sub>. The strong *uC\** feature ensures that E is located on a C head, while the weak *uF* feature can be checked under Agree (Merchant 2004a: 707), because weak features don't need to be checked locally according to

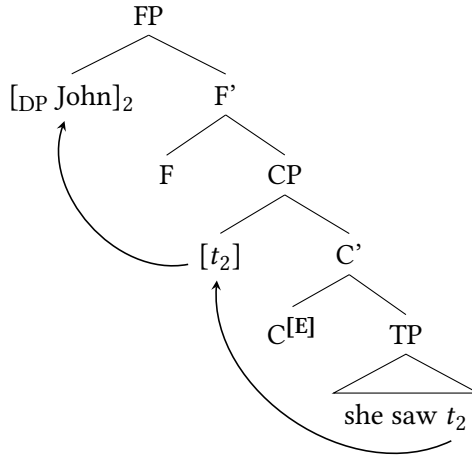


Figure 2.2: Derivation of the fragment answer in (23b) according to Merchant (2004a).

the theory. Otherwise, the derivation is identical to sluicing: After the fragment has been moved, ellipsis applies to the TP.

With respect to the landing site of the fragment in [Spec, FP], Merchant avoids committing himself to an analysis of what kind of projection FP is. However, Merchant (2004a: 675) tentatively suggests that it is a focus projection in the sense of Rizzi (1997).<sup>10 11</sup> Whether or not FP is a focus projection is highly relevant to the theory, because this would provide an explanation for why movement in fragments would occur at all. Since Merchant's theory is embedded in a minimalist framework (Chomsky 1995), movement cannot be optional, but is a *last resort* operation that is mostly driven by the need to check strong features in a local (specifier-head) configuration. In Merchant's (2001) account of sluicing, the *wh*-phrase reaches [Spec, CP] through *wh*-movement, which is driven by uninterpretable features of the *wh*-phrase. Similarly, movement in fragments requires a trigger which the E feature cannot provide: Its syntax, as defined above, contains

<sup>10</sup>Elsewhere (Merchant 2004a: 703) he relates the movement operation that results in fragments to Clitic Left Dislocation (CLLD, Cinque 1990) rather than to focus. See Section 2.4.3 for a discussion.

<sup>11</sup>The idea that FP is a focus projection is further developed by Gengel (2007), who argues explicitly that movement in fragments occurs to check a [+contrastive] feature in [Spec, FP]. This conclusion might be too strong, since in languages like German or English fronting foci is possible yet marked. Specifically, as Weir (2014a) notes and I discuss in greater detail below, object DP fragments are acceptable in situations where fronting objects is definitely not.

only uninterpretable features that determine on which head it can appear. If FP was related to an information-structural concept such as focus or topic, an uninterpretable feature related to this notion could trigger movement in fragments independently from E, just like Merchant (2001) argues for sluicing.

From an empirical perspective, Merchant (2004a) requires evidence that fragments have actually moved. Since he analyzes movement in fragments as regular A'-movement, his theory predicts that the derivation of fragments is subject to movement restrictions that are observed in full sentences: Only those constituents that can be moved to [Spec, FP] and appear in a left-peripheral position in full sentences are predicted to be possible fragments. Merchant (2004a) presents introspective data from different phenomena and languages in support of this prediction, some of which will provide the testing ground for his theory in my experiments.<sup>12</sup>

However, Weir (2014a) shows that the assumption that structures presumably underlying movement and deletion are acceptable across the board is falsified even by simple examples such as (24). The short answer fragment in (24a) is fine despite the ungrammaticality of the presumably underlying fronting structure (24b). The acceptability of left dislocation in a sentence seems not to be necessarily related to the acceptability of the corresponding fragment.

- (24) What did you eat? (Weir 2014a: 168)
- a. Chips.
  - b. \*Chips, I ate *t*.

In order to account for such data while maintaining the idea of movement and deletion, Weir (2014a) claims that movement in fragments is a special type of movement which is restricted to elliptical utterances and which differs from movement in narrow syntax, i.e. before spell out. According to Weir, this *exceptional* movement is triggered by a clash between the prosodic properties of focused expressions, which are marked with a pitch accent, and the ellipsis site, which the E feature requires to be silent. As Weir (2014a) assumes a similar underlying structure as Merchant (2004a) does (see Figure 2.2), that is, a regular sentence whose C head hosts the E feature, the TP is marked for PF-deletion, but still contains the focused DP *John*. This conflict is solved by moving the focused expression(s) out of the ellipsis site and adjoining them to CP.

Exceptional movement differs from narrow syntactic movement. First, it is not driven by feature checking; in fact, Weir (2014a: 195) denies that there is a focus

---

<sup>12</sup>See Section 2.4.4 for details.

feature in English.<sup>13</sup> According to Weir, exceptional movement is nevertheless a last resort operation, because there is no other way of saving the derivation from crashing due to the clash between focus and ellipsis at PF. Second, exceptional movement has no effect on the semantics of the utterance. This is in line with the observation that, unlike Gengel (2007) suggests, fragments are not necessarily contrastive. Weir (2014a: 183) attributes the absence of semantic effects of exceptional movement to its application after spell-out and at PF only. He argues that this also explains why it is restricted to elliptical utterances: The only purpose of exceptional movement is to evacuate focused constituents from the ellipsis site, and because focused constituents can remain in situ in full sentences, exceptional movement is ruled out by economy considerations.

As the discussion in Section 2.4.4 will show, the assumption of exceptional movement notably complicates the empirical evaluation of the movement and deletion account, because the strong correlation between the acceptability of fronting and fragments is no longer predicted. Therefore, the experiments presented below test Merchant's (2004a) version of the theory in the first place, but I also discuss the relevance for the exceptional movement theory whenever its predictions differ from Merchant (2004a).

### 2.2.3 Discourse-initial fragments under sentential accounts

Up to this point, the theoretical discussion has focused mostly on short answer fragments, although I argued in the introduction that the most uncontroversial instances of fragments are discourse-initial ones. Discourse-initial fragments challenge any sentential account of fragments: Given the licensing conditions of ellipsis discussed so far, ellipsis requires an antecedent, and in examples such as (25) no such antecedent seems to be available. Nonsentential accounts do not face this problem, as they derive the propositional meaning of fragments by pragmatic inference. Since some of the experiments presented in this work rely on discourse-initial fragments, in what follows I discuss how sentential accounts can account for these utterances and some of their properties. In particular, I argue

---

<sup>13</sup>Focus fronting is still acceptable in English if the focus is contrastive in the sense of Krifka (2007), i.e., when alternatives to the focused expression are given in context (i).

(i) *Him* I invited, not *her*.

This could still be accounted for by a more specific feature that appears only in contrastive contexts. However, even in cases as (i), focus fronting does not seem to be obligatory, therefore the English data require closer investigation, specifically if movement is to be assumed as non-optional (provided the relevant features are present).

that in those situations where discourse-initial fragments are used, an antecedent that licenses ellipsis can be retrieved from extralinguistic context. This facilitates a unified sentential account of fragments with and without overt antecedents.

- (25) a. [Passenger to taxi driver:] To the university, please!  
 b. [Customer to barista:] A coffee, please!  
 c. [Taking a postcard out of the post box:] From John!

Sentential accounts that assume a QuD-based model of context (Reich 2007, Weir 2014a) can explain the utterances in (25) by assuming implicit QuDs, which are evoked by the extralinguistic context and which are appropriate antecedents (26). For instance, a pedestrian approaching a taxi is likely to ask for a ride and a guest in the coffee shop is very likely to order a drink or some food.

- (26) a. ⟨Where shall I take you?⟩  
~~Take me~~ to the university, please!  
 b. ⟨What would you like to have?⟩  
~~I'd like to have~~ a coffee, please!

Reich (2011: 1852) notes that such discourse-initial fragments are *indeterminate*, that is, there are several possible paraphrases of the missing material. For instance, the answer in (25a) could be understood as *Take me to the university!*, *I'd like to go to the university*, or *Drive to the university!*. Reich takes this to be a defining feature of what he calls situation-based ellipsis (*s-ellipsis*), for the resolution of which the hearer must resort to extralinguistic context. In contrast, antecedent-based *a-ellipses*, which have a linguistic antecedent (e.g. gapping, right node raising and VPE), can be unambiguously resolved (27).

- (27) John goes to the university and Mary ~~goes~~ to the pub.

According to Reich (2011: 1852), indeterminacy suggests that, unlike short answers, discourse-initial fragments are syntactically genuine nonsententials. This implies a non-uniform analysis to fragments: If they have a linguistic antecedent, like an explicit QuD in the case of short answers, they are elliptical, and if they do not, they are nonsentential. However, in Reich (2007), he notes that also in antecedent-based ellipsis like gapping, the focus structure of the second conjunct can vary. If there is wide focus on the first conjunct, different focus structures (28) and henceforth different omission patterns (29) are possible in the second conjunct depending on which implicit QuD is assumed.

- (28) a. [John gave a book to *Sue*]<sub>F</sub>, and John gave [*a baseball*]<sub>F</sub> [to *Bill*]<sub>F</sub>.  
 b. [John gave a book to *Sue*]<sub>F</sub>, and [*Peter*]<sub>F</sub> gave a book [to *Ann*]<sub>F</sub>.  
 (Reich 2007: 478)



- (29) a. [John gave a book to Sue]<sub>F</sub>, and [a baseball]<sub>F</sub> [to Bill]<sub>F</sub>.  
 b. [John gave a book to Sue]<sub>F</sub>, and [Peter]<sub>F</sub> [to Ann]<sub>F</sub>.

(Reich 2007: 478)

Reich (2007: 477) argues that in such cases, “a complete set of possible QuDs [...] is reconstructed, from which the speaker chooses exactly one as the most salient.” The hearer then has to figure out which QuD out of this set is the one that the speaker had in mind. Besides extralinguistic context, a strong cue toward the QuD intended by the speaker is the form of the utterance: If only focused expressions survive gapping, (29a) will accommodate a QuD as *What did John give to whom?* and (29b) *Who gave a book to whom?* This reasoning applies equally to fragments: If the hearer must infer which QuD the speaker had in mind from context and the form of the elliptical utterance in gapping, there is no reason to assume that she is not able to infer the QuD in case of fragments. The set of potential QuDs might often be more restricted in case of gapping than in discourse-initial fragments by the first conjunct, so that there might be a quantitative difference between the size of the set of possible QuDs reconstructed in gapping and fragments. However, there is not necessarily a categorical difference between both constructions.

Furthermore, from a psycholinguistic perspective, the theories of fragments discussed so far are production accounts. What matters primarily to them is the speaker’s perspective: Ellipsis is licensed if there is a QuD in context which the speaker believes to be sufficiently salient. Since the hearer is aware of this, she knows that there must be such a contextually salient QuD as soon as she realizes that the speaker’s utterance is elliptical. If the speaker is cooperative or at least has the intention to get his message across (Grice 1975, Sperber & Wilson 1995), he will only use a fragment when he believes that the QuD is relatively easy to retrieve. For instance, if there is a high risk of being misunderstood due to several equally likely competing QuDs that differ in meaning, the full sentence will be preferred. In fact, this is supported by the experiments on script knowledge in Chapter 5 of this book. Consequently, indeterminacy of the meaning of a QuD does not impede communication if fragments are used only when the QuD is relatively predictable. Even if the meaning of the QuD is retrievable, its lexicalization is not necessarily, as the set of semantically similar QuDs listed above for the taxi example showed. However, communication can succeed even if the hearer fails to recover exactly the lexicalization of the QuD that the speaker had in mind, as long as the recovered QuD causes the hearer to perform the intended action. No matter which of the paraphrases the driver chooses in order to enrich the fragment in (26a), she will still carry the passenger to the university.

The difference between implicit and explicit QuDs is furthermore specifically relevant to the hearer's perspective. If the speaker chooses to produce a fragment, he must have a particular QuD in mind, be it implicit or explicit. Consequently, from his perspective there is no categorical difference between explicit QuDs (in case of short answers) and implicit ones (in case of discourse-initial fragments). Obviously, this changes from the perspective of the hearer, who has to reconstruct the missing material, because this task is facilitated by a QuD. Still though, as I noted above, a speaker who wants to get her message across will only choose to use the fragment if the effort required for the hearer to infer the intended QuD is reasonably small. This is confirmed by my experiments 11 and 12, which show that in particular predictable, that is, easily recoverable, words are omitted.

The assumption that the resolution of ellipsis might require some degree of inference is further supported by research on relatively uncontroversial instances of antecedent-based ellipsis. For instance, both gapping and VPE allow for mismatches between the antecedent and target of ellipsis. The VPE example in (30a) requires the hearer to reconstruct an active VP *look into this problem* given a passive antecedent. Similarly, in gapping, hearers can reconstruct a plural verb given a singular antecedent (30b). Therefore, rather than being a copy-and-paste process (Frazier & Clifton 2001), the reconstruction of elided material seems to be a task that requires retrieving the omitted material from available contextual evidence and that becomes more effortful the more antecedent and target differ (Arregui et al. 2006).

- (30) a. This problem was to have been looked into, but obviously nobody  
did <look into this problem>. (Kehler 2002: 548)
- b. They are going to Chicago, and I <am going> to San Francisco.  
(Pullum & Zwicky 1986: 755)

Taken together, a uniform analysis of discourse-initial and short answer fragments as elliptical sentences is possible. Just like fragments, some antecedent-based ellipses require inferential reasoning about the omitted material, so that the difference between short answers and discourse-initial fragments is a gradual one: In the case of short answers, the missing material is explicitly given, so that its retrieval will be easier than in discourse-initial fragments (on average). If such a uniform account explains the data equally well, it is simpler than a mixed account that distinguishes between discourse-initial fragments and short answers and consequently to be adopted unless there is evidence against it.

Unlike Reich (2007) and Weir (2014a), Merchant (2004a) does not rely on the concept of QuD, but his theory can also account for discourse-initial fragments.

Since e-givenness, which licenses fragments in his theory, operates on semantic representations, fragments in principle require a salient linguistic antecedent. Merchant's account distinguishes between fragments that occur in highly conventionalized contexts like the taxi example (25a) and those which do not, like (31a). In the case of highly predictive contexts, he notes that speakers have strong expectations about what is likely to happen and to be said in such contexts. Merchant (2004a: 730–731) argues that this *script knowledge* in the sense of Schank & Abelson (1977)<sup>14</sup> can make specific linguistic expressions manifest in the sense of relevance theory, that is, “capable [...] of representing it mentally and accepting its representation as true or probably true” (Sperber & Wilson 1986: 39). Such manifest linguistic objects can then serve as antecedents and license omission if they result in parts of the utterance being e-given. For non-conventionalized cases, Merchant (2004a: 722–727) argues that context can still make entities and concepts like the letter and its origin in (31a) manifest. This licenses the ellipsis of very basic deictic expressions, as the predicate *do it* for actions and pronouns for entities in the structure in (31b) which he assumes for the fragment in (30).

- (31) a. [Bill walks into the living room waving a postcard. He says:] The  
postcard is from John!  
b. From John it is.

The deictic expressions are then resolved from context, just like they are in regular sentences. Merchant (2004a: 722) argues that the assumption that the unarticulated structure in such fragments consists in the minimally required and semantically less specified expressions also accounts for the indeterminacy of discourse-initial fragments. Pronouns can be often paraphrased with various complex DPs, so that this apparent property of fragments turns out to be just a more general property of deictic expressions.

Taken together, the assumption that ellipsis can be licensed by extralinguistic context provides an empirically necessary extension of the sentential account of fragments to (apparently) discourse-initial fragments. Nonsentential accounts, which rely on pragmatic inference in any case, do not face this problem, but might resort to the same mechanisms (e.g. script knowledge and implicit QuDs) in order to explain how fragments are interpreted.

---

<sup>14</sup>See Section 5.1 for details on the concept of script and a discussion of its psychological effects.

## 2.3 Fragments as ungrammatical utterances

All of the theories discussed so far agree on the assumption that fragments are grammatical objects, be it by postulating mechanisms that license and trigger ellipsis or by modifying the theory in order to allow for nonsententials to be a well-formed output of syntax. This view is challenged by Bergen & Goodman (2015), who sketch a game-theoretic account of fragment usage and argue that fragments are actually ungrammatical, but that under some circumstances they might still be the preferred means of communicating a message. In simplified terms, Bergen & Goodman argue that hearers use a repair mechanism that helps to figure out the omitted parts of the utterance in order to infer the intended meaning from fragments. Such a mechanism is needed independently of fragments in order to deal with utterances that are corrupted by e.g. acoustic noise. Hearer and speaker are mutually aware of this possibility, so that the usage of a fragment and the subsequent inference about the intended meaning is more economic than that of a full sentence if the missing parts of the utterance are relatively easy to retrieve. Since this theory is highly usage-oriented and closely related to information theory, I discuss it in greater detail in Section 4.2.2. For the time being, what matters is the presumed ungrammaticality of fragments.

Such an account predicts that there are no restrictions on the form of fragments, provided the chosen utterance is the most suitable for communicating a message in a specific context. In contrast, syntactic accounts categorically distinguish grammatical from ungrammatical fragments. Even though the acceptability of a grammatical fragment will also be determined to a large extent by context, there are grammatical principles that cannot be overridden. For instance, the movement restrictions that Merchant (2004a) presents as evidence for his theory, or the requirement for fragments to be maximal XPs according to Barton & Progovac (2005), impose restrictions on the form of fragments that are context-independent.

It is hardly possible to empirically confirm the claim that fragments are ungrammatical objects, but it can be falsified by evidence for context-independent constraints on the form of fragments. The empirical picture is mixed: On the one hand, as my experiment 6 will show, gradual differences in acceptability among ungrammatical fragments are in line with Bergen & Goodman's (2015) account. On the other hand, there are grammatical constraints that override constraints that otherwise license omissions. For instance, Lemke et al. (2017) find that the omission of articles in German newspaper headlines is subject to processing considerations that are related to those in Bergen & Goodman (2015). In standard German however, including newspaper text, article omission is restricted only

to very specific contexts (e.g. predicative and plural nouns). If grammaticality did not matter at all, the same amount of omissions would be expected in similar contexts across text types which are less constrained by normative pressures, like colloquial speech.

This objection concerns only the presumed ungrammaticality of fragments, but not the mechanism that Bergen & Goodman (2015) propose in order to account for the speaker's choice between a fragment and a sentence. Such a mechanism is needed in any case,<sup>15</sup> because also when fragments are assumed to be grammatical the speaker must somehow decide on the form of the utterance. An account of this choice process is beyond the scope of generative syntactic theory, which does not attempt to model production preferences. Taken together, the main difference between the account in Bergen & Goodman (2015) and syntactic theories of fragments concerns the question of whether the set of utterances that are possible in a situation is somehow constrained by syntax or derived by arbitrary omissions.<sup>16</sup>

## 2.4 Testable predictions of theories of fragments

The accounts discussed in the preceding section assume very different derivations of fragments and syntactic structures underlying them. However, all of them are designed to account in principle for the same acceptability data reported in the literature. Therefore, in order to evaluate their predictions empirically, it is necessary to isolate phenomena with respect to which these predictions differ. In this section, I present four such phenomena, which might provide a testing ground for the theories: case connectivity effects, discontinuous fragments, information-structural restrictions, and movement restrictions. As will turn out, only case connectivity and movement evidence will be useful to distinguish between the theories. With respect to constituency and focus marking, the theories either coincide for independent reasons or they do not make precise empirical predictions with respect to them at all.

### 2.4.1 (Anti)connectivity effects: Case marking

Both those authors who defend a sentential analysis of fragments and those supporting a nonsentential account consider case connectivity effects to be an impor-

---

<sup>15</sup>See Section 5.5 for a more detailed sketch of such an account.

<sup>16</sup>Note that Bergen & Goodman (2015) do not discuss how this set of possible utterances is derived and present only a very simple example. Specifically, it is unclear whether they assume that fragments are genuine nonsententials or whether they take grammatical utterances as point of departure and generate the alternative by applying arbitrary ellipses to them.

tant diagnostic for the syntax of fragments. Merchant (2004a: 676–679) presents evidence for connectivity effects from a diverse sample of languages. His general observation is that DP short answers like the German example (32) receive the same case morphology as their counterparts in complete sentences: Just like the *wh*-phrase in the question, the DP in the short answer has to receive accusative case marking (32a), whereas the dative (32b) is ungrammatical. From the sententialist perspective, case connectivity provides indirect evidence for unarticulated structure: In minimalism, at least non-default case must be checked in a local syntactic configuration between the DP and another head, so the grammaticality of such a case marking indicates the presence of an unarticulated licenser. For instance, in (32a) accusative must be checked by a verb, so that the acceptability of case marking in the absence of an overt verb suggests that the structure contains that verb, but that it has been PF-deleted in the course of the derivation. This pattern is reversed in (33), where the verb in the question requires dative. In this case, a dative, but not accusative, DP fragment is acceptable.

- (32)      Wen      sucht Hans?  
              who.ACC seeks Hans  
              ‘Who is Hans looking for?’  
          a.    Den      Lehrer.  
                  the.ACC teacher  
          b.    \*Dem    Lehrer.  
                  the.DAT teacher
- (33)      Wem      folgt Hans?  
              who.DAT seeks Hans  
              ‘Who does Hans follow?’  
          a.    Dem      Lehrer.  
                  the.DAT teacher  
          b.    \*Den      Lehrer.  
                  the.ACC teacher

Case connectivity effects thus support unarticulated structure in fragments. In contrast, as I already noted above, anticonnectivity effects, that is, mismatching case morphology between fragments and full sentences, support nonsentential accounts. An example for anticonnectivity is the ungrammaticality of nominative DP fragments in the English example (34), although this case is required in the corresponding full sentence. Recall that Barton & Progovac (2005) explained these data with the inability of fragments to exhibit structural case morphology under the assumption that nominative is structural case in English.

- (34) Who can eat another piece of cake? (Barton & Progovac 2005: 77)
- a. ?\*I/?\*We/?\*He/?\*She
  - b. Me/Us/Him/Her

The introspective data reported in the literature is contradictory. Barton & Progovac's example (34) exhibits anticonnectivity, but Merchant's German examples in (32) and (33) seem to evidence case connectivity. Both sententialists and nonsententialists present explanations for at least part of the data that seems to contradict their respective theories. In what follows, I first review how the nonsentential account could account for examples like (32a) and conclude that only the sentential account explains such connectivity effects. This rests on the assumption that accusative is structural case in German. This is supported by tests that Progovac (2006) use to show that Serbian accusative is inherent case, but which yield the opposite result when they are applied to German. I then discuss how sentential accounts can deal with anticonnectivity effects as in (33).

#### 2.4.1.1 Nonsentential accounts and connectivity

As discussed above, Progovac et al. (2006) distinguish between default, structural and inherent case. In contrast to structural case features, which are uninterpretable, inherent case features can and must be interpreted by semantics, because inherent case is related to a specific  $\vartheta$ -role. The relatively uncontroversial claim that dative is inherent case by Progovac et al. (2006: 339) explains the acceptability of dative DP fragments as (33a). In contrast, fragments that appear in structural case must be ungrammatical according to the nonsentential account, because structural case features must be checked and structural case-marked DP fragments lack an appropriate verbal head under a nonsentential analysis. This prediction is challenged by the acceptability of accusative DP fragments in German (32a), under the assumption that accusative is structural case in German.

Progovac et al. (2006) however argue that accusative is not necessarily structural case in languages where nominative is the default case. They exemplify this idea for Serbian and present three arguments that attempt to show that Serbian accusative is also inherent, interpretable case. First, they note that an accusative DP "is typically in fact a theme/patient" (Progovac et al. 2006: 339) and exclude a set of other  $\vartheta$ -roles for accusatives (agent, goal/recipient, instrument, locative). As discussed above, the association with a specific  $\vartheta$ -role is the central criterion for categorizing a specific case as inherent case. Second, they argue that in Serbian accusative contributes to semantics and present the contrast in (35) as evidence that accusative has a universal quantificational meaning, while genitive

quantifies existentially. Finally, they state that, in contrast to English, Serbian accusative objects do not always appear adjacent to the verb. This observation is not discussed in greater detail but is probably intended to suggest that no specific syntactic configuration is required for accusative to be licensed.

- (35) a. Dodaj vodu. (Progovac et al. 2006: 340)  
           add water.ACC  
           ‘Add (all) the water.’  
       b. Dodaj vode.  
           add water.GEN  
           ‘Add (some) water.’

Whether these arguments are empirically correct for Serbian is beyond the scope of this book. However, if they applied to German as well, this could derive grammatical accusative DP fragments like (32a) under a nonsentential account and consequently undermine the status of case connectivity effects as evidence for unarticulated structure in fragments.

Nevertheless, the diagnostics used by Progovac et al. (2006) to analyze the Serbian accusative as structural case do not yield the same result for the German accusative. The  $\varnothing$ -role argument that accusative DPs are typically themes or patients probably holds in German too. For instance, it seems reasonable to assume that a random accusative DP drawn from a corpus is more likely to be a patient than a random nominative DP is. However, it is unclear how the gradual likelihood of being a patient can be aligned with the categorical distinction between structural, inherent and default case upon which Barton & Progovac’s (2005) theory relies. This probably concerns the Serbian data as well. The remaining two arguments for analyzing Serbian accusative as inherent case do not hold in German. Quantification relies on the presence of the definite article in German<sup>17</sup> and accusative DPs still appear adjacent to the verb in the unmarked word order in German. This suggests that at least for German the analysis of accusative as structural case is correct. Consequently, if empirically confirmed,

<sup>17</sup>This is evidenced by the German counterparts of (35) in (i). The adjective has accusative case morphology in both examples but the meaning depends on the presence of the definite article.

- (i) a. Geben Sie den frischen Zitronensaft hinzu.  
           give.IMP you the.ACC fresh.ACC lemon.juice to  
           ‘Add (all of) the fresh lemon juice.’  
       b. Geben Sie frischen Zitronensaft hinzu.  
           give.IMP you fresh.ACC lemon.juice to  
           ‘Add (some of the) fresh lemon juice.’



the acceptability of the accusative DP fragments in German challenges Barton & Progovac's (2005) theory because it does indeed predict anticonnectivity effects in that case.<sup>18</sup> Experiment 1 disconfirms this prediction.

#### 2.4.1.2 Sentential accounts and anticonnectivity

Anticonnectivity effects have been presented as evidence against sentential accounts, but at least some of the data can be explained by these theories. With respect to English pronominal short answers discussed above, Merchant (2004a: 700–704) reports similar data for Greek, Dutch, German and French, which, aside from cross-linguistic differences, are characterized by formal differences between the pronouns found in fragments and those in full sentences. For instance, the French data in (36) show that in fragments only the strong, tonic pronoun *moi* is acceptable (36a), whereas the clitic *me* must be used in regular sentences (36b).

However, Merchant notes that in the left periphery only those pronouns that appear in fragments are acceptable. For instance, in Clitic Left Dislocation (CLLD) in French (37a) or hanging topic left dislocation (HTLD) in English (37b) the form of the pronoun matches the one found in short answers.

- (36) Il voulait qui? (Merchant 2004a: 701)  
 he wanted who  
 'Who did he want?'  
 a. Moi/\*Me  
    me.strong/me.weak  
 b. Il me/\*moi voulait.  
    he me.weak/strong wanted  
    'He wanted me'

---

<sup>18</sup>Having defined Serbian accusative as inherent case, Progovac et al. (2006: 340) argue that nominative DP fragments are degraded as answers to questions asking for an accusative (i) for pragmatic reasons. They claim that the accusative short answer is preferred over the nominative one because it is more informative. This cannot explain the German data: If accusative is structural case, the only grammatical short answer is the nominative one. The glosses and the grammaticality judgement in (ib) were added by R.L. based on the text accompanying the example in Progovac et al. (2006).

- (i) a. Koga je Ana posetila? (Progovac et al. 2006: 340)  
       who.ACC is Ana visited?  
       'Who did Ana visit?'  
    b. #Vera!  
       Vera.NOM

## 2 Theories of fragments

- (37) a. *Moi/\*Me, il me voulait.* (Merchant 2004a: 702)  
           *me.strong/weak he me wanted*  
           ‘Me, he wanted’  
       b. *Me/\*I, I watered the plants.* (Merchant 2004a: 703)

Merchant argues that this is empirically in line with his theory, which assumes structures such as (37) to be the source of fragments. However, he is reluctant to assume HTLD as the structure underlying fragments, because, unlike fragments, this construction is insensitive to islands. Instead, he suggests that the formal restrictions on pronominal fragments are due to the fact that weak pronouns cannot be focused. Merchant (2004a) takes this as evidence for the movement component of his theory, but the observation that only pronouns which can be focused are possible fragments is perfectly in line with Reich’s (2007) *in situ* ellipsis account: (38a) shows that the English pronoun *me* can receive a pitch accent marking narrow or contrastive focus, whereas in French, the tonic pronoun (38b) can but the clitic (38c) cannot. The deletion of all non-focused constituents under a semantic identity condition as proposed by Reich (2007) consequently yields the same pattern as movement and deletion.<sup>19,20</sup>

- (38) A: Did they elect Ann?  
       a. B: No, they elected *me*.  
       b. B: Non, ils ont élu *moi*.  
           no they have elected I.strong  
       c. B: \*Non, ils *me* ont élu.  
           no they I.weak have elected

Taken together, the anticonnectivity effects observed by Barton & Progovac (2005) for pronominal DP short answers are explained not only by Barton & Progovac’s (2005) default case hypothesis, but also by both types of sentential accounts. Case connectivity effects concerning inherent case are also expected under both sentential and nonsentential accounts.

<sup>19</sup>Some authors account for mismatches between sentences and fragments with respect to preposition omission by deriving fragments from clefts (see e.g. Rodrigues et al. (2009) for Spanish and Brazilian Portuguese and van Craenenbroeck (2010) for English). Under such an analysis, according to which accusative in fragments evidences a connectivity effect than an anticonnectivity effect, (4) would be derived from a structure like (i).

(i) A: Who can eat another piece of cake?  
       B: It’s ~~me who can eat another piece of cake~~

<sup>20</sup>The French data in (38) are also compatible with the nonsentential account. Clitics like the French *me* appear always adjacent to an inflected verb, which the nonsentential account argues to be absent in fragments. I thank Sebastian Nordhoff for pointing this out.

What remains disputed is the possibility of case marking on full DP fragments, which will serve as a diagnostic for unarticulated structure in my experiments. While sentential accounts predict case connectivity also for structural case, if the nonsentential account is right, DPs that receive structural case marking in full sentences appear in default case in fragments. As noted by Progovac et al. (2006), due to crosslinguistic variation, it is necessary to carefully determine whether a specific case is actually a structural case in a given language. Finally, if fragments are ungrammatical, as suggested by Bergen & Goodman (2015), no grammatical restrictions on case marking in fragments are expected. Still though, case in fragments should tend toward matching that in the corresponding sentence, because it is a strong cue toward the meaning intended by the speaker.

### 2.4.2 Constituency

The theories presented in the previous section disagree on whether fragments must be a single constituent: The nonsentential account by Barton & Progovac (2005) as well as movement and deletion predict this, but in situ deletion and Bergen & Goodman's (2015) approach allow for non-constituent fragments.

#### 2.4.2.1 Constituency under the nonsentential account

Barton & Progovac (2005) argue that any maximal projection XP can be a well-formed output of syntax, therefore a fragment must always consist of a single constituent of arbitrary size and internal complexity. Sequences of constituents that do not have a unifying node cannot be derived. The examples discussed in Barton & Progovac (2005) clearly conform to this restriction. Progovac (2006) discusses more controversial examples like (39), which are harder to analyze as a single constituent, because the predicate *a bargain* does not have a verbal projection, whose specifier could host the subject.

- (39) This a bargain!? (Progovac 2006: 28)

In order to account for such apparent DP-DP sequences too, which in fact are relatively frequent (Reich 2017), Progovac (2006) resorts to a small clause analysis. Small clauses (Williams 1975, Stowell 1981) are expressions that contain a subject and a predicate, but no verbal material. They can appear as root small clauses (39) or be embedded as verbal complement inside a matrix clause (40).

- (40) I consider this a bargain. (Progovac 2006: 41)

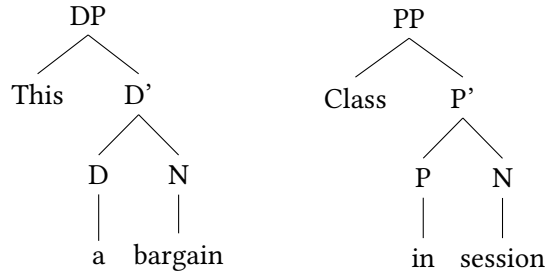


Figure 2.3: DP and PP small clauses according to Progovac (2006: 39).

Progovac (2006: 61) claims that “every sentence/clause is underlyingly a small clause.” The subject is merged into the specifier of a projection whose category is determined by the predicate, as the examples in Figure 2.3 illustrate.<sup>21</sup> If the small clause analysis is correct, it enables the nonsentential account to derive XP-YP sequences as small clauses, as long as they do not contain uninterpretable features. Progovac (2006: 51) argues that the latter condition is met, because the subjects of root and embedded small clauses in English appear in accusative default case (41a), which does not need to be checked according to the nonsentential account. If the small clause serves as input to a regular inflected clause (41b) instead, the subject is selected from the lexicon with nominative case features that are checked as usual in course of the derivation by T. In the embedded (41c), case is checked through ECM by the verb (Progovac 2006: 46).

- (41) a. Her give up?! (Never!) (Progovac 2006: 41)  
 b. She gives up?!  
 c. I never saw her give up.

This analysis works for the (mostly) English examples discussed by Progovac (2006), but there are instances of fragments for which the small clause analysis fails. For instance, the German fragment in (42a) seems to be acceptable just like its English counterpart (39) is, but, unlike in English (40), it cannot be embedded under a verb in a matrix clause (42b). Instead, the complement must be a clause headed by a copula (42c). Reich (2017) presents further arguments against a small clause analysis of German XP-YP fragments, which are based on corpus data from newspaper headlines. He concludes that such cases involve a null copula and consequently analyzes (42a) as underlyingly sentential.

- (42) a. Das ein Angebot?!  
 this a bargain

<sup>21</sup>This analysis can be traced back to Stowell (1981), see Citko (2011) for an overview of alternative structural analyses of small clauses.

- b. \*Ich finde das ein Angebot.  
I find this a bargain  
'I consider this a bargain.'
- c. Ich finde, das ist ein Angebot.  
I find this is a bargain

Furthermore, in question-answer pairs like (19), repeated here as (43), fragments can consist of two non-predicative DPs with an omitted main verb. A small clause analysis is ruled out in this case, because the missing verbal element is not the copula. It is unclear how Barton & Progovac (2005) would account for such fragments without assuming an unarticulated verbal node, ellipsis, or a discontinuous fragment, which their account cannot derive.

- (43) [Waiter to customers in the restaurant:] Who ordered what?  
Customer: She ~~ordered~~ the deep dish pizza.

The nonsentential account hence predicts that only expressions consisting of a single constituent are possible fragments. Otherwise, the grammar assumed by Barton & Progovac (2005) would not be able to generate them. Progovac's (2006) small clause analysis accounts for some apparent non-constituent fragments, but presupposes that small clauses are grammatical in the language. The comparison between German and English suggests that similar fragments are acceptable in both languages even though the small clause analysis does not account for the relevant data in German (Reich 2017). Furthermore, (43) suggests that even in English not all fragments can be traced back to a small clause.

#### 2.4.2.2 Constituency under sentential accounts

Merchant's (2004a) movement and deletion account also predicts that fragments are always single constituents. As their derivation involves movement to [Spec, FP] for feature checking purposes, only one constituent can be moved at a time. The reason for this is that there is only one landing site for the fragment: the specifier of a head carrying the E feature.<sup>22</sup> Of course, there are no restrictions on the internal complexity of this constituent, so that Merchant (2004a) can account for small clause (41a) or VP fragments (44):

- (44) What would you like to do tonight?  
Go to the cinema, ~~I'd like to~~.

<sup>22</sup>If multiple specifiers are assumed, as the minimalist program permits (Chomsky 1995: 285) this restriction obviously does not hold anymore. However, Merchant (2004a) does not seem to resort to multiple specifiers in order to explain apparent non-constituent fragments.

As Merchant (2004a) allows for unarticulated structure in his elliptical sentences, the set of apparent non-constituent fragments that his account can explain is larger than that covered by the nonsentential account. For instance, a sequence of a temporal and a locative adverbial does not form a small clause but seems to be felicitous as a fragment (45). For the movement and deletion account to derive such fragments, they need to form a single constituent at some point of the derivation. The assumption that this is possible in case of such adverbials is supported by German data from Haider (2000). Haider observes that such *clusters* consisting of event-related adverbials, may appear together in the German sentence-initial prefield (46), which is generally considered to host only one constituent (see also the discussion of experiment 10 in Section 3.4.2.1).<sup>23</sup>

(45) Q: When did you last meet him?

A: Last night in the pub.

(46) Gestern im Hörsaal als der Vortrag begann hustete er wie  
yesterday in.the lecture.room when the lecture started coughed he like  
verrückt. (Haider 2000: 104)

mad

‘Yesterday in the lecture room as the lecture started he coughed like mad.’

A similar point is made by Müller (2002) in his analysis of (apparent) multiple prefield constituents in German. As pointed out above, German declarative matrix clauses are verb-second (in what follows, V2), so that only one constituent may precede the inflected verb. In spite of this, Müller (2002, 2003, 2005) reports a diverse sample of corpus data which seem to violate this constraint. For instance, in (47), taken from Müller (2005: 38), a locative and a temporal PP, which do not straightforwardly form a single constituent, can appear together in the prefield.<sup>24</sup> Müller (2002) develops a HPSG account of such data, whose central assumption is that the prefield consists of a single constituent which is headed by an unarticulated verb.

(47) [Vor drei Wochen] [in Memphis] hatte Stich noch in drei Sätzen  
ago three weeks in Memphis had Stich still in three sets  
gegen Connors verloren.

against Connors lost

‘Three weeks ago in Memphis Stich had still lost in three sets against Connors.’

<sup>23</sup>Haider (2000: 104–105) also rules out an alternative derivation that fronts the complete VP after extracting the verb for independent reasons, but cf. Müller (2004) for such a proposal.

<sup>24</sup>Note that, unlike (44), this example does not involve fronting of the complete VP.

The exact predictions of the movement and deletion account thus depend to a large extent on how many and which movement operations and covert elements are assumed in general, therefore it is difficult to derive testable predictions unless there is consensus on these preliminary assumptions. The picture becomes even more complicated if a fine-grained left periphery is assumed. Even in the original sketch of the CP layer in Rizzi (1997) there is a further topic projection above FocP, and Benincà & Poletto (2004) distinguish about half a dozen highly specified functional projections, most of which are located above those related to focus. Depending on where Merchant (2004a) would locate the E feature in a language, the material in these projections could survive ellipsis, so that (if these projections exist in a language) the movement and deletion account predicts the possibility of genuine non-constituent fragments in highly specific information-structural configurations. Finally, the version of movement and deletion by Weir (2014a) also accounts for non-constituent fragments. Weir simply adjoins the moved constituents to CP and, since adjunction is recursive, there is no upper bound limit on the number of constituents extracted with this mechanism.

Taken together, the existence of (apparently) discontinuous fragments as evidence for or against specific accounts of fragments does not seem promising, because theories that in principle predict constituency to be a determining property of fragments can account for some instances of what superficially must be classified as non-constituents. Specifically, movement and deletion requires an extensive set of preliminary assumptions about which syntactic operations and functional projections are available in a given language in order to make clear predictions on non-constituent fragments. Therefore, I do not use constituency directly as a diagnostic for particular theories.<sup>25</sup>

### 2.4.3 Information structure and focus

All sentential accounts of fragments discussed so far (Merchant 2004a, Reich 2007, Weir 2014a) assume that information structure, in particular the focus-background structure of an utterance, determines which words can be omitted in fragments. Since the nonsentential account of fragments does not impose information-structural licensing conditions on fragments, focus-sensitivity could appear to be a promising testing ground to differentiate between sentential and nonsentential accounts. Reich (2007) and Weir (2014a) make the strongest claim on the issue by assuming that fragments are necessarily focused. Reich argues

---

<sup>25</sup> An exception is experiment 10, which compares the acceptability of apparent multiple prefield configurations as fragments and full sentences. However, the experiment does not depend on a specific syntactic analysis of the constructions tested, but on the parallelism between fragments and left dislocations predicted by the movement and deletion account.

## 2 Theories of fragments

that only F-marked expressions survive ellipsis, whereas Weir considers focus to trigger exceptional movement.<sup>26</sup>

As for Merchant (2004a), this is less clear: On the one hand, he tentatively suggests identifying the landing site for fragments as a FocP (Merchant 2004a: 675), on the other hand he emphasizes similarities between fragments and CLLD (Merchant 2004a: 703). However, according to the literature on CLLD, this construction does not involve focus movement but targets a topic position.<sup>27</sup> Further confusion on the status of the presumed left dislocation of fragments comes from the German data in (48). Merchant (2004a: 702) argues that the form of pronouns in the preverbal position (the *prefield*) equals that in the fragments in (49) (judgments are Merchant's).

- (48) Was wolltest du? (Merchant 2004a: 701)  
 what wanted you  
 'What did you want?'  
 a. Das/\*Es wollte ich. (Merchant 2004a: 702)  
 this/it wanted I  
 'This/It I wanted.'  
 b. Das/\*Es.  
 (49) \*Das<sub>i</sub>/\*Es<sub>i</sub> wollte ich ~~t<sub>i</sub>~~.

The structure in (48a) is derived neither by CLLD nor by focus movement, because it is a garden-variety verb-second declarative matrix clause. As I discuss in greater detail in Section 3.4.2.1, the mainstream analysis of German V2 consists in moving the inflected verb to C, whose specifier must be filled by any other constituent (den Besten 1989). Crucially, there are almost no restrictions on the category or information-structural status of the constituent in [Spec, CP], which can be an aboutness or contrastive topic as well as a focus (Frey 2005). If (48a) is a standard verb-second clause and the mainstream analysis of V2 is correct,<sup>28</sup> the

<sup>26</sup>If there is more than one focus, their predictions might differ with respect to ordering. Reich (2007) predicts the same ordering as in regular sentences, but for Weir's theory it matters whether several exceptional movement operations proceed from the top of the syntax tree to its bottom or vice-versa. E is located on the C head dominating the whole TP, so that if the constituents closest to E were evacuated first, both accounts would predict differing orderings.

<sup>27</sup>Benincà & Poletto (2004: 53) distinguish between topic and focus positions in the left periphery by noting that the former "are connected with a clitic or a *pro* in the sentence", while the latter leave a variable which is bound by the moved phrase. Based on mostly Spanish data, Arregi (2003) argues that CLLDed constituents are contrastive topics. Contrastive topics differ from foci both in their syntactic properties and in their prosody (see e.g. Büring 1997, Krifka 2007).

<sup>28</sup>Müller (2004) proposes an analysis of V2 that consists in VP fronting after those constituents that appear in post-verbal positions have been extracted. As he assumes that the VP contains only the finite verb and the prefield when it is fronted, it makes the same predictions as the mainstream account of V2, that is, the verb must survive ellipsis in fragments.



fragment in (49b) cannot be derived from (49a) by assuming an E feature on C: E triggers only PF-deletion of the complement of C, hence the account incorrectly predicts the verb to survive ellipsis:<sup>29</sup> Furthermore, the structure in (48a) is not an instance of CLLD. Even though German declarative matrix clauses are verb-second, a further constituent can be placed left of the prefield (50). DPs appearing there sometimes exhibit no case connectivity at all (50a) and must therefore be analyzed as a hanging topic (Rodman 1974, Vat 1981), but in (50b) the DP *den Peter* resembles CLLD in being case-marked. Unlike (48a), both of these structures require doubling of the left-peripheral constituent by a pronoun.

- (50) a. Der Peter, den habe ich gestern erst getroffen.  
           the.NOM Peter, him have I yesterday just met.  
           ‘(As for) Peter, I just met him yesterday.’  
       b. Den Peter, den habe ich gestern erst getroffen.  
           the.ACC Peter, him have I yesterday just met.  
           ‘Peter, I just met him yesterday.’

However, the structure in (50b) cannot be the source of fragments. If E is located on C, again, both the verb and the pronoun would survive ellipsis. The derivation in Figure 2.4 shows that this would yield the ungrammatical *\*Den Peter, den habe*. Taken together, E cannot be located on C in German.

Merchant (2004a) specifies the syntactic properties of different varieties of E in their lexical entries, so there is no principled reason to assume that the German  $E_F$  must also be  $[uC^*]$ . If it had an  $[uF^*]$  feature and were thus located on F, as Merchant (2004a) initially suggested for English, the verb would always be PF-deleted and the existence of DP fragments straightforwardly explained. The problem with this assumption is that Merchant (2004a) rejected it in order to account for the island sensitivity of fragments based on the PF-deletion of illegal traces (see Section 2.4.4 for discussion). Therefore, if the E feature was located on a F head above CP in German, German fragments should not be island-sensitive, but (51) shows that they are.<sup>30</sup>

<sup>29</sup> A possible explanation would be that T-to-C movement of the verb occurs only in order to satisfy some PF constraint and is therefore not required (which is equivalent to not being allowed in minimalism) in elliptical sentences. I am not aware of any proposal in this direction.

<sup>30</sup> Merchant (2004a) judges the English counterpart of (51b) as ungrammatical when it is interpreted as (ia). If (51b) was interpreted as (ib), he predicts the fragment to be grammatical, because fronting *Charlie* in the matrix clause does not require the extraction out of an island.

- (i) a. Nein, Abby spricht die gleiche Balkansprache, die Charlie spricht.  
           no Abby speaks the same balkan.language that Charlie speaks  
           ‘No, Abby speaks the same Balkan language that Charlie speaks.’  
       b. Nein, Charlie spricht die gleiche Balkansprache, die Ben spricht.  
           no Charlie speaks the same balkan.language that Ben speaks  
           ‘No, Charlie speaks the same Balkan language that Ben speaks.’

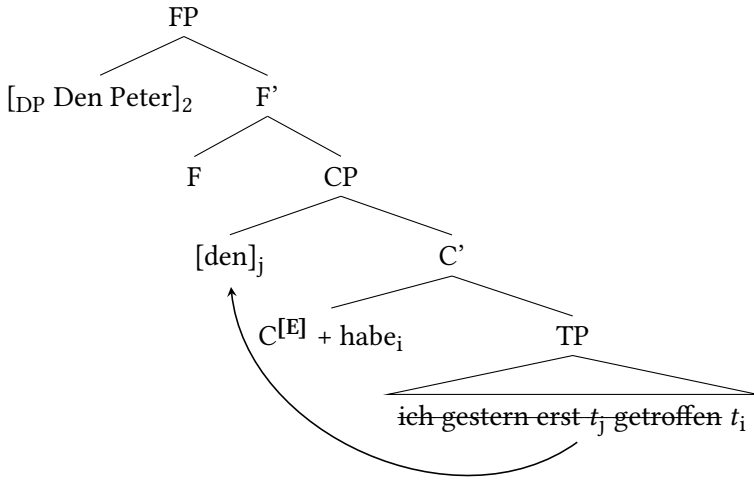


Figure 2.4: The derivation shows that if the E feature was located on [Spec, CP], German finite words are incorrectly predicted to survive ellipsis in fragments.

- (51) a. Spricht Abby die gleiche Balkansprache, die *Ben* spricht?  
 speaks Abby the same balkan.language that Ben speaks  
 ‘Does Abby speak the same Balkan language that *Ben* speaks?’  
 b. \*Nein, *Charlie*. (translated from Merchant 2004a: 708)  
 no Charlie  
 ‘No, Charlie’

Taken together, the predictions of Merchant (2004a) with respect to focus marking are vague, because it is not totally clear whether he assumes fragments to target the focus position [Spec, FP] or whether their landing site is [Spec, CP]. Neither of these versions can account for the full range of the data discussed in this section. The exceptional movement version of the theory (Weir 2014a) does not require information structure-related projections as FocP but simply adjoins fragments to CP, so that it is not affected by these issues.

Nonsentential accounts do not make reference to focus, but the conditions on fragment use that they require are related to information-structural notions. Barton (1990) proposes to “delete up to recoverability” and Stainton (2006) requires that a salient nonlinguistic LF which allows to enrich the fragment to a proposition must be present in context. As foci tend toward being new and consequently not recoverable, both of these ideas make similar predictions with respect to the acceptability of fragments as a focus-based account.

Focus sensitivity therefore does not offer a promising testing ground to distinguish the predictions of the theories of fragments that I investigate. Besides that, the effect of focus is relatively difficult to investigate experimentally. In German and English, focus is often marked prosodically with a H\* pitch accent (Gussenhoven 1983, Pierrehumbert & Hirschberg 1990) and the prominence of prosodic focus marking varies gradually as a function of the size of the focus domain Baumann et al. (2006, 2007). This is hard to manipulate experimentally: As Baumann et al. (2007) report, speakers make use of different strategies to modulate the prosodic prominence of foci, so that items might not be understood as intended. Furthermore, while work in experimental pragmatics has provided evidence for an effect of pitch accents on interpretation of complete sentences (see e.g. Chevallier et al. 2008, Zondervan 2010), it is difficult to apply this to fragments. For instance, in DP fragments consisting only of a noun and an article, the most prominent accent always falls on the noun, hence it is not possible to vary the pitch accents on fragments in order to elicit different focus structures.

### 2.4.4 Evidence for movement

In contrast to any of the other accounts, Merchant's (2004a) theory requires fragments to undergo obligatory movement to a left-peripheral position. From a naïve perspective, this predicts that only expressions that can occur in a left-peripheral position, more specifically, to the left of the head hosting the E feature, are possible fragments. Consequently, whatever might restrict movement to the left periphery in full sentences will also constrain the form of fragments. In the next chapter, I empirically investigate two of the movement restrictions discussed by Merchant (2004a). These restrictions, whose effects on the form of fragments have been first tested in Merchant et al. (2013), concern complement clause topicalization and preposition stranding. The reason for choosing these phenomena is that Merchant et al. (2013) present the first experimental evidence on them, what suggests that they consider them a valid testing ground for the movement and deletion account. Preposition stranding restrictions have the additional advantage that they cannot be overridden by exceptional movement according to Weir (2015) and hence also allow us to test Weir's (2014a) version of the theory.

The idea that movement restrictions constrain the form of fragments is exemplified for complement clause topicalization in (52) and (53), from Merchant (2004a: 690): As has been repeatedly claimed in the theoretical literature (see e.g. Morgan 1973, Chomsky 1981, Stowell 1981, Webelhuth 1992), the complementizer in English non-factive complement clauses is optional when the complement

clause appears in its base position (52a), but becomes obligatory when the complement clause appears in the left-periphery (52b). Merchant (2004a) observes that the same holds for fragments (53). He concludes that this is unexpected under the in situ deletion account, because the complementizer would be optional in fragments too if their underlying structure was (52a) instead of (52b). The other movement restrictions discussed by Merchant (2004a) behave similarly, that is, expressions which cannot appear in the left periphery seem to be unacceptable as fragments.

- (52) a. No one believes (that) I'm taller than I really am.  
       b. \*(That) I'm taller than I really am, no one believes.
- (53) What does no one believe? (Merchant 2004a: 690)  
       #(That) I'm taller than I really am.

Merchant (2004a) interprets such data as evidence in favor of his account, but in order for this to constitute valid evidence for movement in fragments, it is necessary to rule out alternative explanations for the observed pattern, which do not require movement. Throughout this book, some of these data will turn out to be relatively straightforwardly captured by the nonsentential or the in situ deletion accounts. For instance, a construction might have properties that block both movement in full sentences and ellipsis in fragments without having to assume that the latter necessarily undergo movement.

Besides this need for caution when interpreting coincidences between fragments and left dislocation as evidence for movement, acceptable fragments that cannot be left-dislocated potentially constitute counterevidence to Merchant's (2004a) theory. If ungrammatical left dislocations in full sentences always yielded unacceptable fragments, movement and deletion would be falsified even by the most basic examples, such as the unavailability of fronting of a DP which is not contrastive in English (Weir 2014a), or the island-sensitivity of fragments described by Merchant (2004a). This requires them to assume *repair effects* (Merchant 2004a) or exceptional movement (Weir 2014a) in order to conceal the theory with conflicting data.

Repair effects are widely acknowledged in the literature on ellipsis (see e.g. Fox & Lasnik 2003, Merchant 2008, Müller 2011, Lasnik 2015). The general observation is that in some cases ellipses are acceptable even though the presumably underlying nonelliptical structure is not. The idea is that ellipsis can remedy ill-formed structures by deleting those expressions that cause the problem. Since ellipsis is a PF phenomenon, this concerns only degraded PFs, but not derivations which are ungrammatical in the narrow syntax. A prototypical instance

of such repair effects is the island-insensitivity of sluicing. Recall that Merchant (2001) derives sluicing by regular *wh*-movement followed by deletion of the TP in the sluice. (54a) shows that sluicing is fine although the derivation assumed by Merchant involves an ungrammatical island violation: The *wh*-phrase must be extracted across the boundary of the embedded relative clause introduced by *who speaks* (54b).

- (54) a. They want to hire someone who speaks a Balkan language, but I don't remember which. (Merchant 2004a: 705)  
 b. \*They want to hire someone who speaks a Balkan language, but I don't remember [which Balkan language]<sub>i</sub> they want to hire someone who speaks *t<sub>i</sub>*.

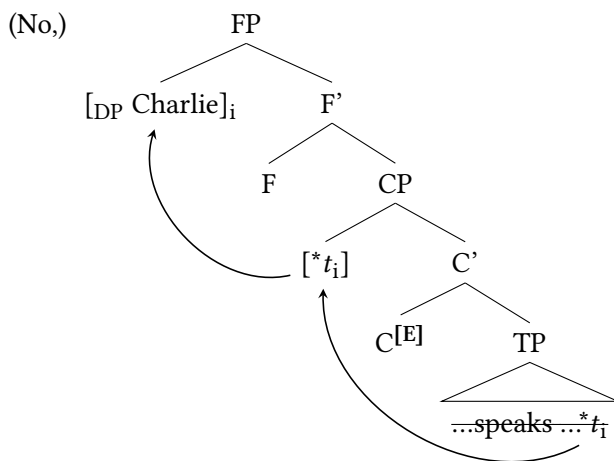


Figure 2.5: Derivation of a fragment answer to (55) according to Merchant (2004a: 708).

Merchant (2004a: 706) accounts for this by assuming that traces of movement across island boundaries have a feature \*, which renders PF representations that contain such features uninterpretable. As Merchant assumes that ellipsis is PF-deletion, it can delete such traces at PF and thus “repair” the defective structure. Merchant notes that his hypothesis can also account for the observation that sluicing is not sensitive to islands but other ellipses, like VPE and fragments, are. In sluicing, the E feature is located on C, so that it deletes all intermediate *\*t* traces within the TP at PF. In contrast, movement in fragments targets [Spec, FP] and, by proceeding cyclically, it leaves a *\*t* in [Spec, CP], as the derivation of

fragment answer to (55) in Figure 2.5 shows. The trace in the specifier survives ellipsis and causes the derivation to crash, because E is placed on C in fragments.

(55) Does Abby speak the same Balkan language that Ben speaks?

In fact, the need to account for the island sensitivity of fragments is what motivates Merchant to reject his initial assumption that E is placed on F in fragments, which is illustrated in Figure 2.6. If this derivation was correct, the PF-deletion of the defective trace would render fragments insensitive to islands, just like Merchant (2001) argues for sluicing. Repair effects complicate the use of paral-

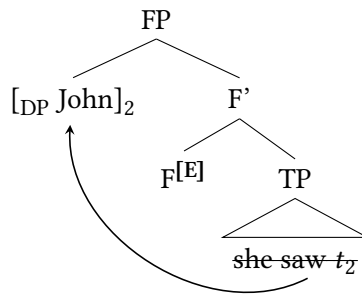


Figure 2.6: Derivation of fragments without an intermediate CP according to Merchant (2004a: 675).

lelisms between fragments and sentences as a diagnostic for movement. As Merchant (2004a: 711) puts it, “[...] the general argument is that parallelisms support a movement and ellipsis analysis, while non-parallelisms reveal repair effects.” Nevertheless, repair effects cannot just be stipulated but require an analysis, like the one involving the \* feature on traces that results from island violation.

Weir’s (2014a) exceptional movement account predicts an even larger set of mismatches between fragments and sentences than Merchant (2004a) because exceptional movement occurs only in fragments and is therefore relatively independent from movement in sentences. Weir only restricts exceptional movement to those movement operations that are “in principle” available in a language. The derivation of empirically testable predictions for his account would require criteria that define which types of movement are available in principle and which are not, but Weir (2015: 10) offers no such criteria and instead proposes that it “is most easily shown by example” how a movement operation is to be classified. He exemplifies this reasoning with the impossibility of fronting NPIs (56a) and bare quantifiers (57a) in English (the judgments are Weir’s) and argues that left

dislocation of such expressions is not blocked due to a syntactic restriction because argument DPs can be fronted in English (58). Consequently, he attributes the ungrammaticality of (56a) and (57a) to some kind of semantic ill-formedness and takes this to illustrate the PF-only character of exceptional movement.

- (56) What didn't John buy? (adapted from Weir 2015: 3)  
 a. ??Any wine, John didn't buy *t*.  
 b. Any wine.
- (57) Who will you talk to? (adapted from Weir 2015: 3)  
 a. \*Someone, I will talk to *t*.  
 b. Someone.
- (58) That guy, I saw *t*. (Weir 2015: 10)

The absence of criteria for which types of movement are in principle available makes it almost impossible to empirically evaluate the exceptional movement account. However, what matters from an empirical perspective is that Weir (2015: 11) explicitly excludes P-stranding (Pullum & Huddleston 2002), and left-branch extraction (Ross 1967, Bošković 2005) in languages which do not allow for either of these operations from the set of movement operations that are available in principle. Therefore, languages which do not lack these phenomena but which will allow for the generation of the corresponding fragments would also provide evidence against exceptional movement.

Repair effects, and in particular exceptional movement, notably complicate the predictions of the movement and deletion account on the correlation between the acceptability of fragments and left dislocation. Specifically, some of the mismatches in acceptability between acceptable elliptical and nonelliptical structures would not provide evidence against movement and deletion if they can be explained by some kind of repair effect. Nevertheless, repair effects increase the complexity of the theory and therefore require independently motivated accounts that justify them. The assumption of defective traces by Merchant (2004a) discussed above might provide such an account, but it still requires independent empirical evidence, for instance, one showing that similar observations can be made for related phenomena. However, and in the first place, empirical perspective however, the movement and deletion account requires evidence *for* movement in fragments that cannot be explained by other theories, like nonsentential approaches or the derivationally simpler in situ deletion account.

## 2.4.5 Summary

In this section I have discussed different potential testing grounds which might allow for the empirical evaluation of the competing accounts of fragments. Their respective predictions are summarized in Table 2.1.

Table 2.1: Overview of the empirical predictions the accounts of fragments make on the phenomena discussed in this section. The predictions of Merchant (2004a) with respect to non-constituent fragments and effects of movement restrictions depend on further theory-internal assumptions. Non-constituent fragment is predicted to be acceptable when multiple movement to the left periphery is possible, e.g. in Italian (Cinque 1990, Rizzi 1997). Under the exceptional movement version of the theory (Weir 2014a), movement restriction effects are only expected for movement which is unavailable *in principle* in a language.

|                          | Merchant<br>(2004a) | Reich<br>(2007) | Barton & Pro-<br>govac (2005) | Bergen & Good-<br>man (2015) |
|--------------------------|---------------------|-----------------|-------------------------------|------------------------------|
| Structural<br>case       | ✓                   | ✓               | ✗                             | ✓                            |
| Non-<br>constituents     | (✗)                 | ✓               | ✗                             | ✓                            |
| Focus<br>sensitivity     | ✓                   | ✓               | ✗                             | ✗                            |
| Movement<br>restrictions | (✓)                 | ✗               | ✗                             | ✗                            |

With respect to the question of whether fragments involve unarticulated structure, case connectivity effects, specifically with respect to structural case marking in fragments, turn out to be the most promising testing ground. Sentential accounts predict strict case connectivity, whereas the nonsentential account predicts that fragments cannot appear in structural case and will exhibit default case morphology instead. As argued above, whether a specific case is structural case in a language or whether it is not might be controversial, but the alternative, focus sensitivity, is even more difficult to test. The nonsentential account does not rely on the concept of focus, but makes similar predictions due to the necessity to retrieve deleted expressions from context. Furthermore, it is difficult to empirically elicit specific focus structures, because focus is mostly encoded prosodically in English and German (see e.g. Zimmermann & Onea 2011: 1658–1660).

Testing whether fragments are derived by movement and deletion, as claimed by Merchant (2004a), obviously requires the investigation of whether movement



restrictions constrain the form of fragments: Restrictions on the form of fragments and left-dislocated expressions would provide evidence for movement and deletion and the corresponding mismatches evidence against it. As for the former, it must also be shown that the nonsentential or the in situ deletion accounts cannot account for the data, while the possibility of repair effects must be considered in case of apparent non-parallelisms. My experiments investigating movement restrictions take the studies by Merchant et al. (2013) on preposition omission and complementizer omission as a starting point. The first of these phenomena will also allow for conclusions on Weir's (2014a) exceptional movement theory, which seems to have a greater empirical coverage than the original version of movement and deletion, but it is also harder to test because of the lack of clear criteria that would distinguish movement operations that can occur in fragments from those that cannot.

As for the assumption that fragments are inherently ungrammatical, I noted above that it is impossible to verify but that it still can be falsified by evidence for linguistic constraints on the form of fragments that cannot be reduced to being the result of game-theoretic reasoning, as Bergen & Goodman (2015) suggest.

In the next chapter, I present a series of experiments that test some of the predictions of the competing theories of fragments which I discussed in this section. Currently there is no consensus on the appropriate syntactic analysis of fragments, and there has been no systematic and empirical investigation of the partially contrary predictions of the theories. My experiments will to some extent fill this gap. The experiments address two main questions that allow us to differentiate between the theories presented in this chapter: First, I test whether fragments are underlyingly sentential. Since the results support a sentential account, I address the question of whether movement restrictions constrain the form of fragments at the case of preposition omission, complementizer omission and multiple prefield constituents in German. Furthermore, the results on the syntax of fragments inform my experiments on their usage in Chapter 5.



### 3 Experiments on the syntax of fragments

In this chapter I present 11 experiments that empirically evaluate the theories of fragments which I introduced in the preceding chapter.<sup>1</sup> The experiments address two main research questions: First, the experiments 1–3 investigate whether fragments are underlyingly sentential or whether they are genuine nonsententials. Since the experiments provide evidence for unarticulated structure in fragments, the experiments 4–10 test whether the generation of fragments obligatorily involves movement or whether fragments are the result of in situ deletion.

The experiments contribute to the theoretical analysis of fragments, by providing further empirical data on the competing theories' predictions. The experiments on default case and multiple prefield constituents are the first attempts to empirically investigate theoretical assumptions that are solely based on introspective data so far and the studies on preposition omission and complement clause topicalization extend previous experimental research. Additionally, these experiments will settle the ground for the experiments on the usage of fragments in Chapter 5, which requires to determine *which* fragments are grammatical. Since the account of fragment usage that I propose in Chapter 4 presupposes that speakers choose between grammatical utterances, so its empirical predictions would be distorted by the inclusion of ungrammatical fragments in the set of utterances among which a speaker can choose. For instance, if fragments were subject to movement restrictions, as Merchant (2004a) argues, only those expressions that can appear in a left-peripheral position would be possible fragments.

This chapter is structured as follows. Section 3.1 presents the experiments that test whether fragments are sentential (experiments 1–3). I use structural case marking on DP fragments as a testing ground for this, because Barton & Progovac (2005) argue that DP fragments cannot appear in structural case. The experiments support a sentential account, since they show that, provided an appropriate context, structural case is preferred over default case in fragments.

---

<sup>1</sup>Experiments 1, 4, 5, 8 and 9 have been published in Lemke (2017). The statistical analyses differ from those reported here, but the conclusions drawn from the data remain the same.

### 3 *Experiments on the syntax of fragments*

Sections 3.2–3.4 present the experiments on three movement restrictions: Preposition omission (experiments 4–7), complement clause topicalization (experiments 8, 9) and multiple prefield configurations in German (experiment 10). In Section 3.2, I present four experiments that investigate whether restrictions on preposition stranding constitute evidence for movement in fragments. The first two experiments support the pattern predicted by Merchant (2004a) for English and German, and the latter experiments investigate potential non-movement accounts of this pattern: Experiment 6 investigates a case checking-based approach to preposition omission in fragments by Progovic et al. (2006), which is disconfirmed. Experiment 7, however, suggests that, based on English data, a nonsyntactic relationship between question and answer can explain the data without necessarily assuming movement. Section 3.3 addresses complement clause topicalization with two experiments. In part, these studies replicate the effect reported by Merchant et al. (2013) for fragments, but crucially not for the corresponding left dislocation structures. The data hence cannot be interpreted as evidence for movement. Finally, in Section 3.4 I test the predictions of the movement and deletion account on fragments that are derived from ungrammatical multiple prefield configurations in German (experiment 10). Again, the experiment does not reveal the pattern predicted by Merchant's (2004a) theory. Section 3.5 summarizes the main results of the experiments.

#### 3.1 **Case marking as evidence for sententiality**

This section presents three experiments that test these predictions of sentential and nonsentential accounts with respect to structural case marking. As I discussed in Section 2.4.1, all sentential accounts predict case connectivity effects: DP fragments always receive the same case morphology as the corresponding DP within a full sentence, because the structure of the full sentence from which they are derived determines their case. In contrast, the nonsentential account by Barton & Progovic (2005) predicts that fragments may not receive structural case marking: Unlike default or inherent case, structural case needs to be checked by a verbal or functional head, which they argue is not present in DP fragments. Instead, they claim that DP fragments that exhibit structural case in full sentences appear in default case in fragments. The pattern is exemplified in (1). Barton & Progovic (2005) predict such anticonnectivity effects only for structural case-marked DPs, because inherent case, e.g. dative or genitive, is interpretable and therefore does not require feature checking.

### 3.1 Case marking as evidence for sententiality

Table 3.1: German inflectional case paradigm for masculine, feminine and neuter definite DPs (*the man, the woman, the book*).

| Case | masc, sg.  | fem, sg. | neut, sg.    | plural                   |
|------|------------|----------|--------------|--------------------------|
| Nom  | der Mann   | die Frau | das Buch     | die Männer/Frauen/Bücher |
| Gen  | des Mannes | der Frau | des Buch(e)s | der Männer/Frauen/Bücher |
| Dat  | dem Mann   | der Frau | dem Buch     | den Männer/Frauen/Bücher |
| Acc  | den Mann   | die Frau | das Buch     | die Männer/Frauen/Bücher |

- (1) Who can eat another piece of cake? (Barton & Progovac 2005: 77)
- a. ?\*I/?\*We/?\*He/?\*She
- b. Me/Us/Him/Her

The case of pronouns is more complex than Barton & Progovac (2005) suggest. The crosslinguistic data on the contrast between strong, weak and tonic pronouns in Merchant (2004a) shows that not only case, but also information structure determines which form of a pronoun is selected. What is more, English has only reduced morphological case marking, therefore it is not the ideal language for testing differences between default, inherent and structural case. For this reason, in experiments 1–3 I investigate the phenomenon in German, a language that has a richer case system than English. In German there are four cases, whose morphological realization on definite DPs is exemplified in Table 3.1.

Case marking occurs most systematically on the article. There are some syncretic forms for plurals as well as for feminine and neuter singular DPs, but for some masculine singular DPs the determiner disambiguates fully between the four cases.<sup>2</sup> Dative and genitive are inherent, because they are related to specific thematic roles or selected by specific lexical items.<sup>3</sup> Given the discussion in the preceding section, accusative must be analyzed as structural case, which marks the direct object of a verb. Nominative marks the syntactic subject and is the default case in German (if such a concept is assumed at all).

<sup>2</sup>For other masculine singular DPs only nominative is morphologically distinct from the remaining cases, as (i) shows. See e.g. Eisenberg (1999: 139–141) for details.

(i) der Student / des Studenten / dem Studenten / den  
the.NOM student.NOM / the.GEN student.GEN / the.DAT student.DAT / the.ACC  
Studenten  
student.ACC

<sup>3</sup>See also Woolford (2006), who distinguishes between lexical case, that is selected by a lexical item and inherent case, which encodes a specific  $\theta$ -role.

### 3.1.1 Experiment 1: Default case, acceptability rating study

#### 3.1.1.1 Background

In experiment 1, I investigate whether structural case connectivity effects occur in German. I test this with an acceptability rating task that compares nominative (2a) and accusative (2b) DP fragments. The nonsentential account by Barton & Progovac (2005) predicts that DP fragments cannot appear in structural case so that accusative DP fragments would be degraded as compared to nominative default case. In contrast, sentential accounts predict case connectivity: If the DP has accusative case marking in the full sentence from which it is derived, accusative is preferred over nominative.

- (2) Jenny and David want to drive to the beach today. While David is packing the picnic basket, he says to Jenny:
- |    |     |                   |            |
|----|-----|-------------------|------------|
| a. | Der | Sonnenschirm!     | Nominative |
|    |     | the.NOM sun.shade |            |
| b. | Den | Sonnenschirm!     | Accusative |
|    |     | the.ACC sun.shade |            |
|    |     | ‘The sunshade!’   |            |

The strength of connectivity effects expected under sentential accounts depends on the QuD or context that is accommodated. For instance, in (2), one could in principle assume either of the structures in (3) as underlying the fragment, but accusative is only licensed in the case of (3a). Therefore, if it was more natural to choose (3b) than (3a), sentential accounts would also predict a preference for nominative, which however would be explained by connectivity effects and not by default case morphology. In order to control the availability of structural case in a complete sentence, the fragments were preceded by a context story that made such a sentence salient. A pretest ensures that e.g. (3a) is accessible in this context. Furthermore, experiment 2 shows that accusative is also more likely to be used in a production task.

- (3)
- |    |                                                    |                                |            |
|----|----------------------------------------------------|--------------------------------|------------|
| a. | Wir müssen noch den                                | Sonnenschirm einpacken!        | Accusative |
|    | we must still the.ACC sun.shade                    | pack.in                        |            |
|    | ‘We still have to load the sunshade (in the car).’ |                                |            |
| b. | Der                                                | Sonnenschirm ist noch nicht im | Nominative |
|    | the.NOM sun.shade                                  | is yet not in.the car          |            |
|    | ‘The sunshade is not yet in the car!’              |                                |            |

### 3.1 Case marking as evidence for sententiality

Table 3.2: Predictions of the nonsentential and sentential accounts on the acceptability of case-marking in fragments.

|                       | Nominative | Accusative |
|-----------------------|------------|------------|
| Nonsentential account | ✓          | ✗          |
| Sentential account    | (✓)        | ✓          |
| Ungrammatical         | (✓)        | ✓          |

However, even if (3a) was more accessible than (3b), sentential accounts do not necessarily predict that nominative is less acceptable than accusative in an acceptability rating task. Under the assumption of case connectivity the hearer must retrieve an antecedent that requires nominative case marking after processing a nominative case-marked fragment. In the event he is able to retrieve such an antecedent, nominative might be perceived as acceptable as well. Table 3.2 summarizes the predictions of the two (families of) theories. Contexts in which a sentence requiring accusative case marking is accessible allow us to distinguish between the predictions of both families of accounts: The nonsentential account predicts a strong preference for nominative, but if a sentential account is correct, accusative must be at least as acceptable as nominative.

#### 3.1.1.2 Materials

All materials follow the pattern in (4) and (5), that is, they consist of a DP fragment preceded by a context story. The context story introduces two characters and a situation, at the end of which one of the two characters utters the fragment. The story ensures that a full sentence that requires accusative case marking is accessible, as was confirmed by a pretest (see Section 3.1.1.3 below for details). All fragments are masculine singular DPs and appear in one of two CASE conditions (accusative/nominative). The restriction to masculine singular nouns excludes case-syncretic forms. Whenever this does not reduce naturalness, the DPs contain an adjective, so that case marking appears twice and is more salient, as in (5). This is particularly important in the case of the accusative indefinite article *einen*, which is often pronounced as *ein* in colloquial speech. I used only discourse-initial fragments because, as discussed in the introduction (Section 1.4), some authors (e.g. Klein 1993, Reich 2011) distinguish short answers, which have a linguistic antecedent, from discourse-initial fragments. Discourse-initial fragments definitely lack an overt linguistic antecedent, hence they are the most uncontroversial instances of fragments.





case, the experiment would not allow for a comparison of the theories' predictions. Therefore, I conducted a pretest in order to select 20 stimuli for the main experiment that made a sentence accessible that requires accusative.

I constructed 40 items following the pattern in (2): Two context sentences introduced two characters and were followed by a target utterance attributed to one of these characters which seemed intuitively likely to be produced in this situation. In the pretest, the target utterance was always a complete sentence containing a transitive matrix verb and an accusative case-marked DP in its post-verbal base position (6). This DP was equivalent to the DP fragment in the main experiment. Subjects were asked to rate the naturalness of the target utterance in the context of the story. In order to present not only (probably) accessible utterances throughout the pretest, a second context story was constructed for each of the target utterances, for which the target utterance was intuitively less accessible, yet not implausible. This yielded an additional unpredictable condition for which I expected worse ratings than for the predictive one. The context story for the unpredictable condition of the sample item (2) is given in (7).

- (6) Wir müssen noch den      Sonnenschirm einpacken!  
we must    still the.ACC sun.shade    pack.in  
'We still have to load the sunshade (in the car).'
- (7) Jenny and David want to spend a relaxing day in their garden.  
As Jenny walks outside with her book, David says:

Twenty-nine voluntary undergraduate students of Saarland University participated in the pretest, which was conducted over the Internet using the LimeSurvey questionnaire software (LimeSurvey GmbH 2012). Subjects were asked to rate the naturalness of the italicized target utterance in the context of the preceding story on a 7-point Likert scale (7 = completely natural). Materials were distributed across two lists so that each subject saw each token set once and only in one condition and each condition equally often. Each subject rated 40 items (20 in the predictive condition and 20 in the unpredictable one), which were mixed with 65 fillers and presented in individual fully randomized order. The fillers resembled the items in consisting of a two-sentence context story and a full sentence uttered by one of the two characters introduced in that story. Five of the fillers included utterances that were grammatically well-formed and not fully implausible, but intuitively unlikely in the described situation. Two participants who rated more than the previously established threshold of 50% of these controls with 6 or 7 points were excluded from further analysis. Since the purpose of the pretest was to establish how accessible an utterance is in context, this ensured that only subjects whose ratings reflected this entered the analysis.

### 3 Experiments on the syntax of fragments

Across all items, utterances were rated as more natural in the predictive condition ( $\mu = 5.43, \sigma = 1.8$ ) than in the unpredictable one ( $\mu = 3.63, \sigma = 2.13$ ). Except for one token set, the target utterance was always rated as more natural in the predictive condition than in the corresponding unpredictable one. An analysis with cumulative link mixed models (in what follows, CLMMs) using the ordinal package (Christensen 2015) in R (R Core Team 2019)<sup>4</sup> reveals a significant main effect of PREDICTABILITY that evidences that target utterances are rated as significantly more natural in the predictive condition ( $\chi^2 = 40.37, p < 0.001$ ). This shows that the intended predictability manipulation was successful. Based on the aggregated rating data by items, the 20 items that received the highest ratings in the predictive condition were selected as materials for the main experiment. The selected materials had a mean rating of 6.03 (range 5.54–6.5), whereas the discarded ones had a mean rating of 4.83 (range 3.0–5.46).

#### 3.1.1.4 Procedure (Main experiment)

Seventy undergraduate students of Saarland University, Potsdam University and Stuttgart University<sup>5</sup> participated in the experiment, which was conducted on the Internet via LimeSurvey. They were compensated with a lottery of 10 times €30 among all participants. Subjects were asked to rate the naturalness of the target sentence, which was highlighted by italic font, on a 7-point Likert scale with labeled extremes (1 = very unnatural, 7 = very natural). Subjects were randomly assigned to one of four lists<sup>6</sup> so that each subject saw each token set once and only in one condition. Each subject rated 20 items (10 per CASE condition), which were presented together with 20 items from experiment 4 and 47 fillers (including nine ungrammatical controls) in individually fully randomized order. Fillers consisted of short stories or dialogues which contained direct speech by at least one of the characters in the story in order to resemble the materials. The target utterance was always a fragment. Among the fillers there were nine ungrammatical controls, which contained e.g. agreement violations or wrong verb inflections. No subject rated more than 50% of the controls as acceptable (6 or 7 points on the scale), so that nobody was excluded from further analysis.

<sup>4</sup>The analysis followed the procedure described for the main experiment in Section 3.1.1.5.

<sup>5</sup>The reason for testing subjects from universities outside Saarbrücken was that dialects spoken in Saarbrücken and the surrounding areas exhibit case syncretism between accusative and nominative and could therefore be insensitive to this distinction. However, the statistical analysis showed that the behavior of subjects from the Saarbrücken region did not differ significantly from that of the subjects from regions without case syncretism.

<sup>6</sup>As experiment 1 has two conditions, each two of the lists were equal with respect to the items from this experiment, but they differed in the items from experiment 4 that were included.

#### 3.1.1.5 Data analysis

I analyzed the data with cumulative link mixed models (CLMMs) (Christensen 2015) in R (R Core Team 2019). CLMMs model the outcome of ordinal dependent variables and take into account the potentially differing distance between the scale items. Unlike linear models, they do not presuppose that the scale is unbound.<sup>7</sup> This is implemented by threshold parameters that quantify this distance between scale items. The ordinal package allows for the modeling of these thresholds as flexible (individual thresholds for each transition between two categories), symmetric (different transitions between scale extremes and mid-range) and equidistant (same distance between all categories). I always started with the most complex structure (flexible thresholds) and subsequently shifted to simpler thresholds whenever this did not significantly worsen model fit, as evidenced by likelihood ratio tests calculated with the anova function in R.<sup>8</sup>

In order to determine the most appropriate model for the data, I used a backward model selection procedure. I started with a full model, which contains all predictors and two-way interactions between them and subsequently excluded effects that do not significantly improve model fit, as evidenced by likelihood ratio tests calculated with the anova function. Following Barr et al. (2013), as long as models converged, I included the full random effects structure, i.e. by-subject and by-item random intercepts as well as by-subject and by-item random slopes for all predictors. *p*-values were calculated with likelihood ratio tests comparing the model fit of the final model to that of models without the specific predictor with the anova function in R. Besides explicitly stated differences, all statistical analysis reported in this work follow this procedure.

#### 3.1.1.6 Results

Figure 3.1 shows the mean acceptability ratings across the CASE conditions and the XPLEASE variation between token sets. On average, accusative fragments ( $\mu = 4.26, \sigma = 2.09$ ) were rated as more acceptable than nominative fragments ( $\mu =$

---

<sup>7</sup>Gibson et al. (2011: 28) note that linear regression can still be used for ordinal data, unless the ratings are close to the endpoints of the scale. In fact, for most of the experiments reported in this book, linear mixed effects regressions conducted with the lme4 package (Bates et al. 2015) yield comparable results.

<sup>8</sup>For most of the analyses reported in this book, the final model had symmetric thresholds, as model fit was significantly worse with equidistant ones, but not significantly improved by additional parameters required for flexible thresholds. This evidences that subjects did not perceive the scale as linear, but that the difference between scale levels in the mid-range was not identical to that closer to the extremes.

### 3 Experiments on the syntax of fragments

3.64,  $\sigma = 2.08$ ). Fragments that could be paraphrased by a XPLEASE construction described above ( $n = 4$ ) were rated as more acceptable ( $\mu = 5.16, \sigma = 1.77$ ) than those that could not ( $\mu = 3.65, \sigma = 2.07$ ). The mean ratings suggest that the preference for accusative is independent from the possible XPLEASE construction.

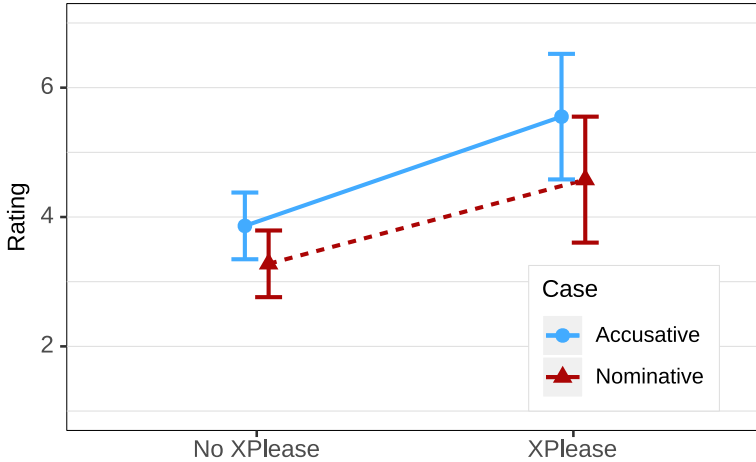


Figure 3.1: Mean ratings and 95% confidence intervals across conditions in experiment 1.

I analyzed the data with CLMMs as described in Section 3.1.1.5. The full model contained fixed effects for CASE (binary), XPLEASE (binary) and the POSITION of the trial in the time-course of the experiment (numeric). The model included by-subject and by-item random intercepts and by-subject random slopes for CASE, XPLEASE and their interaction. For items, I included only random slopes for CASE, as XPLEASE was not varied systematically across token sets. The only fixed effects in the final model were those for CASE and XPLEASE (see Table 3.3): The main effect for CASE shows that nominative fragments were rated significantly as worse than accusative fragments ( $\chi^2 = 7.39, p < 0.01$ ). The XPLEASE effect shows that fragments that could be paraphrased with an XPLEASE construction were significantly more acceptable than the others ( $\chi^2 = 7.3, p < 0.01$ ). There was no significant interaction between both predictors.

#### 3.1.1.7 Discussion

Experiment 1 investigated case connectivity effects in German accusative DP fragments, which would indicate unarticulated structure in fragments and hence provide evidence for a sentential theory of fragments. The data suggest that structural (accusative) case marking is possible in fragments. Even in the absence of

### 3.1 Case marking as evidence for sententiality

Table 3.3: Fixed effects in the final CLMM for experiment 1.

| Predictor | Estimate | SE    | $\chi^2$ | $p$   |    |
|-----------|----------|-------|----------|-------|----|
| CASE      | -0.828   | 0.286 | 7.39     | <0.01 | ** |
| XPLEASE   | 1.875    | 0.632 | 7.30     | <0.01 | ** |

an explicit linguistic antecedent, accusative DPs are rated at least as acceptable as nominative ones. In fact, the significant main effect of CASE shows that accusative is rated even better than nominative if a full sentence requiring accusative is accessible in context. This pattern is unexpected under the nonsentential account, according to which accusative is ungrammatical, but predicted by sentential accounts. From a sentential perspective, the acceptability of accusative is analyzed as a case connectivity effect: Accusative case marking on the DP is also required in the sentential alternative to the fragment, whose salience the pretest confirmed. If the data are to be explained by case connectivity, the lower ratings for nominative indicate that a sentence that requires nominative is unavailable, or at least less accessible in the case of my materials.

This assumption has not been tested in the pretest, though. Even if the full sentence that requires accusative was rated as perfectly natural in the pretest, this does not exclude the possibility that there is an equally accessible sentence that requires nominative. In such a situation, the sentential account would predict no difference in acceptability between accusative and nominative fragments for my materials: No matter in which case the fragment appears, an antecedent for ellipsis resolution is accessible. If this was the case, the preference for accusative in experiment 1 would still challenge the nonsentential account, but neither would be in line with the case connectivity explanation that would provide evidence for sentential accounts of fragments. I address this issue with experiment 2.

The XPLEASE predictor ensures that the overall preference for accusative is not driven by only a few influential data points that might result from a conventionalized usage of accusative fragments in contexts of ordering. In fact, the significant main effect of XPLEASE shows that these items are perceived as more acceptable than the remaining fragments. However, they were rated better in both CASE conditions, and the absence of an interaction between XPLEASE and CASE evidences that the preference for accusative is independent of this construction. Possibly, in the potential XPLEASE construction the QuD was easier to figure out or it is socially more appropriate to communicate with a fragment in such situations.

Table 3.4 summarizes different fragment theories' predictions on structural case-marking. If accusative is structural case in German, i.e. a purely linguistic

### 3 Experiments on the syntax of fragments

device marking a structural relationship between the verb and its complement, the preference for accusative clearly supports a sentential analysis. This holds even in the absence of explicit linguistic context. Consequently, it is not an option to claim that short answers are elliptical and thus exhibit connectivity effects, while discourse-initial fragments are genuine nonsententials and do not.

Table 3.4: Summary of the predictions of fragment theories on structural case marking.

|                       | Nominative | Accusative |
|-----------------------|------------|------------|
| Nonsentential account | ✓          | ✗          |
| Sentential account    | (✓)        | ✓          |
| Ungrammatical         | (✓)        | ✓          |
| Experiment 1          | ✗          | ✓          |

Of course, the conclusion that the data challenge the nonsentential account relies strongly on the theoretical distinction between structural, default and inherent case assumed by Barton & Progovac (2005): If accusative was analyzed as inherent case, the data would not conflict with the nonsentential account. However, in Section 2.4.1 I showed that the diagnostics presented by Progovac et al. (2006) as evidence that the Serbian accusative is indeed inherent case yield the opposite result for German.

So far, the data are in line with Bergen & Goodman's (2015) view that fragments are per se ungrammatical but can be used as long as it is relatively easy to retrieve the omitted material. Any cue that guides the hearer toward the intended meaning will be useful for this purpose, and accusative is definitely such a cue, because parsing the accusative DP rules out all possible meanings that require it to appear in a different case. Furthermore, as Progovac et al. (2006) note, accusative DPs will be relatively likely to be assigned the patient  $\vartheta$ -role. Furthermore, any case functions as a cue toward the associated  $\vartheta$ -role to some extent, as even the relatively unmarked nominative reduces the likelihood of the DP being a recipient (which usually receives dative case marking). However, this reasoning cannot reconcile the data with the nonsentential account by Barton & Progovac (2005), who emphasize the categorical distinction between default, inherent and structural case.

### 3.1.2 Experiment 2: Default case, production study

#### 3.1.2.1 Motivation

Experiment 2 tested whether a sentential alternative that requires accusative is indeed more salient in the context of the materials tested in experiment 1 than one that requires nominative. Only in that case, the preference for accusative in experiment 1 can be attributed to case connectivity and hence be interpreted as evidence for a sentential theory of fragments.

From a sentential perspective, the preference for accusative in experiment 1 is explained by the availability of a linguistic antecedent for ellipsis which contains a verbal node licensing accusative case marking on the fragment DP. The salience of such an antecedent is confirmed by the high naturalness of sentential alternatives to the DP fragments evidenced in the pretest. This line of reasoning, however, implies that a sentence that requires nominative on the DP is less accessible, because nominative was perceived as less acceptable in experiment 1. Since the pretest investigated only sentences containing accusative DPs, it is still possible that a sentence requiring nominative is equally likely and the acceptability ratings call for a different explanation. Experiment 2 addresses this issue with a production study using the same materials as experiment 1. In contrast to the rating study, subjects produced the target utterance themselves. I then quantified the preference for full sentences that required accusative or nominative case marking on the fragment based on the aggregated responses.

#### 3.1.2.2 Materials

The context stories of experiment 1 were used as stimuli in a production task. Instead of reading a fragment, as in experiment 1, subjects saw a hand-drawn image of the object referred to by the DP fragment (see Figure 3.2 for an example). Subjects were asked to read the story and to produce an utterance by the specified character that referred to the depicted object. The use of graphical stimuli avoided the problem that some (but not all) of the nouns in the DP fragments tested in experiment 1 morphologically distinguish accusative and nominative, so that a written presentation would have introduced a case bias.

#### 3.1.2.3 Procedure

The experiment was conducted over the Internet using the LimeSurvey survey presentation software and completed by 38 undergraduate students of Saarland University. They were rewarded with the participation in a lottery of  $5 \times \text{€}30$

### 3 Experiments on the syntax of fragments



Figure 3.2: Sample graphical stimulus used in experiment 2 (CC-BY 4.0 Julia Stark).

among all participants. Subjects read the context story, which was presented above a hand-drawn image that depicted the object referred to by the corresponding DP fragment in experiment 1. They were asked to enter an utterance that was likely to be said by a specified character and that referred to the depicted object into a text field. As there was only one condition, all subjects saw the same materials. The experiment was presented to the same subjects as the acceptability rating study for 10 and the follow-up to experiment 8. As the context stories contained no specifically biasing patterns that might prime subjects, there were no fillers, so that each subject produced 20 responses. The stimuli were presented in individually fully randomized order. The responses were manually annotated for the CASE of the DP referring to the image and the noun used in the DP by the subjects. It was also annotated whether the subjects used the same lemma as tested in experiment 1, a synonym, or whether they did not mention the lemma or a synonym at all. Responses that did not refer to the object depicted by the image (14.96% of the data) were excluded from further analysis.

#### 3.1.2.4 Results

There was a strong overall preference for accusative (88.78% of all trials) over nominative (5.03% of trials). The remaining 6.19% of trials used other constructions that contained a DP in dative or prepositional case.<sup>9</sup> The noun lemma pro-

<sup>9</sup>In this context, prepositional case was always dative or accusative. I distinguish prepositional case from other uses of dative and accusative because e.g. Zwarts (2005) shows that prepositional case is in part determined by the preposition and not related to structure (accusative) or a  $\theta$ -role (dative).



duced by the subjects was the same as tested in experiment 1 in 49.7% of the trials, a synonymous one in another 38.4% and a closely related one (e.g. *salad* instead of *pasta salad*) in 11.9% of the trials.

Table 3.5: Fixed effects in the final GLM for experiment 2.

| Predictor | Estimate | SE    | $\chi^2$ | $p$    |     |
|-----------|----------|-------|----------|--------|-----|
| Intercept | -2.657   | 0.245 | 37.58    | <0.001 | *** |

Since the purpose of the experiment was to investigate the relative likelihood of accusative and nominative when referring to the target DP and there were no predictors, I analyzed the data with an intercept-only logistic regression conducted with the `lme4` (Bates et al. 2015) package in R (R Core Team 2019). The intercept of such a model indicates the likelihood of an observation to fall into one of the two categories, so that a significant intercept will show that these categories are not equally likely. The model in Table 3.5 shows that this is the case ( $\chi^2 = 37.58, p < 0.001$ ).

### 3.1.2.5 Discussion

Experiment 2 investigated whether a full sentence that requires accusative case marking on the DP fragment is indeed more likely than a sentence requiring nominative. If this was the case, the pattern found in experiment 1 would indicate case connectivity and support a sentential theory of fragments. The production study shows that accusative was used significantly more often than nominative in order to name the object referred to by the DP fragment in experiment 1. This is expected if the preference for accusative or nominative in experiment 1 is due to case connectivity: Accusative is preferred, because a full sentence that requires accusative is the most salient antecedent for ellipsis that is made available through the context story. Taken together, experiments 1 and 2 provide clear evidence for unarticulated structure in fragments. First, fragments *can* exhibit structural case morphology, and second, they *do so* preferably when a fully sentential structure requiring structural case is salient in context.

## 3.1.3 Experiment 3: Mixed accounts?

### 3.1.3.1 Background

Experiments 1 and 2 show that in contexts where a DP appears in accusative (structural case) in full sentences in German, accusative DP fragments are also

preferred over nominative (default case) DP fragments. This finding is unexpected under the nonsentential account by Barton & Progovac (2005), but it indicates case connectivity effects, which sentential accounts predict.

This result does not imply that *all* fragments are generated by ellipsis. Barton (2006) raises the possibility of a *mixed* account of fragments, which derives some fragments by ellipsis, whereas others are genuine nonsententials. Such an account would be motivated by the observation that the properties of fragments cannot be captured by a single syntactic mechanism alone. For instance, there might be both instances of case connectivity and anticonnectivity, so that both the nonsentential and a sentential derivation would have to be assumed.

From a theoretical perspective, even that a mixed account might account for the relevant data, Occam's Razor requires us to adopt one of the simpler accounts, unless a mixed captures the empirical picture better. To my knowledge, few serious attempts have been made to work out a mixed account of fragments in detail. Barton (2006) cites Morgan (1989) and herself (Barton 1998) as mixed accounts, but notes considerable differences between both with respect to the scope attributed to the sentential and nonsentential generation mechanisms. According to Barton (2006), Morgan adopts the nonsentential analysis only for fragments that (arguably) cannot be explained by the sentential accounts, such as case-less Korean DPs, whereas Barton (1998) analyzes only those fragments that cannot be derived as genuine nonsententials as elliptical sentences.

No matter how many of the empirically observed fragments are attributed to either derivation mechanism, spelling out a mixed account necessarily involves explaining why the (non)sentential derivation is (not) available or (dis)preferred in a specific context. A straightforward implementation of a mixed account could assume that a trade-off between the effort required to find an antecedent that licenses ellipsis and a cost for pragmatic enrichment (Sperber & Wilson 1986, 1995, Breheny et al. 2006, Chevallier et al. 2008) determines whether ellipsis resolution or pragmatic enrichment is used in order to interpret a fragment.<sup>10</sup> If pragmatic inference is effortful, syntactic ellipsis resolution will be preferred when it is easy to retrieve an antecedent. The more difficult this retrieval becomes, the more promising might be the nonsentential derivation that requires pragmatic inference by the hearer. A mixed account based on this idea predicts case connectivity if there is a salient antecedent for ellipsis resolution and the generation of genuine nonsententials in case no salient antecedent is available.<sup>11</sup>

<sup>10</sup>This is obviously the perspective of the hearer. However, since I assume that the speaker performs audience design (Bell 1984), the speaker will adapt her utterance to expectations about the interpretive behavior of the hearer.

<sup>11</sup>The mixed account would obviously have to explain why a fragment is used at all if there is no salient antecedent that guides toward the meaning intended by the speaker.

### 3.1 Case marking as evidence for sententiality

In experiment 1 I investigated only fragments that have a salient antecedent. Since the mixed account predicts that such fragments are better resolved by ellipsis, it agrees with the sentential account in this case. However, if no salient antecedent is available, the mixed account predicts that fragments are generated as nonsententials. If genuine nonsentential utterances do not exhibit case connectivity, the predictions of sentential and nonsentential accounts diverge here: Sentential accounts predict strict case connectivity, and under the mixed account sketched above structural case should be unavailable.

Table 3.6: Summary of the predictions of the nonsentential, sentential and mixed account with respect to experiment 3.

|                       | Predictable |      | Unpredictable |      |
|-----------------------|-------------|------|---------------|------|
|                       | Nom.        | Acc. | Nom.          | Acc. |
| Nonsentential account | ✓           | ✗    | ✓             | ✗    |
| Sentential account    | (✓)         | ✓    | (✓)           | ✓    |
| Mixed account         | (✓)         | ✓    | ✓             | ✗    |

In experiment 3 I therefore compared the acceptability of accusative and nominative fragments in predictable and unpredictable contexts in a 2×2 design crossing PREDICTABILITY and CASE in an acceptability rating study. The predictions of the sentential, nonsentential and mixed account are summarized in Table 3.6. The nonsentential accounts again predicts that accusative is ungrammatical and that nominative default case is preferred. The sentential account predicts case connectivity: If there is a salient full sentence that requires accusative, accusative will be preferred. Finally, the mixed account matches the behavior of the nonsentential account for unpredictable fragments and that of the sentential account for predictable fragments. Fragments can be generated by ellipsis in predictive contexts, but are genuine nonsententials if no salient antecedent is available.

#### 3.1.3.2 Materials

The materials were derived from those used in experiment 11 (see Section 5.2). The predictability of fragments was manipulated with context stories based on event chains derived from the DeScript (Wanzare et al. 2016) corpus of script knowledge.<sup>12</sup> This ensures that the PREDICTABILITY manipulation is founded on empir-

<sup>12</sup>In experiment 11 some DPs were ambiguous between accusative and nominative case. I replaced these DPs by a semantically similar masculine singular noun that distinguishes accusative and nominative or chose a different event sequence from the same script.

### 3 Experiments on the syntax of fragments

ically observed event probabilities. In very simplified terms, the corpus allows for the estimation of the likelihood of an event in a script-based scenario given the previous events. For instance, a person who is cooking pasta will be likely to pour the pasta into the pot after the water is boiling. Under the assumption that likely events are more likely to be talked about than unlikely ones,<sup>13</sup> the likelihood of utterances in context can be quantified and manipulated based on the corpus data (see Section 5.1.2 for details). The stimuli consisted of short context stories of three sentences each and a fragment uttered by one of the characters introduced in the story (8). The fragments were DPs that referred to an event that was either predictable (8a,b) or unpredictable (8c,d) given the corpus data.

- (8) Today, Marie and Jonas want to cook themselves a large serving of pasta with tomato sauce. As soon as the water started to boil, Jonas has added the pasta. After ten minutes, he says to Marie:
- a. Der Topf mit den Nudeln. Predictable, nominative  
the.NOM pot with the.ACC pasta
  - b. Den Topf mit den Nudeln. Predictable, accusative  
the.ACC pot with the.ACC pasta  
'The pot with the pasta.'
  - c. Der Küchentisch. Unpredictable, nominative  
the.NOM kitchen.table
  - d. Den Küchentisch. Unpredictable, accusative  
the.ACC kitchen.table  
'The kitchen table.'

In both cases, in the corresponding full sentence (9) with default word order the fragment appears in a post-verbal position and exhibits accusative case morphology. In the predictable condition, the underlying sentence (9a) refers to the most likely event to follow the sequence of events underlying the context story, which is *remove the pot from the stove* in the example. The context story refers to the events *water starts to boil*, *add pasta to the water* and *cook pasta*. In the unpredictable condition, the event that the target utterance (9b) referred to (*set table*) was not mentioned in the corpus data, but it is intuitively not fully implausible.

- (9) a. Nimm den Topf mit den Nudeln vom Herd.  
take.IMP the.ACC pot with the.ACC pasta off-the.DAT stove  
'Take the pot with the pasta off the stove.'

<sup>13</sup>See Sections 5.2.5 and 5.3.1 for a discussion.

- b. Deck doch schonmal den Küchentisch.  
 set.IMP PRT already the.ACC kitchen.table  
 ‘Set the kitchen table, please.’

In principle, it would have been desirable to keep the target utterance as constant as possible across conditions and to vary only those properties that are related to the variables that are investigated by presenting a single fragment in an unpredictable and a predictive context. However, if an unpredictable context for the fragment in (8a) that is not based on the corpus data had been constructed from scratch, it would have been impossible to quantify the likelihood of the *remove pot* event in the same fashion as in the predictable condition. This likelihood could also be measured in a norming study, but this would have required considerably more effort than relying on the corpus-based probabilities that were available from experiment 11. Alternatively, the fragment could have been presented with a different corpus-based context story for which it is known that the corresponding event does not happen (e.g. describing the train ride scenario) in the unpredictable condition. In this case though, the target utterance would not only be unlikely, but actually implausible and therefore probably highly degraded in any CASE condition.

#### 3.1.3.3 Procedure

The experiment, which was conducted on the Internet via LimeSurvey, was completed by 47 native speakers of German, who were recruited through the crowdsourcing platform *clickworker*. Subjects were paid €4 for participating in the study. They were asked to rate the naturalness of the italicized target utterance in the context of the story on a 7-point Likert scale (7 = fully natural). Subjects were randomly assigned to one of four lists, to which the materials were distributed according to a 2×2 Latin square design. Each subject saw each token set once and only in one condition. Each subject rated 24 items (six per condition), which were mixed with 16 items from an unrelated experiment and 44 fillers. Materials were presented in a pseudorandomized order that ensured that no two items of the same experiment followed each other. Fillers consisted of context stories followed by an utterance by one of the characters or a question-answer pair. In the latter case, subjects rated the answer, which was presented in italicized font. Six fillers with ungrammatical word order served as attention checks. Three subjects who rated 50% or more of the attention checks as acceptable (6 or 7 points on the scale) were excluded from further analysis.

### 3 Experiments on the syntax of fragments

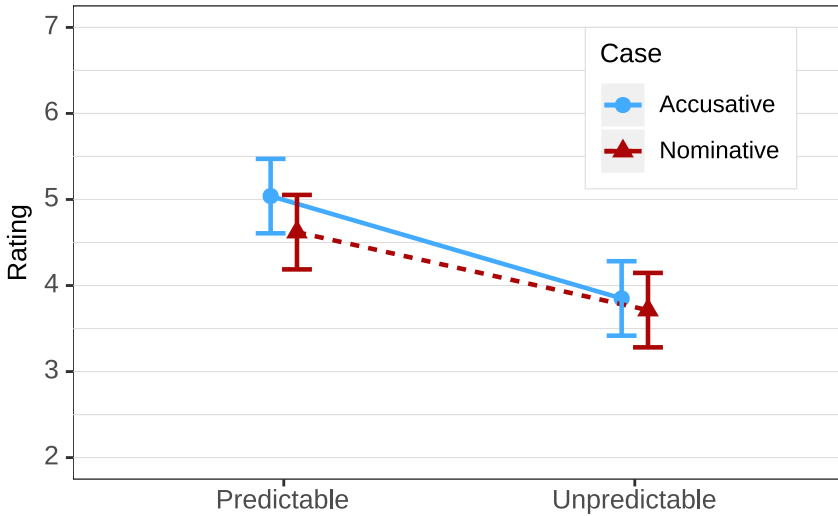


Figure 3.3: Mean ratings and 95% confidence intervals across conditions in experiment 3.

#### 3.1.3.4 Results

Figure 3.3 shows the averaged ratings across conditions. The data were analyzed with CLMMs following the procedure described in Section 3.1.1.5. The full model contained main effects for CASE, PREDICTABILITY and the POSITION of the item in the time-course of the experiment, as well as by-subject and by-item random intercepts and slopes for CASE, PREDICTABILITY and their interaction. Random effects of POSITION and its interaction with the other predictors were removed because models did not converge otherwise. The final model (see Table 3.7) contained only a significant main effect for PREDICTABILITY ( $\chi^2 = 10.04$ ,  $p < 0.01$ ). Neither the main effect of CASE ( $\chi^2 = 2.5$ ,  $p > 0.1$ ) nor its interaction with PREDICTABILITY ( $\chi^2 = 1.78$ ,  $p > 0.1$ ) were significant.

Table 3.7: Fixed effects in the final CLMM for experiment 3.

| Predictor      | Estimate | SE   | $\chi^2$ | $p$   |    |
|----------------|----------|------|----------|-------|----|
| PREDICTABILITY | -1.25    | 0.35 | 10.04    | <0.01 | ** |

#### 3.1.3.5 Discussion

The purpose of experiment 3 was to test a mixed account of fragments. Mixed accounts assume that speakers use two different mechanisms to interpret (and

produce) fragments, which differ in their predictions with respect to case marking: If fragments are generated by ellipsis from full sentences, they can exhibit structural case marking, but if they are base-generated as genuine nonsententials, they cannot. It is a natural assumption that speakers resort to ellipsis when there is a salient antecedent that allows resolution and to pragmatic inference when there is no such antecedent. Consequently, the mixed account predicts that nominative is rated as better than accusative in the unpredictable condition, where it is more difficult to retrieve an antecedent for ellipsis. The absence of a significant interaction between CASE and PREDICTABILITY shows that this prediction is not borne out. Unpredictable fragments are significantly worse than predictable fragments, but this holds independently of case.

In contrast to experiment 1, there was no significant main effect of CASE. This might be due to a reduced accessibility of antecedents that require nominative in the materials for experiment 3 as compared to those used in experiment 1. Whether this is correct could be tested with a further production study. However, this goes beyond the scope of this experiment, which was designed to test whether nominative is relatively more acceptable in unpredictable fragments.

Taken together, experiment 3 provides further evidence against the nonsentential account. Unlike what the nonsentential account predicts, fragments are not rated better in nominative than in accusative. The pattern observed in experiment 3 is unexpected both under the nonsentential and the mixed account.

#### 3.1.4 General discussion: Structural case marking

I presented three experiments that investigated whether fragments are derived from regular sentences by ellipsis or whether they are genuine nonsententials. The experiments used structural case marking on discourse-initial DP fragments as a diagnostic for unarticulated structure. In minimalism, the framework that underlies both the nonsentential and most of the sentential accounts discussed in Chapter 2, structural case must be checked by a verbal head. If, as nonsententialists argue, there is no unarticulated structure in fragments, structural case marking must be unavailable in fragments. In contrast, according to sentential accounts there is an unarticulated verbal head in DP fragments, so that they are expected to appear in structural case whenever they do in the corresponding full sentence. Experiments 1–3 investigated this using the example of German, where accusative is a structural case and nominative the default case, if such a concept is to be assumed at all. Experiment 1 showed that if there is a salient antecedent that licenses accusative, DP fragments can appear in accusative case. In fact, accusative was rated even better than nominative. This disconfirms the prediction of the nonsentential account and in turn supports sentential accounts. Experiment

### 3 *Experiments on the syntax of fragments*

2 showed that in the context of my materials a full sentence requiring accusative on the DP was indeed more likely than one requiring nominative. This finding further strengthens the interpretation of experiment 1 as evidence for connectivity effects. Finally, experiment 3 explored the predictions of a mixed account that assumes that fragment generation by ellipsis is possible, but restricted to contexts where an antecedent is available. Therefore, I tested whether nominative is more acceptable when the corresponding sentence is relatively unpredictable. Since there was no such effect, the data speak against the mixed account that I sketched.

It must be emphasized that this interpretation of the data presupposes that the categorical difference between structural and inherent case assumed by Barton & Progovac (2005) is correct. This distinction is crucial to the nonsentential account, since Barton & Progovac (2005) rely on it to explain crosslinguistic differences between English and Serbian as well as acceptability contrasts within English. Nonsentential accounts that do not rely on the inherent/structural case distinction as heavily as Barton & Progovac (2005) do, like analyses in the simpler syntax framework (Culicover & Jackendoff 2005) or in HPSG (Ginzburg & Sag 2000, Fernández & Ginzburg 2002) might still be able to explain the data. The data are also in line with the idea that fragments are ungrammatical, but can point toward the meaning intended by a speaker, as suggested by Bergen & Goodman (2015). Since accusative case in fragments might indicate that the DP fragment is to be interpreted as patient or object, a preference for accusative is expected under this account, if the DP fragment is the object of a transitive verb in a corresponding complete sentence.

Furthermore, the predictions of the nonsentential account also rely on the classification of a particular case as structural or inherent. Barton & Progovac (2005) argue that nominative is structural case in English, but default case in Serbian, so that nominative DP fragments are predicted to be ungrammatical in English, but not in Serbian. The conclusions that I draw from my data rely on the assumption that accusative is indeed structural case in German. Progovac et al. (2006) argued that accusative is inherent, semantically interpretable, case in German, and if this is correct, the nonsentential account would correctly predict German accusative DP fragments to be acceptable. I argued above that, given the tests used by Progovac et al. (2006) for Serbian, accusative must be classified as structural case in German, so that the nonsentential account makes incorrect predictions.

Even though the experiments did not explicitly test the predictions of a construction-based account, they suggest that the acceptability of accusative case cannot be explained by the assumption of a conventionalized construction like



e.g.  $\langle DP \rangle_{Acc}$ : Both in experiments 1 and 3, the acceptability of accusative does not vary as a function of the predictability or degree of conventionalization of the fragment. Of course this does not rule out the possibility that some fragments are stored as constructions in the mental lexicon, but this cannot explain the acceptability of accusative DP fragments in my experiments.

Taken together, the experiments presented in this section support a sentential analysis of fragments against the nonsentential account by Barton & Progovac (2005). They furthermore empirically confirm two properties of fragments that have to be taken into account in the investigation of fragment usage: Fragments exhibit case connectivity and that they can appear in structural case.

Having established that fragments are underlyingly sentential, this leads to my second main research question on the syntax of fragments, that is, whether the generation of fragments involves movement to the left periphery (Merchant 2004a) or whether fragments result from in situ deletion (Reich 2007).

### 3.2 Movement restrictions: Preposition omission

The experiments in Section 3.1 provided evidence for unarticulated syntactic structure in fragments, in particular for a silent verbal head that checks accusative structural case in DP fragments. In the remainder of this chapter I investigate which kind of structure underlies fragments. The main controversy with respect to this question is whether the derivation of fragments involves obligatory movement to the left periphery. In Sections 2.2 and 2.4.4 I outlined the positions: On the one hand, there is Merchant's (2004a) influential movement and deletion account, which has been more recently adopted and partially modified by other researchers (see e.g. Aelbrecht 2009, Sato 2011, Weir 2014a, Döring 2016, Saab & Lipták 2016, Murphy 2018) and assumes that the derivation of fragments involves obligatory movement to the left periphery. On the other hand, in situ deletion accounts (Reich 2007, Ott & Struckmeier 2016, Griffiths et al. 2018) derive fragments from regular sentences without assuming this obligatory movement step.

Since the in situ deletion approach does not require an additional movement operation in fragments, it is derivationally simpler. Therefore, the movement and deletion account requires additional evidence for movement in fragments. Such evidence consists in effects of movement restrictions on the form of fragments, which are then taken to indicate that left dislocation is a necessary step in the generation of fragments. Merchant (2004a) discusses a series of such similarities between fragments and dislocation constructions, but even the brief discussion in Section 2.4.4 showed that not all of these parallelisms constitute genuine

evidence for movement, because some of them can be explained under a non-movement account as well. Since the in situ deletion account is derivationally less complex than movement and deletion, I assume that it is the null hypothesis.<sup>14</sup> Consequently, the movement and deletion account is only superior to in situ deletion if there are parallelisms between fragments and left dislocation that cannot be explained by the in situ deletion account. In my experiments I focus on three movement restrictions that have been claimed to or that might constrain the form of fragments. This section investigates restrictions on preposition stranding and omission in German, in Section 3.3 I address complement clause topicalization and in Section 3.4 multiple prefield constituents in German.

#### 3.2.1 Preposition omission as evidence for movement

##### 3.2.1.1 The P-Stranding Generalization

Among the movement restrictions that Merchant (2004a) presents as evidence for his theory, the most compelling one is his *P-Stranding Generalization* (PSG, in what follows). In a nutshell, the PSG states that only in those languages that allow for P-stranding under regular *wh*-movement it is possible to omit prepositions under sluicing and in fragments. Merchant takes this as evidence that P-stranding is a necessary step in the derivation of preposition-less answers to questions where the *wh*-phrase corresponding to the answer is embedded within a PP. English is a typical example of a P-stranding language (10).

- (10) Who was Peter talking with? (Merchant 2004a: 685)  
Mary.

Originally, Merchant (2001) introduced the PSG as evidence for his movement and deletion account of sluicing. He presents introspective data from a relatively large and typologically diverse sample of languages<sup>15</sup> that overall show the predicted pattern. The contrast between languages that allow for P-stranding and those with obligatorily pied-piping is exemplified in (11) and (12) for English and German, respectively: English allows for P-stranding in *wh*-questions (11a) and preposition omission under sluicing (11b), whereas in German the preposition is always pied-piped in *wh*-questions (12a) and realized in sluices (12b).

<sup>14</sup> Actually the nonsentential account is even simpler, but I rejected it in the preceding section because it does not capture the empirical picture correctly.

<sup>15</sup> The data are from Arabic, Basque, Czech, Danish, English, Frisian, German, Greek, Hebrew, Icelandic, Irish, Norwegian, Polish, Russian, Serbo-Croatian, Slovene, Swedish and Yiddish.

- (11) a. Who was he talking with? (Merchant 2004a: 666)  
b. Peter was talking with someone, but I don't know (with) who.
- (12) a. \*Wem hat sie mit gesprochen? (Merchant 2004a: 667)  
Who.DAT has she with talked  
'With whom was she talking?'  
b. Anna hat mit jemandem gesprochen, aber ich weiß nicht,  
Anna has with somebody.DAT spoken but I know not  
\*(mit) wem.  
with who.DAT  
'Anna was talking with somebody, but I don't know (with) who.'

The reasoning behind the PSG is that, if sluicing is generated by regular *wh*-movement as discussed above, the preposition cannot be omitted in German, because this would require P-stranding to occur during the derivation (13). As P-stranding is available in English (11a), but not in German (12a), the corresponding sluice (11b)/(12b) is grammatical in English, but not in German.<sup>16</sup>

- (13) Peter was talking with someone, but I don't know who<sub>i</sub> ~~Peter was talking with  $t_i$~~ .

In his 2004a article, Merchant extends this analysis to fragments. He argues that only languages that have P-stranding allow for the omission of the preposition in short answers to questions whose *wh*-phrase is the complement of a preposition. Again, the observation is supported by crosslinguistic data.<sup>17</sup> The contrast is exemplified here for English (14) and German (15).

- (14) Who was Peter talking with? (Merchant 2004a: 685)  
(With) Mary.

<sup>16</sup>In German, there are a few constructions which are similar to P-stranding. In (i), the particle *her* contained in the complex *wh*-phrase *woher* can be stranded, but in contrast to the PPs providing evidence for the PSG, the moved element appears to the left of *her* if no extraction occurs. See e.g. van Riemsdijk (1978) for an analysis of such extractions in Dutch.

- (i) a. Wo hast du das Buch denn her?  
where have you the book PRT *her*  
b. Woher hast du das Buch denn?  
where.from have you the book PRT *her*  
'Where did you get the book from?'

<sup>17</sup>The data are from Bulgarian, Czech, Danish, English, German, Greek, Hebrew, Icelandic, Norwegian, Russian, Swedish and Yiddish (Merchant 2004a: 685–687).

- (15) Mit wem hat Anna gesprochen? (Merchant 2004a: 686)  
 with whom has Anna spoken?  
 ‘Who was Anna talking to?’
- a. Mit dem Hans.  
 with the.DAT Hans  
 ‘With Hans.’
- b. \*Dem Hans.  
 the.DAT Hans  
 ‘Hans.’

Merchant’s account of this pattern is parallel to his analysis of sluicing. He assigns the underlying structures in (16) to the fragments in (14) and (15b). Since he derives fragments by movement to the left periphery, the derivation of the DP short answer requires P-stranding, which is available in English (16a), but not in German (16b). Note that fronting a non-contrastive object (16a) is marked in English, so that a movement-based account probably has to assume exceptional movement (Weir 2014a): The acceptability of the fragment does not pattern with the degradedness of the corresponding left dislocation.

- (16) a. Mary<sub>i</sub> Peter talked to t<sub>i</sub>.  
 b. \*[Dem Hans]<sub>i</sub> hat Anna mit t<sub>i</sub> gesprochen.

Before going into further detail, it must be noted that the crosslinguistic correlation between the availability of P-stranding and the acceptability of preposition omission under ellipsis is not perfect: There are languages which seem to allow for preposition omissions even though they lack P-stranding and vice versa. However, for these data to constitute evidence against the PSG, it is necessary to rule out alternative sources for these fragments and sluices, which do not require P-stranding. For instance, Rodrigues et al. (2009: 176) show that in Spanish preposition omission under sluicing can be acceptable (17a) even though P-stranding in questions is not (17b).

- (17) a. Juan ha hablado con una chica pero no sé cuál.  
 Juan has talked with a girl but not know which  
 ‘Juan has talked with a girl but I don’t know which one’
- b. \*¿Qué chica ha hablado Juan con?  
 what girl has talked Juan with  
 ‘Which girl did John talk to?’

Since (17b) suggests that extraction out of PPs is ungrammatical, the movement account cannot derive the grammatical sluice *cuál* ‘which one’ by extraction out

of a PP *con cuál*. Rodrigues et al. (2009: 178) argue that (17a) can be derived from a cleft structure like (18), which does not require the ungrammatical extraction out of a PP. Similar accounts have been proposed by Szczegielniak (2008) for Polish, van Craenenbroeck (2010) for English and Sato (2011) for Indonesian. More recently, it has been controversially debated whether *all* of the data conflicting with the PSG can be explained by the cleft hypothesis.<sup>18</sup>

- (18) Juan ha hablado con una chica pero no sé cuál es la chica con  
 Juan has talked with a girl but not know which is the girl with  
 la que ha hablado Juan.  
 the that has talked Juan  
 ‘Juan has talked with a girl but I don’t know which one was the girl with  
 whom Juan talked.’

For German, Lemke (forthcoming) shows that preposition omission in German are rated as more acceptable when the DP is a proper noun, which is not headed by an overt article (19) than when the DP has an article, like (15).

- (19) ?(Mit) Hans.  
 with Hans  
 ‘(With) Hans.’

Given the discussion in the literature it is at least questionable whether the derivation from clefts is a crosslinguistically plausible explanation for the empirically observed apparent exceptions to the PSG. Furthermore, a movement-based theory of fragments that assumes that their derivation is possible both from regular leftward movement and clefts must explain why and when speakers pursue each of these strategies and hence produce the corresponding fragment. Providing an answer to these questions is beyond the scope of this book, since they presuppose a movement-based account, which I do not adopt *a priori*. However, the possibility that preposition-less fragments or sluices have been derived from clefts must be obviously taken into account. Following the line of reasoning that I pursue in order to investigate unarticulated structure, fragments that are derived by ellipsis from clefts must exhibit the morphosyntactic properties of the corresponding phrase in that cleft. In the case of preposition omission, this concerns morphological case marking on the DP fragments that result from preposition omission. For instance, in German, DPs that are the complement of a preposition exhibit prepositional case marking (accusative, dative or genitive) (20a), whereas in a cleft the DP appears obligatorily in nominative case (20b). The cleft account could thus potentially explain why proper nouns, which are

<sup>18</sup>See e.g. Stigliano (2018, 2019) for Spanish and Nykiel (2013) for Polish.

### 3 Experiments on the syntax of fragments

not marked for case and can hence be interpreted as nominative are acceptable in (19).<sup>19</sup> Consequently, if a DP fragment cannot be derived from (20a) due to the unavailability of P-stranding in German, but from the cleft in (20b), it must always appear in nominative case.<sup>20</sup>

- (20) Für wen ist das Päckchen?  
 for who<sub>ACC</sub> is the package  
 'For whom is the package?'

- a. Es ist für meinen Vater.  
 it is for my.<sub>ACC</sub> father  
 'It is for my father.'

PP fronting

<sup>19</sup>This prediction is specific to German and does not necessarily hold for all languages with morphological case marking. For instance, Szczegielniak (2008) derives Polish preposition-less DP fragments from a cleft construction where the DP appears in prepositional case. German however has no such construction.

<sup>20</sup>A similar empirical prediction is implied by the discussion on *that is*-ellipsis in Merchant (2004a), who argues that the copula and a pronoun like *it* are inherently given and can be omitted in the absence of a corresponding *linguistic* antecedent. Therefore in (20) a DP fragment answer could be also derived from the sentence (i). Like the cleft account, this analysis also predicts nominative case on the DP fragment.

- (i) Mein Vater war es.  
 my.<sub>NOM</sub> father was it  
 'My father it was.'

However, at least for sluicing it seems doubtful that such a *that is* construction is the structure underlying preposition omission data in German. Lemke et al. (forthcoming) show that in a production task subjects frequently produce constructions like (iia), but not their sluiced counterparts in (iib). In contrast, when the potentially reduced phrase was introduced by a PP that matches the category of the antecedent *mit jemandem*, like in (iii), sluicing was relatively frequent. In an acceptability rating study Lemke et al. (forthcoming) find that (iib) is heavily degraded as compared to the sluice derived from (iii). This provides further evidence that such instances of preposition omission examples in German are not derived from *that is* constructions.

- (ii) a. Hans hat mit jemandem getanzt, aber ich weiß nicht, wer das war.  
 Hans has with somebody danced, but I know not who that was  
 'Hans danced with somebody, but I don't know who that was.'  
 b. \*Hans hat mit jemandem getanzt, aber ich weiß nicht, wer.  
 Hans has with somebody danced, but I know not who  
 'Hans danced with somebody, but I don't know who.'
- (iii) Hans hat mit jemandem getanzt, aber ich weiß nicht, mit wem (Hans getanzt hat).  
 Hans has with somebody danced, but I know not with who Hans danced has  
 'Hans danced with somebody, but I don't know with whom Hans danced.'

- b. Es ist mein Vater, für den das Päckchen ist. Cleft  
it is my.NOM father for who.ACC the package is  
‘It is my father for whom is the present.’

I return to this issue in experiment 6, which compares the acceptability of PP short answers to that of nominative and prepositional case-marked DP short answers. Before turning to the empirical investigation of the predictions made by the PSG, I briefly discuss how the PSG can be theoretically motivated in a generative framework rather than just postulating it as a descriptively appropriate generalization. This is particularly important because some implementations of the PSG explain the relevant data without assuming movement at all, and this would weaken the status of the PSG as evidence for movement in fragments.

### 3.2.1.2 P-stranding and pied-piping in minimalism

Merchant (2004a) leaves open the question of why P-stranding is possible in English, but not in German. In Merchant (2001), he makes a tentative suggestion in terms of an analysis of prepositions as case markers, but the idea is not spelled out in detail. However, as I argued above, the processes underlying a movement restriction that is taken to support movement and deletion are not trivial, because they might also facilitate non-movement explanations for the data (which might also explain as well, but for independent reasons, why movement is blocked). In what follows, I briefly illustrate how the viability of specific accounts of fragments relies on the analysis that P-stranding/pied-piping itself receives.

Pied-piping is specifically problematic in minimalism, because in this framework movement is in general assumed to be a feature-driven last resort operation. This implies that (i) movement is never optional (because it is the last resort to save derivations from crashing) and (ii) that the constituent that is moved hosts the feature that needs to be checked. However, in pied-piping not only this constituent itself, but a superordinate constituent is moved, which *contains* the constituent hosting the feature. This is exemplified in Figure 3.4 for *wh*-movement in a question like (21): Not only the *wh*-phrase hosting the uninterpretable *wh*-feature, but the complete PP appears in [Spec, PP]. Unless exceptional movement (Weir 2014a) is assumed, which is not feature-driven, this concerns the generation of fragments under the movement and deletion account, which needs to assume some kind of (probably focus-related) feature in order to motivate movement to [Spec, FP].

- (21) For whom is the present?

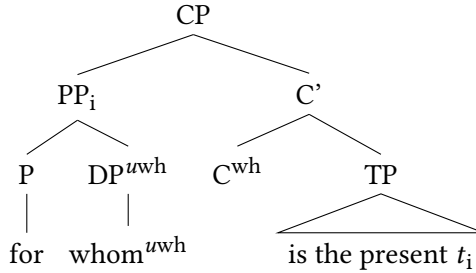


Figure 3.4: Derivation of (21) under a pied-piping analysis. In this analysis, only the DP hosts the *wh*-feature.

One possible solution to this problem is the assumption of *feature percolation* (Chomsky 1973, Grimshaw 2000). The idea is that under certain conditions syntactic features can ‘percolate’ to projections outside the maximal projection whose head hosts the feature. In (21), for instance, the *wh*-feature would percolate to the PP level, so that the complete PP is marked [*uwh*] and must move to [Spec, CP]. As Figure 3.5 shows, technically speaking, under this analysis there is no pied-piping in the proper sense, because the PP as a whole just behaves like any other *wh*-phrase. This idea has been explicitly applied to prepositional pied-piping in *wh*-questions (Trissler & Lutz 1992, Grimshaw 2000, Trissler 2000, Yoon 2001, Lasnik 2006, Sato 2011). As there is no reason to assume that a concept such as feature percolation is restricted to *wh*-questions, it can be immediately applied to the movement operations that derive fragments according to Merchant’s theory. If movement in fragments is driven by some focus-related feature and the choice between P-stranding and pied-piping depends on whether this feature percolates to PP, the structures underlying P-stranding and pied-piping would only differ in whether the complete PP, or only the DP has the focus feature (22).

- (22) a. ~~This package~~ is [for my father]<sub>F</sub>.  
 b. ~~This package~~ is for [my father]<sub>F</sub>.

Now, if the preposition is obligatorily pied-piped for whatever reason (see the references below and footnote 23 for some hypotheses), this indicates that the focus structure licensing P-stranding, i.e. focusing the DP only, is not available in German (23). This pattern however is exactly what the in situ deletion account requires in order to explain the contrast between English and German. Since it assumes that only the non-focused parts of the utterance may be deleted, the ellipsis of the preposition is never licensed in German, because the corresponding focus structure (23b) is unavailable in this language. Taken together, if it



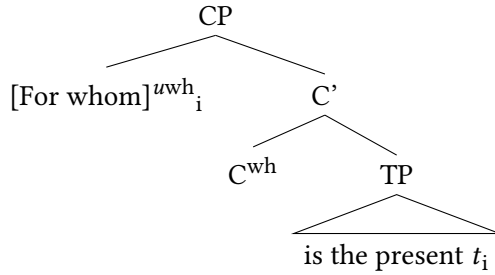


Figure 3.5: Derivation of (21) under a feature percolation analysis. In this analysis, the complete PP is marked as  $[+wh]$  as the result of feature percolation.

is assumed that in such examples the PP is marked with the feature triggering movement, the movement and deletion account makes exactly the same prediction as in situ deletion and therefore does not gain larger explanatory power. This holds for any version of the theory that assumes feature percolation.

- (23) a. Das Päckchen ist [für meinen Vater]<sub>F</sub>.  
b. \*Das Päckchen ist für [meinen Vater]<sub>F</sub>.

If the preposition omission data are taken to be evidence for movement, there must be a genuine movement restriction, which has no effect on the potential form of fragments if the PP remains in situ. Such a movement restriction could be the assumption that PP is an island for extraction in German, but not in English, as has been first suggested by van Riemsdijk (1978). In simplified terms, he argues that the preposition can be reanalyzed in English as forming a syntactic unit with the verb, so that it can be separated from the noun by movement.<sup>21</sup> More recently, Abels (2003) has picked up the idea of structural differences between the PP in languages that have and those that lack P-stranding. Abels argues that PP is a phase in German, but not in English. A phase can only be evacuated through its left edge, i.e. the specifier of the highest projection within this phase. From this perspective, extraction of the complement of P out of the PP is possible in English, but not in German, because it would have to be first moved to [Spec, PP] in German.

<sup>21</sup>This also predicts an asymmetry between prepositional objects and adjuncts, because the preposition is subcategorized only in the case of the former. Pullum & Huddleston (2002) observe that indeed extraction out of adjunct PPs seems to be dispreferred as compared to prepositional objects. This is also empirically supported by my experiment 7, where the overall preference for omitting prepositions in short answers was reduced for adjunct PPs as compared to argument PPs.

### 3 Experiments on the syntax of fragments

Movement of the complement of a phrase head to its specifier is not possible, because movement occurs only for feature checking purposes and features can be checked in a head-complement relation.<sup>22</sup>

Such structural differences between the German and English PP might explain why extraction out of the PP is blocked in German, but it does not explain why the PP can be moved as a whole. Under the assumption of feature-driven movement, if feature percolation is ruled out, this approach requires an additional explanation for why the PP can be moved as a whole in case extraction is not possible. This could be modeled by integrating optimality-theoretic violable constraints on locality in the minimalist framework, as Heck (2008) suggests, but such additional assumptions clearly complicate the overall theoretical framework. The phase-based account is also compatible with Weir's (2014a) assumption of exceptional movement. Weir's account in principle requires fronting the focused phrase in order to evacuate it from the ellipsis site, but according to the theory, it can carry along "any material which it might need to pied-pipe" (Weir 2014a: 186). If the complement of the preposition cannot be extracted for independent reasons in languages which, like German, do not have P-stranding, pied-piping saves the derivation from crashing. Recall that this movement occurs only for PF reasons, specifically, the incompatibility of the pitch accent marking focus with ellipsis. The fact that the fragment is focused does not imply the existence a focus feature that needs to be checked, hence Weir does not need to assume feature percolation.

---

<sup>22</sup>In a later version of his theory, Abels (2012) argues that PP is always a phase. In this theory, the preposition is the head of the phase, which can be evacuated only through its edge, [Spec, PP]. The complement of P can never be moved to this position, because movement must be triggered by feature checking requirements in minimalism. The complement is already in a local configuration with the P head that allows for feature checking, so that no movement is required and consequently licensed. Abels (2012: 245–268) argues that the critical difference between languages that allow P-stranding and those that do not is that the former involve additional structure (a projection headed by an empty morpheme) between the preposition and the DP. Consequently, this morpheme, but not the DP is the complement of P. Under this analysis, P-stranding involves movement out of a more deeply embedded position within the PP, which is also available in languages without P-stranding, like German (i). Abels motivates the assumption of the empty morpheme in languages that do not have P-stranding with the argument that it has overt counterparts in some languages.

- (i) Was (für Bücher) hast du (für Bücher) gelesen (für Bücher)? (Abels 2012: 255)  
what for books have you for books read for books  
'What (kind of books) have you read?'

The brief discussion in this section showed that the theoretical analysis of P-stranding and pied-piping<sup>23</sup> determines whether the correlation between preposition omission and P-stranding in fragments evidences a movement restriction or whether it is expected under non-movement accounts as well. Advocates of movement and deletion would have to commit to one of these analyses but as for now there is no consensus about which of the analyses discussed so far is correct. The picture becomes even more complicated when the difference between acceptable and unacceptable instances of P-stranding *within* a P-stranding language (see footnote 27 below for English) is taken into account. Finally, if the availability of P-stranding and preposition omission in fragments is attributed to a structural difference between English and German PPs, the apparent optionality of P-stranding in contexts that allow for it in English also calls for an explanation, because in minimalism syntactic operations are not optional. Consequently, before interpreting the P-stranding generalization as evidence for movement, there should be a more explicit account of pied-piping and P-stranding.

Solving this intricate theoretical problem is beyond the scope of this work. The discussion in this section showed that if pied-piping is explained through feature percolation or when the structured propositions approach by Reich (2002a) (see Section 3.2.5.1.1) is pursued, movement and deletion and in situ deletion make the same empirical predictions. However, it also showed that if feature percolation is rejected and the PSG is explained with a ban on extraction out of PPs in German, as suggested by Abels (2003, 2012), the PSG provides evidence for the movement and deletion account. Therefore, in what follows I investigate first whether the general pattern that the PSG predicts for fragments holds in German and English (experiments 4 and 5), before I explore potential non-movement explanations for the data. If the pattern found in experiments 4 and 5 can be explained without the assumption of movement, the assumption of the additional movement step is empirically unmotivated. Experiment 6 tests a hypothesis that has been first mentioned in Barton & Progovac (2005), who attribute the impossibility of omitting the preposition in German to its involvement in prepositional case checking. This prediction is not borne out. Experiment 7 provides evidence for a nonsyntactic parallelism between question and answer, which can be accounted for in terms of processing (Levelt & Kelter 1982, Nykiel 2017) or a structured proposition analysis of question semantics (Reich 2002a, 2007).

---

<sup>23</sup> Besides the accounts referred to so far, among the syntactic accounts, Sato (2011) assumes that in German PPs, but not in English ones, the determiner in D is head-moved to P. Other researchers, as Tokizaki (2010) and Philippova (2014), claim that P-stranding is never syntactically blocked, but ruled out when it yields prosodically ill-formed structures.

### 3.2.2 Experiment 4: Preposition omission in German

#### 3.2.2.1 Background

Experiments 4 and 5 empirically test the predictions of the PSG for German and English respectively. If the PSG indeed is the result of a movement restriction, the movement and deletion account predicts that preposition omission in fragments is possible in English but not in German. Furthermore, experiment 4 establishes a baseline for the follow-up experiment 6.

For German, Merchant et al. (2013) conducted a similar experiment that confirms the introspective judgments for question-answer pairs like (15), which I repeat here as (24) for convenience: Just like the movement and deletion account predicts, in this context, PP short answers are rated significantly better than DP short answers ( $\mu_{PP} = 5.99$  vs.  $\mu_{DP} = 4.76$  on a 7-point Likert scale where 7 = fully acceptable).<sup>24</sup>

- (24) Mit wem hat Anna gesprochen?  
with whom has Anna spoken  
'Who was Anna talking to?'
- a. Mit dem Hans.  
with the.DAT Hans  
'With Hans.'
- b. \*Dem Hans.  
the.DAT Hans  
'Hans.'

Merchant et al. (2013) tested negative short answers to polar questions like (25) (Merchant et al. 2013: 34), where fragments were always contrastive foci in the sense of Krifka (2007), i.e. they overtly negate a contextually given alternative. The intention was to improve the acceptability of fronting objects in sentential structures and to obviate the objection that the presumed underlying sentences involving fronting should be rejected across the board.

- (25) a. Willst du auf den *Torhüter* verzichten?  
want you on the goalkeeper do.without  
'Do you want to do without the goalkeeper?'

---

<sup>24</sup>Given that the movement and deletion accounts predicts DP short answers to be ungrammatical, their ratings are surprisingly high. The authors explain this with the fact that case-marked DP fragments are not ungrammatical across the board in German, but that they cannot be derived when the question asks for a PP.

- b. Nein, \*(auf) den Stürmer.  
no (on) the.ACC striker  
'No, (without) the striker.'

Like the experiment by Merchant et al. (2013) I contrasted PP and prepositional case-marked DP fragments in an acceptability rating task. The movement and deletion account predicts that PPs should be rated significantly better than DPs.

### 3.2.2.2 Materials

A sample item is given in (26). Unlike in the study by Merchant et al. (2013), the short answers were always information foci, that is, answers to questions that do not belong to a limited set of contextually given alternatives. Information foci are more similar to the non-contrastive instances of short answer fragments discussed in the literature, like (24). All items contained a question-answer pair. In order to increase their naturalness, they were introduced by a two-sentence context story.

- (26) Martin is sitting at the kitchen table in his shared apartment and is wrapping a present.
- a. Sein Mitbewohner Nils fragt ihn: Für wen ist das Päckchen?  
his roommate Nils asks him for who.ACC is the package  
'His roommate Nils asks him: For whom is the package?'
- b. Martin sagt: (Für) meinen Vater.  
Martin says for my.ACC father  
'(For) my father.'

The preposition was always given in the question, because givenness is a requirement for ellipsis according to both families of sentential accounts of fragments.<sup>25</sup> Otherwise, the omission of the preposition in the answer would be blocked for this independent reason in English as well. Finally, like in the study

<sup>25</sup>This is not the case for all *wh*-phrases that require a PP answer in German, as the locative example in (i) shows.

- (i) a. Wo ist Anna?  
where is Anna  
'Where is Anna?'
- b. Auf einer Party / In der Disco / Bei einer Freundin / ...  
on a party / in the disco / at a friend  
'At a party / In the disco / At a friend's / ...'

### 3 Experiments on the syntax of fragments

by Merchant et al. (2013), all DP fragments appeared in the case required by the corresponding preposition in the context question ( $n_{Acc} = 7$ ,  $n_{Dat} = 11$ ,  $n_{Gen} = 2$ ).

#### 3.2.2.3 Procedure

Materials were presented together with experiment 1 to 70 undergraduate students who were self-reported native speakers of German. The study was conducted on the Internet using the LimeSurvey questionnaire presentation software. Subjects were compensated with the participation in a lottery of  $10 \times \text{€}30.00$  among all participants. Their task consisted in rating the naturalness of the answer, which was italicized, in the context of the preceding material on a 7-point Likert scale (7 = fully natural). Each subject rated 20 items (10 PP and 10 DP short answers), which were presented with 20 items from experiment 1 and 47 fillers. The fillers were also short dialogues with fragment answers. The fragment answer was never a PP or a prepositional case-marked DP. See Section 3.1.1.4 for details on presentation, randomization and exclusions.

Table 3.8: Mean ratings (standard deviation) across conditions in experiment 4 and in the corresponding experiment by Merchant et al. (2013).

| Condition                | Exp. 4      | Merchant et al. (2013) |
|--------------------------|-------------|------------------------|
| Preposition present (PP) | 6.61 (0.99) | 5.99 (1.64)            |
| Preposition omitted (DP) | 4.42 (2.05) | 4.76 (2.03)            |

#### 3.2.2.4 Results

Like in Merchant et al. (2013), PP fragments ( $\mu = 6.62$ ,  $\sigma = 0.99$ ) were rated as more acceptable than DP fragments ( $\mu = 4.42$ ,  $\sigma = 2.05$ ). Table 3.8 shows that even the absolute ratings were relatively close to those in Merchant et al. (2013). This replicates their findings with non-contrastive short answers.

The data were analyzed with CLMMs in R following the procedure described in Section 3.1.1.5. The full model contained fixed effects for PREPOSITION, the prepositional CASE (Accusative/Dative/Genitive) and the POSITION of the trial in the time-course of the experiment as well as all two-way interactions thereof. I included by-subject random intercepts and slopes for PREPOSITION, CASE and the interaction thereof, as well as by-item random intercepts and slopes for PREPOSITION. As CASE was not varied within items, there were no by-item random effects

for this predictor. The final model (see Table 3.9) contained only fixed effects for POSITION, CASE and PREPOSITION. The POSITION effect ( $\chi^2 = 12.5$ ,  $p < 0.001$ ) shows familiarization with the task, which is factored out by including it in the model and is not of theoretical interest. The CASE effects indicate acceptability differences between items, but the absence of PREPOSITION:CASE interactions indicates that CASE had no specific effect on the acceptability of preposition omission. Therefore, and because prepositional case was not balanced and varied systematically across materials in this experiment, I consider it a further control predictor. The highly significant effect of PREPOSITION ( $\chi^2 = 44.26$ ,  $p < 0.001$ ) replicates the effect found by Merchant et al. (2013): Omitting the preposition in short answers to questions asking for a PP is strongly dispreferred in German.

Table 3.9: Fixed effects in the final CLMM for experiment 4.

| Predictor        | Estimate | SE    | $\chi^2$ | $p$    |     |
|------------------|----------|-------|----------|--------|-----|
| PREPOSITION      | -4.601   | 0.474 | 44.26    | <0.001 | *** |
| CASE (DAT v ACC) | -1.247   | 0.398 | 8.21     | <0.01  | **  |
| CASE (GEN v ACC) | -3.134   | 0.653 | 15.61    | <0.001 | *** |
| POSITION         | 0.009    | 0.003 | 12.5     | <0.001 | *** |

### 3.2.2.5 Discussion

Experiment 4 replicates the effect for information focus that Merchant et al. (2013) report for contrastive focus. Table 3.8 shows that the average ratings across PREPOSITION conditions were relatively close to those reported by Merchant et al. (2013) even in absolute terms (6.61 vs. 5.99 in the PP condition and 4.42 vs. 4.76 in the DP condition). This clearly confirms the introspective acceptability judgments for examples like (15). From the perspective of the movement and deletion account, it might seem surprising that DP fragments, which should be impossible to derive, are rated as relatively acceptable. Merchant et al. (2013) attribute this to the fact that they are not ungrammatical per se in other contexts, such as the ones I tested in experiment 1. Furthermore, Merchant et al. (2013) emphasize that in acceptability rating studies only the differences between conditions but not the absolute ratings can be interpreted, because there is no absolute reference level of (un)grammaticality on the 7-point scale. I return to this issue in the discussion of experiment 6, which compares the acceptability of PP short answers to prepositional case and nominative DP short answers. For the time being, what

matters is that the data show a strong preference for realizing the preposition in German short answers to PP questions and hence replicate the effect found by Merchant et al. (2013).

### 3.2.3 Experiment 5: Preposition omission in English

#### 3.2.3.1 Background

Experiment 4 shows that DP short answers to PP questions are degraded in a language which lacks P-stranding, like German, just like the PSG predicts. Experiment 5 investigates the pattern for English. Introspective examples from the literature like (14), which I repeat for convenience as (27), and corpus data (Nykiel 2017) suggest that it is possible to omit the preposition in English, but Merchant et al. (2013) present no evidence on how subjects respond to this in an acceptability rating paradigm. The prediction of the PSG is that omitting the preposition in fragments as (27) should be (i) relatively more acceptable than in German and (ii) at least as acceptable as realizing it.

- |      |                             |                         |
|------|-----------------------------|-------------------------|
| (27) | Who was Peter talking with? | (Merchant 2004a: 685)   |
| a.   | With Mary.                  | Preposition realization |
| b.   | Mary.                       | Preposition omission    |

In addition to testing these predictions, experiment 5 investigates whether the acceptability of fragments is reflected in the acceptability of the left dislocation structures from which fragments are derived according to the movement and deletion account. This is predicted by the “original” version of the movement and deletion account of fragments (Merchant 2004a), which underlies the reasoning of the experiments by Merchant et al. (2013). In the case of preposition omission, the preference for omitting the preposition in a fragment should reflect the preference for P-stranding in the corresponding full sentence: If the bare DP *Mary* was preferred over the PP *to Mary* in (27), so should be P-stranding (28a) over pied-piping (28b).

- |      |                                  |             |
|------|----------------------------------|-------------|
| (28) | a. Mary, Peter was talking with. | P-stranding |
|      | b. With Mary, Peter was talking. | Pied-piping |

Experiment 5 tests this in a 2×2 design crossing SENTENTIALITY with PREPOSITION omission/realization (in fragments) or P-stranding/pied-piping (in full sentences). Since the left dislocation structures in (28) contain a redundant matrix clause and the unmotivated fronting operation, I expected them to be overall



degraded as compared to fragments. In order to avoid a potential floor effect, SENTENTIALITY was tested as a between subjects IV. In this setting, Merchant (2004a) predicts the absence of a significant interaction between the two predictors, since he expects that expressions that are preferred in a left-peripheral position will be more acceptable to the same extent as fragments.

### 3.2.3.2 Materials

The stimuli were mostly identical to those used in the German experiment 4, which were translated into American English by a native speaker.<sup>26</sup> Since English allows for P-stranding and P-stranding has been argued to be preferred over pied-piping specifically in colloquial speech (Pullum & Huddleston 2002: 628),<sup>27</sup> the preposition in the context question was always stranded. This ensured that P-stranding is not blocked due to semantic, pragmatic or processing reasons, such as the congruence between question and answer (see also Section 3.2.5.1). All materials were introduced by the same context story as in experiment 4. For the reasons discussed above, I tested not only fragments, but also the full sentences which underlie the fragments according to the movement and deletion account (29). In German this was not required, because P-stranding is generally assumed to be ungrammatical.

### 3.2.3.3 Procedure

54 native speakers of American English were recruited via the *Prolific Academic* crowdsourcing platform for a web-based acceptability rating study, which was conducted using the LimeSurvey presentation software. Each participant received £2 for participation.<sup>28</sup> Subjects were asked to read the materials and then

<sup>26</sup>One item had to be replaced because its English counterpart did not involve a preposition.

<sup>27</sup>Pullum & Huddleston (2002: 628–631) also note that the choice between pied-piping and P-stranding in English is not always totally unconstrained, as some contexts strongly favor or require either of the variants. For instance, they argue that P-stranding is favored in case of prepositional verbs (i), while pied-piping is in case of adjunct PPs (ii).

- (i) a. What<sub>i</sub> are you asking for *t<sub>i</sub>*? (Pullum & Huddleston 2002: 629)
- b. ?[For what]<sub>i</sub> are you asking *t<sub>i</sub>*?
- (ii) a. \*[What circumstances]<sub>i</sub> would you do a thing like that under *t<sub>i</sub>*?
- b. [Under what circumstances]<sub>i</sub> would you do a thing like that *t<sub>i</sub>*? (Pullum & Huddleston 2002: 631)

<sup>28</sup>Since *Prolific Academic* is a British platform, payments are made in pounds and transferred to the participants' PayPal accounts.

### 3 Experiments on the syntax of fragments

rate the naturalness of the italicized target utterance in the context of the question on a 7-point Likert scale (7 = fully natural). They were assigned to one out of four lists. SENTENTIALITY was tested between subjects, because the markedness of the non-contrastive topicalization structures in (27) could yield a floor effect otherwise. Consequently, the materials were distributed across four lists according to a 2×2 Latin square design, so that each subject saw each token set in one PREPOSITION condition, half of the subjects saw only sentences and half only fragments. Just like in experiment 4, each subject rated 20 items (10 per PREPOSITION condition). Materials were mixed with 20 materials from experiment 9 and 45 fillers. All fillers consisted of context stories followed by a dialogue. The target utterance was always the last utterance in that dialogue, and fillers were adapted so that participants assigned to the fragment lists rated only fragments and those working on the sentential lists only sentences. Materials were presented in individually fully randomized order. Fillers included five ungrammatical controls, which contained e.g. wrong auxiliaries or voice. Four subjects who rated more than 50% of these with 6 or 7 points on the scale, were excluded from further analysis.

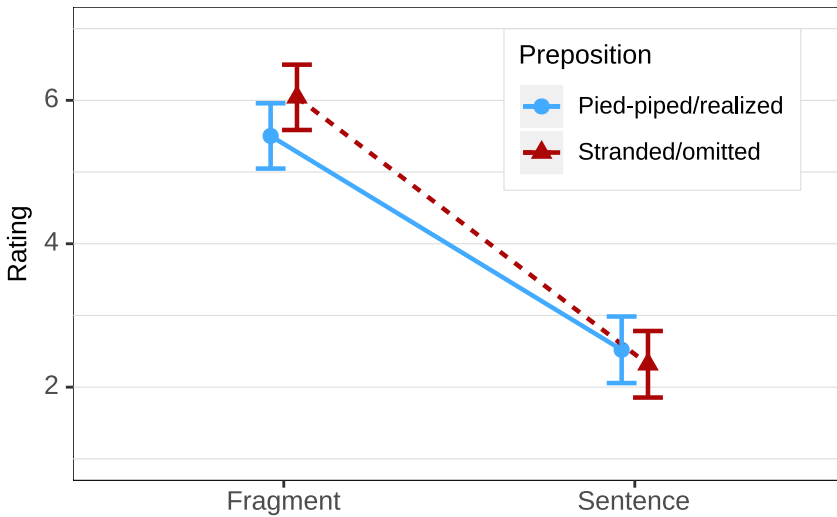


Figure 3.6: Mean ratings and 95% CIs for experiment 5.

#### 3.2.3.4 Results

Figure 3.6 provides a summary of the average ratings across conditions. The fragment data show that the omission of the preposition was slightly more acceptable

in English ( $\mu = 6.04, \sigma = 1.55$ ) than its realization ( $\mu = 5.5, \sigma = 1.67$ ). The sentential left dislocation constructions were heavily degraded across the board no matter whether the preposition was stranded ( $\mu = 2.32, \sigma = 1.43$ ) or pied-piped ( $\mu = 2.52, \sigma = 1.52$ ). Like in the previous rating studies, the data were analyzed with CLMMs in R following the procedure described in Section 3.1.1.5. The full model contained fixed effects for PREPOSITION, SENTENTIALITY and POSITION of the trial in the experiment as well as all two-way interactions thereof. It also contained by-item random intercepts and slopes for PREPOSITION, SENTENTIALITY and the interaction thereof and by-subject random intercepts and slopes for PREPOSITION. By-subject random effects for SENTENTIALITY were not included, because this IV was tested between subjects.

Table 3.10: Fixed effects in the final CLMM for experiment 5.

| Predictor                 | Estimate | SE    | $\chi^2$ | $p$    |     |
|---------------------------|----------|-------|----------|--------|-----|
| PREPOSITION               | 0.258    | 0.138 | 3.83     | 0.062  |     |
| SENTENTIALITY             | 2.997    | 0.322 | 54.94    | <0.001 | *** |
| PREPOSITION:SENTENTIALITY | 0.499    | 1     | 11.73    | <0.001 | *** |

The final model (see Table 3.10) contains a significant main effect of SENTENTIALITY ( $\chi^2 = 54.94, p < 0.001$ ), which shows that short answer fragments are rated better than the presumably underlying left dislocation structures. The marginal main effect of PREPOSITION shows that across the board there is no significant difference between P-stranding/PP fragments and pied-piping/DP fragments ( $\chi^2 = 3.38, p < 0.06$ ), but the significant interaction between the two predictors ( $\chi^2 = 11.73, p < 0.001$ ) suggests that specifically in the fragment condition DP fragments are more acceptable than PP fragments.

### 3.2.3.5 Discussion

Experiments 4 and 5 had the purpose of empirically testing the pattern that the PSG predicts with respect to preposition omission in languages with and without P-stranding. Taken together, the experiments empirically confirm the pattern that the PSG predicts for German and English short answer fragments. In German, omitting the preposition in the answer is strongly degraded. In contrast, in English both DP and PP fragments are rated as relatively acceptable, and the omission of the preposition is actually preferred over its realization. This is in line with the prediction that omitting the preposition in short answers is degraded in languages that lack P-stranding, but at least as acceptable as realizing it in languages that allow for this syntactic operation.

### 3 Experiments on the syntax of fragments

The English data from experiment 5, however, also show that all of the left dislocation structures that underlie fragments according to movement and deletion are strongly degraded. Furthermore, there is a significant interaction between SENTENTIALITY and PREPOSITION: Contrary to what Merchant (2004a) predicts, the preference for omitting the preposition in fragments does not match the acceptability of the corresponding left dislocation structures. This observation can be reconciled with the PSG if the exceptional movement version of the theory (Weir 2014a) is assumed, which, however, does not predict a strict parallelism between fragments and left dislocation. Given the discussion on pied-piping in the introduction to this section, Weir (2014a) would have to explain pied-piping as the result of a ban on extracting PPs which are the complement of a preposition (Abels 2003, 2012). In that case, fronting the complete PP is the only way to evacuate focused constituents from the ellipsis site (Heck 2008).

The PSG does not explain why omitting the preposition is preferred in my materials in English. A possible reason for this could be that this is due to P-stranding in the question. This could either reflect a general preference for omitting the preposition whenever it is given in the question (this is not possible in German), or be an effect of question-answer congruence, as already hinted at above. I return to this question in experiment 7.

The data also have implications for the other theories of fragments discussed in Chapter 2. For the nonsentential and in situ deletion accounts, the challenge is to provide an explanation for the data from experiments 4 and 5 that does not rely on movement. Experiments 6 and 7 test such non-movement accounts of the preposition omission data, which are based on case checking (experiment 6) and a nonsyntactic parallelism between question and answer (experiment 7). Empirical evidence for either of these hypotheses would leave movement-based accounts without an explanatory benefit over the simpler in situ deletion account.

The results also tentatively speak against the claim by Bergen & Goodman (2015) that fragments are ungrammatical. Since the missing preposition (as well as the other omitted material) was unambiguously retrievable from the question, their account does not predict a crosslinguistic difference between English and German with respect to the acceptability of DP fragments. In fact, if all that matters is whether the hearer can retrieve the omitted material, preposition omission might be expected to be *more* acceptable in German than in English, because the German DP has prepositional case marking which restricts the set of possible prepositions. For instance, a hearer who encounters a dative DP fragment can figure out that the missing preposition must be among those requiring dative, whereas such a cue is not available in English. This prediction is therefore disconfirmed by the data. This argument of course does not neglect the relevance

of information-theoretic and processing-based factors to the acceptability and usage of fragments. The experiments in Chapter 5 show that predictability plays an important role in the choice between omitting and realizing words.

### 3.2.4 Experiment 6: Preposition omission and case

#### 3.2.4.1 Background

Experiment 6 investigates how acceptable nominative DP short answers are as compared to PP and prepositional case-marked DP short answers in German. This tests two predictions by movement and deletion and the nonsentential account of fragments: First, movement and deletion predicts that nominative DP fragments are grammatical as answers to PP questions when they can be derived from a cleft. Second, the nonsentential account also predicts nominative fragments to be acceptable in such contexts: Nominative is the default case in German and does not need to be checked by a preposition. Under both of these lines of reasoning it is expected that PP fragments are grammatical, that nominative DP fragments are relatively acceptable because they are grammatical, and that prepositional case-marked DP fragments are ungrammatical. According to movement and deletion, their derivation involves ungrammatical P-stranding and according to the nonsentential account, prepositional case-marked DP fragments contain a strong uninterpretable case feature.

Barton & Progovac (2005) do not explicitly discuss the status of the PSG as evidence for movement, but they observe a crosslinguistic difference between English and Serbian with respect to the possibility of omitting prepositions in telegraphese utterances. At the example of (29), Barton & Progovac (2005: 88,89) show that prepositions can be omitted in English sentences like (29a), whereas they are obligatory in the Serbian example (29b). As Serbian lacks P-stranding (Merchant 2004a: 667–668), this pattern resembles the PSG. However, preposition omission occurs *in situ* in (29), so the pattern cannot be explained by a restriction on extracting a DP out of the PP, like Merchant (2004a) proposes for preposition omission in short answers. Whatever blocks the omission in Serbian must be independent from movement. Note also that these data are highly relevant to the *in situ* deletion account, because they show that omitting a preposition when the remnant is a prepositional case-marked DP can be ruled out even *in situ*.

- (29) a. Please pick me up (at) Summerside Motel.  
 b. Vidimo se                   \*(na) JFK aerodrom-u.  
     see.1PL Reflexive   on JFK airport.LOC  
     ‘See you (at) JFK airport.’

Barton & Progovac (2005) argue that the independent factor that blocks the omission of the preposition in Serbian is case checking: English case features are weak and can remain unchecked, whereas Serbian has strong case features, which must be checked. According to Barton & Progovac, the strength of Serbian case features is evidenced by the morphological marking of case. Note that this requires an analysis of prepositional locative case in (29b) as structural case, because inherent case can always remain unchecked according to the nonsentential account. I return to this issue below. In contrast to the ungrammatical (29b), Barton & Progovac observe that the preposition can be omitted in Serbian when the DP appears in nominative case, which they argue is default case (30). As they pursue a nonsentential account, they do not assume that examples like (30) or the English (29a) involve the deletion of the preposition, but analyze the noun phrase as a bare NP, which simply appears in default case. Omitting the preposition is licensed because it is recoverable “from the verb and the rest of the clause” (Barton & Progovac 2005: 89).

- (30) Vidimo se,                   JFK aerodrom.                   (Barton & Progovac 2005: 89)  
       see.1PL Reflexive JFK airport.NOM  
       ‘See you, JFK airport.’

If the contrast in (29) could be generalized crosslinguistically, these data would provide a non-movement explanation for the PSG: The preposition cannot be omitted in languages that have prepositional case marking, but it can in languages that do not, because prepositional case must be checked by an overt preposition. Interestingly, most of the languages that lack P-stranding according to Merchant (2004a), such as German and Slavonic languages, have prepositional case marking.<sup>29</sup> I anticipated above that this reasoning requires that prepositional case is analyzed as structural case.<sup>30</sup> At least in German, this seems to be correct. In Section 2.4.1 I defined structural case as a purely linguistic device that makes no significant contribution to meaning. In contrast, inherent case makes such contributions by encoding a specific  $\theta$ -role. Whether a specific prepositional case encodes aspects of meaning is an empirical issue. In German for instance, there is a tendency for dative prepositional case to mark locations or sources

<sup>29</sup>This is not true for all of the languages that allow DP short answers to PP questions according to Merchant (2004a). Icelandic has morphological case marking and still allows for P-stranding and omission in short answers, whereas Hebrew has no morphological case marking and no P-stranding. In the case of Hebrew this could be due to the incorporation of the preposition by the noun.

<sup>30</sup>This has been explicitly claimed by den Dikken (2013: 24). In his minimalist approach, prepositional case is checked by the head of a functional projection in the PP layer and not by P itself. The layer relies crucially on the presence of the preposition, hence prepositional case is licensed only if the preposition is present.

and for accusative to mark goals (Zwarts 2005: 8). Still, this relationship is not systematic, because a dative DP can encode a source (31a) as well as a goal (31b) and location (31c) depending on the preposition it appears with. Taken together, just like I argued in Section 2.4.1 for accusative, prepositional case behaves rather like structural than like inherent case, because it is not strictly associated with a specific  $\vartheta$ -role. Even though in German the dative prepositional case might be likely to mark the location of an event, it does not always convey this aspect of meaning, unlike inherent dative, which marks the recipient.

- (31) a. Er rannte aus dem Park. (Zwarts 2005: 5)  
he ran out the.DAT park  
‘He ran out of the park.’  
b. Er rannte zum Park. (Zwarts 2005: 5)  
he ran to-the.DAT park  
‘He ran to the park.’  
c. Er spielte im Park Badminton.  
he played in-the.DAT park badminton  
‘He played badminton in the park.’

If prepositional case is thus not semantically interpretable, it requires an overt licenser just like other instances of structural case do in Barton & Progovac’s framework.<sup>31</sup> The nonsentential account can hence explain at least a large part of the preposition omission data without assuming unarticulated linguistic structure, actually *because* of the assumption that there is no unarticulated material in fragments that could check the uninterpretable case features.<sup>32</sup>

<sup>31</sup>Progovac et al. (2006: 342) argue that a specific prepositional case, like accusative in the Serbian example (i), is acceptable in fragments if it can also be used as inherent case encoding a  $\vartheta$ -role. Nevertheless, non-prepositional accusative is associated with the PATIENT  $\vartheta$ -role, so that the fragment is assigned a misleading interpretation if the preposition is omitted.

- (i) a. Na čega je Stefan seo? (Progovac et al. 2006: 342)  
on what did Stefan sit  
‘What did Stefan sit on?’  
b. \*(Na) stolicu.  
on chair.ACC  
‘(On) a chair.’

Note however that this accounts only for prepositional case that is otherwise inherent, such as dative and genitive in German. As I argued above in Section 2.4.1, there are good reasons not to analyze German accusative as inherent (possibly in contrast to Serbian), so that this explanation does not immediately concern my experiments.

<sup>32</sup>Barton & Progovac (2005) do not discuss sluicing, therefore it remains open whether they would make a similar prediction for the phenomenon that originally motivated the PSG.

### 3 Experiments on the syntax of fragments

Experiment 6 tests this prediction by collecting acceptability ratings for three types of short answer fragments in German: PP fragments, prepositional case-marked DP fragments and default case-marked DP fragments. German does not allow for P-stranding, it has prepositional case marking (accusative, dative and genitive) as well as nominative default case that never appears as prepositional case. The nonsentential account makes in principle the same predictions as the PSG does for PP and prepositional case-marked DP fragments: PP fragments (32a) are expected to be acceptable and prepositional case-marked DP fragments (32b) to be degraded. However, the theories disagree on the acceptability of default case DP fragments (32c). The nonsentential account predicts them to be acceptable, because default case does not need to be checked and the preposition can be easily retrieved from the question. Any sentential account however predicts in principle that the form of the answer matches that of the question, so nominative DP fragments will be degraded as compared to PP fragments.

- (32) Für wen ist das Päckchen?  
for who.ACC is the package  
'For whom is the package?'
- a. Für meinen Vater. PP  
for my.ACC father  
'For my father.'
- b. Meinen Vater. DP, prepositional case  
my.ACC father  
'My father.'
- c. Mein Vater. DP, nominative case  
my.NOM father  
'My father.'

This does neither imply that the nonsentential account predicts nominative DP fragments to be as acceptable as PP fragments, nor that sentential accounts predict nominative DP fragments to be as degraded as prepositional case-marked DP fragments. As for the nonsentential account, Progovac et al. (2006) argue that even though default case DP fragments are grammatical, they might still be dispreferred for pragmatic reasons. They exemplify this with (33), for which they claim that nominative is degraded, because the speaker could have chosen the matching accusative fragment (recall that Progovac et al. (2006) claim that the Serbian accusative is inherent case). Therefore, nominative DP fragments might be worse than PP fragments, but the nonsentential account predicts them to be better than structural case-marked DP fragments, which are blatantly ungrammatical.



- (33) a. Koga je Ana posetila? (Progovac et al. 2006: 340)  
           who.ACC is Ana visited  
           ‘Who did Ana visit?’  
       b. Vera!  
           Vera.NOM

In contrast, according to the sentential account, the acceptability of a fragment depends on the availability of a matching antecedent. For instance, nominative would be acceptable if subjects formed a (highly marked) clefted structure, that requires nominative on the DP, as implicit antecedent. Anticipating the results of the experiment, nominative is rated as even worse than accusative, so this theoretical possibility does not need to be further discussed.

- (34) Es ist mein Vater, für den das Päckchen ist.  
       it is my.NOM father for who.ACC the package is  
       ‘It’s my father, for whom the package is.’

#### 3.2.4.2 Materials

The stimuli were identical to those used in experiment 4 except for the additional nominative DP fragment condition. The three conditions are exemplified in (32) above. As compared to experiment 4, one further item was added in order to present each of the three conditions equally often to the participants.

#### 3.2.4.3 Procedure

The experiment was conducted over the Internet using the LimeSurvey presentation software and completed by 48 participants recruited on the *clickworker* crowdsourcing platform. Each participant was paid € 4.00 for their participation. Subjects were asked to rate the naturalness of the italicized short answer fragment in the context of the question on a 7-point Likert scale (7 = fully natural). Materials were mixed with 24 items from experiment 11 and 44 fillers. Both the materials from experiment 11 and the fillers resembled the items of experiment 6 in having a context story and an italicized target utterance which subjects rated. Subjects were assigned to one of six lists, to which materials were allocated by a Latin square so that each subject saw each token set only once and each condition equally often. Each two of these lists contained the same materials for experiment 6, but differed with respect to the materials from experiment 11. All lists were presented in an individually pseudo-randomized order that ensured that no two items or fillers of the same category immediately followed each other. Three

### 3 Experiments on the syntax of fragments

subjects rated more than the previously established threshold of two out of five ungrammatical controls as natural (6 or 7 points) and thus were excluded from further analysis.

#### 3.2.4.4 Results

Table 3.11 summarizes the ratings for the three conditions. The ratings for PP and prepositional case-marked DP fragments replicate the previous studies. Nominative DP fragments are perceived as even less acceptable than prepositional case-marked DPs.

Table 3.11: Mean ratings (standard deviation) by condition in experiment 6, in experiment 4 and in the P-stranding study by Merchant et al. (2013). *Structural case* refers to prepositional case in experiment 6 and accusative in the other two studies.

| Condition           | Exp. 6      | Exp. 4      | Merchant et al. (2013) |
|---------------------|-------------|-------------|------------------------|
| PP                  | 6.31 (1.18) | 6.61 (0.99) | 5.99 (1.64)            |
| DP, Structural case | 4.81 (1.99) | 4.42 (2.05) | 4.76 (2.03)            |
| DP, Nominative      | 3.57 (1.96) | –           | –                      |

The data were analyzed with CLMMs in R following the procedure described in Section 3.1.1.5. In this case, the procedure slightly differed from that applied to previous studies, because the IV was a nominal-scaled factor with three levels. The likelihood ratio tests used for model selection only allow for the comparison of two models containing or lacking a factor and but not for pairwise comparisons between the individual levels. Therefore, I created two subsets from the complete data set in order to compare the factor levels pairwise. I first compared only the PP to the prepositional case-marked DP fragments, thus replicating experiment 4. Then I tested whether the nominative and prepositional case-marked DP fragments differed significantly in acceptability by analyzing these conditions only. This procedure allows for pairwise comparisons between factor levels and not only for testing whether including a factor as a whole improves model fit. For both subsets I started with a full model containing fixed effects for FRAGMENTTYPE, the POSITION of the trial in the experiment and their interaction and by-item and by-subject random intercepts and random slopes for each predictor. The final models (see Tables 3.12 and 3.13) show that all contrasts between the levels of FRAGMENTTYPE are significant. PPs are rated significantly better than case-marked DPs ( $\chi^2 = 29.18, p < 0.001$ ) and nominative case-marked DPs are

even worse than prepositional case-marked DPs ( $\chi^2 = 15.37$ ,  $p < 0.001$ ). In the model that compared PPs to prepositional case-marked DP fragments there was also a significant POSITION effect, which did not interact with FRAGMENTTYPE.

Table 3.12: Fixed effects in the final model comparing PP fragments to prepositional case-marked DP fragments.

| Predictor    | Estimate | SE    | $\chi^2$ | $p$    |     |
|--------------|----------|-------|----------|--------|-----|
| FRAGMENTTYPE | 2.58     | 0.423 | 15.37    | <0.001 | *** |
| POSITION     | 0.01     | 0.004 | 4.34     | <0.05  | *   |

Table 3.13: Fixed effects in the final model comparing prepositional case-marked DP fragments to nominative DP fragments.

| Predictor    | Estimate | SE    | $\chi^2$ | $p$    |     |
|--------------|----------|-------|----------|--------|-----|
| FRAGMENTTYPE | 1.838    | 0.414 | 29.18    | <0.001 | *** |

#### 3.2.4.5 Discussion

Experiment 6 tested whether nominative DP fragments are more acceptable than prepositional case-marked DP fragments, and how they are rated in comparison to PPs. Both the nonsentential account (Barton & Progovac 2005) and the movement and deletion account predict a preference for nominative DPs over case-marked ones, but for independent reasons. According to the nonsentential account, prepositions can be omitted in short answer fragments only if they are not required for checking prepositional case. The nonsentential account predicts fragments in nominative default case to be acceptable under such circumstances. According to the movement and deletion account, DP short answers might be grammatical as answers to PP questions in German when they can be derived from a cleft. In that case, the DP must exhibit nominative case morphology.

The experiment disconfirms these predictions: Nominative DP fragments are significantly less acceptable than prepositional case-marked ones. PP short answers are even more strongly preferred than prepositional case-marked DPs. This finding challenges the cleft-based analysis of apparent preposition omission under ellipsis in languages that lack P-stranding. In a language with overt case marking, like German, such an account predicts that DP fragments derived

### 3 *Experiments on the syntax of fragments*

from clefts by ellipsis appear in nominative, just like in the corresponding full sentences. Even though the movement and deletion account does not predict that the resulting nominative DP fragments are as acceptable as PPs (they might be degraded for pragmatic reasons), they are expected to be more acceptable than prepositional case-marked DPs, which it analyzes as being derived only by ungrammatical P-stranding. This prediction is clearly disconfirmed by the experiment, since this predicted acceptability pattern is inverted in the data.

From the perspective of the nonsentential account, the contrast to the English data from experiment 5, which revealed a preference for preposition omission, is particularly striking. If the explanation for the English pattern is that the preposition can be omitted because it is given in the question and the resulting nominative DP fragment does not require case checking, the same pattern is expected for German default case (nominative) DP fragments. Therefore, experiment 6 strongly suggests that the nonsentential case checking account should be rejected. Sentential accounts predict connectivity effects between question and answer, and consequently are in line with the preference for realizing the preposition in the answer as well.

The gradual acceptability cline between the three short answer types is in line with the idea that fragments are ungrammatical but can be interpreted after applying a probabilistic repair mechanism, as Bergen & Goodman (2015) suggest. The prepositional case-marked DP fragments are clearly dispreferred in all three experiments and are therefore degraded in the context of PP questions in German. Nonetheless, prepositional case can function as a probabilistic cue that points toward the preposition that is missing. Since prepositional case is determined by the preposition, processing a dative DP will restrict the range of possible prepositions to those requiring dative. Nominative DP fragments could be rated worse because they lack this cue. Under this perspective, both case-marked and default case DP fragments are ungrammatical, because they lack an appropriate antecedent. The acceptability difference between these ungrammatical utterances could be explained by the effort required to figure out which part of the utterance is missing. However, the comparison between English and German in experiments 4 and 5 shows that the recoverability of the preposition cannot be the whole story. Its omission is less acceptable in German despite the fact that English lacks prepositional case as a cue toward in the omitted preposition. Consequently, the preference for prepositional case in comparison to default case might be due to differences in recoverability of the omitted preposition, but the inverted preference for DP and PP short answers between German and English must receive a different explanation.

Taken together, experiment 6 disconfirms the case checking-based account that follows from the discussion on preposition omission in Barton & Progovac (2005). The experiment also showed that nominative DP fragments, which might be derived from cleft structures according to movement and deletion are more strongly degraded than presumably ungrammatical DPs in prepositional case. The relative acceptability of these DPs as compared to nominative ones, which all of the theories can derive, suggests that they might not be fully ungrammatical but dispreferred for independent reasons. Experiment 7 addresses this issue.

#### 3.2.5 **Experiment 7: Question-answer parallelism**

##### 3.2.5.1 **Background**

Experiment 7 tests the hypothesis that the availability of P-stranding and of preposition omission in a language often cooccur due to a nonsyntactic relationship between question and answer. If this hypothesis were confirmed, the data that Merchant (2004a) interprets as evidence for movement in fragments could be explained without assuming that movement is required as an explanatory link between the acceptability of fragments and the availability of P-stranding.

The data discussed so far are in line with the PSG: Experiments 4 and 5 confirm its predictions for English and German and experiment 6 rules out the alternative nonsentential account based on case checking. However, parallelisms between fragments and movement constructions support the movement and deletion account only if derivationally simpler theories, like the nonsentential and in situ deletion accounts, cannot explain the data as well. In the case of the PSG, there are at least three possible explanations for the data: First, as Merchant (2004a) argues, they could of course evidence a genuine movement restriction. Second, there could be an independent factor that blocks the omission of the preposition in languages that disallow P-stranding. The nonsentential case checking hypothesis that I tested in the previous section is an example for this line of reasoning. Even though I rejected this explanation, the impossibility of omitting prepositions in German (as compared to English) in in situ contexts might still evidence such a different independent factor other than the one I tested in experiment 6.<sup>33</sup>

---

<sup>33</sup>For instance, Zwarts (2005: 21) claims that preposition and case are “not two semantically independent elements” in German, but that they are interpreted together. Another such factor could be crosslinguistic differences with respect to feature percolation. If a focus feature percolated from the DP to PP obligatorily in German but not in English, e.g. due to PP-internal movement operations required for case checking, the in situ deletion account could explain why the omission of the preposition is blocked in languages that do not allow for P-stranding.

### 3 Experiments on the syntax of fragments

A third possibility, which experiment 7 explores, is that there is no structural difference between English and German PPs that blocks the omission of the preposition in fragments, but that the form of the *wh*-phrase in the question constrains that of the answer: If the preposition is pied-piped in the question, it must be realized in short answers, and if it is stranded in the question, it is omitted in the answer. From this perspective, preposition omission in German short answers is not blocked by properties of the answer, but dispreferred because of the impossibility of stranding the preposition in the question. In principle, both forms of the answer are possible, but one of them is strongly preferred for non-syntactic reasons.

Experiment 7 tests this hypothesis by eliciting answers to questions with pied-piping and P-stranding in a production study in English, which allows for both forms of the answer. If subjects adapt their answer to the question, they must produce a higher ratio of preposition omission when the preposition is stranded the question and realize it more often in the answer when it is pied-piped.

This raises the question of why such a mechanism would be observed at all. In the theoretical literature, there are at least two possible explanations that predict a nonsyntactic question-answer parallelism: First, a structured propositions account of question semantics (von Stechow 1981, Reich 2002b) relates the focus structure of question and answer. Second, structural persistence (Nykiel 2017) between question and answer could result from speakers' tendency to reuse structure from previous discourse (Levelt & Kelter 1982, Nykiel 2017).

#### 3.2.5.1.1 Structured propositions

The central idea of the structured propositions account of question-answer parallelism is that the focus structure of the answer is determined by that of the question (Reich 2002a,b, 2007).<sup>34</sup> If the preposition in the question is focused, it must also be focused in the answer and therefore cannot be omitted there. If the preposition is not focused in the question, it also not in the answer, and consequently it can be targeted by ellipsis.<sup>35</sup>

<sup>34</sup>Note that this might also be expected under a movement and deletion account. However, if it were the case, the prediction of movement and deletion and in situ deletion with respect to preposition omission would be fully aligned: Both theories predict that only words which belong to the focus survive ellipsis, and no matter whether they are previously moved, the outcome is identical. In that case, the PSG would not provide genuine evidence for movement. As I showed in Section 3.2.1.2, it does only if the focus structure of pied-piping and P-stranding questions (and the corresponding) answers is identical and pied-piping occurs because extraction out of PP is ungrammatical.

<sup>35</sup>See Griffiths (2019) for a similar account of the PSG data.

Reich (2002b) models the semantics of questions as a set of structured propositions, which are sensitive to focus structure (von Stechow 1981). For instance, Reich (2002b: 82) defines the semantics of (35a) as denoting the set of propositions in (35b), which is summarized as (35c). The idea is that, instead of defining the semantics of (35a) as a set of unstructured propositions (36), the focus, i.e. the *wh*-phrase in (35), is separated from the background of the question. Congruent answers must match this focus-background structure.

- (35) a. What did John drive? (Reich 2002b: 82)  
b.  $\{\langle \text{Mary's red convertible}, \lambda x. \text{John drove } x \rangle, \langle \text{Peter's Porsche}, \lambda x. \text{John drove } x \rangle, \dots \}$   
c.  $\lambda p \exists x [\text{thing}'(x) \ \& \ p = \langle x, \lambda y. \text{John drove } y \rangle]$
- (36)  $\lambda x. \text{John drove } x$

Reich (2002b) does not address pied-piping, but in Reich (2002a) he develops a structured propositions analysis of complex *wh*-phrases. He assumes that the pied-piped material belongs to the focus in complex *wh*-phrases like *whose book* in (37a), whose semantics he defines as (37b). He argues that an answer that does not match this focus-background structure is incongruent, because it is not included in the denotation of the question. Applied to pied-piping of prepositions, the semantics of a question like (38a) would be defined as (38b), whereas that of the corresponding P-stranding question is given in (39). This requires congruent answers to (38) and (39) to differ in their focus structure: (40a) is a congruent answer to (38), but (40b) is not. In the case of a P-stranding question like (39), the opposite holds.

- (37) a. Whose book did you read? (Reich 2002a)  
b.  $\{\langle \langle x, \lambda z. z's \text{book} \rangle, \lambda y. \text{you read } y \rangle; x \text{ is a person} \}$
- (38) a. With whom did you talk?  
b.  $\{\langle \langle x, \lambda z. \text{with } z \rangle, \lambda y. \text{you talk } y \rangle; x \text{ is a person} \}$
- (39) a. Who did you talk with?  
b.  $\{\langle \langle x, \lambda y. \text{you talk with } y \rangle; x \text{ is a person} \}$
- (40) a. I talked [with John]<sub>F</sub>.  
b. I talked ~~with~~ [John]<sub>F</sub>.

The assumption that relatively subtle differences in meaning can impact on the form of fragments has recently been reinforced by Weir (2018). He observes that despite the questions in (41) and (42) being relatively meaning-equivalent, fragments that do not match the semantics of the *wh*-phrase are heavily degraded.

### 3 Experiments on the syntax of fragments

- (41) Q: How many signals did the machine send? (Weir 2018: 1289)
- a. (i) It sent *two* signals.
  - (ii) ?It sent a signal *twice*.
  - b. (i) *two* (signals).
  - (ii) \**twice*.
- (42) Q: How many times did the machine send a signal? (Weir 2018: 1289)
- a. (i) ?It sent *two* signals.
  - (ii) It sent a signal *twice*.
  - b. (i) \**two* (signals).
  - (ii) *twice*.

The structured propositions approach offers a straightforward explanation for the contrast between German and English with respect to the acceptability of preposition omission in fragments. Unlike English, German lacks P-stranding in questions, hence questions always involve pied-piping and only answers where the complete PP is focused are congruent. As the central assumption of the in situ deletion account is that focused expressions survive ellipsis, the preposition can never be omitted in those contexts. With respect to English, where both P-stranding and pied-piping are available, the structured propositions approach predicts a relatively strict congruence between the form of the question and that of the answer. As the contrast between (41) and (42) suggests, there should be a strong preference for the answer to match the form of the question: Pied-piping questions are expected to require PP short answers, whereas P-stranding questions require DPs.

#### 3.2.5.1.2 Structural persistence

The second account of question-answer parallelism is based on processing. The idea is that both DP and PP fragments can be derived by the syntax in the context of PP questions in English and German, but that a tendency for speakers to reuse syntactic structure from previous discourse explains why short answer fragments often match the form of the preceding question. This has been proposed by Nykiel (2017), who traces back the observation of a tendency to reuse syntactic structure from previous discourse to Levelt & Kelter (1982). Levelt & Kelter (1982) conducted a series of experiments that investigated how and why speakers reuse structure in the example of the optional omission of prepositions in Dutch questions and answers like (43), from Levelt & Kelter (1982: 80). For this language, they argue that the preposition *aan* ‘to’ can be freely omitted in the question and the short answer fragment without changing its meaning.



- (43) a. (Aan) wie laat Paul zijn viool zien?  
to whom allows Paul his violin see  
'Who allows Paul to see his violin?'
- b. (Aan) Toos.  
to Toos  
'Toos.'

Throughout their experiments, Levelt & Kelter (1982) find an effect of the form of the question on the form of the answer: The preference for omitting the preposition in the answer is stronger when it is been omitted in the question and vice versa. Note that (43) does not involve P-stranding, but the general idea straightforwardly applies to P-stranding data. If speakers reuse the structure given in a P-stranding question, they should prefer DP fragments, whereas PP fragments would be preferred in case of pied-piping questions.

This prediction is supported by English corpus data (Nykiel 2014, 2017). Nykiel (2017) shows that despite an overall preference for omitting the preposition in the remnant,<sup>36</sup> the rate of DP fragments (90%) is significantly higher when the preposition is stranded or omitted than when the preposition is pied-piped or appears adjacent to its object (58.8%). Even though her corpus studies are concerned only with English data, Nykiel (2017: 41–42) argues that her observations for English also provide an explanation for the crosslinguistic pattern: If the form of the antecedent (e.g. the PP in the question) determines the form of the fragment, DP short answers to questions where the *wh*-phrase is the complement of a preposition are only possible when there are DP antecedents. In German there are never such antecedents because German has no P-stranding. Consequently, the preposition is never omitted in the answer.

### 3.2.5.1.3 Predictions of question-answer parallelism

Both the structured propositions and the structural persistence accounts provide a non-movement explanation for the data that Merchant (2001, 2004a) presents as evidence for the PSG. Besides explaining the crosslinguistic data, such an approach predicts that within a language that allows for an alternation between P-stranding and pied-piping, the form of the answer will match that of the question, like Levelt & Kelter (1982) showed for preposition omission in Dutch.

Since the goal of experiment 6 is to investigate whether question-answer parallelism can explain the preposition omission data without having to assume

<sup>36</sup>Nykiel (2017) investigated a more extensive range of antecedents, instead of looking only into question answer pairs. In the case of the latter, the remnant is the short answer.

### 3 Experiments on the syntax of fragments

movement in fragments, the experiment does not need to differentiate between the semantic and the processing accounts. However, the structured propositions account predicts a relatively strict match between question and answer, whereas this relationship might be looser if the structural persistence account is correct. From a semantic perspective, if focusing the preposition in the question blocks its omission in the answer, the form of the question would strictly determine that of the answer. The tendency to reuse structure might interact with and be overridden by other constraints, such as the tendency to be brief and to omit redundant material.<sup>37</sup> The conclusions on this issue will only be tentative.

In contrast to question-answer parallelism, the movement and deletion account does not necessarily predict a correlation between the form of the question and the answer: As I discussed above, if the alternation between DP and PP short answers is traced back to different focus structures, it does not provide evidence specifically for movement. Consequently, movement and deletion implies that P-stranding and pied-piping questions do not differ with respect to their focus structure. It is important to note that a syntactic theory like the movement and deletion account does not conflict with the assumption of processing constraints, like the structural persistence account. Processing constraints can determine the choice for a particular utterance when grammar allows for various options. However, if an independently evidenced processing constraint can explain the data that the syntactic theory was designed to account for, this syntactic theory yields no explanatory benefit over simpler theories. In that case, the preposition omission data would lose their status as evidence for movement in fragments.

#### 3.2.5.1.4 Approach

Experiment 7 uses a production task to investigate whether subjects adapt answers to the form of the question. I conducted the study in English, which allows for both P-stranding and pied-piping. There is some previous experimental evidence that points into this direction: The corpus studies by Nykiel (2014, 2017) and the experiments by Levelt & Kelter (1982) suggest that there is question-answer parallelism, but Levelt & Kelter (1982) investigate a related yet different phenomenon and Nykiel (2014, 2017) considers very diverse antecedents and remnants, such as interrogative fragments and elliptical questions. An experimental study allows for controlling this variance and for testing only pied-piping and P-stranding questions, which are the relevant antecedents given the above discussion on question-answer parallelisms.

---

<sup>37</sup>See Section 4.2.2 for discussion.

In the experiment, subjects read a context story followed by a question with a stranded (44a) or a pied-piped (44b) preposition and are asked to produce a natural answer to that question. Question-answer parallelism predicts a (relatively) higher rate of DP fragments for P-stranding questions (44a) and a higher rate of PP fragments for questions with pied-piping (44b).

- (44) Molly and Cooper are colleagues and talk about football during a break. Because this evening there is an important match, Cooper asks Molly:
- a. Who are you rooting for? P-stranding
  - b. For whom are you rooting? Pied-piping

If subjects provide sentential answers, I expected them to follow the unmarked SVO word order in (45a), because experiment 5 suggests that non-contrastive object fronting (45b,c) is at least highly marked, if not ungrammatical, in English.

- (45) a. I'm rooting for the Packers.  
 b. \*For the Packers, I'm rooting.  
 c. \*The Packers, I'm rooting for.

#### 3.2.5.2 Materials

All materials followed the pattern given in (44): A short context story consisting of two sentences introduced two characters and was followed by a question asked by one of these characters. This question was always a *wh*-question, where the *wh*-phrase was the complement of a preposition. The question was presented in one of two conditions, P-stranding (44a) and pied-piping (44b).

I investigated three different types of questions, which differ in the status of the PP with respect to the verb. The reason for this is that the choice between the two constructions is not fully unconstrained, but depends on syntactic properties of the PP. For instance, van Riemsdijk (1978: 26) argues that P-stranding is not possible in adjunct PPs, and Nykiel (2017) shows that pied-piping is less frequent the stronger the semantic connection between verb and preposition is. Investigating the reason underlying these contrasts is beyond the goal of the experiment,<sup>38</sup> but if P-stranding was blocked or triggered by some syntactic property of the PP and this remained uncontrolled it could mask effects of the form of the question. Therefore, I tested questions with non-locative PPs which are sub-categorized by the verb (44), locative complement PPs (source / goal / location) (46a) and adjunct PPs (46b). This will (i) show whether the type of the PP affects

<sup>38</sup>But see van Riemsdijk (1978), Chomsky (1981), Pullum & Huddleston (2002) and Nykiel (2017).

the preference for P-stranding or pied-piping, and (ii), if this was the case, these preferences can be taken into account in the statistical analysis.

- (46)    a.    Where do you come from?                                  Locative  
             b.    For whom did you buy it?                              Adjunct

Finally, note that the difference between adjunct and complement is also potentially relevant to the movement and deletion account. If there are movement restrictions on some PPs in English, Merchant (2004a) predicts the fragments derived via illicit movement operations to be less acceptable and therefore to be only rarely produced.

### 3.2.5.3 Procedure

53 self-reported native speakers of American English were recruited on the *Pro-lific Academic* crowdsourcing platform. The experiment was conducted over the Internet and presented using the LimeSurvey survey presentation software. Subjects were rewarded £2 for participating. The task consisted in reading the context story and the question and entering the answer that the subject considered to be most natural into a text field. Form and meaning of the answer were totally unconstrained apart from this; specifically, subjects were told to produce an utterance, but there was no restriction with respect to sententiality. Subjects were assigned to one of two lists, to which the materials were distributed with a Latin square. There were 24 items, eight of which had locative PPs, eight adjunct PPs and eight argument PPs. Materials were balanced across lists, so that each subject saw 12 items per QUESTION (P-stranding/pied-piping) condition and within those 12 items per condition there were four of each of the PP types (argument / adjunct / locative). Materials were mixed with 19 items of an unrelated experiment and 25 unrelated fillers and presented in individually pseudo-randomized order that ensured that no two items of the same experiment immediately followed each other. All fillers resembled the items in requiring subjects to produce an answer to a question. The *wh*-phrase was the complement of a preposition in these questions.

The answers were annotated manually. First, it was recorded whether the answer was a direct answer to the question or not. An answer was classified as a direct answer when it contained or consisted in a DP or PP that corresponded to the *wh*-phrase in the question. This excludes cases such as (47).

- (47) Who are you rooting for?  
a. I don't really care, I just go for the excitement of it all.

- b. I'm a big Pats fan.
- c. Not sure.

The restriction to direct answers excluded a total of 20.1% of the data. Direct answers were then annotated for two further features. First, I tracked whether the answer was a complete sentence or a fragment. Second, it was annotated whether the preposition was omitted or realized in fragments and whether it was realized in situ, pied-piped or stranded in sentences.

### 3.2.5.4 Results

Figure 3.7 gives an overview of the complete data set. Within the direct sentential answers there was no structural variation at all: As I expected, the PP always occurred in its postverbal base position and there was no instance of pied-piping or P-stranding. Therefore, I restricted the further analysis to fragments. Across all conditions, direct fragment answers that could be statistically analyzed made up 55.3% of the complete data.

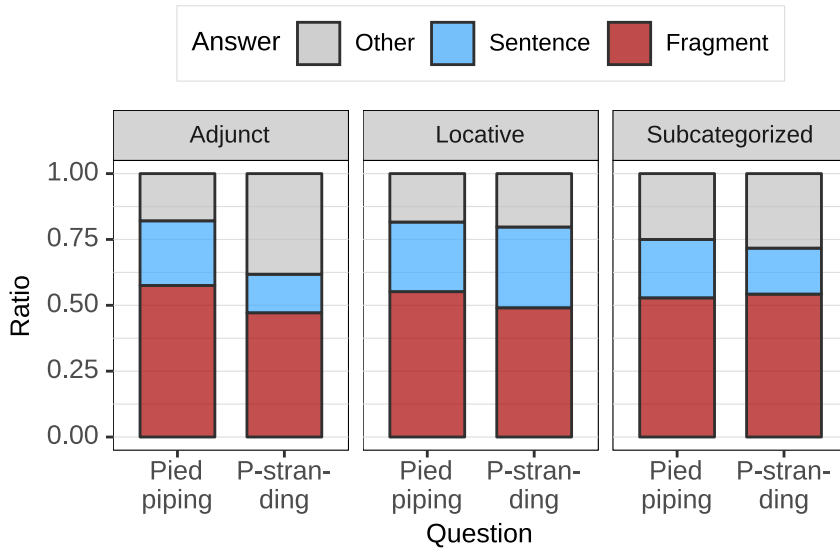


Figure 3.7: Ratios of answer categories in exp. 4. “Other” indicates indirect answers or constructions not involving P-stranding/Pied-piping.

Figure 3.8 gives an overview of the direct fragment answers by condition and answer type. Across all conditions the preposition was omitted more often in the answer (82.4%) than it was realized, both when the answer preposition was

### 3 Experiments on the syntax of fragments

stranded (84.1%) in the question and when it was pied-piped (80.8%). In order to test whether the form of the *QUESTION* and the *PPTYPE* (locative / adjunct / subcategorized) had an effect on the likelihood of preposition omission in the answer, the data were analyzed with logistic mixed effects regressions fitted with the *lme4* (Bates et al. 2015) package in R following the procedure described in Section 3.1.1.5. The regressions predicted the likelihood of the *OMISSION* of the preposition in the answer. Ten subjects who produced less than five direct fragment answers were excluded from the analysis, this resulted in the loss of a further 3.3% of the remaining data.

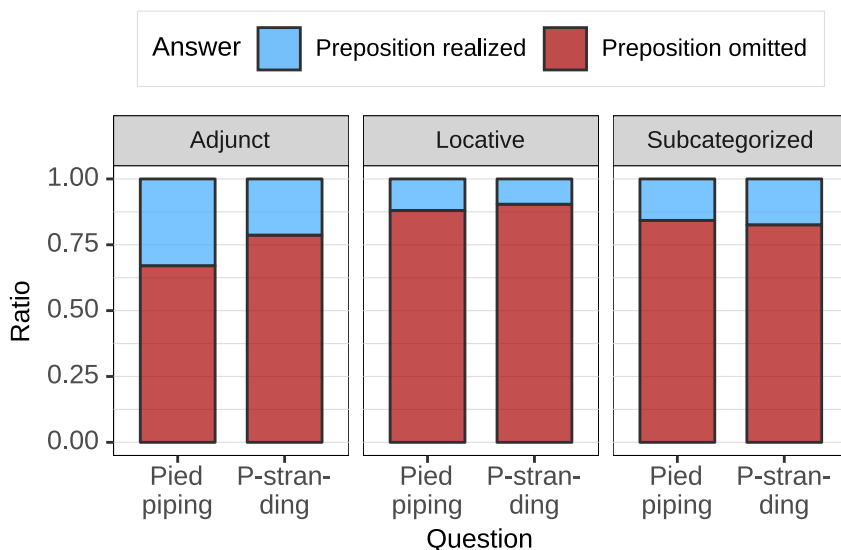


Figure 3.8: Ratio of preposition omission/realization across conditions.

*PPTYPE* was a ternary factor, therefore the data had to be analyzed by conducting pairwise comparisons between factor levels just like in experiment 6. A first analysis compared the two types of complement PPs, i.e. locative and subcategorized PPs, which each other. The initial model contained main effects for *QUESTION*, *PPTYPE*, *POSITION* (numeric) and all two-way interactions between these predictors. The model had only random intercepts for subjects and items, because it did not converge with a more complex random effects structure. As the difference between locative and subcategorized PPs did not turn out to be significant ( $\chi^2 = 0.57$ ,  $p > 0.5$ ), these conditions were pooled for further analysis and hence compared only adjunct to complement PPs.

After pooling locative and subcategorized PPs, the complete data set could be analyzed at once. The initial model contained main effects for *QUESTION*, *PPTYPE*

(now binary), POSITION and all two-way interactions between these predictors. The model had only random intercepts for subjects and items and a by-subject random slope for PPTYPE, because it did not converge with a more complex random effects structure. The final model (see Table 3.14) contains significant effects for both IVs: The preposition in the short answer fragment is more likely to be omitted when the PP is a complement than when it is an adjunct ( $\chi^2 = 5.23$ ,  $p < 0.05$ ). What is more important with respect to the goal of the experiment is that the preposition is also more likely to be omitted in the answer when it is stranded in the question ( $\chi^2 = 4.85$ ,  $p < 0.05$ ). There is no significant interaction between the IVs ( $\chi^2 = 0.86$ ,  $p > 0.3$ ).

Table 3.14: Fixed effects in the final GLMM for experiment 7.

| Predictor | Estimate | SE    | $\chi^2$ | $p$   |   |
|-----------|----------|-------|----------|-------|---|
| QUESTION  | 0.719    | 0.308 | 4.85     | <0.05 | * |
| PPTYPE    | 1.822    | 0.727 | 5.23     | <0.05 | * |

### 3.2.5.5 Discussion

Experiment 7 had the purpose of testing the nonsyntactic explanation for the crosslinguistic coincidence between the availability of preposition omission in short answers and of P-stranding in questions: Speakers tend to match the form of the answer with that of the question, therefore, if a language has no P-stranding in questions, preposition omission in the answer will be degraded. The experiment supports this hypothesis: If the preposition is stranded in the question it is significantly more likely to be omitted than when it is pied-piped. This is the result that question-answer parallelism predicts.

In absolute terms, the effect seems to be less pronounced than the one found by Nykiel (2017) in her corpus study. This could be in part due to the restriction to P-stranding and pied-piping questions as antecedents, whereas Nykiel (2017) investigated a larger variety of antecedents and remnants. Furthermore, the experimental design might have contributed to reducing the effect of the antecedent. First, although they were not told to do so, a few participants noted that some of the questions were not totally natural by adding a comment like “the question sounds odd” in the text field.<sup>39</sup> As I discussed above, P-stranding is preferred in

<sup>39</sup>These responses were not analyzed.

### 3 Experiments on the syntax of fragments

colloquial speech (recall that the experimental stimuli were presented as informal dialogues) for complement PPs and dispreferred for adjunct PPs. Therefore, in each condition one of the questions is not perfectly natural, and in natural settings there is probably a higher ratio of P-stranding in questions for complement than for adjunct PPs. Since the form of the question affects that of the answer, in corpus data this would probably increase the difference in counts of each of the answer variants as compared to the more controlled setting of the experiment. Second, even though the POSITION effect is not significant, subjects produced a relatively large amount of answers throughout the experiment and probably got used to the task and matched their answers to a lesser degree to the form of the question due to familiarization. This is tentatively supported by the observation that parallelism is most pronounced for the first item that each subject saw: Subjects who saw a complement PP in the question almost always omitted the preposition in the answer (90% omission rate for pied-piping and 90.9% for P-stranding), but only 16.7% of those who saw a pied-piped adjunct PP in the question and all of those who saw a P-stranded one did so. These numbers are not significant due to the reduced number of observations, but they suggest that a crowdsourced experiment with only one trial per subject might be a promising option to prevent such a familiarization effect.

The preference for preposition omission in all conditions is in line with the English corpus data in Nykiel (2017). She also reports that the preposition is omitted more often than it is realized both when it is pied-piped and stranded in the antecedent. This overall preference for preposition omission is unexpected under the structured propositions account of question-answer parallelism. If pied-piping occurs because the preposition belongs to the focus of the question, and the focus structure of the answer is determined by that of the question, one would theoretically expect a perfect match between both. Such an effect can be reduced in experimental settings, but pied-piping in the question does not result in an inversion of participants' preferences in any of the conditions. The data are therefore tentatively more in line with the structural persistence account. From a processing perspective, competing constraints introduce a probabilistic bias that can be overridden by others. Speakers might tend to omit the preposition because it is given, and this constraint might cancel out part of the effect of the tendency to reuse material given in context. Furthermore, the effect of constraints that are specifically relevant for oral on-line communication might be less prominent in experimental settings.<sup>40</sup> Recall also that the relative acceptability of proper noun

---

<sup>40</sup>See e.g. Zhan et al. (2017), who did not find effects of audience design in a production study, even though such effects had been attested in related previous work.



DP fragments as short answers to PP questions (Lemke forthcoming) also suggests that some deviation from question-answer congruence is possible even in German. This is unexpected under a semantic account.

The main effect of PPTYPE is in line with the observation by van Riemsdijk (1978) that P-stranding is less acceptable in adjunct than in complement PPs. However, this tendency does not override the preference for omitting the preposition in the experiment: The ratio of omitted prepositions is indeed the lowest observed throughout the experiment if the PP is an adjunct and the preposition pied-piped in the question (see Figure 3.8). Still, even in this case, 67% of prepositions are omitted. This might suggest that the syntactic relationship between the preposition and the verb at least affects preposition omission in fragments.

Taken together, the production study supports a nonsyntactic question-answer parallelism. If the form of short answers follows that of questions, the absence of P-stranding in German questions explains straightforwardly why DP fragments are degraded in such contexts. The experiment does not allow for a conclusion on why we observe this parallelism, but the strong overall preference for omitting the preposition in the answer seems to be more in line with a processing account than with the structured propositions approach. This finding does not falsify the movement and deletion account. Movement and deletion is neither incompatible with the assumption of differing focus structures between pied-piping and P-stranding questions nor with structural persistence. However, both of these accounts explain the pattern observed for fragments without assuming movement as an obligatory step in the derivation of fragments. This clearly weakens the status of the PSG as genuine evidence for movement.

#### 3.2.6 **General discussion: Preposition omission**

In section 3.2 I presented four experiments that tested the predictions of the PSG, which is taken to be one of the central pieces of evidence for movement in fragments, and the viability of non-movement explanations for the data. Experiments 4 and 5 empirically support the predictions of the PSG for German, where prepositions are obligatorily pied-piped in questions, and for English, which allows for P-stranding. Just like the PSG predicts, in German there is a strong preference for realizing the preposition, whereas in English its omission is acceptable, and actually preferred in the context of questions with P-stranding.

A further result of experiment 5 is that fronting the PP in a complete sentence in English is heavily degraded, as has been already noted by Weir (2014b). This questions some of the arguments by Merchant (2004a), which are based on the

### 3 Experiments on the syntax of fragments

idea that the acceptability of fragments patterns with that of left dislocation structures. The low ratings for both pied-piping and P-stranding in the answer suggest that a movement and deletion account is viable only if exceptional movement is assumed, as Weir (2014a) proposes. The exceptional movement account however requires an explanation for why the preposition is sometimes pied-piped in English. If extraction out of the PP is possible, and only the DP complement of P is focused, there is no need to front the complete PP in English. For German this is not a problem, because pied-piping is the only way to extract the focused DP out of the ellipsis site if extraction out of PP is blocked for independent reasons in this language.

Although the experiments 4 and 5 are in line with the PSG, they only provide evidence for movement if explanations under derivationally simpler accounts, such as the nonsentential and in situ deletion accounts, must be ruled out. Experiments 6 and 7 tested two of these alternative explanations.

Experiment 6 investigated the hypothesis that, as suggested by Barton & Progovac (2005), the preposition cannot be omitted in languages with strong case features because prepositional case is structural case and must always be checked. According to their theory, prepositional case cannot be checked in fragments because there is no unarticulated syntactic material that could do so (in this case, a preposition). Instead, they expect DPs to appear in default case. The data clearly disconfirm this prediction: Default case was rated even worse than the significantly degraded prepositional case-marked DPs. Consequently, the case checking hypothesis, at least in the version suggested by Barton & Progovac (2005), must be discarded. Experiment 6 also provides further evidence against the nonsentential account, because presumably ungrammatical prepositional case-marked DP fragments are rated as more acceptable than grammatical, yet possibly pragmatically odd, nominative DP fragments. Furthermore, the experiment questions the cleft-based analysis of preposition-less fragments in languages that lack P-stranding (Szczeżeliński 2008, Rodrigues et al. 2009). In German, fragments derived from clefts must exhibit nominative case morphology, but the experiment shows that nominative DP fragments are degraded not only as compared to PPs but also to presumably ungrammatical prepositional case-marked DP fragments.

Experiment 7 addressed the hypothesis that the crosslinguistic variation found in fragments is the result of a tendency for answers to structurally match the corresponding questions. I discussed two possible explanations for this, one in terms of a structured propositions account of question semantics (Reich 2002a, Griffiths 2019) and one based on a tendency to reuse syntactic structure from previous discourse (Levelt & Kelter 1982). Question-answer parallelism provides

a straightforward account of the P-Stranding Generalization: If the preposition is pied-piped in the question, it is necessarily, or at least preferably, realized in the answer. Since the preposition is always pied-piped in languages like German, it is never omitted in short answers. For languages that allow for both pied-piping and P-stranding, the form of an answer would tend to match that of the question: Pied-piping in the question would yield a relatively higher ratio of PP short answers, and P-stranding more DP short answers. Experiment 7 provides evidence for such a preference using a production task. The experiment could hence replicate the effect observed in a corpus study by Nykiel (2017) and the evidence for structural parallelism by Levelt & Kelter (1982) in a controlled experiment that investigated specifically effects of P-stranding/pied-piping in question on the form of the answer. However, despite a significant effect of the question's form, preposition omission was preferred in all conditions in absolute terms. The parallelism between question and answer seems to be weaker than expected under a structured propositions account, hence an explanation in terms of structural persistence tentatively seems to fit the observed pattern better.

The evidence for question-answer parallelism does not contradict the movement and deletion account, because syntactic theories do not neglect effects of processing constraints but restrict the set of alternative expressions on which such constraints operate. However, experiment 7 evidences that the form of the question affects the form of the answer in English. If it does so in German too, this provides a straightforward non-movement explanation for the preposition omission data: DP short answers are not degraded in German because they are ungrammatical, but because they never match the form of the question. As I argued above, it depends on the viability of non-movement accounts of the preposition omission data whether they constitute evidence for movement or not. The parallelism between question and answer evidenced in my experiment and in Nykiel (2017) provides such an explanation and consequently undermines the status of the preposition omission data as evidence for movement.

Taken together, preposition stranding does not uniquely support the movement and deletion account as strongly as claimed by Merchant (2004a). Specifically, the data can be equally well explained in terms of question-answer parallelism under the assumption of and in situ deletion account. The nonsentential account is of course also compatible with processing constraints, however it predicts that prepositions cannot be omitted if the DP is case-marked. This has been disconfirmed by experiment 6. In the next section I investigate a further movement restriction on complement clause topicalization, which has been argued to hold crosslinguistically in Germanic languages (Webelhuth 1992) and that has already been empirically investigated by Merchant et al. (2013).

### 3.3 Movement restrictions: Complementizer omission

#### 3.3.1 Complementizer omission as evidence for movement

##### 3.3.1.1 Movement restrictions on complement clauses

The second movement restriction that I investigate empirically is the (im)possibility of fronting complementizer-less complement clauses (in what follows, CCs).<sup>41</sup> This restriction is particularly relevant to the movement and deletion account because Merchant (2004a) argues that it constrains the form of fragments and Merchant et al. (2013) present empirical evidence in support of this prediction.

According to Merchant (2004a), it was Stowell (1981) who first noted that only CCs headed by an overt complementizer can appear in a sentence-initial position.<sup>42</sup> Stowell (1981: 396f) observes that this holds for both subject CCs (48a) and topicalized object CCs (48b). (48c) shows that omitting the complementizer in object position is possible, hence the ungrammaticality must be attributed to the sentence-initial position of the CC.<sup>43</sup>

- (48) a. \*(That) the teacher was lying was hardly obvious.  
 b. \*(That) the teacher was lying, Ben already knew.  
 c. Ben knew the teacher was lying.

Merchant (2004a) cites Morgan (1973) with the observation that the same restriction holds for CP fragments, under the condition that the speaker “does not believe or subscribe to [the meaning of the fragment, R.L.]” (Merchant 2004a: 690), i.e. when it is non-factive (49a). The contrast between (49b) and (49c) shows that the complementizer is obligatory only when the CC is topicalized (49c), but that it can be omitted when the CC remains in situ.

- (49) a. What does no one believe? (Merchant 2004a: 690)  
       #(That) I’m taller than I really am.  
 b. No one believes (that) I’m taller than I really am.  
 c. \*(That) I’m taller than I really am, no one believes.

<sup>41</sup>Webelhuth (1992: 83–85) argues that a similar pattern holds across all Germanic languages, the difference being that some do not allow for complementizer omission in situ.

<sup>42</sup>Stowell (1981) in turn attributes the observation to Kayne (1981), but Morgan (1973: 744) makes a similar point even before that.

<sup>43</sup>Stowell (1981: 396) explains the data by arguing that complementizer-less CCs are headed by a phonetically null element, which c-commanded by the verb according to the Empty Category Principle (Chomsky 1981). This is possible only in the complement position, but not when the clause is base-generated in the subject position (47a) or moved to the topic position (47b).

Merchant argues that this challenges in situ deletion accounts of fragments, which must explain why the complementizer cannot be omitted in fragments even though this is possible in full sentences when the CC appears in situ.

Taken together, according to the literature the pattern seems to be robust for non-factive CCs of verbs: Only CCs with overt complementizers may be fronted. If the derivation of fragments involves regular A'-movement, as Merchant (2004a) claims, this predicts fragment CCs to always require overt complementizers. As I repeatedly noted above, if exceptional movement is assumed, it is crucial to determine whether this movement is available "in principle" or not, but Weir (2014a) does not provide criteria that determine whether this is true for a specific movement operation. Since the literature on the phenomenon cites no acceptable instance of this beyond parenthetical uses,<sup>44</sup> it is probably to be classified as not available in principle. Non-movement accounts in turn require an independent explanation for why the complementizer cannot be omitted. However, the pattern is currently only partially empirically supported. Therefore, before any conclusions can be drawn it must be empirically investigated whether it actually holds. This is the goal of the experiments in Section 3.3.

#### 3.3.1.2 Previous experimental evidence

Merchant et al. (2013) present first experimental evidence for an apparent parallelism between the movement restriction on complementizer-less CCs and the form of fragments. In their experiment 1, they tested short answer fragments like (49a) in an acceptability rating study. They find that fragments headed by an overt complementizer ( $\mu = 4.25$  on a 5-point scale, where 5 = perfect) are rated significantly better than complementizer-less ones ( $\mu = 3.73$ ) and interpret this as reflecting a movement restriction on complementizer-less CCs. Since movement and deletion predicts that complementizer omission is ungrammatical, the ratings for these fragments are surprisingly high. Merchant et al. (2013) suggest that this is due to the possibility of interpreting complementizer-less CCs as indirect answers. With an indirect answer, the speaker does not give a congruent answer to the question (as discussed in Section 2.2.1), but provides any piece of information that might help his interlocutor to figure out an answer. For instance,

<sup>44</sup>Webelhuth (1992) argues that the following example is fine, provided an intonational break after the first clause:

- |     |                                                                                                                         |                      |
|-----|-------------------------------------------------------------------------------------------------------------------------|----------------------|
| (i) | [Hans ist krank gewesen] hat Peter gemeint.<br>Hans is sick been has Peter meant<br>'Peter thought Hans had been sick.' | (Webelhuth 1992: 89) |
|-----|-------------------------------------------------------------------------------------------------------------------------|----------------------|

### 3 Experiments on the syntax of fragments

in (50) Mary does not commit herself to the claim that the defeat is the reason for John being angry, but she suspects that the defeat might be the reason for John's anger.

(50) Bill: Why is John so angry?

Mary: (I'm not sure, but) Barcelona lost to Liverpool yesterday.

The study by Merchant et al. (2013) leaves open several issues: (i) Acceptability ratings were collected only for fragments, (ii) some items include CCs of prepositions, (iii) some of the matrix verbs are factive, and (iv) they do not explore alternative explanations that do not imply movement for the data. In what follows I review these issues in greater detail.

First, Merchant et al. (2013) tested the acceptability of fragment CCs but not of corresponding topicalization structures. The authors assume that the introspective contrast between (49b) and (49c), which has been repeatedly cited in the literature since Stowell (1981), accounts for the measured acceptability of the corresponding fragments. However, some native speakers of American English that I consulted do not share the grammaticality judgments that Merchant et al. (2013) assign to the topicalized CCs. Furthermore, Featherston (2007) notes that introspective data sometimes do not generalize to a larger population and withstand empirical investigation despite of being widely agreed on and repeatedly cited in the theoretical literature. The validity of the pattern in (49) however is crucial to the experiment by Merchant et al. (2013): If it was not confirmed, the judgments for fragments could not be attributed to those for the presumably underlying left dislocation structures. This calls for an empirical investigation of both the fragments and corresponding left dislocation structures.

Second, half ( $n = 8$ ) of the items tested by Merchant et al. (2013) involve CCs which are embedded under a PP, like (51). These CCs are always ungrammatical in situ (51a), whereas they are acceptable when the complementizer is present in a fronted position (51b) and as a fragment (51c). Even though Merchant (2004a) argues that this speaks against the in situ deletion account, actually it does not: In situ deletion takes regular grammatical sentences as the input for ellipsis and analyzes ellipsis as a post-spellout phenomenon. If movement of the CC is obligatory for whatever reason, it has to occur before ellipsis, so that the only grammatical input for in situ deletion is (51b) with an overt complementizer.

(51) What are you ashamed of?

- a. \*I am ashamed of (that) I ignored you.
- b. \*(That) I ignored you, I am ashamed of.
- c. \*(That) I ignored you.

Third, among the remaining eight items, some contained factive matrix verbs, like *to regret* in (52). Factive verbs, which presuppose the truth of their complement, as *to regret* or *to conceal* do, are widely assumed to require, or at least strongly prefer, CCs with overt complementizers (see Kiparsky & Kiparsky 1970; Hegarty 1992). Merchant (2004a: 689) himself cites a related observation by Morgan (1973) that complementizers may not be omitted when the speaker “does not believe or subscribe” to the content of the CC. Consequently, if the verb disprefers or even disallows complementizer-less CCs in general, any structure derived from it will be degraded, independently of whether the CC is moved to the left periphery or whether the matrix clause is PF-deleted in situ.

- (52) What did John regret? (Merchant et al. 2013: 31)  
(That) he joined the Navy.

Finally, there are potential non-movement explanations for both the empirically observed fragment data and the introspective judgments on topicalization. As for fragments, a complementizer-less answer can be interpreted both as a direct and as an indirect answer to a question, whereas a complementizer unambiguously marks the answer as direct. Therefore, the ratings reported by Merchant et al. (2013) possibly do not reflect grammaticality, but usage preferences. In what refers to complementizer omission in full sentences, realizing the complementizer in fronted clauses could facilitate processing. When the hearer parses a complementizer-less CC, the complement clause can be incorrectly analyzed as a matrix clause until the matrix verb is encountered. In contrast, the initial complementizer requires that the CC is parsed as the complement of a verb, and this will facilitate processing. This is also in line with corpus data by Jaeger 2010, who shows that the likelihood of a CC in a specific context predicts whether the complementizer will be omitted or not.<sup>45</sup>

In Section 3.3 I present experiments based on the study on complement clauses by Merchant et al. (2013) in German and English (experiments 8 and 9), which address these issues. In the I collect ratings for both topicalized and fragment CCs, test only CCs that are acceptable in situ and use only non-factive matrix verbs. The goal of the experiments is two-fold. First, the data for fragment an-

<sup>45</sup>The possibility of a processing account has already been discussed by Stowell (1981: 397), who argues against the processing account based on data as (i). The argument is that the CC is very likely at the point where it occurs due to the subcategorization preferences of the predicate, and still the complementizer cannot be omitted.

(i) It surprises me \*(that) you have heard about Roger. Stowell (1981: 397)

swers will show whether the preference for realizing the complementizer in fragments is replicated when controlling for factivity. Second, the data for left dislocation answers might provide empirical evidence for the movement restriction on complementizer-less CCs on which the experiment by Merchant et al. (2013) is based. Anticipating the results, the experiments suggest that the effect reported by Merchant et al. (2013) does not evidence a movement restriction on complementizer-less CCs but results from independent factors. The German data replicate the preference for realizing the complementizer in fragments, but this effect is not reflected in the acceptability of the corresponding left dislocation structures. In English, there is no significant difference between fragments at all and null complementizers are even preferred in full sentences.

#### 3.3.2 Experiment 8: CC topicalization in German

##### 3.3.2.1 Background

Experiment 8 replicates experiment 1 in Merchant et al. (2013) in German under more controlled conditions. As compared to the study by Merchant et al. (2013), there were four main modifications: First, the experiment was conducted in German, second, I used a context story to preclude the possibility of an indirect answers interpretation, third, I tested both fragments and full sentences, and fourth, I added a third condition, subjunctive mood verb-second CCs.

As for the first modification, in German I expected a similar pattern to the English data in Merchant et al. (2013). Webelhuth (1992: 83) claims that subject CCs require overt complementizers in all Germanic languages.<sup>46</sup>

<sup>46</sup>Webelhuth presents (i) as evidence in favor of this claim. However, omitting the complementizer in verb-last CCs is also unacceptable in situ and cannot be attributed to the prefield position (iia). In the case of the example, its verb-second counterpart (iib) seems to be degraded as well. A further shortcoming of the example is that the predicate is factive, so that complementizer omission seems to be dispreferred in situ (iic). (iic) is only acceptable with a break in intonation between the first and the second clause that marks that each clause is an independent sentence.

- (i) \**(Daß) Hans nicht kommt, ist schade.* (Webelhuth 1992: 83)  
       that Hans not comes is pity  
       ‘It is a pity that Hans does not come.’
- (ii) a. *Es ist schade, \*(dass) Hans nicht kommt.*  
        it is pity that Hans not comes  
       b. \**Hans kommt nicht, ist schade.*  
           Hans comes not is pity  
       c. \**Es ist schade, Hans kommt nicht.*  
           it is pity Hans comes not  
           ‘It is a pity (that) Hans does not come.’



The target sentence was introduced by a context story and a short dialogue with two turns per speaker (53). The context story had the purpose of making an indirect answer less likely. Merchant et al. (2013) suggest that complementizer-less CCs were rated as relatively acceptable in their experiment due to their interpretation as indirect answers. In my experiment, the context stories ensured that this was not possible: For instance, in (53), it is very unlikely that the reporter has sufficient knowledge about the crime to give an indirect answer. The CCs were tested not only as fragments, but also in a left-peripheral position within a complete sentence in order to investigate whether the assumed movement restriction on complementizer-less CCs is empirically confirmed. Furthermore, acceptability differences between the sentential conditions allow me to test the prediction of the movement and deletion account that the acceptability of left dislocations matches that of the corresponding fragments.

Finally, besides the German counterparts of the CCs investigated by Merchant et al. (2013), I added a third type of complement clause, verb-second complement clauses in subjunctive mood, to test the hypothesis that verb-second CCs are rated as acceptable because they are interpreted as indirect answers.

- (53) [Context story] This weekend a famous painting has been stolen from the museum. The newscaster is reporting on the investigation of the robbery.  
 [Newscaster:] What's the news about the art robbery?  
 [Reporter:] The investigators are discussing how the burglar got into the building.
- a. [N:] Was glaubt Kommissar Wagner?  
           what believes inspector Wagner  
       '[N:] What does inspector Wagner believe?'
  - b. [R:] Dass der Täter durch das Fenster eingestiegen ist.  
           that the thief through the window entered is  
       '[R:] That the thief entered through the window.'

The reason for this is that CCs headed by *dass*, the German equivalent to *that*, are verb-last, whereas complementizer-less CCs are always verb-second. Therefore, verb-second complement clauses are ambiguous between a direct answer (a CC fragment) and an indirect answer (a matrix clause). Since subjunctive mood encodes reported speech in German, it enforces the interpretation as a direct answer. If indicative complementizer-less CCs were rated as more acceptable only because they can be interpreted as indirect answers, they would be more acceptable than subjunctive ones. Subjunctive might also provide some insights on the viability of the pragmatic and processing accounts of complementizer omission

### 3 Experiments on the syntax of fragments

that I sketched above. If complementizer-less fragments are dispreferred due to their ambiguity, subjunctive should improve their acceptability because it excludes the possibility of an indirect answer. Similarly, in fronted CCs the likelihood of an embedding matrix verb will increase as soon as the inflected verb in the CC is encountered. Since the CC is verb-second, this happens relatively early, hence processing it should be easier in comparison to the indicative verb second clause. This setting yields a 2×3 design that crosses SENTENTIALITY and CCTYPE.

The movement and deletion account predicts that, if complementizer-less CCs are degraded as full sentences, they will be as fragments, too. If the movement restriction on complementizer-less CCs was not empirically supported, Merchant (2004a) predicts that whatever pattern is observed, there should be no interaction between SENTENTIALITY and CCTYPE. The in situ deletion account makes no specific predictions with respect to this, because it does not assume that left dislocation structures are the input to the derivation of fragments. In situ deletion might predict a parallelism between the acceptability of complement clauses as fragments and in their base position, but if both the realization and the omission of the complementizer is grammatical for non-factive matrix verbs it is unlikely that this results in a strong acceptability contrast.

#### 3.3.2.2 Materials

All materials follow the pattern in (53). The target utterance is preceded by a two-sentence context story and a dialogue between two characters. The context story is always such that the character who produces the target utterance is not in the epistemic position to produce an indirect answer. The target utterance is the second turn by the second character. In the sentential conditions, the matrix clause in the target utterance (*glaubt er*, ‘he believes’) is given in the immediately preceding question. Table 3.15 provides an overview of the conditions for the target sentence in (53).

The issue that factive verbs prefer CCs with overt complementizers according to the literature (Kiparsky & Kiparsky 1970, Hegarty 1992) was addressed by testing only three non-factive matrix verbs (*glauben* ‘to believe’, *meinen* ‘to mean’ and *sagen* ‘to say’). A corpus search on the German newspaper corpus TüBa-D/Z (Telljohann et al. 2004) confirms that each of the three verbs occurs with each of the CC types investigated, but they seem to differ quantitatively in their subcategorization preferences, as Table 3.16 shows. In order to factor out possible subcategorization preferences of specific verbs I included the MATRIXVERB as a predictor in the statistical analysis. However, the analysis showed that the verb has no significant effect on acceptability.

### 3.3 Movement restrictions: Complementizer omission

Table 3.15: Target sentences in experiment 8. See example (53) for glosses. The subjunctive is distinguished from indicative only by the copula (*sei* instead of *ist*).

| Target utterance                                              | Condition      |
|---------------------------------------------------------------|----------------|
| Dass der Täter durch das Fenster eingestiegen ist, glaubt er. | VL, sent.      |
| Der Täter ist durch das Fenster eingestiegen, glaubt er.      | V2, ind, sent. |
| Der Täter sei durch das Fenster eingestiegen, glaubt er.      | V2, ind, sent. |
| Dass der Täter durch das Fenster eingestiegen ist.            | VL, frag.      |
| Der Täter ist durch das Fenster eingestiegen.                 | V2, ind, frag. |
| Der Täter sei durch das Fenster eingestiegen.                 | V2, ind, frag. |

Table 3.16: Counts (ratio) in TüBa-D/Z of each type of CC for the three matrix verbs tested in experiment 8.

| verb    | VL CC     | V2 indicative CC | V2 subjunctive CC |
|---------|-----------|------------------|-------------------|
| glauben | 50 (0.49) | 37 (0.36)        | 15 (0.15)         |
| sagen   | 32 (0.27) | 34 (0.28)        | 54 (0.45)         |
| meinen  | 8 (0.11)  | 10 (0.14)        | 52 (0.74)         |

#### 3.3.2.3 Procedure

The experiment was completed by 38 undergraduate students of Saarland University who were rewarded with the participation in a lottery of  $10 \times € 30.00$ . All of them were native speakers of German. The experiment was conducted over the Internet using the LimeSurvey presentation software. Subjects were asked to rate the naturalness of the critical utterance, which was italicized, in the context of the previous discourse on a 7-point Likert scale with labeled extremes (7 = totally natural). Subjects were assigned to one of six lists. As the topicalization structures are probably highly marked due to the redundant matrix clause and the heavy CC in the initial position, that is dispreferred for processing reasons (Hawkins 2004), SENTENTIALITY was tested as a between subjects IV. Half of the subjects saw only topicalized CCs and the other half only the corresponding fragments. Each subject thus rated 21 items (7 per CCTYPE condition), which were mixed with 24 items from an unrelated experiment and 40 fillers. The materials were presented in individually fully randomized order. The fillers and items of the unrelated experiment had a similar structure as those of the current experiment. Subjects always rated an italicized target utterance that appeared at the

### 3 Experiments on the syntax of fragments

end of a shot dialogue which was preceded by a context story. Three subjects who rated more than two out of five ungrammatical controls as acceptable (6 or 7 points on the scale) were excluded from further analyses.

#### 3.3.2.4 Results

Figure 3.9 shows the averaged ratings across conditions. Fragments ( $\mu = 5.85$ ,  $\sigma = 1.54$ ) were overall rated as better than sentences ( $\mu = 4.77$ ,  $\sigma = 1.75$ ). Recall that SENTENTIALITY was tested between subjects, so that these figures cannot be directly compared to each other. Rather, they show that a floor effect for the topicalization conditions could be avoided by the between subjects design. Ratings for the verb-second conditions were very close to each other both within sentences ( $\mu_{\text{indicative}} = 4.86$ ,  $\mu_{\text{subjunctive}} = 4.9$ ) and fragments ( $\mu_{\text{indicative}} = 5.72$ ,  $\mu_{\text{subjunctive}} = 5.64$ ). Verb-last CCs were rated slightly better than verb-second CCs in the fragment ( $\mu = 6.21$ ,  $\sigma = 1.23$ ) and slightly worse in the sentence condition ( $\mu = 4.56$ ,  $\sigma = 1.76$ ).

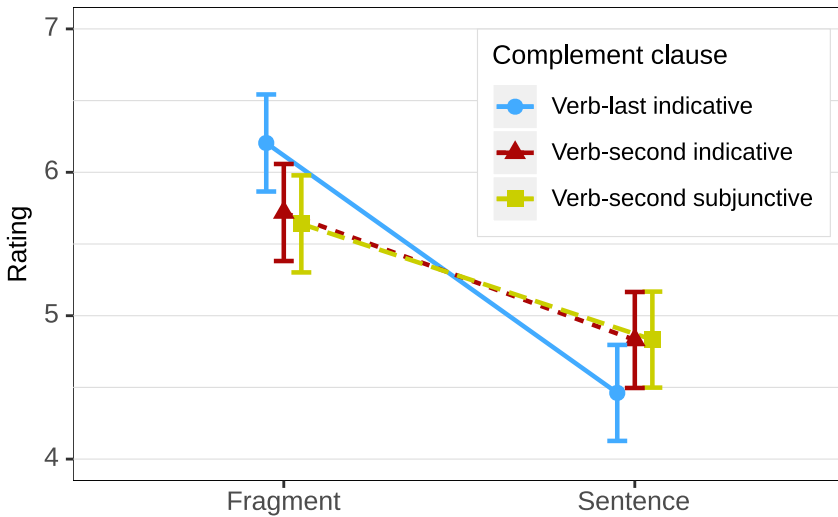


Figure 3.9: Mean ratings and 95% confidence intervals across conditions in experiment 8.

The data were statistically analyzed with CLMMs in R following the procedure described in Section 3.1.1.5. Since CCTYPE was a ternary factor, I first conducted an analysis on the data for verb-second CCs only in order to test the conjecture that they did not differ significantly from each other. If this was confirmed,

the two verb-second conditions could be pooled for further analysis and CCTYPE treated as a binary predictor. The initial model for this analysis contained fixed effects for CCTYPE, SENTENTIALITY, POSITION of the trial in the experiment and MATRIXVERB in order to account for an effect of the verb's subcategorization preferences, as would be evidenced by a MATRIXVERB:CCTYPE interaction. As for random effects, I included by-subject and by-item random intercepts and random slopes for SENTENTIALITY, MATRIXVERB, CCTYPE and the interactions thereof.<sup>47</sup> Since the analysis of verb-second CCs revealed no significant difference between subjunctive and indicative ( $\chi^2 = 0.19, p > 0.6$ ) verb-second CCs, the verb-second conditions were pooled for further analysis. The full model fitted to the complete data set after pooling had the same effects structure as the model for verb-second CCs.

The final model is summarized in Table 3.17. It contained significant main effects of both IVs and a significant interaction between them. The main effect of SENTENTIALITY ( $\chi^2 = 14.57, p < 0.001$ ) confirms that fragments are rated better than left dislocations averaging over CCTYPE conditions.<sup>48</sup> The main effect of CCTYPE ( $\chi^2 = 6.02, p < 0.05$ ) indicates that verb-last clauses were preferred over verb-second ones, and the significant SENTENTIALITY:CCTYPE ( $\chi^2 = 25.75, p < 0.001$ ) interaction shows that this preference is specifically large for fragments. The MATRIXVERB had no significant effect on acceptability and neither participated in any significant interaction. The significant main effect of POSITION ( $\chi^2 = 16.76, p < 0.001$ ) shows that items were perceived as more acceptable the later they appeared in the experiment, while the interaction with SENTENTIALITY ( $\chi^2 = 12.73, p < 0.001$ ) suggests that this effect was stronger for fragments than for left dislocations. These predictors are not theoretically interesting, but their inclusion factors out the familiarization effects.

In addition to the analysis of the complete data set, I performed an analysis of the sentential left dislocation conditions only in order to test the movement restriction on complementizer-less CCs. The full model had the same effects structure like the one in the main analysis, with the obvious exception that all effects for SENTENTIALITY were removed. The final model contained only a significant main effect of CCTYPE ( $\chi^2 = 10.68, p < 0.01$ ) evidencing that clauses headed by a complementizer were rated as significantly worse than their complementizer-less counterparts.

<sup>47</sup> A by-subject random slope for SENTENTIALITY and the corresponding interactions were omitted as it would make no sense for a between subjects IV. Similarly, there was no by-item slope for MATRIXVERB and interactions thereof, because MATRIXVERB was not varied between items.

<sup>48</sup> The significance of main effects of predictors that significantly interacted with others was assessed by the procedure described in Levy (2018). I sum-coded the other predictor participating in the interaction and then compared a model that contains the predictor to be tested to one that does not with a likelihood ratio test.

### 3 Experiments on the syntax of fragments

Table 3.17: Fixed effects in the final CLMM for experiment 8.

| Predictor              | Estimate | SE    | $\chi^2$ | $p$    |     |
|------------------------|----------|-------|----------|--------|-----|
| SENTENTIALITY          | 0.834    | 0.231 | 14.57    | <0.001 | *** |
| CCTYPE                 | -0.19    | 0.067 | 6.02     | <0.05  | *   |
| POSITION               | 0.009    | 0.002 | 16.76    | <0.001 | *** |
| SENTENTIALITY:CCTYPE   | -0.447   | 0.068 | 25.75    | <0.001 | *** |
| SENTENTIALITY:POSITION | 0.008    | 0.002 | 12.73    | <0.001 | *** |

#### 3.3.2.5 Discussion

Experiment 8 investigated (i) whether fronting complementizer-less CCs is degraded as compared to CCs headed by overt complementizers, and (ii) whether this movement restriction is reflected in the acceptability of fragments.

The significant SENTENTIALITY:CCTYPE interaction shows that the preference for realizing the complementizer is stronger for fragments than for left dislocated CCs. The analysis of the sentential conditions only shows that the preference for realizing the complementizer is even inverted in full sentences: In contrast to the claim by Webelhuth (1992), who argues that this movement restriction is a property of “all finite declarative argument clauses in the Germanic languages” (Webelhuth 1992: 83), omitting the complementizer is preferred in the full sentence. At least for German, this undermines the reasoning upon which the interpretation of the data in Merchant et al. (2013) is based, since there is no evidence for a movement restriction on complementizer-less CCs. Furthermore, the sentential conditions seem to be grammatical across the board, as their mean ratings (for all conditions  $\mu > 4.5$ ) are clearly above those for ungrammatical controls ( $\mu = 2.4$ ).

The fragment data alone resemble the pattern reported by Merchant et al. (2013): Fragment CCs are more acceptable when they are introduced by a complementizer than when they are not. However, since there is no evidence for the movement restriction stated in the literature, this effect cannot be attributed to constraints on movement. Furthermore, all three fragment conditions seem to be relatively acceptable, and even the dispreferred complementizer-less fragments were rated much better than ungrammatical controls ( $\mu = 2.18$ ). Unlike in the experiment by Merchant et al. (2013), the explanation that they were interpreted as indirect answers is ruled out by the context story. Furthermore, there was no significant difference at all between subjunctive and indicative mood, even though the indirect answer interpretation is only possible in the indicative conditions. Since subjunctive encodes reported speech, it cannot be interpreted as an

indirect answer and they would have been preferred over indicative ones if the indirect answer explanation for the relatively high ratings for complementizer-less CCs in Merchant et al. (2013) was correct.

Taken together, the data indicate that, at least in German, the preference for realizing the complementizer in short answer fragments observed by Merchant et al. (2013), which my experiment replicates, seems not to be caused by a movement restriction on complementizer-less CCs. If this finding is robust and generalizes to English, it would question the reasoning behind the experiment by Merchant et al. (2013): If the movement restriction in question does not hold, no conclusions with respect to this can be drawn from the acceptability of the corresponding fragments. Experiment 9 will replicate this study in English to investigate whether the difference between the results in Merchant et al. (2013) and my German data evidences a cross-linguistic difference or whether it must be attributed to properties of the experiment design.

#### 3.3.2.6 Follow-up study with shorter contexts

##### 3.3.2.6.1 Background and materials

The stimuli tested in experiment 8 were relatively long and complex as compared to those used by Merchant et al. (2013). Instead of a question-answer pair, they consisted of a context story and two turns per character. This might have biased subjects to base their ratings rather on the naturalness of the target utterance in discourse than on its grammaticality alone. In order to address this concern, I conducted a follow-up experiment using similar materials, but slightly longer context stories and two-turn dialogues which consisted only of the critical question-answer pair (54).

- (54) [Context story] This weekend a famous painting has been stolen from the museum. The newscaster is reporting on the investigation of the robbery. The investigators are currently discussing how the burglar got into the building.  
[Newscaster:] Was glaubt Kommissar Wagner?  
[Reporter:] Der Täter ist durch das Fenster eingestiegen(, glaubt er).

##### 3.3.2.6.2 Method

The experiment was presented on the Internet using LimeSurvey. Originally, it was conducted in the same session as the production study on case marking (experiment 2), where subjects were asked to produce utterances referring to

### 3 Experiments on the syntax of fragments

graphical stimuli. The experiment was completed by 38 undergraduate students of Saarland University, who were rewarded with the participation in a lottery of  $5 \times \text{€ } 30.00$  among all participants.<sup>49</sup> The task and assignment to lists was identical to experiment 8 and SENTENTIALITY was tested between subjects again. Each subject rated 21 items (7 per CCTYPE condition). Materials were presented together with 35 items of experiment 10 and 25 unrelated fillers including four ungrammatical controls in individually pseudo-randomized order. Pseudo-randomization ensured that no two items of the same experiment followed each other. One participant rated 50% or more of the ungrammatical attention checks as acceptable (6 or 7 points on the scale) and was therefore excluded from further analysis.

#### 3.3.2.6.3 Results

Figure 3.10 shows the aggregated ratings across conditions. The pattern is similar to experiment 8, despite the slightly different ratings in absolute terms, which might be due to the differing materials that were tested together with the items in experiments 8 and the follow-up study. The statistical analysis followed the same procedure as for experiment 8. First, pairwise analyses compared two of the three levels of CCTYPE at a time. In all three analyses there were significant effects of CCTYPE or significant interactions of CCTYPE and SENTENTIALITY, therefore I did not pool the data. In all of the analyses, the full model contained main effects for SENTENTIALITY, CCTYPE and MATRIXVERB as well as all two-way interactions. I also included by-subject random intercepts and slopes for CCTYPE, MATRIXVERB and their interaction, as well as by-item random intercepts and slopes for CCTYPE, SENTENTIALITY and their interaction. By-item random effects for MATRIXVERB were not considered because the matrix verb was not varied between items. The same holds for by-subject SENTENTIALITY random effects.

First, I analyzed only the data for the indicative (verb-second and verb-last) complement clauses. The final model is summarized in Table 3.18. A significant effect of SENTENTIALITY ( $\chi^2 = 12.64, p < 0.001$ ) evidences an overall preference for fragments over sentences with sentence-initial CCs. The SENTENTIALITY:CCTYPE interaction ( $\chi^2 = 7.16, p < 0.01$ ) shows that in the case of fragments

---

<sup>49</sup>Due to a technical problem, two out of the six lists (which included 10 subjects) were assigned experimental materials in the incorrect conditions. The corresponding participants were replaced by subjects recruited on the *clickworker* crowdsourcing platform. These subjects rated the correct materials, which were mixed with the stimuli from experiment 10 and the same fillers as in the original lists. Each subject on the replacement lists was paid € 3.00 for participating. Since the distribution issue does not concern the stimuli of experiment 10, in the case of experiment 10 I report the original data.



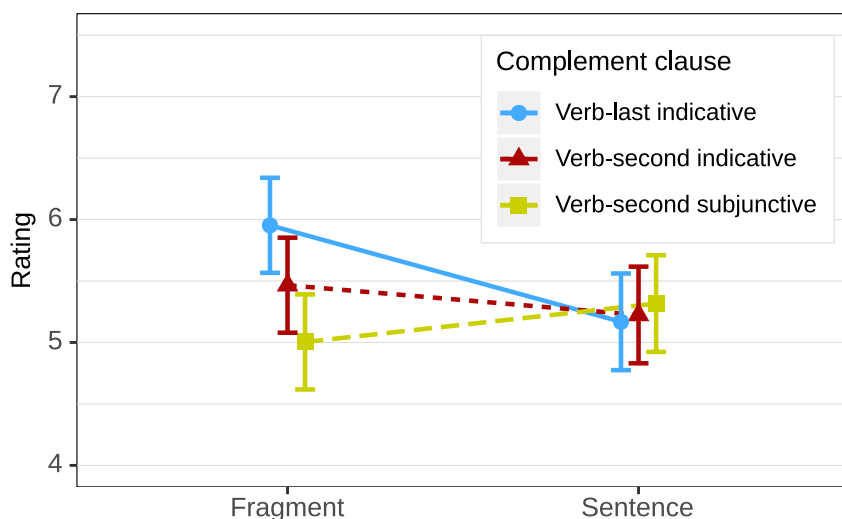


Figure 3.10: Mean ratings and 95% confidence intervals across conditions in the follow-up to experiment 8.

there is a preference for verb-last CCs with overt complementizers, which however is not observed for sentences, since the main effect of CCTYPE is not significant ( $\chi^2 = 1.67, p > 0.1$ ). The MATRIXVERB did neither have a significant main effect nor did it interact with any of the other predictors.

Table 3.18: Fixed effects in the final CLMM for the verb-second and verb-last indicative conditions in the follow-up to experiment 8.

| Predictor            | Estimate | SE    | $\chi^2$ | $p$    |     |
|----------------------|----------|-------|----------|--------|-----|
| SENTENTIALITY        | 1.315    | 0.37  | 12.64    | <0.001 | *** |
| CCTYPE               | -0.347   | 0.263 | 1.67     | >0.1   |     |
| SENTENTIALITY:CCTYPE | -1.289   | 0.472 | 7.16     | <0.01  | **  |

In a second analysis I compared only the indicative and subjunctive verb-second conditions. In this case, the only significant effect in the final model (see Table 3.19) is an interaction of SENTENTIALITY and CONDITION ( $\chi^2 = 4.81, p < 0.05$ ): Subjunctive verb-second CCs are significantly degraded as compared to indicative ones as fragments as compared to the left dislocation conditions. The effects of CONDITION ( $\chi^2 = 1.681, p > 0.1$ ) and SENTENTIALITY ( $\chi^2 = 0.01, p > 0.9$ ) are not significant but kept in the model due to the significance of the interaction. Again, there were no effects of MATRIXVERB.

### 3 Experiments on the syntax of fragments

Table 3.19: Fixed effects in the final CLMM for the indicative and subjunctive verb-second conditions in the follow-up to experiment 8.

| Predictor            | Estimate | SE    | $\chi^2$ | $p$     |
|----------------------|----------|-------|----------|---------|
| SENTENTIALITY        | 0.035    | 0.533 | 0.01     | >0.9    |
| CCTYPE               | -0.45    | 0.343 | 1.681    | >0.1    |
| SENTENTIALITY:CCTYPE | -1.201   | 0.535 | 4.81     | <0.05 * |

Finally, I compared only the data for the subjunctive and the verb-last CCs, which cannot be interpreted as indirect answers. The final model is summarized in Table 3.20. The significant effect of CCTYPE ( $\chi^2 = 7.24, p < 0.01$ ) shows that subjunctive CCs are overall dispreferred as compared to verb-last CCs. In this analysis, the overall preference for fragments is only marginal ( $\chi^2 = 2.8, p > 0.05$ ). Like in the analysis of the indicative data, the SENTENTIALITY:CCTYPE interaction ( $\chi^2 = 22.33, p < 0.001$ ) shows that verb-last CCs are particularly preferred as fragments. Again, there were no effects of MATRIXVERB.

Table 3.20: Fixed effects in the final CLMM for the verb-second indicative and the verb-last subjunctive conditions in the follow-up to experiment 8.

| Predictor            | Estimate | SE    | $\chi^2$ | $p$        |
|----------------------|----------|-------|----------|------------|
| SENTENTIALITY        | 0.579    | 0.345 | 2.8      | >0.05      |
| CCTYPE               | -0.971   | 0.351 | 7.24     | <0.01 **   |
| SENTENTIALITY:CCTYPE | -2.543   | 0.53  | 22.33    | <0.001 *** |

#### 3.3.2.6.4 Discussion

The follow-up study finds relatively similar results to experiment 8 with shorter contexts. Again, there is evidence for the preference for verb-last CCs with overt complementizers that Merchant et al. (2013) report in fragments, but not in complete sentences. Unlike in experiment 8, the subjunctive verb-second fragments were slightly degraded as compared to indicative verb-second fragments, but this effect is also not reflected in left dislocation structures in full sentences.

### 3.3.3 Experiment 9: CC topicalization in English

#### 3.3.3.1 Background

Experiment 8 replicated the pattern reported by Merchant et al. (2013) for fragments but found no difference between the corresponding topicalized CCs. This could either indicate a crosslinguistic difference between German and English or be due to the absence of the movement restriction. In order to distinguish between these explanations I conducted an English version of experiment 8 that would show whether there is evidence for the presumed movement restriction on complementizer-less CCs or whether English patterns with my German data. Since there were no meaningful differences between the main experiment 8 and the follow-up, the shorter stimuli tested in the follow-up were used.

The predictions of movement and deletion are identical to experiment 8: If the data in Merchant et al. (2013) are due to a movement restriction, complementizer-less CCs should be degraded both as fragments and in a left-peripheral position.

#### 3.3.3.2 Materials

The materials were in principle identical to those from experiment 8 and were translated into American English by a native speaker. A sample item is given in (55) (context story) and (56) (target utterances). The most important difference to the German experiment was the omission of the subjunctive conditions, because in English subjunctive does not have the same status as a marker of reported speech that it has in German. This resulted in a 2×2 design that crossed CCTYPE (*that* vs. null complementizers) with SENTENTIALITY. Again, SENTENTIALITY was tested as a between subjects IV. Besides reducing the likelihood of a floor effect, this allowed for a comparison with my German studies and the original experiment by Merchant et al. (2013). The English verbs used in the matrix clause that embedded the CC in the sentential conditions were *to believe*, *to think*, *to say* and *to mean*.

- (55) [Context story] This weekend a famous painting has been stolen from the museum. The newscaster is reporting on the investigation of the robbery. The investigators are currently discussing how the burglar got into the building.  
[Newscaster:] What does inspector Wagner believe?

- (56) [Reporter:]
- a. That the criminal entered through the window.
  - b. The criminal entered through the window.
  - c. The criminal entered through the window, he believes.
  - d. That the criminal entered through the window, he believes.

### 3.3.3.3 Procedure

The experiment was completed by 54 native speakers of American English, who were recruited on the *Prolific Academic* crowdsourcing platform. The study was run over the Internet using the LimeSurvey presentation software. Each participant received £2 for participation. Subjects were asked to rate the naturalness of the italicized target utterance in the context of the question. They were assigned to one out of four lists. As SENTENTIALITY was tested between subjects, two of the lists contained only fragment and two only sentential target utterances. The materials were distributed across four lists according to a Latin square, so that each subject saw each token set once and each CCTYPE condition equally often. Each subject rated 20 items (10 per CCTYPE condition).<sup>50</sup> The stimuli were mixed with 20 items from experiment 5 and 45 fillers. All fillers consisted of context stories which were followed by a dialogue. The target utterance was always the last utterance in that dialogue, and fillers were adapted so that participants assigned to the fragment lists rated only fragments and those assigned to the sentential lists only sentences. The stimuli were presented in individually fully randomized order. The fillers included five ungrammatical controls, which contained e.g. wrong auxiliaries or voice. Four subjects who rated more than 50% with 6 or 7 points on the scale were excluded from further analysis.

### 3.3.3.4 Results

Figure 3.11 shows the aggregated ratings across conditions. The ratings for fragments are almost identical independently of the presence of a complementizer ( $\mu_{that} = 5.69$ ,  $\mu_{null} = 5.8$ ). This suggests that the effect reported by Merchant et al. (2013) was not replicated. Furthermore, topicalized CCs without overt complementizers ( $\mu_{null} = 3.98$ ), appear to be more acceptable than those headed by *that* ( $\mu_{that} = 3.08$ ), in contrast to the introspective data reported in the literature.

The data were analyzed with CLMMs in R following the procedure described in Section 3.1.1.5. I first fit a full model to the complete data set. The full model

---

<sup>50</sup>One of the German materials ( $n = 21$ ) was not used in order to obtain an even number of materials.

### 3.3 Movement restrictions: Complementizer omission

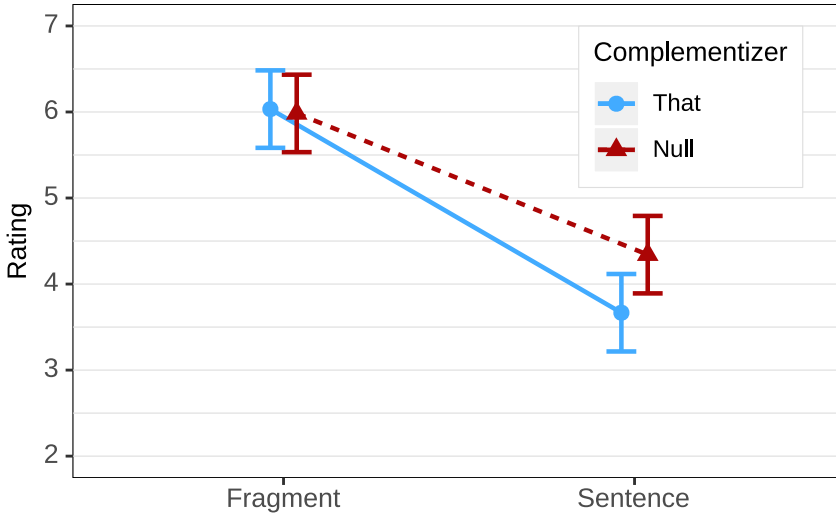


Figure 3.11: Mean ratings and 95% confidence intervals across conditions in experiment 9.

contained fixed effects for SENTENTIALITY, CCTYPE, the POSITION of the trial in the time-course of the experiment and the MATRIXVERB as well all two-way interactions between the IVs. The models had by-item random intercepts by-item random slopes for SENTENTIALITY, CCTYPE and their interaction, as well as by-subject random intercepts and by-subject random slopes for CCTYPE. By-subject random slopes for SENTENTIALITY were not considered, because SENTENTIALITY was tested as a between subjects IV.

The final model (see Table 3.21) contains significant main effects of SENTENTIALITY, CCTYPE, POSITION, MATRIXVERB and the SENTENTIALITY:CCTYPE interaction. The main effect of SENTENTIALITY ( $\chi^2 = 29.85, p < 0.001$ ) confirms that fragments are preferred over topicalized CCs. The main effect of CCTYPE ( $\chi^2 = 10.58, p < 0.01$ ) shows that, unlike it has been argued in the theoretical literature based on introspective data, complementizer omission is overall preferred. Finally, the significant interaction between SENTENTIALITY and CCTYPE ( $\chi^2 = 6.05, p < 0.05$ ) shows that the preference for complementizer omission is specifically strong for sentences. The POSITION effect ( $\chi^2 = 11.64, p < 0.001$ ) reveals a slight overall improvement of ratings over time, but the absence of interactions with other predictors shows that this does not affect any condition in particular. The MATRIXVERB main effect ( $\chi^2 = 7.6, p < 0.01$ ) reveals a preference for materials with the matrix verb *believe* as compared to other matrix verbs, for the other matrix verbs there was no such effect. Since the verb was not

### 3 Experiments on the syntax of fragments

varied systematically across all items, this might be due to properties of individual materials, hence I consider this a control predictor. Just like in the German experiments, I addressed the movement restriction with an analysis of the data for the full sentences only, following the same procedure as for the main analysis. The final model contains a POSITION effect ( $\chi^2 = 5.63, p < 0.05$ ) and a main effect of CCTYPE ( $\chi^2 = 13.7, p < 0.001$ ) that confirms the preference for complementizer-less clauses in a left-peripheral position.

Table 3.21: Fixed effects in the final CLMM for experiment 9.

| Predictor            | Estimate | SE    | $\chi^2$ | $p$    |     |
|----------------------|----------|-------|----------|--------|-----|
| SENTENTIALITY        | 1.886    | 0.302 | 29.85    | <0.001 | *** |
| CCTYPE               | 0.443    | 0.13  | 10.58    | <0.01  | **  |
| MATRIXVERB (BELIEVE) | 0.441    | 0.15  | 7.60     | <0.01  | **  |
| POSITION             | 0.009    | 0.003 | 11.64    | <0.001 | *** |
| SENTENTIALITY:CCTYPE | -0.323   | 0.128 | 6.05     | <0.05  | *   |

#### 3.3.3.5 Discussion

Like experiment 8, experiment 9 investigated an assumed movement restriction on complementizer-less CCs that constrains the acceptability of the corresponding fragments according to Merchant et al. (2013). The data do neither provide evidence for the assumed movement restriction nor for the effect that Merchant et al. (2013) report for fragments. The overall pattern in my English data is similar to that found for German: Short answers are preferred across the board and the overall difference in acceptability between CCTYPE conditions is rather small in absolute terms. In English, there was no difference in acceptability between both types of fragment CCs. This contrasts with the study by Merchant et al. (2013), who found such an effect, and suggests that the preference for CCs with overt complementizers in their experiment was due to the use of factive matrix verbs, which disprefer complementizer-less CCs. Once more, there is no evidence for the movement restriction on which the study by Merchant et al. (2013) is based: Complementizer-less CCs were even rated as more acceptable than in the sentential conditions, whereas the opposite pattern has been repeatedly assumed in the theoretical literature based on introspective data.

#### 3.3.4 General discussion: Complementizer omission

Experiments 8 and 9 investigated whether the presumed movement restriction on complementizer-less CCs that Merchant et al. (2013) investigated indeed provides evidence for movement in fragments. The results suggest that it does not: First, there is no empirical evidence for the presumed movement restriction on complementizer-less CCs. Second, complementizer-less CC fragments are slightly degraded in German, but this pattern is not reflected in the corresponding full sentences. In English, there is no difference between fragments, and in full sentences the complementizer-less CC is even preferred. This suggests that Merchant et al. (2013) interpret their data based on incorrect assumptions and that the preference for realizing the complementizer in their materials is not related to movement restrictions.

The study by Merchant et al. (2013) differs from my experiments in three main aspects: First, the restriction to CCs that are acceptable *in situ*, that is complements of non-factive verbs, ensures that a possible preference for complementizer realization cannot be explained by *in situ* deletion. Degraded ratings for CCs which are ungrammatical even *in situ* are expected by *any* ellipsis account. Second, the use of context stories ruled out the possibility of indirect answers, and third, I collected ratings for the left dislocation structures that Merchant (2004a) assumes to be the source of fragments. Merchant et al. (2013) simply relied on the movement restriction on complementizer-less CCs reported in the literature.

In both experiments, topicalized CCs were rated as worse than fragments across the board, even though CCTYPE was tested within subjects in order to attenuate this effect. None of the experiments confirmed the pattern reported in the literature with respect to topicalized CCs: Both in English and German there were significant SENTENTIALITY:CCTYPE interactions showing that the preference for realizing the complementizer is stronger in fragments than in left dislocation structures. The analyses of the sentential conditions in isolation show that complementizer-less clauses were rated as more acceptable in the left-peripheral position in English, and at least as acceptable in German (in the follow-up to experiment 8 there was no significant difference). This is the opposite pattern to the one reported in the theoretical literature that underlies the interpretation of the data in Merchant et al. (2013). If complementizer-less CCs are not dispreferred in a left-peripheral position, the preference for realizing the complementizer in fragments cannot be attributed to a movement restriction: No matter why we observe this pattern in fragments, it does not provide evidence for movement.

Rebecca Woods (p.c.) pointed out the possibility that the ratings for topicalized complementizer-less CCs improved because they were interpreted as paren-

### 3 Experiments on the syntax of fragments

theticals rather than as matrix clauses. Parentheticals do not subcategorize the embedding clause but can be inserted into regular verb-second clauses, hence it should be possible to omit the complementizer in that case. However, in the German subjunctive conditions, which did not differ in acceptability from indicative, subjunctive mood marks the utterance as indirect speech. Since indirect speech is in general embedded under a matrix verb, a parenthetical reading of the matrix clause seems less appropriate in that case than with regular indicative verb-second clauses. If the indicative verb-second fragments were improved by the parenthetical interpretation, subjunctive should be to a lesser extent and hence rated worse. This is clearly not the case. In order to be completely sure however, the items would have to be modified in a way that rules out the parenthetical reading, for instance by including a negation in the matrix clause.

When only fragments are taken into account, the German, but not the English, data resemble those reported by Merchant et al. (2013), even when controlling for factivity. In both German experiments realizing the complementizer was preferred, whereas there was no significant effect of the complementizer on English fragments. The difference in acceptability between conditions is smaller than in the experiment by Merchant et al. (2013) in absolute terms even though I used a 7-point scale instead of a 5-point scale. This might be due to the restriction to non-factive verbs, which are known to allow for complementizer omission. Even though the results on sentence-initial CCs do not evidence a movement restriction, a possible difference between German and English with respect to complementizer omission might be investigated in future research.

Like the preposition omission data discussed in the previous section, my results do not *falsify* the movement and deletion account. However, they challenge the interpretation of a relatively weak preference for realizing complementizers in short answer fragments when the experiment is replicated under more carefully controlled conditions as evidence for movement. The movement and deletion account can of course derive both fragment CCs with and without complementizers from the corresponding left dislocations, which, unlike as has been argued in the literature, seem to be well-formed. However, there is no empirical evidence for the movement restriction assumed by Merchant (2004a) and Merchant et al. (2013) that constrains the form of fragments. Therefore, even if the pattern predicted by Merchant (2004a) is observed in some fragments (as it was by Merchant et al. (2013) and in experiment 8), this cannot be traced back to a movement restriction on complementizer-less complement clauses. Since the movement and deletion account is derivationally more complex than in situ deletion, the absence of evidence for movement forces us to stick to the simpler in situ deletion account. If there is no movement restriction on complementizer-less



CCs, this phenomenon may simply be the wrong testing ground for the movement and deletion account. For this reason, in Section 3.4 I explore whether a well-documented restriction on German prefield configurations constrains the form of fragments.

## 3.4 Movement restrictions: The German prefield

### 3.4.1 The German prefield and movement in fragments

The replications and extensions of the experiments by Merchant et al. (2013) question the assumption that preposition omission and complement clause topicalization provide evidence for movement. In the case of preposition omission, the data can be explained without assuming movement by question-answer parallelism, and for complement clause topicalization, there is no evidence for the presumed movement restriction itself. This section presents a last experiment on the syntax of fragments that investigates a well-documented movement restriction concerning the prefield position in German. In a nutshell, the idea is that if the preverbal position in German verb-second sentences is analyzed as the landing site for fragments, as Merchant (2004a) suggests, only those constituents that can appear in this position are possible fragments.

#### 3.4.1.1 The prefield position in German verb-second clauses

The German declarative matrix clause is generally assumed to be strictly verb-second, so that only one constituent can precede the inflected verb. Traditionally, this is modeled with the *topological field model* (Drach 1937) of the German sentence, which divides the sentence into three regions, the so-called *fields*. These fields are delimited by two positions hosting verbal elements, the left and right *brackets*. Table 3.22 shows how these fields are filled in declarative matrix clauses: The left bracket hosts the inflected verb and the right bracket the participle or infinitive, if the sentence contains such. The region left to the left bracket is called prefield and contains exactly one constituent, yielding the obligatory verb-second order. By default, all other constituents appear in the middle field, the region between the brackets. In case of extraposition, constituents can be located in the postfield, i.e. in the region following the right bracket. (57) shows that, unlike in SVO languages, all arguments (including the subject) appear in the middle field if the prefield is filled by an adverbial.

- (57) Montag will Peter [einen Kuchen] backen, [der glutenfrei ist].  
monday wants Peter a cake bake that gluten.free is  
‘On Monday, Peter wants to bake a cake that is gluten-free.’

### 3 Experiments on the syntax of fragments

In generative terms, the standard syntactic analysis of German verb-second sentences assumes head movement of the verb from T to C followed by movement of the prefield constituent to [Spec, CP], as sketched in Figure 3.12.<sup>51</sup>

Table 3.22: German topological fields model (LB = left bracket; RB = right bracket).

| Prefield | LB    | Middle field |        | RB     | Postfield |             |     |
|----------|-------|--------------|--------|--------|-----------|-------------|-----|
| Peter    | will  | einen        | Kuchen | backen | der       | glutenfrei  | ist |
| Peter    | wants | a            | cake   | bake   | that      | gluten.free | is  |

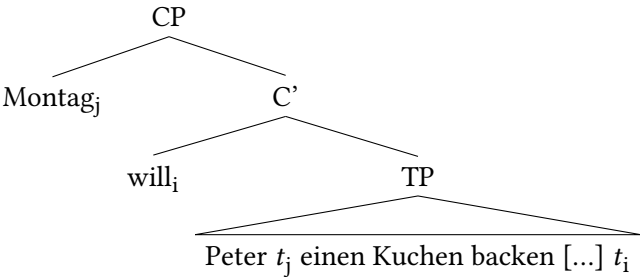


Figure 3.12: Following den Besten (1989), the verb-second word order of the German declarative matrix clause is the result of moving the inflected verb to C and another constituent to [Spec, CP].

<sup>51</sup>This description is highly simplified. Furthermore, both the order of the movement operations and their motivation and casual connection (does one of them trigger the other, and if so, which?) have been controversially discussed (see Brandner (2004) for an overview of competing analyses of V2). In fact, Gereon Müller (2004) has argued that verb-second order is derived by remnant movement of the whole *vP* to [Spec, CP] after the other constituents than the verb and the prefield constituents have been moved out for independent reasons. The resulting structure is given in (i), taken from (Müller 2004: 181).

(i) [CP [vP<sub>5</sub> Das Buch<sub>2</sub> t<sub>1</sub> t<sub>4</sub> hat<sub>3</sub>] [C' C [TP Fritz<sub>1</sub> [T' [vP<sub>4</sub> t<sub>2</sub> gelesen] [T' t<sub>5</sub> T]]]]]

Müller’s account is not compatible with Merchant’s (2004a) version of movement and deletion. The E feature on C would always cause the complete *vP* to survive ellipsis, so that there would be no way of generating DP fragments in German. The corpus data by Reich (2017) and my previous experiments disconfirm this prediction. Note that this does not speak against Müller’s analysis, unless movement and deletion is assumed.

## 3.4.1.2 The prefield position in the movement and deletion account

In order to draw conclusions on the validity of the movement and deletion account from prefield configurations, it is crucial to assume that movement in fragments really targets the prefield according to Merchant's account. The structure in Figure 3.12 differs from the one that Merchant (2004a) assumes for English because C and [Spec, CP] are always filled in regular German declarative matrix sentences. In contrast, English declarative matrix sentences are TPs, and the C head hosting E is phonologically empty. Therefore, it is not immediately clear which position Merchant identifies as the landing site for fragments in German. In principle there are three options: First, like Merchant suggested for English, there could be an FP above CP and the E feature could be located on C. Second, there could be an FP above CP and the E feature be hosted by F. Finally, there could be no FP in German, but an E feature located on C that triggers movement of fragments to [Spec, CP]. All of these options are compatible with the theory, because Merchant (2004a) suggests to account for crosslinguistic differences with respect to the availability and properties of ellipses by postulating differences in the specifications of the lexicon entries of the E feature.

For the German verb-second clause, as the discussion in Section 2.4.3 showed, any analysis that locates the E feature on C incorrectly predicts that the inflected verb survives ellipsis, because it is moved to C and only the complement of C is PF-deleted.<sup>52</sup> In contrast, assuming an FP above CP and locating E on F has the advantage of deleting the verb and thus being able to generate DP fragments. However, it incorrectly predicts fragments to be insensitive to islands, because the defective trace in [Spec, CP] would be deleted along the way. As for the third option, the assumption of an FP in German lacks empirical support, because the prefield hosts only one constituent and fronted foci appear in the regular prefield (58a) instead of preceding other prefield constituents (58b). If no FP is assumed in German, this rules out the first two options listed above.

- (58) a. *Einen Topf* musst du nehmen, keine Pfanne!  
       a.ACC     pot    must you     take no     pan  
       b. \**Einen Topf* du musst nehmen, keine Pfanne!  
       a.ACC     pot you   must    take no     pan  
       'A pot you must take, not a pan!'

<sup>52</sup>Note that this problem also concerns the exceptional movement version of the theory by Weir (2014a). Weir can account better for the non-constituent fragments discussed in this section, because he simply adjoins fragments to CP and there is no upper bound on the number of adjuncts. There might be constraints on the order of constituents in fragments, depending on whether the most deeply embedded focused constituents or the closest ones to the E feature are fronted first. However, Weir places the E feature on C and consequently falsely predicts that finite verbs (which are also in C), as Figure 3.12 shows, always survive ellipsis.

Furthermore, the identification of the landing site for fragments as [Spec, CP] is also implicitly adopted by Merchant (2004a) himself. As I discussed in Section 2.4.3, Merchant (2004a: 702) presents parallelisms between the form of fragments in the prefield and in fragments as evidence for his account. This implies that the presumed landing site for fragments is the regular prefield, that is, [Spec, CP].

#### 3.4.2 Experiment 10: Multiple prefield constituents

##### 3.4.2.1 Background

The German prefield is a promising testing ground for movement and deletion: Since Merchant (2004a) assumes it to be the landing site for fragments, only those expressions that can appear there in regular sentences can be grammatical fragments. Experiment 10 tested this hypothesis using the same method as in the experiments on CC topicalization, i.e. by comparing the acceptability of fragments to that of the corresponding left dislocation structures.

Testing this prediction empirically requires establishing which expressions can and which ones cannot appear in the prefield. Since the prefield can be filled with most of the syntactic categories,<sup>53</sup> the main restriction is that only a single constituent can precede the verb, because there is only one landing site in [Spec, CP]. In fact, the possibility of fronting an expression in verb-second clauses is often used as a constituency test in German.

Despite this, a corpus-based collection of multiple prefield constituents by Stefan Müller (Müller 2002, 2003, 2005) shows that, at least superficially, the requirement of German declarative matrix clauses to be verb-second is frequently violated. Three examples from Müller (2003: 32, 38, 59) are given in (59). Some of these examples can probably be analyzed as single constituents, depending on the theoretical background that is assumed. However, in (58a), it seems odd to adjoin a sentential adverb to a DP<sup>54</sup> and is it also not immediately clear that the locative and temporal adverbials in (58b) are simply adjoined to each other, as suggested by Haider (2000). (58c) can be analyzed as VP fronting following movement of the verb to T or C, depending on where the adverbial *des öfteren* is

<sup>53</sup>Only few expressions, such as modal particles, cannot appear there, as (i) shows.

(i) \*Wohl<sub>i</sub>/Ja<sub>i</sub> hat Peter *t<sub>i</sub>* ein paar Leute eingeladen. (Ott & Struckmeier 2016: 227)  
      PRT       has Peter   a   few people invited  
      ‘Peter has probably invited a few people.’

<sup>54</sup>But see Bogal-Allbritten (2013) for an account of how modal adverbials modify DPs.

placed.<sup>55</sup> In fact, Müller (2005: 13–22) himself argues that such prefield configurations constitute a single constituent headed by a phonologically null verb.

- (59) a. [Vermutlich] [vom gleichen Täter] wurden zwei Tankstellen in  
probably of.the same criminal were two gas.stations in  
Hemsbach und Heidelberg überfallen.  
Hemsbach and Heidelberg assaulted  
‘Probably by the same criminal two gas stations in Hemsbach and  
Heidelberg were assaulted.’
- b. [Vor drei Wochen] [in Memphis] hatte Stich noch in drei  
before three weeks in Memphis had Stich still in three  
Sätzen gegen Connors verloren.  
sets against Connors lost  
‘Three weeks ago in Memphis Stich had still lost in three sets against  
Connors.’
- c. [Studenten] [einem Lesetest] unterzieht er des öfteren.  
students a reading.test submits he the more.frequent  
‘Students a reading test he submits frequently.’

From the perspective of an empirical investigation of movement and deletion however, it is irrelevant how apparent multiple prefield constituents are analyzed: If the German prefield is the landing site for fragments, the movement and deletion account predicts that only those expressions that can somehow be moved there are possible fragments. Experiment 10 tests this prediction by comparing three instances of multiple prefield configurations that Müller (2003) classifies as acceptable to two of those he argues that are ungrammatical. Again, these prefield configurations are tested both as fragments and in sentences.

<sup>55</sup> Müller excludes some apparent multiple prefield constituents from the set of problematic cases. For instance, Müller (2003: 21) argues that in (i) *den Wagen* and *zu reparieren* are not independent from each other, as accusative on *den Wagen* has to be licensed by the verb *reparieren*. Therefore, what is fronted is a complete verbal projection and not two independent constituents.

(i) Den Wagen zu reparieren wurde versucht. (Müller 2003: 21)  
the car to repair was tried  
‘It was intended to repair the car.’

### 3.4.2.2 Materials

Experiment 10 investigates five different prefield configurations in an acceptability rating study. Since multiple prefield configurations are restricted to specific information-structural contexts (Müller 2005, Bildhauer 2011), all critical utterances were preceded by a context that elicited the appropriate information structure. For instance, in (60), where both the direct and indirect object are fronted, Müller (2003: 59) notes that the first prefield constituent *seinem Chef* ‘his boss’ must be a contrastive topic (Büring 1997) and the second one *eine E-Mail* ‘an e-mail’ a contrastive focus. The context question in (60a) intended to rule out the possibility that the utterance is inappropriate in out-of-the-blue contexts by information-structurally licensing the marked word order. Each item was tested as a fragment (60b) and in the prefield of a full sentence (60c).

(60) Direct and indirect object

- a. Hätte er der Personalabteilung ein Fax schicken sollen?  
has.SBJV he the HR.department a fax send shall  
‘Should he have sent a fax to the HR department?’
- b. Nein, seinem Chef eine E-Mail.  
no his boss an e-mail  
‘No, his boss an e-mail.’
- c. Nein, seinem Chef eine E-Mail hätte er schicken sollen.  
no his boss an e-mail has.SBJV he send shall  
‘No, his boss an e-mail he should have sent.’

The experiment tested three prefield configurations that are grammatical according to Müller (2005) (direct + indirect object (60), local + temporal adverbial (61) and argument + sentential adverb (62)) and two that are not (subject + other argument DP (64) and non-clausemates (65)). Some of these conditions might be analyzed as involving a single constituent in the prefield: The cooccurrence of a locative and a temporal adverbial (61) could be analyzed as adjoined to each other, but semantically both modify the remaining clause and not one of them the other one. In the sentential adverb and argument condition (62), the adverb *angeblich* ‘allegedly’ takes scope over *in seiner Stammkneipe* ‘in his favorite pub’ only. This might indicate that it forms a constituent with the noun or DP. Müller (2003: 31) nevertheless cites (63) by Jacobs (1986: 112) for evidence that sentential adverbs cannot occur inside a PP. This indicates that they may be semantically associated with the noun, but not syntactically. After all, the purpose of the experiment is not to isolate prefield configurations that are equally acceptable but to investigate whether acceptability differences among them are reflected in the acceptability of fragments.

- (61) Locative and temporal adverbial
- a. Wann hast du Hans denn getroffen?  
when have you Hans then met  
'So when did you meet Hans?'
  - b. [Heute morgen in der U-Bahn] habe ich ihn getroffen.  
today morning in the subway have I him met  
'This morning in the subway I met him.'
- (62) Sentential adverb and argument XP
- a. Wo war Herr Veit zum Tatzeitpunkt?  
where was Mr Veit to.the time.of.crime  
'Where was Mr Veit at the time of the crime?'
  - b. [Angeblich in seiner Stammkneipe] war er zum Tatzeitpunkt.  
allegedly in his favorite.pub was he to.the time.of.crime  
'Allegedly in his favorite pub he was at the time of the crime.'
- (63) \*Peter träumt von (vermutlich/ sogar/ nicht) (ihr/ Luise/ Geld).  
Peter dreams of probably/ even/ not her/ Luise/ money  
'Peter dreams of probably/even/not her/Luise/money.'

The two ungrammatical prefield configurations are given in (64) and (65). As for (64), Müller (2003: 59) notes that a preverbal subject and an additional argument cannot appear together in the prefield. The ungrammatical configuration in (65) involves the extraction of two constituents from different clauses to the prefield. This violates the requirement of prefield constituents to be clause-mates (Fanselow 1993). In (65), *den Hund* 'the dog' is the direct object of the embedded verb *ärgern* 'to bother', while *Paul* is the indirect object of the matrix verb *verbieten* 'to forbid'. Note that, unlike in (64), there is no subject involved in the multiple prefield sequence in (65). The prefield configuration itself should thus be acceptable if the constituents were not extracted from different clauses, as in (66), because it consists of the direct and the indirect object like the presumably grammatical (60).

- (64) Subject and argument XP
- a. Wer möchte welche Aufgabe übernehmen?  
who wants which task take.on  
'Who wants to take on which task?'
  - b. \*[Ich die Spülmaschine] möchte übernehmen.  
I the dishwasher want take.on  
'I want to take on the dishwasher.'

### 3 Experiments on the syntax of fragments

- (65) Extraction out of different clauses
- a. Wem hast du verboten, wen zu ärgern?  
whom have you forbidden who to bother  
'Who did you forbid to bother who?'
  - b. \*[Paul den Hund] habe ich verboten, zu ärgern.  
Paul the dog have I forbidden to bother  
'I forbade Paul to bother the dog.'
- (66) Paul den Hund habe ich geschenkt.  
Paul the dog have I given.as.present  
'I gave Paul the dog (as present).'

It shall be noted that Merchant (2004a: 710–711) briefly discusses the (introspective) observation that short answer fragments to multiple *wh*-questions are relatively acceptable even though the corresponding multiply filled prefield is heavily degraded. Without going into details, he suggests that this also evidences repair effects by ellipsis. I address this in the discussion of this experiment.

#### 3.4.2.3 Procedure

The experiment was completed by 38 undergraduate students of Saarland University. All were native speakers of German. They were rewarded with the participation in a lottery of  $5 \times € 30.00$  among all participants. The experiment was conducted in the same session as the production study on case marking (experiment 2), where subjects were asked to produce utterances referring to graphical stimuli. Subjects were asked to rate the naturalness of the italicized target utterance in the context of the question on a 7-point Likert scale (7= fully natural). In order to prevent possible floor effects, SENTENTIALITY was tested as a between subjects variable. Each subject rated 35 items (7 per PREFIELD configuration). The materials were presented together with 21 items of the follow-up to experiment 8 and 25 unrelated fillers including four ungrammatical controls in individually pseudo-randomized order. A pseudo-randomized presentation ensured that no two items of the same experiment followed each other. Fillers were adapted so as to match the SENTENTIALITY of the materials in the list, so that subjects saw either only sentential or only fragment target utterances. This also matched the design of the follow-up to experiment 8, which had the same manipulation. One participant rated 50% or more of the ungrammatical controls with 6 or 7 points on the scale and was therefore excluded from further analysis.



## 3.4.2.4 Results

Figure 3.13 shows that across all conditions fragments were rated better than sentences and that there was a large extent of variation between conditions. Specifically, the presumably ungrammatical conditions and grammatical conditions do not behave uniformly: The grammatical SAdv, XP and the grammatical LocAdv, TEMPADV are both almost equally acceptable as fragments, but strongly differ as prefield configurations. Due to these differences, it would not be appropriate to pool the presumably ungrammatical and grammatical conditions. Instead, I conducted pairwise comparisons between each two of the conditions, yielding a series of 10 2×2 contrasts (SENTENTIALITY×PREFIELD) that I analyzed separately. For each of these contrasts I statistically analyzed a subset of the data containing only those data points belonging to the respective condition with CLMMs according to the procedure described in Section 3.1.1.5.

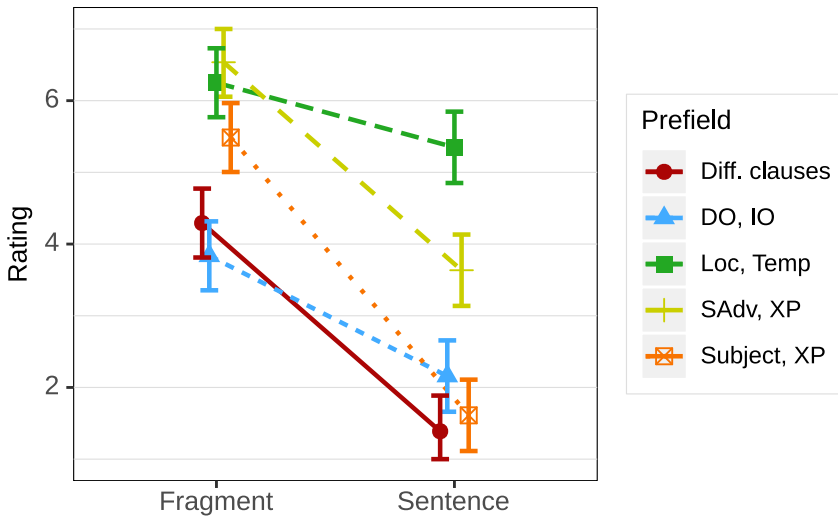


Figure 3.13: Mean ratings and 95% confidence intervals across conditions in experiment 10.

The initial model for each data set contained main effects for SENTENTIALITY and PREFIELD as well as the interaction between these predictors. As for by-subject random effects I included only the intercept and a slope for PREFIELD, since SENTENTIALITY had been tested between subjects. Similarly, I included by-item random intercepts and slopes for SENTENTIALITY, because PREFIELD was not varied between items. The crucial predictor is the SENTENTIALITY:PREFIELD interaction, which indicates that the difference between conditions cannot be explained solely by a theoretically uninteresting overall preference for fragments

### 3 Experiments on the syntax of fragments

Table 3.23: Significance of the pairwise comparisons between prefield configurations.  $p$ -values were Bonferroni-corrected, i.e. multiplied by the number of comparisons ( $n = 10$ ).

|           | SUBJ, XP                      | SADv, XP                      | LOC, TEMP                     | DO, IO                        |
|-----------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| NON-CM    | n.s.                          | n.s.                          | $(\chi^2 = 11.52, p < 0.001)$ | $(\chi^2 = 15.77, p < 0.001)$ |
| DO, IO    | $(\chi^2 = 15.73, p < 0.001)$ | $(\chi^2 = 10.95, p < 0.001)$ | n.s.                          |                               |
| LOC, TEMP | $(\chi^2 = 14.66, p < 0.001)$ | $(\chi^2 = 11.32, p < 0.001)$ |                               |                               |
| SADv, XP  | n.s.                          |                               |                               |                               |

or the markedness of a specific construction. Since the movement and deletion account predicts that only those prefield configurations which are acceptable as such yield grammatical fragments, it predicts no such interactions.

Table 3.23 summarizes the pairwise comparisons. Due to multiple comparisons, the reported  $p$ -values were Bonferroni-corrected, that is, multiplied by the number of comparisons ( $n = 10$ ). First of all, in most of the pairwise comparisons there are significant SENTENTIALITY:PREFIELD interactions. The pattern in Table 3.23 suggests that there is a split between conditions, but that this split does not occur between those prefield configurations that are grammatical and those that are ungrammatical according to Müller (2003). The SENTENTIALITY:PREFIELD interactions are significant for any comparison between LOC, TEMP and DO, IO and the remaining three predictors, but not between these two predictors. There are no significant interactions in the comparisons between the other three predictors. This suggests that the preference for the fragment is stronger in the two presumably ungrammatical prefield conditions and in SADv, XP.

#### 3.4.2.5 Discussion

Experiment 10 tested whether movement restrictions on German multiple prefield configurations, which have not been investigated from this perspective previously, constrain the form of fragments. The idea underlying the experiment is that, if the movement and deletion account is correct, only those expressions that may appear in the preverbal position in German verb-second clauses yield

acceptable fragments. Statistically, this would be reflected in the absence of SENTENTIALITY:PREFIELD interactions between conditions.

As Table 3.23 shows, however, there are significant interactions in six out of ten pairwise comparisons. These interactions are unexpected under a movement and deletion account, but they do not ultimately falsify it, because some of the data points are close to the extremes of the scale and thus could reflect ceiling and floor effects, respectively. However, there are three conditions which are close to the mean rating for ungrammatical controls ( $\mu = 1.79$ ) for sentences and ( $\mu = 2.25$ ) for fragments,<sup>56</sup> and among these specifically the SUBJ, XP condition yields acceptable fragments.

These interactions arise specifically when comparing the presumably grammatical prefield conditions (which involve the fronting of at least one adverbial) to the ungrammatical ones (which involve the fronting of more than one DP) and show that fragments that Merchant (2004a) would derive from ungrammatical prefield configurations are rated as better than expected based on the main effects only. This contradicts the prediction of the movement and deletion account and is specifically pronounced for the SUBJ, XP condition, which is as degraded as ungrammatical fillers in the sentence condition, but more acceptable than the grammatical DO, IO if it appears as a fragment. The SAdv, XP condition seems to be somehow special, as it patterns with the ungrammatical conditions with respect to the relative preference for fragments, but is based on a possibly grammatical prefield configuration. Taken together, the experiment shows that some prefield configurations which are clearly ungrammatical are the presumed underlying structure of well-formed fragments.

Although this seems to challenge the movement and deletion account, Merchant (2004a: 710–711) argues that fragments derived from multiple prefield constituents evidence repair effects. The mechanism he proposes for the repair of island violations could potentially explain the data from the non-clausemates condition. If the embedded clause in (65), repeated here for convenience as (67), is an island, extraction of the DP *den Hund* out of it will leave a defective trace, in the [Spec, CP] of the embedded clause. This trace is deleted by ellipsis on PF, so that the derivation is saved. Still though, it is unclear how *Paul* and *den Hund* would be merged to a single constituent that can appear in the prefield.

- (67) \*[Paul den Hund] habe ich verboten, zu ärgern.  
       Paul the dog have I forbidden to bother  
       'I forbid Paul to bother the dog.'

<sup>56</sup>SENTENTIALITY was tested between subjects, so that I present separate mean ratings for ungrammatical controls in each of the two groups.

### 3 Experiments on the syntax of fragments

The case of the SUBJ, XP condition is even harder to explain from a movement and deletion perspective, because deriving fragments from a verb-second sentence is intricate, as the discussion in Section 3.4.1.2 showed. Under a standard analysis of German verb-second word order, fragments cannot target the prefield in [Spec, CP] because this would falsely predict that the finite verb also survives ellipsis after raising to C. Furthermore, if the subject is fronted in a matrix clause, as it is in the SUBJ, XP conditions, there is no such defective trace that would be deleted by ellipsis. The only solution in terms of movement and deletion is the stipulation of a functional projection above CP, whose head hosts the E feature and whose specifier is the landing site for fragments. However, such a projection (i) would be specific to ellipsis, (ii) would predict fragments to be island-insensitive, and (iii) is not in line with Merchant (2004a), who identifies the regular prefield as the landing site for fragments. The more stipulations lacking independent evidence have to be made in order to reconcile the data with the predictions of the movement and deletion account, the less explanatory adequate is the theory.

If the movement and deletion account fails to account for the data, the obvious question is whether the other theories discussed so far do better. Reich's (2007) *in situ* deletion account can explain the observed pattern because the close relationship between question and answer in terms of focus structure is central to his theory. The constituents that survive ellipsis correspond to *wh*-phrases in the context question and are therefore focused, or at least not given, as in case of the adverbials which are not included in the question. In contrast, the part of the sentence that is omitted in the fragment is backgrounded by the question. *In situ* deletion hence predicts all fragments tested in experiment 10 to be grammatical. The observation that some of the prefield configurations are more acceptable than others does not concern *in situ* deletion, because it derives fragments from regular sentences with only one prefield constituent. Weir's (2014a) exceptional movement account in principle makes a similar prediction, because he argues that all focused constituents are adjoined to CP when they are moved out of the ellipsis site.

The nonsentential account might be able to account for those of the fragments that can somehow be analyzed as forming a constituent. This concerns the conditions involving adverbials, if adverbials are analyzed as adjoined to each other in the LOC, TEMP condition and to the DP in the SADV, XP condition. It is unclear how Barton & Progovac (2005) would account for the DP-DP fragments in three out of the five conditions. This is particularly relevant for the NONCLAUSEMATES and SUBJ, XP conditions, for which a small clause analysis seems to be impossible, because the base positions of the involved constituents are definitely distant

from each other. Taken together, the multiple prefield data contradict the movement and deletion account and the nonsentential account, but they are in line with in situ deletion.

### 3.5 The syntax of fragments: Discussion

This chapter presented a total of 10 experiments that empirically investigated predictions of the competing theories of fragments discussed in Chapter 2. The experiments addressed two main research questions that differentiate between the theories: First, whether fragments are sentential, and second, whether their derivation requires obligatory movement. The first question was investigated by using structural case marking on DP fragments as evidence for an unarticulated verbal head. Experiments 1–3 provide evidence for such unarticulated structure. The second question was investigated at the case of three movement restrictions which the movement and deletion account predicts to affect the acceptability of fragments: A ban on preposition omission in German (experiments 4–7), a crosslinguistic restriction on topicalizing complementizer-less CCs (experiments 8 and 9), and restrictions on multiple prefield constituents in German (experiment 10). Taken together, these experiments find no clear evidence for movement and deletion and in conjunction with the experiments on sententiality support the in situ deletion account. In what follows I discuss the main results and their relevance to my experiments on the usage of fragments in Chapter 5.

#### 3.5.1 Fragments are sentential

The experiments in Section 3.1 investigated case connectivity as potential evidence for unarticulated structure in fragments at the example of accusative structural case in German. Accusative is assigned to the complement of transitive verbs, hence accusative case marking on DP fragments evidences the existence of an unarticulated verbal head in fragments. Experiment 1 showed that fragments in German can appear in accusative structural case even in the absence of linguistic context. Experiment 2 confirmed the availability of a salient antecedent in the materials. Since the experiments show that accusative structural case is acceptable, they support a sentential account of fragments. This interpretation relies on the assumption that the German accusative is structural case. Progovac et al. (2006) question this for Serbian, but I showed that their diagnostics yield the opposite result for German.

### 3 Experiments on the syntax of fragments

Some of the experiments designed to test the movement and deletion account also speak against the nonsentential account. Experiment 6 tested a potential explanation for the impossibility of omitting prepositions in German short answers if the *wh*-phrase in the question is embedded within a PP. If prepositional case is structural case, which must be checked by the preposition, Barton & Progovac (2005) predict that DP fragments in prepositional case are ungrammatical. In contrast, nominative default case-marked DPs are expected to be grammatical. The experiment clearly disconfirms this prediction, because nominative is rated as even worse than prepositional case in the absence of a preposition. Progovac et al. (2006) argue that default case might be sometimes degraded for pragmatic reasons, but the theory would still predict a pragmatically odd expression to be more acceptable than an ungrammatical one.

Experiment 10 showed that some apparently discontinuous fragments, DP-DP sequences like *Ich die Spülmaschine* ‘Me the dishwasher’ are acceptable in appropriate contexts. Since fragments have to be maximal projections according to the nonsentential theory, it is unclear how it would account for non-constituent fragments. The small clause analysis that Progovac (2006) proposes fails to explain the data, because (i) the second DP is not the complete predicate of the first one, and (ii) Reich (2017) discards a small clause analysis in German for theoretical and empirical reasons. Another possible account is Müller’s (2003) suggestion of an empty verbal head, but the idea of the nonsentential account is obviously to do without unarticulated structure.

Finally, experiment 3 addressed the possibility of a mixed account of fragments that restricts the nonsentential derivation of fragments to contexts where no antecedent for ellipsis (resolution) is available. Such an account would predict that accusative on DP fragments is licensed only in contexts where a salient antecedent is available and that fragments appear in nominative default case otherwise. The experiment disconfirms this prediction, as accusative is preferred to the same extent in both types of contexts.

Taken together, the experiments on the syntax of fragments presented so far disconfirm the central predictions that distinguish the nonsentential account from its competitors: Neither is structural case unavailable in fragments, nor is default case preferred over ungrammatical alternatives. I take this as evidence that fragments are elliptical sentences.

#### 3.5.2 Fragments are not obligatorily moved

In Sections 3.2–3.4 I presented a series of experiments that investigated how the unarticulated structure in fragments looks like. These experiments tested

potential evidence for the influential movement and deletion account by Merchant (2004a), which assumes that the derivation of fragments involves obligatory movement to the left periphery before ellipsis applies. I pursued the approach that the alternative, *in situ* deletion, is the null hypothesis because this theory is derivationally simpler. Movement will only be assumed if there is empirical evidence that the *in situ* deletion account cannot explain. All experiments on movement follow the line of reasoning that movement would be evidenced by effects of movement restrictions on the form of fragments, as Merchant (2004a) suggests. I investigated three instances of movement restrictions: obligatory pied-piping of prepositions (experiments 4–7), complement clause topicalization (experiments 8 and 9) and restrictions and different configurations in the German prefield (experiment 10).

The first series of experiments addressed the P-Stranding Generalization (Merchant 2001, 2004a), which states that only those languages that have P-stranding allow for the omission of prepositions in short answers. Experiments 4 and 5 confirm the introspective pattern reported in the literature: In German, the preposition cannot be omitted in short answers to questions where the *wh*-phrase is the complement of a preposition. In contrast, in English the preposition can be stranded in the question, and when it is, omission is possible and actually preferred. Experiment 7 tested whether this pattern can be explained without the assumption of movement in short answers by a structural parallelism between questions and answers. Structural parallelism provides a non-movement explanation for the PSG: In German, DP short answers to PP questions are unavailable because the preposition always appears as the complement of P in the question. This approach makes the testable prediction that the form of the answer matches that of the question in languages where P-stranding is available. There are two possible accounts of question-answer parallelism: a semantic one based on the idea that questions denote structured propositions (Reich 2002a), and a psycholinguistic one, which builds on the observation that interlocutors reuse structure from previous discourse (Levelt & Kelter 1982). Under both of these approaches, there is a nonsyntactic relationship between the availability of P-stranding in the question and the preference for preposition omission in the answer. Experiment 7 revealed an effect of the form of the question on that of the answer, just like question-answer parallelism predicts, which is line with the corpus data in Nykiel (2017). The evidence for question-answer parallelism does not *falsify* the movement and deletion account but it provides an explanation for the data under the derivationally simpler *in situ* deletion account.

As for complement clause topicalization, Merchant et al. (2013) showed that an apparently well-established movement restriction on the topicalization of com-

### 3 *Experiments on the syntax of fragments*

plement clauses that lack an overt complementizer is in line with the acceptability of the corresponding fragments. Experiments 8 and 9 replicate their experiment in English and German under more controlled conditions. First, only complement clauses which are grammatical *in situ* are tested, second, ratings for the corresponding left dislocation structures are collected as well, and third, context stories ruled out the possibility of indirect answers. Experiment 8 confirms the effect reported by Merchant et al. (2013) for German fragments, but the data for left dislocations provide no evidence for the alleged movement restriction. In English (experiment 9), both types of fragments were equally acceptable and the pattern for left dislocations resulted to be the opposite of that reported in the literature. Hence, any effects found in fragments must be attributed to other parameters than a movement restriction.

Finally, experiment 10 tested whether restrictions on multiple prefield constituents in German constrain the acceptability of the corresponding fragments. Again, this is not the case. Specifically, the ungrammatical prefield configuration of a subject and a further argument DP is acceptable as a fragment and there is no obvious explanation for this under the movement and deletion account.

None of the three potential sources of evidence for movement that I investigated provides evidence for movement in fragments: The P-stranding data are in line with the movement and deletion account, but can also be explained by question-answer parallelism. The alleged movement restriction on topicalization could not be empirically demonstrated, and multiple prefield configurations that are ungrammatical result in acceptable fragments. Furthermore, from a more theoretical perspective, the derivation of fragments from German verb-second sentences turned out to entail unsolved theoretical issues that concern the location of the E feature in the German left periphery and empirically false predictions resulting from that.

#### 3.5.3 **Conclusion and outlook**

Taken together, experiments 1–10 favor the *in situ* deletion account by Reich (2007). Specifically, the experiments that investigate movement restrictions find no evidence for the parallelism between the acceptability of left dislocation and the corresponding fragments that underlies the arguments in Merchant (2004a). This may be accounted for by the exceptional movement version of the movement (Weir 2014a), but since the derivationally simpler *in situ* deletion account can explain the data as well, the assumption of an additional movement step is unmotivated and does not improve the empirical coverage of the theory. As for the nonsentential account, I focused on the version proposed by Barton & Progovac (2005). The predictions of this theory are disconfirmed by experiments 1–3,



which show that DP fragments can appear in structural case, and experiment 6, which rules out the nonsentential account's case checking-based account of the PSG. Nonsentential accounts operating in different frameworks (e.g. Ginzburg & Sag 2000, Fernández & Ginzburg 2002, Culicover & Jackendoff 2005) might still be able to explain the data.

From a more agnostic theoretical perspective, the experiments presented so far evidence some properties of fragments which have been controversially discussed in the literature. First, since fragments are derived by ellipsis, they receive the same morphological marking as in the corresponding sentences. Second, an in situ deletion account can generate discontinuous fragments, e.g. DP-DP sequences, even when they do not form a single constituent. Third, fragments are not subject to movement restrictions, hence sequences of words that cannot appear in a left-peripheral position might be nevertheless well-formed fragments.

Syntactic accounts of fragments explain how these expressions are generated by grammar, but not why speakers use fragments at all, and when they do so. Some of the accounts state that omissions in fragments must be licensed, e.g. by a salient linguistic (Merchant 2004a, Reich 2007) or nonlinguistic (Stainton 2006) antecedent, but even if an ellipsis is licensed, it does not always occur. Furthermore, if different fragments are possible in a context, it is unclear why specific words are omitted, and why others are not. In the second part of this book, I propose an information-theoretic account that explains (i) why fragments are sometimes preferred over full sentences, and (ii) why specific words are preferably omitted in fragments. In simplified terms, I hypothesize that fragments are preferred over full sentences when they make the most efficient use of the hearer's limited processing resources. Empirically testing this account requires not only to model the choice between a fragment and a full sentence but also to predict which words are omitted.

Since the experiments on the syntax of fragments suggested that fragments are grammatical objects, in order to determine usage preferences, it is necessary to take the syntactic properties of fragments into account. In this respect, the results on the syntax of fragments inform the investigation of their usage. This prevents models of fragment usage to incorrectly predict that fragments which are ruled out by grammar are the most suitable utterance to perform a speech act.



## 4 An information-theoretic account of fragment usage

The second part of this book addresses the question of why speakers use fragments at all, and more specifically under which circumstances they prefer to utter a (particular) fragment rather than a full sentence. The theoretical literature discussed in Chapter 2 is not concerned with this issue. In generative syntax, usage preferences are often considered to be irrelevant to syntactic theory (see e.g. Newmeyer 2003), since historically the goal of this line of research is to develop a model that generates grammatical utterances but not ungrammatical ones.<sup>1</sup> Some theories of fragments make predictions with respect to the licensing conditions of fragments. For instance, according to the in situ deletion account (Reich 2007), fragments are licensed by a salient implicit or explicit QuD. However, licensing is only a necessary but not a sufficient condition for ellipsis to actually occur: Expressions which are given in the QuD and not focused *can* be omitted, but they do not need to be omitted, as the felicitousness of sentential answers to *wh*-questions trivially shows. Therefore, in order to explain usage preferences, there needs to be an additional mechanism that determines whether a speaker chooses a full sentence or a fragment in order to get her message across.

In this chapter I propose an account of such a mechanism, whose predictions are tested with a rating and a production study in Chapter 5. I pursue the idea that the choice between two or more ways of encoding a message is constrained by information-theoretic (Shannon 1948) processing principles. Throughout the remainder of this book, the term *signal* refers to an individual utterance that can be used to communicate a proposition. The term *message* refers to the proposition that the speaker wants to communicate. I understand the term *message* as including those pragmatic inferences that contribute to truth-conditional meaning, which are called *explicatures* in relevance theory (Sperber & Wilson 1995). This concept comprises e.g. the resolution of deixis and anaphors, but not further pragmatic inferences, like implicatures. A message may therefore be conveyed

---

<sup>1</sup>Bergen & Goodman (2015) provide an account of fragment usage, which is theoretically relatively closely related to my research, but as I discuss below in Section 5.5, their study is based on a very small and artificial data set.

by different signals, but a signal can also convey different messages. The latter point is particularly relevant in case of fragments, whose interpretation depends on how the omissions are resolved.

The central prediction of my account of fragment usage is that, given a set of grammatical utterances that can be used to communicate a message, the utterance that is most well-formed with respect to the information-theoretic principle of Uniform Information Density (UID, Levy & Jaeger 2007) is chosen to communicate this message. Information theory is a promising framework for modeling omissions in fragments, because it has been shown to account for the distribution of omission and reduction phenomena at different levels of linguistic analysis. Two constraints on linguistic variation follow from information theory: First, more frequent messages will receive shorter signals on average. Second, the internal structure of the message will be optimized by distributing information uniformly at the maximal relatively error-free transmission rate. This imposes an upper bound on the densification of the utterance and can provide an explanation for why omission is sometimes dispreferred even when it is licensed. I pursue the idea that a major reason for this distribution of optional reductions is that the Shannon information of an expression indexes its processing effort (Hale 2001) and that distributing processing effort uniformly makes the most efficient use of the limited cognitive resources that are available to the hearer.

This chapter is organized as follows. Section 4.1 provides a brief overview of information theory as it was originally introduced by Shannon (1948), before I outline the specific predictions that information-theoretic constraints make on the usage and form of fragments in Section 4.2.

### 4.1 Information theory

Information-theoretic concepts have been applied to diverse linguistic phenomena, but when Shannon (1948) developed the theory in the middle of the twentieth century, it was not intended to explain phenomena of natural language production and comprehension. Shannon was concerned with efficient communication across a noisy channel from an engineering perspective, for instance, he lists several techniques, such as telegraphy, telephony, radio or television that he assumed his theory would apply to. In this section I sketch the fundamental aspects of the theory that are relevant to its application to (psycho)linguistic questions in Section 4.2.

From a linguistic perspective, the *information* conveyed by a linguistic expression might be intuitively thought of as related to its meaning: For instance, processing an utterance like (1) modifies the Common Ground (Stalnaker 2002) by

adding the proposition that the sentence encodes, a set of presuppositions and possibly further pragmatic inferences. One might think that utterances are more informative the more information they add to the Common Ground.

- (1) The pub at the corner serves burgers and chicken wings.

The information-theoretic definition of *information*, however, is actually simpler. As Shannon (1948: 379) himself puts it, “semantic aspects of communication are irrelevant to the engineering problem” of getting a message across the channel. Instead, Shannon’s notion of information is solely determined by the probability of a message to appear in context.<sup>2</sup> The less likely a message is, the more informative it is, and vice versa. When applied to the sentence level, this idea is relatively intuitive, because unlikely messages require a larger update of the hearer’s assumptions about the state of the world, or of the Common Ground. For instance, a sentence that describes stereotypical situations, like (1), or even (2a), will appear to be less informative than one that describes surprising situations (2b). If a default hearer does not know anything about this pub in particular, she will assume that it is almost certainly true that they serve beer, very likely that they serve regular pub food, but unlikely that they serve Japanese cuisine.

- (2) a. The put at the corner serves beer.  
b. The pub at the corner serves tempura and ramen.

In principle, Shannon information could be quantified on a scale between 0 and 1 encoding the probability of a message given a probability distribution over all messages that are possible in the situation. In that case, a lower value on the scale would be equivalent to higher information. A message that is the only option to be uttered in a context has a probability of 1 and an impossible one has a probability of 0. Instead of the absolute likelihood, Shannon proposes to use the negative logarithm of the event probability, which he argues to be more suitable for various reasons, such as mathematical and practical usefulness.<sup>3</sup> Shannon uses the base 2, so that information is measured in bits according to the formula

<sup>2</sup>Bar-Hillel & Carnap (1953) proposed a semantic extension of the theory, but I restrict myself to Shannon’s version because this is in line with the current research in the field.

<sup>3</sup>In linguistics, given the large set of possible outcomes (e.g. possible sentences), the probability of an individual sentence, word or morpheme often turns out to be very low. In statistical analyses, such variables are often highly skewed and can be transformed into a (more) linear relationship by log-transformation, that e.g. linear mixed effects models (Bates et al. 2015) presuppose. Furthermore, Smith & Levy (2013) observe that the relationship between corpus frequency (i.e. probability) and reading time is logarithmic (but cf. Brothers & Kuperberg (2019)). This empirically supports the log-transformation of bare probabilities.

in Equation 4.1. Inverting the polarity has two effects: First, information is never negative, because  $p(\text{message}|\text{context})$  cannot be negative or larger than 1; and second, the amount of information is larger the less likely a message is. In the psycholinguistic literature, this concept of information is often referred to as *surprisal*, and I will also use both terms interchangeably in what follows.<sup>4</sup>

$$I = \log_2 \left( \frac{1}{p(\text{message} | \text{context})} \right) = -\log_2 (p(\text{message} | \text{context})) \quad (4.1)$$

I illustrated the relationship between probability and information on the basis of sentences, but the definition in 4.1 can be straightforwardly applied to expressions on any level of linguistic representations. For instance, on the word level, the information of *beer* in (2a) can be calculated as shown in Equation 4.2. Similarly, the information of a phoneme within a word or the likelihood with which of a specific part of speech follows another one can be quantified.

$$I(\text{beer}) = -\log_2 (p(\text{beer} | \text{the pub at the corner serves})) \quad (4.2)$$

The specific predictions that information theory makes with respect to the well-formedness of linguistic expressions result from an interaction of this probabilistic notion of information with the assumption that communication occurs through a *noisy channel*, which is a crucial part of the communication system that Shannon (1948) assumes. Figure 4.1 illustrates this system. In Shannon's original framework of communication through a technical device, he defines the components of the system roughly as follows: The *information source* produces the message to be sent, whose form is determined by the modality of communication. For instance, it can range from a sequence of letters in telegraphy to functions over time of different complexity, like an acoustic signal in telephony or spatial coordinates and color in the case of television (Shannon 1948: 380–381). The *transmitter* encodes the message into a format that allows it to be sent over the channel. The encoded message is termed the *signal*, which can consist of electric impulses in telephony or sequences of dots, dashes and spaces in telegraphy (Shannon 1948: 382). The signal is sent to the *receiver* over the *channel*. In Shannon's examples, the channel can be the wire or cable the signal is sent across. The receiver has to decode the incoming signal, that is, to convert it back into the original format. The message is then interpreted by the *destination*, which is the intended recipient.

<sup>4</sup>The term *Surprisal* was introduced by Hale (2001), who in turn attributes it to Attneave (1959) and accounts for the fact that unexpected messages appear as surprising to the hearer and require more processing effort (see Section 4.2.3 for details).

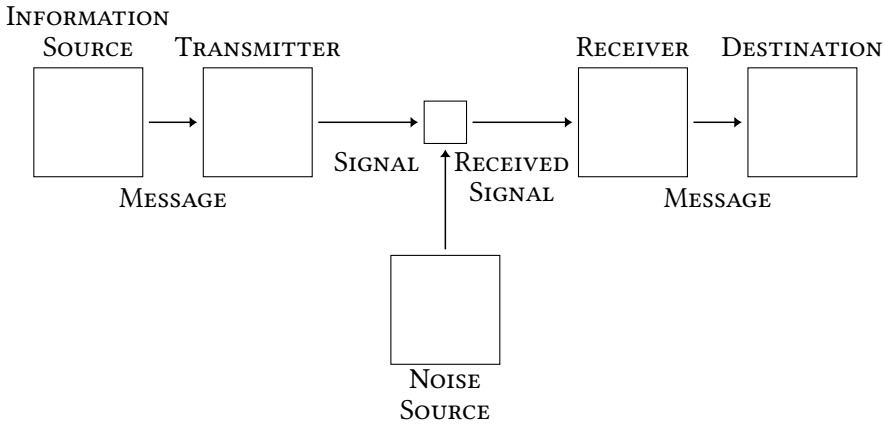


Figure 4.1: Shannon's model of communication (Shannon 1948: 381).

On an abstract level, encoding consists in assigning a signal to each possible message, and depending on which signal is assigned to which message, communication can be more or less efficient. Shannon distinguishes two properties of the channel that constrain the optimal form of signals to be sent through it, which must be considered by an encoding strategy in order to communicate efficiently.

First, the channel can be (and in practice most of the time is) noisy: Random noise can corrupt the signal during the transmission process, so that the signal passed to the recipient can differ from that sent by the source. For instance, if the signal consists in a sequence of letters A, B, and C, noise could transform a sent signal ABCA into ABBB. Shannon (1948: 410) observes that noise potentially constitutes a problem for communication, but that “by sending the information in a redundant form the probability of errors can be reduced.” An example of redundant encoding is to send each letter four times. For the above signal, this yields AAAABBBBCCCCAAAA. If now only one of the four repetitions of each letter was corrupted by noise on average, the intended letter could still be recovered by assuming that the most frequent letter in each substring is the one intended by the sender. Of course, this encoding strategy makes communication less efficient: If the signal length is increased by the factor  $n$ , sending it will take  $n$  times longer than sending the short signal. Therefore, efficient coding will involve a trade-off between the transmission of as much information as possible in a given interval of time and minimizing the probability of errors by including additional redundancy. Second, the channel has a limited *channel capacity*, which is measured in bits transmitted per unit of time. Shannon (1948: 401–413) shows that, given an appropriate coding system, information can be transmitted with a very low er-

ror rate unless the transmission rate does not exceed channel capacity. However, when channel capacity is exceeded, the likelihood of errors increases faster than the gain in intended transmission rate. Hence, attempts of increasing the transmission rate above channel capacity will never yield an advantage, but further reduce the actual transmission rate.

Taken together, in order to communicate efficiently across a noisy channel, the best choice is to communicate at a rate close to but not exceeding channel capacity: Not making use of the available bandwidth would be inefficient and more time-consuming, while exceeding channel capacity harms the purpose of communication due to the increased likelihood of errors and, as Shannon shows, will not yield an effectively higher transmission rate. In simplified terms, this requires interlocutors to allow for a certain degree of redundancy in their signal whenever channel capacity would be exceeded otherwise. As long as this is not the case, they should densify their utterance as much as possible in order to maximize efficiency.

On an abstract level, the idea that underlies information-theoretic research on language is that these general constraints on communication can explain optional variation in language. Grammar often provides a variety of signals that can be used to communicate a message, but does not explain why speakers choose a particular one in a specific situation. Specifically in the case of ellipsis, grammar determines whether an omission is licensed, but not all omissions that are licensed necessarily occur. From an information-theoretic perspective, a perfectly grammatical utterance might be dispreferred as compared to another one, for instance, because it is too redundant or because it exceeds channel capacity. This idea is worked out in detail in what follows.

## 4.2 Information-theoretic constraints on language

The model of communication in Figure 4.1 that Shannon (1948) assumes can easily be translated to any communicative situation between two interlocutors. Just like in the model, the speaker first has to encode her message, which can be thought of as a proposition, into an acoustic or written signal. This signal is sent across an acoustic or visual channel to the hearer, who has to decode, i.e. parse and interpret it. During the transmission process, the signal can be corrupted by noise. As I discuss in greater detail below, noise can be thought of as acoustic noise in the environment or as any other factor that results in a difference between the message sent by the speaker and the message received by the hearer. In principle, utterances can be optimized with respect to the properties of the



communication system on any level of linguistic analysis, be it phonemes, morphemes, words, more abstract syntactic constructions or complete sentences.

The account of fragment usage that I propose assumes that interlocutors optimize their utterances with respect to the goal of communicating efficiently through a noisy channel, as has been shown by previous research on information-theoretic constraints on language. Applied to natural language, the encoding process consists in assigning a linguistic signal, i.e. an utterance, to the intended message, that is, the proposition to be communicated. In the case of the choice between fragment and sentential utterances this optimization can involve the optional omission of words in the utterance, which might result in a preference for the fragment.

For the purpose of efficient communication, the encoding process must consider the properties of the components of the communication system. Previous research has identified two components of the model with respect to whose properties the signal is optimized: the source and the channel. Most of the recent work on information theory in psycholinguistics, in particular studies on the UID hypothesis (Levy & Jaeger 2007), focus on adaptation to properties of the channel. Even before that, Zipf (1935) observed a relationship between the frequency and length of linguistic expressions which suggests that statistical properties of the source also constrain encoding preferences. As Pate & Goldwater (2015) show, the optimization of the signal to properties of the source, which they term *channel coding* and *source coding* make partially differing predictions. In order to specify testable predictions of information theory on the form and usage of fragments, effects of source and channel coding have to be teased apart.

In what follows I discuss the predictions of source and channel coding on the form of utterances. It will become evident that both of these strategies predict that more frequent messages are more likely to be reduced, but only channel coding predicts the insertion of additional redundancy, the adaptation to the communicative situation and hence to the hearer's expectations.<sup>5</sup> Furthermore, in the case of fragments, only channel coding makes explicit predictions about *which* words are omitted and which ones are realized. These are the predictions which I test in the two experiments in Chapter 5.

### 4.2.1 Source coding

In Shannon's (1948) terminology, the source is the part of the communication system that generates messages, which then have to be encoded in order to be

---

<sup>5</sup>My experiments do not investigate the adaptation to the communicative situation. For evidence with respect to this see e.g. Pate & Goldwater (2015).

sent over the channel. Source coding focuses on the probability distribution over possible messages. In applications of information theory to linguistic phenomena the possible messages will differ in their likelihood most of the time. For instance, in a pub scenario, it is relatively likely that the customer will order drinks and food (in particular specific types thereof), but less likely that he wants to tell the waiter about an interesting linguistic paper that he recently read. As I discussed in Section 4.1, encoding consists in assigning a signal to each message, and most of the time the expressions that are available in the set of possible signals differ in length. In this situation, an efficient source coding strategy reserves the shorter signals for likely messages and assigns longer signals to unlikely ones (Shannon 1948: 402). This reduces the average length of an actually produced signal on average, because it ensures that shorter signals are sent more often.

Source coding effects have been reported specifically on the word level, since Zipf (1935) observed that more frequent words tend to be shorter on average in English, Latin and Chinese.<sup>6</sup> This motivates his Law of Abbreviation, which states that “as the relative frequency of a word increases, it tends to diminish in magnitude.” He argues that this principle results from a tendency toward “saving of time and effort” (Zipf 1935: 38). The number of possible short words in a language is naturally restricted by the limited inventory of phonemes and the syllabic structure of that language. Assigning shorter signals to more frequent messages reduces the average length of a random word as compared to a hypothetical language where there is no correlation between frequency and length. Consequently, the length-frequency correlation allows for the transmission of more words in less time, or, in the case of written speech, space.<sup>7</sup> More recently, Zipf’s (1935) observation has been replicated for a larger sample of languages by Piantadosi et al. (2011) and for semantically similar words that differ in length (e.g. *math* vs. *mathematics*) by Mahowald et al. (2013).<sup>8</sup>

The idea that shorter signals are assigned to more frequent, i.e. likely, messages can be extended to the sentence level. Source coding then predicts that more frequent messages are preferably encoded as a fragment. Unlikely messages will rather be encoded as a full sentence if all of the shorter fragments

---

<sup>6</sup>Zipf himself refers to Kaeding (1897) and Eldridge (1911) for previous tentative evidence in favor of this hypothesis. Zipf (1935: 23–25) however argues that for methodological reasons neither of these studies ultimately confirms the hypothesis.

<sup>7</sup>Zipf (1935: 30–36) relates diachronic changes, like the shortening of *gasoline* to *gas* and the substitution of *automobile* by *car*, to an adaption to the increased frequency of these words.

<sup>8</sup>Mahowald et al. (2018) furthermore show that across a variety of languages, even when word length is controlled, more frequent words have more phonotactically probable forms than less frequent ones.

have already been allocated to more likely messages. For instance, consider an extremely simplified taxi scenario, which models the communicative situation after a pedestrian (the speaker) hailed a taxi. In the scenario, there are only two messages that differ in their probability (3)<sup>9</sup> and three signals that differ in length (4), including a fragment (4c). Since the fragment can be derived from both (4a) and (4b), it can encode both messages in (3). A speaker who performs source coding will assign the fragment to the more probable message (3a), so that the only way of encoding the less informative message in (3b) is the less informative full sentence in (3b).

### (3) Messages

- a. The pedestrian wants a ride to the university.  $p = a; a > 0.5$
- b. The pedestrian wants to know how to get to the university.  $p = 1 - a$

### (4) Signals

- a. Take me to the university, please!  $length = 6$
- b. Explain me the way to the university, please!  $length = 8$
- c. To the university, please!  $length = 4$

Source coding consequently predicts that more likely messages are more often encoded as fragments, but this does not necessarily imply that less likely messages are preferably encoded as full sentences: In a hypothetical situation with very few possible messages, as long as there is a sufficient number of fragments to encode all messages, none of the messages will be encoded as a sentence. Source coding thus predicts no upper bound on densification: If unpredictable messages receive longer signals, this is only a by-product of the assignment of short signals to predictable messages, as the taxi example illustrated.

Source coding on the sentence level also makes no predictions on the internal form of the signal. It can explain why fragments are more often used to communicate predictable messages, but it cannot explain why specific words are omitted in these fragments. As will become evident throughout the next section, this is predicted only by channel coding accounts like Uniform Information Density.

### 4.2.2 Channel coding

Focusing on the source alone disregards the idea that the signal is transmitted over a noisy channel with limited capacity in Shannon's model: Speakers should

---

<sup>9</sup>For expository purposes, I do not provide semantic representations in this case, but simply a paraphrase of the speaker's communicative intention.

not only optimize their utterance with respect to properties of the source, but also with respect to those of the channel. In particular, they should avoid exceeding or repeatedly underutilizing the available capacity. This requires them to keep track of the distribution of information across the signal. Channel coding predicts a preference for those signals that ensure a transmission rate below but close to channel capacity. This is the main difference between the prediction between source and channel coding, because source coding imposes no such upper bound on the densification of the signal.

##### 4.2.2.1 Uniform Information Density

In the literature, channel coding has been discussed under different labels. In what follows, I sketch the general idea, its psychological reality and methods used to investigate its empirically testable predictions.

The idea that speakers adapt their utterance to the channel has been applied to linguistic phenomena for the first time by Fenk & Fenk (1980), Fenk-Oczlon (1989) and Fenk-Oczlon (1990) even before the more recent rise of information-theoretic accounts of language processing in psycholinguistics. Fenk & Fenk (1980) proposed a principle of *Constant Flow of Information*, which states that the amount of information sent by unit of time varies only weakly around a constant mean (Fenk-Oczlon 1990: 38). Fenk & Fenk (1980: 403) argue that increasing the rate of transmission too far above this mean would exceed the processing capacity of the hearer, whereas falling below the mean would be inefficient. This approach models channel capacity as an upper bound on the cognitive resources of the hearer, an idea that I also adopt in this book (see Section 4.2.3 for details).

More recently, the idea that speakers tend toward distributing information uniformly across the speech signal has been reformulated for different levels of linguistic analysis. For instance, the *Smooth Signal Redundancy* hypothesis by Aylett & Turk (2004) predicts that speech is smoothed on the phonetic level by modulating the length of syllables. On the level of complete sentences, Genzel & Charniak (2002, 2003) propose an *Entropy Rate Constancy* principle, which states that the average entropy of sentences throughout a text is constant. Whereas these principles are related to specific levels of analysis, Levy & Jaeger (2007: 24) explicitly claim that their *Uniform Information Density* (UID) hypothesis holds on any level of analysis. Work relying on the notion of UID has mostly focused on morphosyntactic variation, such as contractions (Frank & Jaeger 2008) and omissions of function words (Levy & Jaeger 2007, Jaeger 2010) as well as grammatical markers (Kurumada & Jaeger 2015, Norcliffe & Jaeger 2016), which will turn out to be particularly relevant to the investigation of omissions in fragments.

For this reason, I will use the notion of UID in order to refer to the principle of distributing Shannon information as uniformly as possible across the utterance.

On an abstract level, UID predicts that if there are several signals that can encode a message, everything else being equal, the speaker will choose the signal which comes closest to the ideal distribution of information, which approximates channel capacity without exceeding it. At least in the original version of the theory, the set of possible signals is restricted to grammatical expressions, since Jaeger (2010: 25) argues that optional variation occurs only within “the bounds defined by grammar.” This assumption is crucial to the empirical investigation of fragments, because it implies that UID favors the most well-formed *grammatical* signal and does not take into account utterances that might distribute information even more uniformly but that cannot be derived by grammar.

Following Levy & Jaeger (2007: 849), the *information density* of an utterance is defined as “the amount of information per unit comprising the utterance”, that is, the sum of the information carried by each individual word within the utterance. This total information density mass is to be distributed as uniformly as possible across the words comprising the utterance. Distributing information uniformly implies that speakers “avoid peaks and troughs in information density” (Levy & Jaeger 2007: 849), which result from transmitting too little or too much information per unit of time. In that sense, troughs are local information minima that result in an inefficient use of channel capacity, whereas peaks are information maxima that exceed channel capacity and therefore hamper communication.

### 4.2.2.2 UID effects on omissions

In order to illustrate how omissions in fragments may contribute to the optimization of utterances with respect to UID, consider again the taxi example that I discussed above. In this situation, a pedestrian hails a taxi, because he needs a ride to the university. In this simplified example, he can in principle choose between a full sentence (5a) and a fragment (5b) to communicate this message.<sup>10</sup>

- (5) a. Take me to the university, please.
- b. To the university, please.

In the taxi scenario it will be in general very likely that the passenger wants to go somewhere, so the material that is omitted in the fragment (*take me*) is very predictable. In contrast, it is unlikely that the driver knows the passenger’s

---

<sup>10</sup>Of course, this is highly simplified, because he could make use of a wide variety of different syntactic constructions and lexicalizations of fragments and sentences.

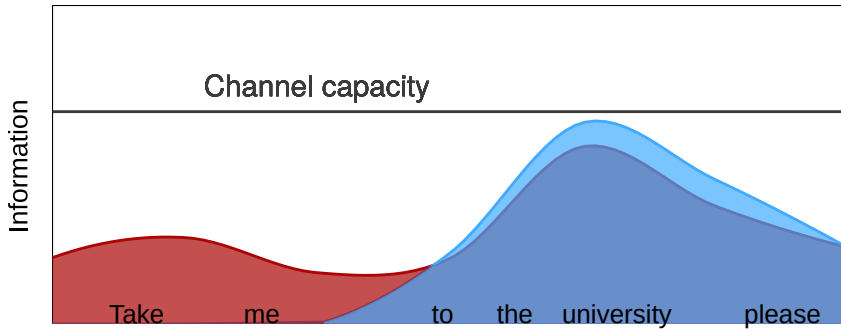


Figure 4.2: Hypothetical ID profile for the predictable sentence *take me to the university* and the meaning-equivalent fragment *to the university* in the taxi scenario. The blue area illustrates the distribution of information in the fragment and the red area that in the full sentence.

destination, this destination is unpredictable and relatively informative. Figure 4.2, which shows the distribution of information over time,<sup>11</sup> illustrates this idea with hypothetical information density (ID) profiles for the fragment and the corresponding full sentence.

If the pedestrian wants the driver to tell him the way to the university instead, he has to choose between the fragment in (5b) and the sentence in (6). In that case, *tell me the way* is probably less predictable than *take me*, as Figure 4.3 suggests. Of course, whether a word is predictable depends on properties of the utterance context. For instance, when an utterance like (6) is not produced by a passenger approaching the taxi, but by the driver of another car with a foreign license plate, it might be more likely that he would ask the local taxi driver for the way than that he wants to go somewhere. Similarly, when the passenger is brought to the university by the same driver on every Wednesday, or he wears a Denver Nuggets hat and shirt an hour before the match starts, the destination might be more likely and the utterance possibly even further reduced.

(6) Tell me the way to the university, please.

Figure 4.2 shows how the local information minimum, or *trough*, in the ID profile that is caused by the redundant *take me* is smoothed by omitting this expression.

<sup>11</sup>The variable on the abscissa in principle is time, because channel capacity and transmission rates are defined as an amount of information transmitted per unit of time, e.g. in bits per second. In practice, however, specifically corpus-based work (see e.g. Levy & Jaeger 2007, Frank & Jaeger 2008, Jaeger 2010) simplifies this to the amount of information transmitted per word, because duration measures for words or appropriate transcriptions into phonemes are not available in the corpora or would complicate the statistical analysis.

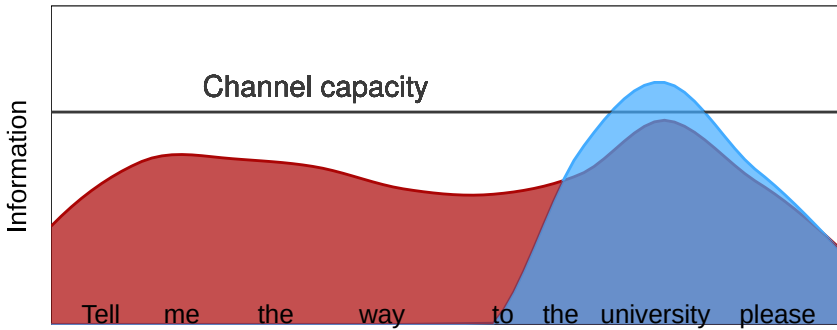


Figure 4.3: Hypothetical ID profile for the unpredictable sentence *tell me the way to the university* and the meaning-equivalent fragment *to the university* in the taxi scenario. The blue area illustrates the distribution of information in the fragment and the red area that in the full sentence.

From a UID perspective, omitting such predictable words optimizes the signal. If these omissions target words that are obligatory in full sentences, this results in the preference of the fragment over the full sentence. In contrast, given the density profiles in 4.3, *tell me the way* does not yield a trough in the profile, hence there is no pressure to omit these words. Furthermore, its omission would result in a *peak* in the density profile that exceeds channel capacity.<sup>12</sup> Therefore, from the UID perspective, it is not beneficial to omit these words. Actually, even if *tell me* was redundant, its insertion is preferred as long as it contributes to reducing the peak on *to the university*.

The tendencies to (i) to omit predictable expressions and (ii) to realize expressions that reduce peaks on upcoming material are the central predictions of UID on the well-formedness of linguistic expressions. Both of them are empirically supported by previous research, as Frank & Jaeger (2008) show for contraction in English, Kurumada & Jaeger (2015) for Japanese case markers, Levy & Jaeger (2007) for relative pronouns in English, Jaeger (2010) for complementizers, Norcliffe & Jaeger (2016) for relative pronouns in Yucatec Maya, Asr & Demberg (2015) for discourse markers and Lemke et al. (2017) for articles. With the exception of Kravtchenko (2014), who investigates the omission of subjects in Russian, these studies investigate semantically relatively vacuous function words. It is therefore reasonable to assume that UID constrains omissions in fragments too,

<sup>12</sup>Note that Figure 4.3 is not fully accurate, because I assigned the identical fragment *to the university* different density profiles in the left and right panel for the purpose of illustration. See below in this section for a discussion of this issue.

but this does not necessarily follow from previous research. The finding by Tily & Piantadosi (2009) that more predictable nouns are more likely to be pronominalized, i.e. reduced, points in a similar direction, if ellipsis as a more radical form of reduction of given material. Furthermore, the relationship between predictability and reduction has also been evidenced by studies which find that predictable expressions are more likely to be reduced in terms of duration and/or articulatory effort both on the word level and on that of individual syllables (see e.g. Bell et al. 2003, Aylett & Turk 2004, Bell et al. 2009, Tily et al. 2009, Demberg et al. 2012, Kuperman & Bresnan 2012, Seyfarth 2014, Pate & Goldwater 2015, Brandt et al. 2017, 2018, Malisz et al. 2018).

Even though the concept is labeled *Uniform Information Density*, at least in the version adopted in current psycholinguistics, the property of uniformity is an artifact of the assumptions made and not a goal of the encoding strategy in its own right. Uniformity only follows from the approximation of the transmission rate to the channel capacity, but a uniform distribution far below channel capacity will still be dispreferred compared to less uniform signals that make a more efficient use of channel capacity. This leads to an important distinction between the effect of peaks and troughs with respect to the choice between alternative ways of encoding a message. Troughs are inefficient and therefore always to be avoided, if possible. In contrast, peaks only dispreferred if they exceed channel capacity. In what follows, I will imply this interpretation of uniformity when stating that a signal is more or less compliant to UID, unless stated otherwise.

As for now, there have been no attempts to quantify channel capacity. This would not be a promising endeavor, because, as Shannon (1948) showed, channel capacity is not a constant, but varies as a function of the noise rate in the channel. This indeterminacy of channel capacity is expected if UID is interpreted as a psycholinguistic constraint on communication. UID implies that speakers are engaged in audience design and adapt their utterances to expected properties (e.g. preferences and cognitive abilities) of the hearer, so this will necessarily involve inferences under uncertainty about channel capacity. Furthermore, from a methodological perspective, the absolute information estimate depends on the corpus used for this purpose. Since information is based on probabilities, a larger lexicon will result in lower average probabilities of individual items. What matters for empirical research on UID is that, even if channel capacity is unknown, on average, more informative words are more likely to yield a peak and more uninformative words are more likely to cause a trough in the ID profile.

Taken together, UID predicts that, given a set of possible signals, i.e. sentential and nonsentential utterances that can be used to encode a message and that com-



ply with grammar,<sup>13</sup> the preferred utterance is that which distributes information most uniformly across the utterance. This leads to the two specific predictions in (7a,b), which in turn imply (7c): If omissions occur more often in predictive contexts, because average words are more likely, the signal will on average be shorter in such situations.

### (7) *Predictions of UID on fragments*

- a. *Avoid troughs*: The more likely a word is in context, the more likely it is to be omitted (within the limits of grammar).
- b. *Avoid peaks*: Uninformative words can be inserted before very informative words in order to lower the surprisal of the latter (within the limits of grammar).
- c. *Densification*: Shorter signals, like fragments, are preferred in predictive contexts.

#### 4.2.2.3 UID effects on word order

The distribution of information across the signal can also be optimized by re-ordering the expressions that it comprises. Effects of word order on UID are not as central to my research question as omissions, because they are not unique to fragments, but my experiment 12 will also show that there is evidence for UID effects on word order. Word order effects are predicted by UID because a word's information depends on the context in which it occurs and variation of word order changes this context. In general, the larger the context of a word is, the more predictable this word tends to become, because preceding material narrows the range of possible continuations of the utterance. This has also been shown by Levy (2008) for verb-final contexts in German based on reading time data by Konieczny & Döring (2003). The more arguments of the final verb the hearer parses, the smaller does the set of potential completions of the sentence become. This surprisal reduction is reflected in reduced reading times for words that occur later in the clause. Fenk-Oczlon (1983) argues that this also accounts for the general tendency for given or topicalized expressions to precede new or focused ones (Chafe 1976): Given expressions are on average more predictable than new

---

<sup>13</sup>Most of the time, this set of possible signals will be too large to empirically evaluate the alternatives. Even if no fragments were considered at all, a theoretically infinite number of sentences can be used to convey a single message. Therefore, when it comes to evaluating the alternative set empirically, I will assume that there is one sentence equivalent to each message from which a set of fragment alternatives is derived by ellipsis. This concerns in particular the evaluation of the production study in Section 5.3.

ones, and therefore this ordering reduces the information of the new ones and yields a more uniform ID profile.<sup>14</sup> More recently, Sikos et al. (2017) report an effect of UID on the choice between pre- and postnominal modification of German nouns, and Speyer & Lemke (2017) observe an effect of the aggregated surprisal of relative clauses on their extraposition in historic stages of German.

Even though UID predicts effects on word order, fragments often do not allow for word order variation. For instance, in a German DP or PP, word order is relatively fixed and determined by grammar, and UID only determines the choice between grammatical expressions. I return to this issue in the discussion of experiment 12, because the data set that I collected in this study is suitable for the investigation of UID effects on optional word order variation, too.<sup>15</sup>

---

<sup>14</sup>Most of the time, topicality and high predictability will probably cooccur. However, these concepts operate on different levels of analysis. Topicality determines what an utterance is about (Reinhart 1981, Krifka 2007) but predictability determines only the likelihood of a word to be mentioned at a particular point within the utterance. Expressions can be predictable, but not topical, if they have are likely in context and they are focused. For instance, in (i), taken from Kuperberg et al. (2020), the target noun *swimmers* has a high cloze probability as compared to *trainees* or *drawer* – all of these three words appear in the comment and the topic of the sentence are the lifeguards.

- (i) The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence, they cautioned the (swimmers/-trainees/drawer).

<sup>15</sup>Exceptions to the fixed word order within DPs and PPs are e.g. DPs that contain multiple adjectives (i) and a few prepositions that can also appear postnominally (ii). I leave such variation aside, because in the first case there are semantic and phonological constraints that strongly bias the ordering of adjectives (Martin 1969, Dixon 1977, Cinque 1994, Wulff 2003), and postposition (ii) is restricted to single lexical items (Di Meola 2003).

- (i) a. das neue schnelle Fahrrad  
the new fast bike
- b. das schnelle neue Fahrrad  
the fast new bike  
'the fast new bike'
- (ii) a. wegen des Regens  
because the.GEN rain.GEN
- b. des Regens wegen  
the.GEN rain.GEN because  
'because of the rain'

### 4.2.2.4 Summary: Predictions of UID on the form of fragments

UID makes empirically testable predictions with respect to the preferred way of encoding of a message. Speakers use omissions of optional words in order to modulate the information density profile in two ways: On the one hand, omitting uninformative words can smooth the profile by avoiding troughs. On the other hand, the insertion of redundant words before otherwise unpredictable ones reduces peaks in the profile. Such effects, which will be observed on the word level, are more specific than the general preference to assign shorter utterances to more predictable messages that follows from source coding. This relationship between the probability of a message and the length of the signal that source coding predicts also results from UID. For predictable messages, the individual words will also be on average more likely, so that there will be a higher ratio of omissions, which in turn results in a shorter signal. While UID and source coding share the prediction that shorter signals are assigned to more likely messages, only UID predicts *which* words are omitted in fragments.

### 4.2.3 UID as efficient distribution of processing effort

The studies cited above support the basic prediction of UID, that is, the tendency of distributing information uniformly across the utterance. However, in the literature two different ways of mapping the abstract concepts in Shannon's model of communication to natural language have been suggested. These interpretations differ particularly with respect to the channel. On the one hand, specifically in phonetic research, the channel is interpreted rather literally as the space through which the signal is sent (see e.g. Aylett & Turk 2004). On the other hand, from a psycholinguistic perspective, the channel has been related to the processing resources available to the hearer, and channel capacity interpreted as an upper bound to these resources (see e.g. Fenk & Fenk 1980). Before returning to UID effects on omissions, I briefly review these approaches and argue why I adopt the second possibility and interpret Shannon information as a measure of processing effort, as has been suggested e.g. by Hale (2001) and Levy (2008).

The interpretation of the channel which is more closely related to the communicative situation modeled by Shannon (1948) conceptualizes the channel as the medium between speaker and hearer. From this perspective, the message can be corrupted by noise during transmission and UID ensures "robust information transfer in a potentially noisy environment while conserving effort", as Aylett & Turk (2004: 32) put it. Noise can be acoustic, but it can also consist in other modifications of the signal, like hearers being distracted by other tasks

(Häuser et al. 2019). As Shannon (1948) showed, an increased likelihood of noise reduces channel capacity, because the potential corruption of the message has to be counterbalanced by inserting additional redundancy. In particular on the phonetic level and in case of high noise ratios this is a reasonable assumption, because the prediction of information theory that speakers adapt their utterances in the presence of acoustic noise is a well-established finding, known as the *Lombard effect* (Lombard 1911). Experimental research has shown that this adaptation concerns a variety of parameters, including an increase in F0, speech level and vowel duration (Summers et al. 1988, Junqua 1994, 1996). This is in line with information-theoretic studies that find effects of predictability on the articulation and duration of words and phonemes (Bell et al. 2003, Aylett & Turk 2004, Bell et al. 2009, Brandt et al. 2017, 2018, Malisz et al. 2018).

However, it is unclear whether the assumption that UID effects are related to the presence of environmental noise holds to the same extent for higher levels of linguistic analysis, such as words or complete sentences. In regular face-to-face communication, in the absence of a significant source of acoustic noise, and specifically if the word level is concerned, it seems relatively unlikely that complete are misheard. Words that are similar to each other, like *Harry* and *Mary* might be misunderstood if a part of the word is corrupted by noise, but it is less likely that *Harry* is misunderstood as *Susan* for this reason.

The link between predictability and processing effort allows for an interpretation of UID as a strategy to communicate efficiently even in the absence of (perceptual) noise. Levy & Jaeger (2007: 850) note that “independently of whether linguistic communication is viewed as a noisy channel, UID can be seen as minimizing comprehension difficulty.” This is based on the insight that the effort required to process an expression is proportional to its predictability in context (Hale 2001, 2016, Levy 2005, 2008). In psycholinguistics it is a well-established finding that, everything else being equal, more predictable words are read faster (see e.g. Ehrlich & Rayner 1981, McDonald & Shillcock 2003, Demberg & Keller 2008, Smith & Levy 2013). Levy & Jaeger (2007: 850) relate UID and processing effort by suggesting that a uniform distribution of information minimizes the total processing effort of an utterance, which they define as the sum of the processing effort of all the words within this utterance.<sup>16</sup> From this perspective, the concept

<sup>16</sup>There is some disagreement in the literature on the scale on which processing effort and word probability are related. Levy & Jaeger (2007) note that this conclusion presupposes that the relationship between surprisal and processing effort is superlinear, but this assumption has been questioned more recently. For instance, Smith & Levy (2008, 2013) conclude that the relationship between surprisal and processing effort (as quantified by reading times in eye tracking and self-paced reading experiments) is linear, and more recently, Brothers & Kuperberg (2019)

of channel capacity in Shannon’s model can be interpreted as delimiting the upper bound of the processing resources available to the hearer for language comprehension within a given amount of time.<sup>17</sup> I follow this reasoning and therefore interpret channel capacity as an unknown and variable upper bound to the cognitive resources that are available to the hearer for processing within a fixed interval of time. Therefore, the results of the experiments on fragment usage presented below do not hinge on a specific (linear or logarithmic) relationship between the likelihood of a word and the effort required for processing it, but on the assumption that the cognitive resources available to the hearer are limited and on the insight that predictable words require less processing effort.

But *why* would processing effort be correlated to the probability of words or constructions in the first place? Following Hale (2001) and subsequent work (Levy 2005, Hale 2006, Levy 2008), the central idea is that processing effort is proportional to the work done by the human parser. Under the assumption of a fully parallel parser, this work consists in discarding those parses that are incompatible with an input.<sup>18</sup> In Hale’s (2001) model, the information, and consequently the processing effort, of a word is higher, the larger the cumulated probability mass of the parses that it disconfirms is. Formally, Hale (2001: 162) derives the surprisal of a word as shown in Equation 4.3, where the prefix probability  $\alpha_n$  is the cumulated probability mass of all parses that are compatible with the input at the corresponding word and  $\alpha_{n-1}$  is the cumulated probability mass of the parses compatible with the previous word.

$$S(w_n) = \log_2 \left( \frac{\alpha_{n-1}}{\alpha_n} \right) \quad (4.3)$$

This measure is equivalent to Shannon’s (1948) definition of information, because the higher the probability mass of the parses that are compatible with a

---

argue that raw corpus frequency is a better predictor of reading times than surprisal. Despite these concerns, even Smith & Levy (2013: 311), who argue against the superlinear relation, note that, if surprisal indexes processing effort, speakers should not overload their interlocutors’ working memory. Similarly, Jaeger (2010: 51) argues that this relationship “might be expected from any system that has access to limited resources.”

<sup>17</sup>This predicts effects of the situational context on channel capacity even in the absence of strong noise sources. For instance, if competing tasks that require a share of the cognitive resources which are otherwise available for language processing, this will also reduce channel capacity (Engonopoulos et al. 2013, Häuser et al. 2019). The prediction of UID is that if speakers are aware of that the hearer’s resources are allocated otherwise, they will also reduce the information density of their utterance by making their utterance more redundant.

<sup>18</sup>In contrast to Hale (2001), Levy (2008) uses Kullback-Leibler divergence between probability distributions over parses before and after processing an input. Levy’s approach is also sensitive to gradual changes in probability that do not result in the rejection of a parse.

word is, the more predictable this word is. Since all  $\alpha \leq 1$  and  $\alpha_n \leq \alpha_{n-1}$ , the larger the probability mass of the parses that are compatible with  $w_{n-1}$  but not  $w_n$  is, the higher is the surprisal of  $w_n$ . Surprisal equals 0 in case  $w_n$  excludes no parse that is compatible with  $w_{n-1}$ .

Taken together, there are two ways of interpreting the channel with respect to natural language: one based on the presence of noise in the channel and one relating Shannon information and processing effort. It is beyond the scope of this work to test whether the noisy channel-based or the processing-based interpretation of UID is correct, and they are not mutually exclusive. However, the processing effort version of UID seems intuitively more plausible to account for omissions in fragments.

#### 4.2.4 UID vs. other accounts of predictability-driven reduction

In the introduction to this chapter, I noted that currently there is no comprehensive theory of why specific omissions in fragments occur. However, there are two potential alternative explanations for part of the predictability effects on omissions in fragments that UID predicts. First, Ferreira & Dell (2000) analyze the optional omission of function words as driven solely by properties of language production. Second, information-theoretic measures like surprisal are probably often correlated to information-structural concepts like givenness, focus or topicality. Therefore, I dedicate the remainder of this chapter to distinguishing the predictions of these approaches from the information-theoretic one that I pursue.

##### 4.2.4.1 Availability-based production

Availability-based production (e.g. Bock 1987, Ferreira & Dell 2000) explains part of the data that I interpreted above as evidence for UID as the result of properties of language production.<sup>19</sup> This approach relies on the difficulty of retrieving a lemma from memory. The idea is that speakers intend to produce speech fluently, and that the effortful retrieval of infrequent words delays speech production and thus results in disfluencies. These disfluencies are counterbalanced by inserting optional words that keep speech production fluent. As Ferreira & Dell (2000: 299) suggest, the insertion of such words has a similar effect as an “um”.

The main prediction of this approach is that insertions of optional words occur before unpredictable words, as Ferreira & Dell (2000) show for complementizers in English. UID predicts this too, but for a different reason: Realizing words before unpredictable ones can reduce the surprisal of the latter and hence smooth peaks in the ID profile. However, availability-based production neither implies

---

<sup>19</sup>See also Jaeger & Buz (2017) for an overview and a comparison to UID.

that words that are themselves more predictable are more likely to be reduced, nor that predictable words tend to appear toward the beginning of the sentence. Therefore, if they were empirically confirmed, these two predictions will provide evidence for UID. Since UID and availability-based production are theories about different aspects of language, as Jaeger & Buz (2017) note, they do not mutually exclude each other, but what matters in the context of my experiments is that data that cannot be explained by production preferences alone will support UID.

### 4.2.4.2 Information structure

Even though there is no fully worked-out information-structural account of fragment usage, information-structural and information-theoretic concepts are probably often related. This might raise the question of whether surprisal is actually an artifact of information-structural notions like givenness or topicality. In what follows I show that the information-theoretic approach has explanatory, empirical and methodological advantages over a purely information-structural one.

Specifically sentential accounts of fragments assume a close relationship between information-structural concepts such as focus, background, givenness or topicality and ellipsis: For instance, Merchant (2004a) requires elided expressions to be e-given and Reich (2007) and Weir (2014a) argue that only foci survive ellipsis. The observation that in only expressions which are given can be elided reminds of the finding that given referents tend to be prosodically less prominent (Féry & Ishihara 2009), and is in line with the analysis of ellipsis as an extreme form of reduction of prosodic prominence (Tancredi 1992).

This raises the question of whether information structure alone can explain the distribution of omissions or whether information-theoretic considerations are required in addition. From an information-structural perspective, the omission of predictable material might result from a tendency for predictable words to be given, or highly salient, whereas foci are less predictable.<sup>20</sup> For instance, in the

---

<sup>20</sup>The question of whether surprisal is sometimes an artifact of information-structural concepts (which goes beyond the scope of this book) might be addressed with appropriate experimental studies, for instance by comparing focused expressions that differ in the number and likelihood of focus alternatives. While information theory predicts gradual effects of predictability, discrete concepts of focus and givenness predict a categorical difference between expressions that are focused and those that are not. Similarly, not all given expressions are equally likely to be talked about in upcoming discourse. For sluicing, Lemke et al. (forthcoming) show that even though in both contexts in (i) the person referred to by *somebody* is contextually given, participants are more likely to complete (ia) with a question referring to this referent (*with whom*).

- (i) a. Mary was making out with somebody, but I don't know ...
- b. Mary painted her room with somebody, but I don't know ...

taxi example discussed above, a salient implicit QuD like *Where do you want to go?* might license ellipsis of everything but the focus, which corresponds to the *wh*-phrase in the answer *Take-me to the university*. Since foci is defined by the presence of alternatives (Rooth 1992), they are necessarily less predictable than given constituents.

From a theoretical perspective, the main problem for a purely information-structural account of fragment usage is that information structure might license ellipsis, but it does not trigger it. Concepts like *e-givenness* determine which words *can* be omitted, but obviously *e-given* words are not always omitted. Therefore, information structure can only explain why certain expressions cannot be omitted. In contrast, UID provides an account of why predictable words are preferably omitted. Furthermore, unlike UID, an information-structural account of fragment usage does not predict the insertion of redundancy before unpredictable words: The omission of a target word is licensed only by its own information-structural status (like e.g. *(e-)givenness*). UID additionally predicts that the likelihood of the word that follows a target word also determines whether the target word is omitted. This does not neglect that information structure can contribute to the predictability of a word being omitted, but information structure alone does not explain all of the effects that UID predicts.

Taken together, there is probably a high degree of overlap between the givenness and surprisal of an expression, but only an information-theoretic account can explain why an expression whose omission is licensed is sometimes overtly realized. Nevertheless, it might be an interesting line of research to tease apart the predictions of an information-theoretic and an information-structural account in a controlled experimental setting.



## 5 Evidence for UID effects on omissions in fragments

This chapter presents two experiments which investigate the predictions of the UID-based account of fragment usage.<sup>1</sup> This account makes the three testable predictions in (1). (1a) and (1b) are specific to UID, whereas (1c) can be analyzed either as an implication of (1a) and (1b) or as the result of efficient source coding.

- (1) Predictions of UID on fragments
  - a. *Avoid troughs*: The more likely a word is in context, the more likely it is to be omitted (within the limits of grammar).
  - b. *Avoid peaks*: Uninformative words can be inserted before very informative words in order to lower the surprisal of the latter (within the limits of grammar).
  - c. *Densification*: Shorter encodings, like fragments, are preferred in predictive contexts.

The experiments investigate these issues at the case of discourse-initial fragments, which are the most uncontroversial instances of fragments. Since these fragments lack linguistic antecedents, the predictability of words within them is mostly constrained by extralinguistic context. In order to quantify effects of extralinguistic context on the predictability of utterances and words within them, both experiments rely on script knowledge (Schank & Abelson 1977) as an approximation to extralinguistic context. Scripts trigger expectations about upcoming events and can be used to modulate the predictability of utterances that are related to these events. Furthermore, there is a crowdsourced corpus of script knowledge available that can be used to precisely quantify this predictability.

Both experiments manipulate the likelihood of utterances with context stories like (2), which are based on event probabilities extracted from the DeScript corpus of script knowledge (Wanzare et al. 2016). For instance, in context of this story, the most likely event to follow is that of pouring the pasta into the boiling

---

<sup>1</sup>Experiment 11 has been published in Lemke, Schäfer & Reich (2021) and experiment 12 has been published in Lemke et al. (2020) and Lemke, Reich, et al. (2021).

water, hence I assume that utterances that refer to this event, like (2a) are more likely than those referring to events which are unpredictable in the script corpus.

- (2) Annika and Jenny want to cook pasta. Annika put a pot with water on the stove. Then she turned the stove on. After a few minutes, the water started to boil.
- |                                 |               |
|---------------------------------|---------------|
| a. Pour the pasta into the pot. | Predictable   |
| b. Set the table.               | Unpredictable |

Experiment 11 compares the acceptability of the sentences in (2a,b) to that of DP fragments derived from these sentences. Given clause (1c), UID predicts a relatively stronger preference for fragments in case of the predictable utterance (2a) than in case of (2b). Experiment 12 uses the same context stories to elicit a data set with a production task that is suitable for investigating the more fine-grained predictions in (1a) and (1b). The presence of ellipses in the data collected in experiment 12 requires a new method to estimate surprisal. My method extends the surprisal estimation technique proposed by Hale (2001) to elliptical data by allowing for an arbitrary number of omissions between words.

This chapter is organized as follows. In Section 5.1 I propose scripts as an approximation to extralinguistic context and describe how I created experimental materials based on the DeScript corpus (Wanzare et al. 2016). Sections 5.2 and 5.3 present the experiments, and Section 5.4 summarizes the main results.

## 5.1 Scripts as a model of extralinguistic context

In information-theoretic research on language, the surprisal of words is most frequently estimated from corpora using statistical language models.<sup>2</sup> Previous studies in the field relied mostly on  $n$ -gram models, which model the context of a word  $w_i$  as the  $n-1$  words that precede it. Unigram models consider only the overall frequency of  $w_i$ , bigram models return the conditional probability  $p(w_i | w_{i-1})$ ,

<sup>2</sup>Other methods include approximating the surprisal of function words with the likelihood of particular constructions, such as relative clauses (Levy & Jaeger 2007) or complement clauses (Jaeger 2010). This can be done either by calculating e.g. the subcategorization preferences of verbs from previously parsed corpora (Jaeger 2010) or by using probabilistic parsers, which operate on part of speech annotations and calculate the likelihood of the syntactic construction investigated (Levy & Jaeger 2007). Furthermore, other authors have simply stipulated that, everything else being equal, expressions that occur later in a sentence or text are more predictable, because previous material narrows the range of possible continuations. See e.g. Fenk-Oczlon (1989), Fenk-Oczlon (1990) and Genzel & Charniak (2002) for studies that (partially) relied on this assumption and Levy (2008: 1147) for empirical evidence.

trigram models use  $p(w_i | w_{i-2} w_{i-1})$ , and so on. By restricting context to a few words at most,  $n$ -gram models are only very coarse approximations to the models of context that human interlocutors probably construct. Even though there are currently more sophisticated language modeling techniques,<sup>3</sup> they are also not suitable to estimate surprisal in discourse-initial fragments: Since there is no or only little context in these utterances, the likelihood of words within them is determined by extralinguistic context to a large extent. Text corpora do not contain information about extralinguistic context, so the models cannot quantify its effect on the likelihood of words. Therefore, investigating predictability effects on fragments place requires a model of extralinguistic context. As I anticipated in the preceding section, I use script knowledge for this purpose.

Scripts (Schank & Abelson 1977) are stereotypical representations of everyday situations, which contain information about the default ordering of events as well as participants and objects involved (Bower et al. 1979). Three properties of fragments make scripts particularly suitable as an approximation<sup>4</sup> to extralinguistic context: First, scripts are accessed during text comprehension in order to retrieve implicit material, second, at least some scripts are shared by most speakers of a language, and third, people predict upcoming events based on script knowledge. In experimental settings, it can be assumed that scripts which are shared by a majority of the population trigger similar expectations for most of the participants. This allows for controlling and manipulating context-driven expectations: In the taxi scenario an utterance like *take me to ...* will be likely. Furthermore, there are script corpora available which consist in descriptions of the stereotypical time-course of scripts provided by a large number of participants. Based on these corpora, it is possible to build probabilistic models of context which can be used to precisely estimate event probabilities.

This section describes the model of extralinguistic context based on the DeScript script knowledge corpus (Wanzare et al. 2016) that underlies the stimuli for experiments 11 and 12. Section 5.1.1 introduces the concept of script as defined by Schank & Abelson (1977) and briefly discusses previous psychological

<sup>3</sup>Some models take into account hierarchical structure (Stolcke 1995, Hale 2001, Roark 2001, Levy 2008) or even material contained in previous sentences (Iyer & Ostendorf 1996, Oualil et al. 2016, 2017, Singh et al. 2016, Grave et al. 2017, Khandelwal et al. 2018, Devlin et al. 2019).

<sup>4</sup>It shall be noted that scripts are an approximation to extralinguistic context rather than a complete models thereof. Context is not only determined by script knowledge, since non-conventionalized, visual or other sensory information will also have an effect on expectations about upcoming events and utterances. In the taxi scenario, the pedestrian might be wheeling a bike, therefore it becomes relatively unlikely that he would ask for a taxi ride. However, the effect of such properties of context is difficult to quantify and in my stimuli I avoid the mention of such unexpected referents.

evidence that scripts indeed prime upcoming events. Section 5.1.2 presents the approach I used for estimating event probabilities from the DeScript corpus of script knowledge (Wanzare et al. 2016).

### 5.1.1 Script knowledge

The concept of *script* has been established by Schank & Abelson (1977) as an extension of the idea of *frames* developed by Minsky (1974). In principle, a script can be defined as the mental representation of a stereotypical everyday activity, such as grocery shopping, visiting a doctor, eating in a restaurant or attending a lecture (Bower et al. 1979). Schank & Abelson (1977) attribute scripts a central role in text comprehension, which consists in filling the gap between what is explicitly said and what is understood. They exemplify this at the case of a short story about visiting a restaurant (3).

- (3) John went to a restaurant. He asked the waitress for coq au vin. He paid the check and left. (Schank & Abelson 1977: 38)

Although the story omits many details, for instance that John sat down at a table, read the menu, ordered something to drink, and even the central act of eating the dish that he ordered, a hearer will infer these events from knowledge about the stereotypical time-course of eating at a restaurant as well as about the people and objects involved. Events that are highly predictable at some point in the script can remain implicit and will nevertheless be integrated in the hearer's representation of the described situation.

#### 5.1.1.1 The structure of scripts

In order to quantify the likelihood of an event at a specific point in the script it is crucial to know how its internal structure looks like. This concerns the hierarchical structure of the script as well as the ordering of events: If scripts were fully ordered sequences of events, each event would deterministically indicate what happens next. In what follows, I briefly sketch the structure of scripts as described in Schank & Abelson (1977), which differs in some aspects from the representations of scripts on which my stimuli are based. Figure 5.1 shows the structure of (a part of) the restaurant script according to Schank & Abelson (1977).<sup>5</sup>

First, each script has a *header* ("Restaurant"), a set of *roles* identifying the participants involved in the script, and a set of *props*, i.e. objects that typically appear

---

<sup>5</sup>Note that I replaced the original conceptual dependency theory (Schank 1975) representations of scripts by their natural language counterparts for expository purposes.

## 5.1 Scripts as a model of extralinguistic context

*Script: Restaurant*

Track: Coffee shop

*Props*

Tables, Menu, Food,

Check, Money

*Entry conditions*

customer is hungry

customer has money

*Roles*

Customer, Waiter, Cook,

Cashier, Owner

*Results*

customer has less money

owner has more money

customer is not hungry

customer is pleased (optional)

---

*Scene 1: Entering*

customer enters restaurant

customer looks at tables

customer decides where to sit

customer goes to table

customer sits down

---

*Scene 2: Ordering*

(menu on table) (waiter brings menu) (customer asks for menu)

customer takes menu

customer signals waiter

waiter goes to table

customer asks waiter for menu

waiter takes menu

waiter goes to table

waiter gives menu to customer

customer reads menu

customer chooses food

customer signals waiter

waiter goes to table\*

customer orders food

waiter goes to cook

waiter tells cook choice of food

cook tells waiter 'no food'

waiter goes to customer

waiter tells customer 'no food'

(go back to \*) or

(go to scene 4 [leaving])

cook does [prepare food script]

(go to scene 3 [eating])

Figure 5.1: Extract of the restaurant script based on Schank & Abelson (1977: 43). For expository purposes, I replaced the CDT representations by natural language counterparts and simplified Scene 2.

in that script. A script can have several *tracks*, for instance, Schank & Abelson (1977: 40–41) distinguish a “fancy restaurant track” and a “fast-food track” in order to account for differing sets of props, roles, events, and ordering thereof depending on the type of restaurant. Scripts are activated by their necessary *entry conditions* and lead to a set of *results*, some of which might be optional. If the entry conditions are not satisfied, e.g. when a customer who is not hungry or who has no money goes to the restaurant, the events will not follow their stereotypical time-course, so that applying the script will not yield a benefit in comprehension. The results follow from the application of a script or a particular version thereof.

Schank & Abelson (1977) assume that script events are hierarchically grouped into *scenes*. For instance, they divide the restaurant script into entering, ordering, eating and exiting scenes.<sup>6</sup> Each scene in turn comprises a set of (partially) ordered events. Although many events in the restaurant follow each other obligatorily, the ordering of events within a scene is neither complete, nor is the path to be followed through each scene fully linear. The *entering* scene is described as fully linear, but the *ordering* scene can develop in different ways depending on whether the menu is on the table when the customers sit down. If the waiter does not bring food to the customer, it is either possible to return to the food choice event or to skip the *eating* scene and proceed with *leaving*. I address this partially nonlinear ordering of events by estimating their likelihood in the context of the previous one(s) from a script corpus.

### 5.1.1.2 Scripts as primes for upcoming material

This brief sketch of Schank & Abelson’s (1977) view on scripts suggests that scripts are a promising approximation to extralinguistic context. Since scripts about everyday events are shared by a wide part of the population, and their representation is relatively homogeneous between individuals (Bower et al. 1979),

---

<sup>6</sup>There is some experimental evidence that scripts are indeed represented as hierarchical structures in memory. For instance, Bower et al. (1979) report a segmentation task on script data suggesting that subjects agree to a large extent on the placement of boundaries between script events. They argue that this indicates a natural segmentation of scripts into scenes. Abbott et al. (1985) present a memory recall task that evidences a distinction in asymmetric priming between script events and scene headers: While the former facilitate the recall of the latter, this does not hold vice versa. More recently, a similar structure has been assumed by Cooper & Shallice (2000) in order to model errors in script-based behavior, but see Botvinick & Plaut (2004) for a non-hierarchical account. As all of my materials involve a sequence of three consecutive script events on the same granularity level, both flat and hierarchical script models predict that the next event will be activated. I hence remain agnostic with respect to the precise representation of script knowledge in memory.

it is reasonable to assume that a script evokes similar predictions within at least most of the participants in an experiment. As I discuss in greater detail in the next section, the basic idea underlying my experiments is to manipulate the likelihood of a target event with a script-based context story. (4) exemplifies the structure of a sample item used in experiment 11 for the pasta cooking script. The context story consists of a sentence referring to the script title (4a) and a sequence of the three immediately preceding events (4b–d).<sup>7</sup> Given this context story, I expect script events (4e) that are likely in that context to be more predictable than non-script events (4f). This setting allows for the investigation of the hypothesis that utterances referring to predictable events are more likely to be reduced.

|     |    |                             |                  |
|-----|----|-----------------------------|------------------|
| (4) | a. | cook pasta                  | Script title     |
|     | b. | put pot with water on stove | Context event 1  |
|     | c. | turn stove on               | Context event 2  |
|     | d. | water boils                 | Context event 3  |
|     | e. | pour pasta into pot         | Target event     |
|     | f. | set table                   | Non-script event |

Based on Schank & Abelson's (1977) theory, it seems natural to assume that (4e) is predictable in the context of (4a–d), however, it needs to be empirically shown that this is indeed the case. Fortunately, a large bulk of experimental studies indicates that text comprehension involves the generation of predictive inferences about upcoming material (see e.g. Bower et al. 1979, McKoon & Ratcliff 1986, van den Broek 1994, van der Meer et al. 2002, Nuthmann & van der Meer 2005, Camblin et al. 2007, Otten & Van Berkum 2007, Hare et al. 2009, Bicknell et al. 2010, Matsuki et al. 2011, Metusalem et al. 2012, Delogu et al. 2018). For instance, Bower et al. (1979) find that sentences referring to script events are read faster when they follow the immediately preceding event in the script, as compared to contexts where one or two events in between them are omitted. This suggests that subjects generate expectations that constrain processing as they read script-based stories. More recently, van der Meer et al. (2002) show that the priming effect of script knowledge is stronger for upcoming events. Taken together, previous research on script knowledge predicts that the context story in (4a–d) will indeed prime the target event in (4e) as compared to the unrelated (4f). In what follows, I explain how the likelihood of an event in context was calculated based on the DeScript corpus of script knowledge.

---

<sup>7</sup>For details on how this structure is generated and why other potential script events such as *grab a large pot* or *open the pasta package* are not included, see Section 5.1.2.3.

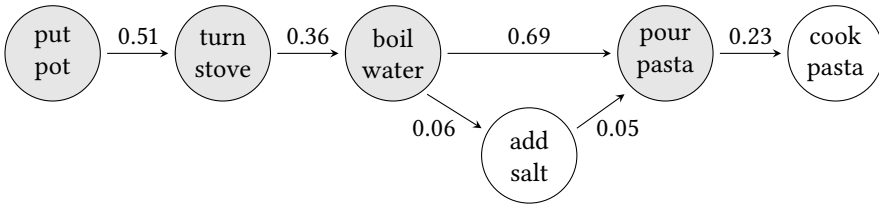


Figure 5.2: Sample event chain with transition probabilities between events estimated from the preprocessed DeScript data. The four events marked in grey were used in the item for the pasta scenario in experiment 11.

## 5.1.2 Estimating event surprisal from script corpora

### 5.1.2.1 Scripts as probabilistic event chains

The script representations underlying the materials for experiments 11 and 12 model scripts as probabilistic networks rather than as linear event sequences, as most of the previous work on scripts did. In such a network, each event is assigned a state  $e_i$  which has a transition probability to another state  $e_j$ . The transition probability  $p(e_j | e_i)$  indicates the likelihood of  $e_j$  to follow  $e_i$  and can be estimated for each pair of states  $\langle e_i, e_j \rangle$  in the total set of states that is determined by the script. The transition probabilities can range from 0, i.e.  $e_j$  never follows  $e_i$ , to 1 in case  $e_j$  always follows  $e_i$ . Figure 5.2 shows a part the abstract representation of the pasta scenario. Based on the transition probabilities it is possible to extract a linear sequence of the most likely events to follow each other even though the underlying structure itself is nonlinear. For instance, the four events marked in grey in Figure 5.2 were used to build the item for the pasta scenario given in (4).

The usage of probabilistic structures has conceptual and methodological advantages over hand-crafted script structures like the restaurant script sketched in Figure 5.1 above. First, probabilistic structures might be more empirically appropriate descriptions of differing representations of the same script within the population. Intuitions of an individual researcher might deviate from the overall most likely time-course of events, and averaging over data from about 100 contributors to the corpus data for each script is a better approximation to the general expectations about a script. Second, probabilistic event chains model the uncertainty about upcoming events in the script. Even in the detailed sketch of the restaurant script by Schank & Abelson (1977) there is no fully linear order but several branchings, for instance the waiter might serve the customer the intended dish or not. Speakers might have probabilistic expectations about which of these continuations is more likely, and these expectations must be quantified



when it comes to estimating the likelihood of upcoming events. This uncertainty is a property of scripts even if a single speaker has a fully deterministic view on it. For instance, a person who always prepares scrambled eggs in the same way has knowledge about how others do it. If scripts are used in the comprehension of new stories, it seems reasonable to rely not only on individual preferences: A hearer will expect with a certain probability that bacon, vegetables or spices are added. Finally, transition probabilities between events can be estimated from script corpora. In contrast, if the ordering between two events is reverted only once in the corpus, a linear order cannot be established. Taken together, all of these issues favor the resort to probabilistic representations of scripts that I use in my studies. In the remainder of Section 5.1, I describe the procedure that I used to extract context stories like that given above from the DeScript corpus of script knowledge (Wanzare et al. 2016).

### 5.1.2.2 Selection of scripts from the DeScript corpus

The experiments presented in this chapter are based on data extracted from the DeScript corpus (Wanzare et al. 2016). DeScript has been crowdsourced on Amazon Mechanical Turk and comprises about 100 event-sequence descriptions (ESDs) by native speakers of English for each of 40 scripts. The corpus is freely available in XML format and is partially annotated for coreferences between script events. The scripts contained in the corpus differ both in their internal complexity (e.g. *washing dishes* and *going to a funeral*), that is, variation between (the number of) events and their order as well as with respect to the number of participants involved.

The corpus also contains both what Schank & Abelson (1977) termed situational and instrumental scripts, between which Schank & Abelson (1977) assume a categorical distinction. Instrumental scripts “usually” have only one participant (Schank & Abelson 1977: 65), and the order of events in the script is more fixed than in situational ones. As examples of instrumental scripts, they list scripts like *lighting a cigarette*, *starting a car* or *frying an egg*. According to Schank & Abelson (1977: 66), a consequence of this (apparently gradual) structural difference is that instrumental scripts do not make use of “powerful predictive mechanisms” and that their details are forgotten faster than those of situational scripts as the story unfolds. As Schank & Abelson (1977: 66) put it, “[w]e simply don’t expect that ‘I fried an egg’ is the beginning of a story about an interesting thing that happened in the process of egg frying.” If the predictive potential of both script types differed, this could potentially affect the outcome of the experiments.

Since my experiments investigate encoding preferences for utterances, it was necessary that (at least) two characters participate in the script. This was not the case for the instrumental scripts in DeScript, leaving at most 17 scripts that involved a second participant. In order to test six trials per condition in experiment 11 in a 2×2 design, I required 24 scripts that involved at least two participants. Therefore, I adapted some of the scripts that originally did not contain a second participant, but for which it is reasonable to assume that the script would not be significantly changed by the introduction of such an additional character. For instance, in the pasta cooking scenario, it is very plausible that a couple, roommates or friends prepare a meal together and talk in the meantime. Consequently, I chose 24 scenarios from DeScript that best met this requirement.

Therefore, I included a binary control predictor `SCRIPTTYPE` in the statistical analysis of experiment 11 that (i) shows whether one of these script types is more predictive than the other one, and that (ii) if so, allows me to factor out this effect. Anticipating the results, there is no significant difference between script types with respect to their predictive potential.

### 5.1.2.3 Estimating event probabilities

Estimating the likelihood of an event in context of the preceding one(s) requires transforming the representations provided by the contributors to the corpus into event chains. After that, the likelihood of an event can be estimated with *n*-gram language models. Since in this case the primitive expressions are event labels instead of natural language words, I refer to his procedure as *event sequence modeling*, even though it is technically identical to the modeling of natural language data. Event sequence modeling requires that each event in the relevant corpus data is assigned a unique label that distinguishes it from other events. In what follows I describe how I preprocessed the DeScript data for the selected 24 scenarios in order to construct materials for experiments 11 and 12.

Following Manshadi et al. (2008), event labels consisted of the main verb of each event description and the post-verbal noun, which is its direct object in case of transitives.<sup>8</sup> The corpus data were therefore preprocessed in order to obtain event representations like (5), based on which event probabilities were

---

<sup>8</sup>More sophisticated methods for representing script events can take into account the semantic role of a character with respect to the verb (Chambers & Jurafsky 2008), use skip-grams (Jans et al. 2012) or include multiple arguments for each verb (Chambers & Jurafsky 2009, Pichotta & Mooney 2014). These approaches outperform simpler approaches in evaluation tasks in computational linguistics, but for my purpose of assigning each event a distinctive label taking the verb and post-verbal noun as event representation was sufficiently accurate.

estimated. Note that for the purpose of event sequence modeling, it does not matter whether e.g. *turn stove* is the most accurate description of the event of turning the stove on: As long as the same label is assigned to all instances of the corresponding event and to no instance of any different event, the model will correctly determine the likelihood of the corresponding event.

(5) put pot turn stove boil water pour pasta

The event descriptions in DeScript are diverse in various respects. First, script knowledge differs between individuals, who might perform the same script, e.g. cooking pasta or scrambled eggs, in a different fashion. Second, descriptions that do not differ in the nature and time-course of events sometimes do so in precision and granularity. Some subjects mention that they turn on the stove or take the pan out of the cupboard, while others begin the ESD with breaking up the eggs into the pan. Sometimes these omissions concern events that are necessary conditions for the following events: Even if picking a pan is not mentioned, this must have happened at the point where the eggs are broken inside it. Finally, descriptions of the same event differ with respect to the lexical items chosen, pronominalizations and ellipses, as the examples from DeScript in (6) show.

- (6) a. Pour eggs into the pan
- b. Put contents of bowl in pan
- c. Pour them into a pan
- d. Pour in pan

To some extent, this diversity is a property inherent to script knowledge, specifically with respect to different stereotypical orders of events between speakers. Since my UID account of fragment usage implies that speakers engage in audience design, whenever this adaption concerns script knowledge, the speaker must adapt her utterance to the (inferred) script knowledge of the hearer rather than to her own. Consequently, she must infer which expectations about the script the hearer has. Under the assumption that the sample of script representations for a given scenario in DeScript comes close to being representative for an average hearer, differences between the probability of events in the DeScript data will reflect relevant differences in likelihood of events given a generic hearer. Therefore, modeling the likelihood and ordering of events reflects psychologically relevant aspects of script knowledge. The opposite arguably holds for differences in lexical choice or syntactic constructions when describing the individual events. All of the descriptions in (6) refer to the same event of pouring the eggs into the pan, consequently they should be treated as the same event

in event sequence modeling. This requires a notable amount of preprocessing, that I describe in greater detail below. Differences in granularity are probably a case somewhere between actual diversity between script representations, which needs to be reflected in the event chains and linguistic variance in the corpus-based descriptions. On the one hand, it could be argued that in a sequence like (7) a stove and a pan are necessarily involved, and that the pan must have been put on the stove and heated in order to cook the eggs. On the other hand, I use the event chains as an approximation to the likelihood of events being referred to by an utterance, and events that are considered irrelevant enough to be omitted in an ESD might not be likely enough to be talked about. Therefore, I did not assimilate the ESDs with respect to granularity. Furthermore, doing so would involve a high degree of arbitrariness when it comes to deciding whether an event is necessary in the time-course of the script or not.

- (7) a. Break two eggs in a bowl
- b. Get a whisk
- c. Whisk eggs together till they are light and fluffy
- d. Add a bit of milk
- e. Start cooking eggs

The lexical and syntactic variation within the event descriptions in DeScript requires the assimilation of these descriptions, so that a single label is assigned to each event. For this purpose, the corpus data were pre-processed using a semi-automatic approach that is summarized in Figure 5.3 and described in what follows. After preprocessing, each instance of each event is assigned a unique label, so that event sequence models can be used to estimate its probability to occur in context. The labels for events were generated by first extracting the main verb and its complement noun from the event descriptions in DeScript. For this purpose, the raw DeScript data were Part of Speech-tagged with the Stanford parser (Klein & Manning 2003) for English contained in the Python Natural Language Toolkit (NLTK) (Loper & Bird 2002). The data were then dependency-parsed using the Stanford dependency parser contained in the NLTK. The parser was often misguided by the high ratio of elliptical event descriptions, subject omissions and verb-first imperatives that are infrequent in the written corpora on which it was trained. In such situations, it interprets e.g. initial verbs as nouns, specifically when there are homonymous with nouns like *set* and then assigns wrong POS tags to following words. This was addressed by using a language model file trained by Micaela Regneri and Ines Rehbein on a modified set of training corpora

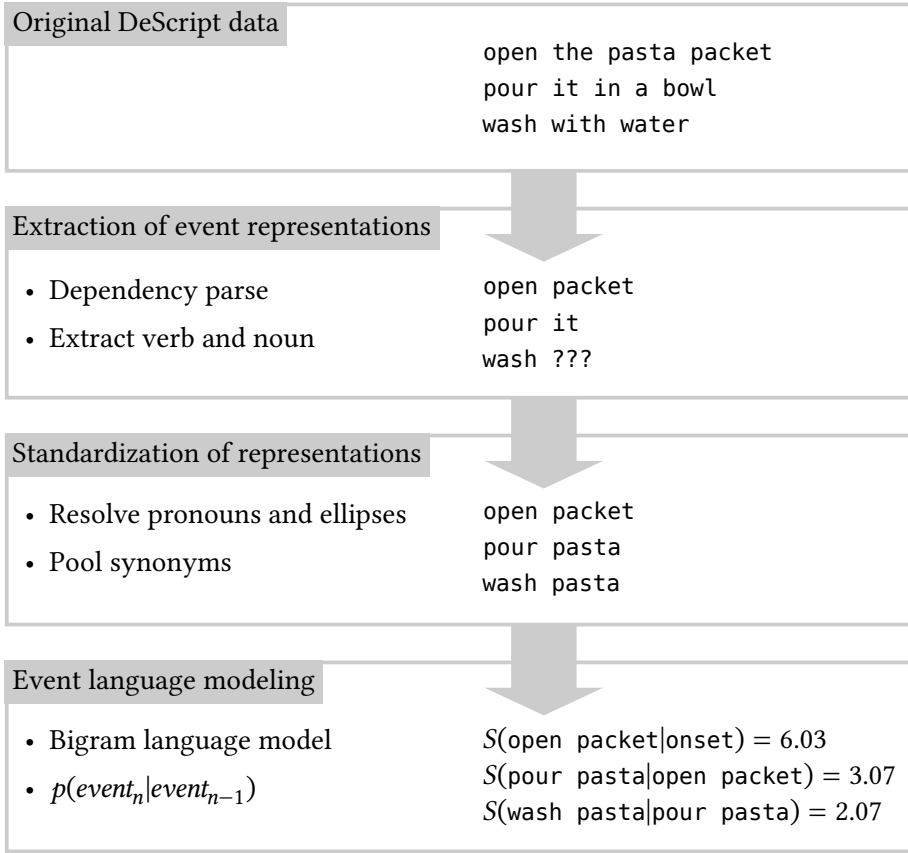


Figure 5.3: Overview of the preprocessing procedure for a sample sequence of events from DeScript.

from which some of the sentence-initial noun phrases had been removed.<sup>9</sup> This method allows the parser to analyze English SVO structures with missing subjects as such instead of analyzing initial verbs as nouns and results in a higher accuracy of the parser. After parsing, the main verb and its direct object were extracted using Python scripts. In case there was no direct object, a placeholder was inserted and reviewed manually.<sup>10</sup> The resulting verb noun event representations for each scenario were further manually preprocessed in order to pool synonym words and syntactically differing descriptions of the same event. The

<sup>9</sup>See Regneri (2013: 49–50) for details. I thank Simon Ostermann for suggesting this approach and sharing the model file trained on the modified corpora.

<sup>10</sup>I thank Lisa Schäfer for her suggestions and ideas that significantly influenced the methods described in this section.

rationale for this procedure was that (i) each script should involve a set of mutually exclusive participants (both animate and inanimate, i.e. roles and props in the terminology in Schank & Abelson (1977)), that there should be a unique label for each participant, and that (ii) the same held for events, so there should be a unique label for each event within the script.<sup>11</sup> The first requirement ensures that synonyms, such as *pan* and *skillet*, were pooled to a single lemma, whereas the second one requires the same label to be assigned to different descriptions of the same action, like those given in (6). This is crucial for interpreting the event sequence models calculated on these representations, because otherwise the probability mass of e.g. the event referring to pouring the eggs into the pan would be split among the events *pour egg*, *put content* and *pour in*. In order to obtain unique labels for each event, it was also necessary to resolve ellipses and the reference of pronouns. Finally, the data for each scenario were screened using an R script in order to ensure the uniqueness of each participant, each action, and consequently each event within the script.

After preprocessing, the likelihood of each event was estimated with bigram event sequence models with Good-Turing discounting using the SRILM toolkit (Stolcke 2002). In contrast to the language modeling approaches discussed so far, the primitives are not words, but events, and the models return the probability of an event given the previous one (or the script onset) based on representations like (7). The usage of higher order *n*-grams would not have been reasonable given the relatively small amount of data of about 100 ESDs per scenario. Even after preprocessing, relatively homogeneous scenarios such as *train ride* had a vocabulary size (the number of different primitive events) of 121, more diverse scenarios, such as e.g. *making scrambled eggs* even had a unigram vocabulary size of 192. As there is often more than one possible successor for each event, this yields a vocabulary of 351 bigrams for the train and of 672 bigrams for the eggs scenario.

Preprocessing the DeScript data for 24 scripts using automatized and manual procedures yielded a high-quality data set that I used to estimate the likelihood of

---

<sup>11</sup>This idea of preprocessing elicited script data is not fully new. Bower et al. (1979) started their series of experiments on script knowledge by collecting natural data on knowledge about five scripts. Subjects provided list descriptions of the events involved in the stereotypical time-course of each script, thus yielding data relatively similar to current script corpora. The data provided for each script (there were between 24 and 37 subjects and consequently descriptions per script) were preprocessed by unifying “paraphrases and synonyms” (Bower et al. 1979: 181) and then used to build ordered event lists comprising events mentioned by more than 25% of the respective subjects. Except for the smaller number of scripts and participants per script, this procedure anticipates the collection of script knowledge in more recent corpora of script knowledge (see Section 5.1.2) and the preprocessing approach that I apply to such data.

script-based events. The method described in this section ensures that the probability mass of an event is not split among alternative lexicalizations, and that speakers' script knowledge is a probabilistic estimate of how people represent a particular script, including differences in the events involved, their ordering and granularity. I used these probabilistic representations of script knowledge to construct the materials for experiments 3, 11 and 12.

## 5.2 Experiment 11: Script knowledge, rating

### 5.2.1 Background

Experiment 11 tests the densification prediction in clause (1c) that fragments are more strongly preferred in predictive contexts, which according to UID, results from the tendency to omit predictable words in order to avoid troughs in the ID profile. Empirical evidence for this prediction will also show that the script-based predictability manipulation works at all. This is a requirement for using the same stimuli in the production task in experiment 12 that investigates the predictions of UID on omissions on the more fine-grained word level.

### 5.2.2 Materials

Experiment 11 compares the acceptability of predictable and unpredictable DP fragments (8a,b) to that of corresponding full sentences (8c,d) in 2×2 design (PREDICTABILITY × SENTENTIALITY) in a rating study.<sup>12</sup>

- (8) Annika and Jenny want to cook pasta. Annika put a pot with water on the stove. Then she turned the stove on. After a few minutes, the water started to boil. Now Annika says to Jenny:

- |    |                                                                                                                         |               |
|----|-------------------------------------------------------------------------------------------------------------------------|---------------|
| a. | Die Nudeln, bitte.<br>the pasta please<br>'The pasta, please.'                                                          | Predictable   |
| b. | Den Küchentisch, bitte.<br>the.ACC kitchen.table please<br>'The kitchen table, please.'                                 | Unpredictable |
| c. | Schütte die Nudeln ins Wasser, bitte.<br>pour the pasta in.the water please<br>'Pour the pasta into the water, please.' | Predictable   |

<sup>12</sup>The experiment was conducted in German, but I provide an English translation of the context story here for convenience.

- d. Deck schon mal den Küchentisch, bitte. Unpredictable  
 set already PRT the.ACC kitchen.table please  
 ‘Set the kitchen table, please.’

The script-based context story is identical for all conditions. The target utterance in the predictable conditions always refers to the most likely event in the context of the story (pour pasta in the example). In the unpredictable conditions, it refers to an event that did not appear in the script data at all, or that has a probability of 0 in this context, but that is intuitively not implausible to be talked about in the situation described by the context story. The idea that underlies this approach is that probable events are more likely to be referred to with an utterance than unpredictable ones.<sup>13,14</sup> Each context story consists of four sentences, the first of which introduced to the script and mentioned its title, e.g. *cook pasta*.<sup>15</sup> The remaining three sentences represent a high-probability event chain based on the bigram event probabilities in the preprocessed DeScript data. This sequence ensures that the event in the target sentence (8a) (pour pasta) refers to the most likely event to follow the previous one (boil water). The preceding two events in the context story are selected by the same criterion, so that the event that follows them is the most likely one in the script representations derived from DeScript. Events that were overall rare ( $n < 8$ ) in the processed data for each script were

<sup>13</sup>Using the probability of an event as a proxy for that of an utterance is a simplification, and very likely events might actually *not* be talked about because they are so obvious. However, experiment 12 confirms that predictable events are more likely to be talked about.

<sup>14</sup>In order to rule out the possibility that differences between the PREDICTABILITY conditions concern other factors than predictability, it would have been desirable to test the *same* utterance in a predictive and an unpredictable context. However, due to the usage of script corpus data, the only possible method was to construct one story per item and to vary the utterance between PREDICTABILITY conditions. If new (unpredictable) contexts were constructed from scratch, it would have been impossible to estimate the likelihood of the target utterance in the same way as for the corpus-based stories. Therefore, I used one context story by item and varied the target utterance between PREDICTABILITY conditions. Potential differences between both utterances are furthermore accounted for by by-item random slopes for PREDICTABILITY in the statistical analysis.

<sup>15</sup>Schank & Abelson (1977) argue that scripts are only accessed for text comprehension after their *activation* and *instantiation*. They assume that it were implausible that possibly hundred of scripts are always active at the same time, but that scripts are only activated when their *header* is encountered. In the sample item, the header is *cook pasta* in the first sentence of the context story. According to Schank & Abelson (1977: 47–48), instantiation consists in copying the complete script to working memory, so that its components can be accessed during the comprehension of linguistic input. In the theory, instantiation occurs when a subsequent sentence fits the structure of an activated script, i.e. the hearer encounters a script event. Even though this distinction is controversial (see e.g. Rabs et al. 2017), in the experimental stimuli each script should be active and instantiated after processing the introductory sentence of the context story and the following sentence, which refers to the first event.



not considered in this process. Otherwise, for instance an event that was mentioned by only one out of 100 participants would be taken to represent the script knowledge of the complete population. The context story ends with an introductory sentence like *now Annika says to Jenny* in (8), that determines which of the characters produces the target utterance. This target utterance differs between the four conditions in (8a–d).

In the predictable conditions, the target utterance refers to the event that is most likely to follow the last event in the context story. The event referred to by the target utterance in the unpredictable conditions has a probability of 0, either because it is not contained in the data for that script, or because it never appears at this point of the script in DeScript. All sentential target utterances have a transitive main verb, whose direct object DP is equivalent to the target utterance in the fragment conditions. I added a *please* to all materials from a token set whenever this made the utterances sound more natural.

Since there were only 17 scripts involving verbal communication (*situational scripts* in the terminology of Schank & Abelson (1977)) in DeScript, I adapted seven of the remaining *instrumental* scripts by introducing a second participant. The pasta scenario is an example of such an adapted instrumental script. The types of scripts differ in the method of generating the target utterance. In situational scripts, the target utterance occurs at the point at which the characters in the script perform a speech act, such as e.g. telling the employee the choice of food in the fast food script. In this case, the utterance in the predictable condition is a stereotypical order (9). The context story then consists of the three events that are most likely to precede the target event order meal (10).

- (9) I'll have the cheese burger with fries.
- (10) John wants to eat something in a fast food restaurant at the train station. After entering the restaurant, he approached the counter. Then he carefully read the menu above the counter. The choice for a meal was easy.

In instrumental scripts, the predictable utterance always refers to a target event that is likely in the context of the three events in the context story. I intended to select only those events for which it seemed natural that one of the participants would tell the other what to do in this situation. An example is the pasta scenario in (8) above. In adapted instrumental scripts, both characters are introduced in the first sentence. In the situational scripts, the second character, e.g. the fast food restaurant employee, is not mentioned until the point where (s)he appears in the script. In order to factor out potential effects of script type, a corresponding predictor was included in the statistical analysis.

The main prediction of UID with respect to the materials is that fragments are relatively more strongly preferred in the predictable condition, as a significant interaction between SENTENTIALITY and PREDICTABILITY would indicate. Since UID presupposes audience design, it predicts that production preferences are in line with the perceived well-formedness of utterances. If fragments are preferred in predictive contexts, they should also be perceived as more acceptable than the corresponding sentence in this situation, the opposite being the case for unpredictable utterances. However, this inversion is not necessarily expected in the experiments, because independent factors might result in an overall preference for sentences or fragments. For instance, fragments might be perceived as impolite or there might be a pressure to be brief in some situations.

### 5.2.3 Procedure

The experiment was conducted over the Internet using the LimeSurvey survey presentation software and completed by 48 self-reported native speakers of German recruited on the *clickworker* crowdsourcing platform. Each participant was rewarded with € 4.00 for participation. Subjects were asked to rate the naturalness of the target sentence, which was presented in italics, in the context of the context story on a 7-point Likert scale (7 = fully acceptable). They were assigned to one of four lists, to which materials were assigned by a 2×2 Latin square, so each subject saw each of the 24 token sets once and 6 items in each of the four conditions. Materials were mixed with 21 items from experiment 6 and 44 unrelated fillers. Both the fillers and the materials for experiment 6 resembled the items in containing a context story and an italicized target utterance which subjects rated. In the materials of experiment 6 and in 18 out of the 44 fillers, the target utterance was a fragment, in the remaining 26 fillers it was a sentence. This ensured that sententiality was almost balanced throughout the experiment. Materials were presented in individual pseudo-randomized order that ensured that no two items or fillers of the same category immediately followed each other. Three subjects who rated more than two out of five ungrammatical controls with 6 or 7 points on the scale were excluded from further analysis.

The main experiment was followed by a questionnaire that measured the participants' familiarity with the scripts on which the materials were based, in order to account for potential individual differences between subjects. Since PREDICTABILITY is manipulated through script knowledge, the predictable conditions should be predictable only to subjects who possess the relevant script knowledge, for which I consequently expect a larger effect of script knowledge. Some of the

scripts in DeScript describe situations about which probably every German subject will have knowledge, such as grocery shopping or cooking pasta, but this may not be the case for e.g. fixing a bicycle tire, going to the sauna or borrowing a book at the library. In the script knowledge questionnaire, subjects were asked to check on a 5-point scale how familiar they were with the script scenarios (5 = very familiar).<sup>16</sup> In the instructions for this questionnaire, familiarity was defined as “knowing how these situations typically develop” and not restricted to the participants’ own experiences, but also to knowledge reported by others or gained through the media. The scenarios were described by nonsentential phrases, such as “train ride” or “baking a cake”, which were equivalent to the script titles. The z-transformed script knowledge scores were used as a predictor in the statistical analysis. If the acceptability of fragments is conditional on script knowledge, the expected contrast between predictable and unpredictable utterances, particularly fragments, should increase the more familiar subjects are with the scenario.

#### 5.2.4 Results

Figure 5.4 summarizes the aggregated rating data by condition. The data were analyzed with CLMMs following the procedure described in Section 3.1.1.5. The full model contained main effects of SENTENTIALITY, PREDICTABILITY, SCRIPTTYPE (situational/instrumental), the POSITION of the item in the experiment, and the z-transformed SCRIPTKNOWLEDGE score from the questionnaire that followed the main experiment. I also included all two-way interactions and the three-way interaction between SENTENTIALITY, PREDICTABILITY and SCRIPTKNOWLEDGE. The three-way interaction could show whether the SENTENTIALITY:PREDICTABILITY interaction predicted by information theory is stronger the more familiar subjects are with the scenario. The model contained by-item random intercepts and slopes for SENTENTIALITY, PREDICTABILITY and SCRIPTKNOWLEDGE and by-subject random intercepts and slopes for these predictors and all interactions between them, including the three-way interaction. Backward model selection maintaining only those effects significantly improving model fit, as evidenced by likelihood ratio tests, yielded the final model summarized in Table 5.1.

The final model contains significant main effects for both experimentally manipulated IVs. The main effect of SENTENTIALITY ( $\chi^2 = 30.05, p < 0.001$ ) reveals a general preference for sentences over fragments, and the main effect of PREDICTABILITY ( $\chi^2 = 10.49, p < 0.01$ ) shows that predictable utterances are

---

<sup>16</sup>Due to a technical problem, only the script knowledge scores for 22 out of the 24 scenarios were recorded. Since regression modeling is robust to missing data, it allows for the inclusion of SCRIPTKNOWLEDGE as a predictor in the analysis despite this issue.

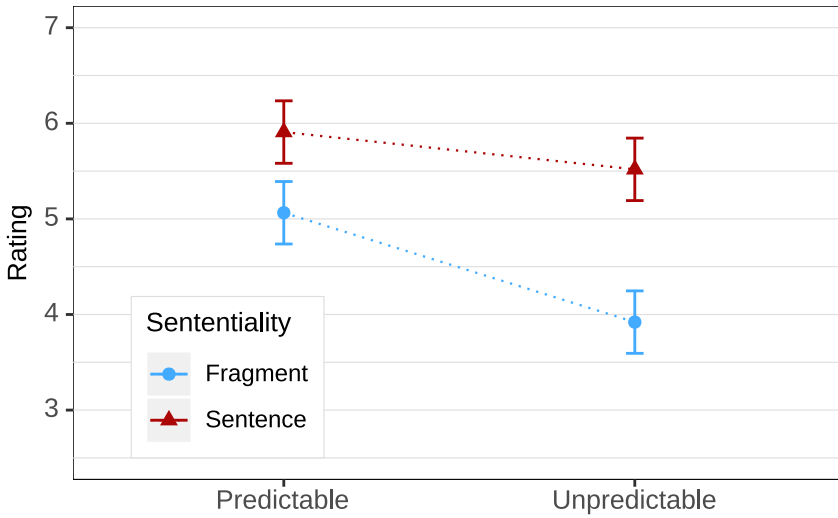


Figure 5.4: Mean ratings + 95% CIs for experiment 11.

also preferred overall. The significant interaction ( $\chi^2 = 9.61, p < 0.01$ ) between both predictors confirms the prediction of UID that the relative preference for sentences is weaker in the predictable condition: Fragments are more acceptable when they refer to a predictable event. SCRIPTKNOWLEDGE does not seem to constrain this interaction given the model, as the three-way interaction was not significant ( $\chi^2 = 2.36, p > 0.1$ ). There is also no significant main effect of SCRIPTKNOWLEDGE ( $\chi^2 = 0.03, p > 0.8$ ), but a significant interaction with PREDICTABILITY ( $\chi^2 = 5.08, p < 0.05$ ): The more people know about a scenario, the more they distinguish between the predictable and unpredictable conditions. The absence of any significant effect of SCRIPTTYPE or interaction with other predictors suggests that the situational and adapted instrumental scripts trigger equally strong predictability effects. Finally, there is a theoretically uninteresting POSITION main effect that however does not interact with any of the other predictors and shows that ratings improve in the course of the experiment.

### 5.2.5 Discussion

Experiment 11 shows that fragments are more acceptable when they refer to a message that is predictable in context. This shows that the manipulation of utterance predictability through script knowledge works and that information-theoretic considerations determine the preferred form of utterances.

Table 5.1: Fixed effects in the final CLMM.

| Predictor                          | Estimate | SE    | $\chi^2$ | $p$    |     |
|------------------------------------|----------|-------|----------|--------|-----|
| SENTENTIALITY                      | -0.958   | 0.143 | 30.5     | <0.001 | *** |
| PREDICTABILITY                     | 0.554    | 0.162 | 10.49    | <0.01  | **  |
| SCRIPTKNOWLEDGE                    | 0.012    | 0.109 | 0.03     | 0.86   |     |
| POSITION                           | 0.021    | 0.002 | 21.57    | <0.001 | *** |
| SENTENTIALITY:<br>PREDICTABILITY   | 0.22     | 0.073 | 9.61     | <0.01  | **  |
| PREDICTABILITY:<br>SCRIPTKNOWLEDGE | 0.206    | 0.094 | 5.08     | <0.05  | *   |

#### 5.2.5.1 Script-based event chains as an approximation to context

First of all, the experiment confirms the suitability of the script-based manipulation of predictability, which is empirically founded on event probabilities estimated from the DeScript corpus of script knowledge (Wanzare et al. 2016). This is evidenced by the main effect of PREDICTABILITY, which confirms that utterances that refer to predictable events are perceived as more natural across the board. The interaction between PREDICTABILITY and the SCRIPTKNOWLEDGE scores collected with the questionnaire following the main experiment shows that this effect is particularly strong when subjects are familiar with a scenario. This confirms the utility of the materials as triggers for script-based expectations.

#### 5.2.5.2 Evidence for information-theoretic constraints

From the information-theoretic perspective, the most important result of the experiment is the significant SENTENTIALITY:PREDICTABILITY interaction. As information theory predicts, fragments are more acceptable when they encode a predictable message than when they encode an unpredictable one. Experiment 11 thus provides first empirical evidence that the perceived acceptability of a fragment depends on the likelihood of the message they encode.

It might seem surprising that sentences were rated as more acceptable than fragments across all conditions. Fragments are frequently used, and if information-theoretic constraints determine the choice of an utterance there must be situations where a fragment is the most well-formed encoding for a message. For most of my materials however, the sentence was also preferred in the pre-

dictable condition.<sup>17</sup> Even though this is unexpected, UID does not make predictions about a main effect of SENTENTIALITY that could be investigated with experiment 11. UID does also not predict whether the interaction is strong enough to invert a potential preference for fragments or sentences in one of the PREDICTABILITY conditions, and information-theoretic constraints are certainly not the only factor that determines the choice of an encoding. Depending on which other factors are at work in the context of the relatively diverse materials, these can potentially override effects of information-theoretic constraints. Furthermore, as I anticipated above, the script knowledge-based predictability manipulation that I adopted implies that likely events are more likely to be talked about. Even if this assumption is correct (the production data collected in experiment 12 confirm it), quantitative predictions about *how much* more or less well-formed an encoding is can only be made if it is known how likely the corresponding message is. From this perspective, all that the experiment can show, and that it does show, is that predictability has an effect on the acceptability of fragments. Even if the likelihood of a message in the context of my materials is still unknown, the probability that a message is likely enough to license ellipsis will be higher the more likely that event is.

#### 5.2.5.3 Why are sentences more acceptable?

The observation that sentences are preferred over fragments both in the predictable and the unpredictable conditions is unexpected, even though as I discussed above, this does not challenge the interpretation of the data as evidence for my UID-based account. In what follows I discuss several potential reasons for this preference for sentences: (i) the written presentation modality, (ii) politeness considerations, (iii) pragmatic considerations of the choice for a specific fragment, and (iv), a potential mismatch between the likelihood of an event and that of a corresponding utterance.

First, the written presentation might yield a preference for sentences, because fragments seem to be less appropriate in written speech. I attempted to counteract this by explicitly presenting the target utterance as spoken by one of the characters, but effects of the written presentation cannot be ruled out. Alternatively, materials could be presented auditorily, but this would increase the complexity of the experiment, because prosody must be controlled.

Second, in some situations the use of fragments might be perceived as too informal. I addressed this issue with a questionnaire on the formality of the script-

---

<sup>17</sup>In three of the 24 items, the fragment was rated as more acceptable than the sentence in the predictable condition on average. In the unpredictable condition, this was never the case.

based situation underlying my materials which was presented after experiment 12. The questionnaire was presented in the same way as the script knowledge questionnaire presented after experiment 11. Participants used a 5-point Likert scale (5 = very formal) to rate how formal they perceived the situation described by each script. I then investigated a potential effect of the mean formality rating with linear models, which show that the effect of formality on the relative preference for fragments is far from being significant ( $F(1) = 0.26, p > 0.6$ ).<sup>18</sup> This suggests that the preference for sentences cannot be attributed to formality.

Third, the average acceptability difference between fragments and sentences might be not due to surprisingly high ratings for sentences, but because some of the fragments were not the optimal encoding in the predictive condition. Recall that in order to control the form of fragments I always tested DP fragments that appeared in the postverbal position in the full sentence. This ensured that materials were as similar as possible to each other in length and internal structure, but this type of fragments is not necessarily the most optimal one with respect to UID. As I observed in the introduction to this section, fragments can have different syntactic categories, or consist of a sequence thereof. Therefore, a wide range of fragments can be derived from a single sentence like (11a), some of which are given in (11b–d). In the specific case of the pasta scenario, the DP (11b) might be rather uninformative, because cooking pasta involves various events that concern the pasta. Following UID, in this case, *the pasta* might be omitted due to its predictability rather than the PP in (11d), which would survive ellipsis because it is less predictable. If that was the case, at least some of the fragments tested in the experiment might not (only) be degraded because a sentence is preferred in the situation, but because there are other fragments that are more well-formed and can communicate the same message.

- (11)    a.    Pour the pasta into the pot, please!  
           b.    The pasta!  
           c.    The pasta into the pot!  
           d.    Into the pot!

Finally, the assumption underlying the design of the materials, that a likely event is more likely to be talked about, so far has not been empirically evidenced. As I

---

<sup>18</sup>The IV in the analysis consisted in the mean formality ratings per scenario. The DV was the difference between the mean rating for the sentence and the mean rating for the fragment per scenario. The full model contained only a main effect of FORMALITY and the final model was an intercept-only model. As for previous analyses, models were fit with the `lme4` package (Bates et al. 2015) and model comparisons were conducted with the `anova` function in R.

noted above, information theory might even predict the opposite, because referring to an event that follows necessarily from context is extremely redundant.<sup>19</sup> If the messages in the predictable conditions were not as likely as I assumed them to be, this would also explain the overall preference for sentences. This does not undermine the conclusions that I draw so far from the experiment: If utterances that refer to predictable events were unlikely in the context of my materials, the interaction would be expected to go in the opposite direction. Even if both the fragment and the sentence were highly redundant, UID still predicts that the less redundant fragment will be preferred over the full sentence.<sup>20</sup> This calls for an empirical estimation of the message probability in my materials, which is one of the goals of experiment 12. Anticipating the results that I discuss in greater detail in Section 5.3.6.6, the message underlying my materials in the predictable condition is indeed more likely (19.1%) than that in the unpredictable condition (0.4%). However, there is a large degree of variation between items. For instance, in the train scenario the message used in the predictable condition was produced in 96.6% of the trials, whereas there were three scenarios for which the target utterance tested in the predictable condition was never produced.

### 5.2.5.4 Summary and outlook

Taken together, experiment 11 shows that fragments are more acceptable when they refer to a message that is predictable in context. This supports my hypothesis that predictable utterances are more likely to be reduced. The overall preference for sentences is unexpected but does not contradict the predictions of information theory. However, this finding is expected both under UID and a source coding approach, which predicts a general tendency to assign shorter codes to more likely messages. UID additionally predicts that specifically those words that

---

<sup>19</sup>Such considerations also seem to be reflected in the original DeScript data. Some contributors to the corpus provide very brief descriptions of the cooking pasta scenario, for instance they omit events like pouring salt into the water or even the action of turning the stove on, which is a necessary condition for events that occur later in the script.

<sup>20</sup>Schäfer et al. (2021) present data on verb phrase ellipsis that are in line with this prediction. In a self-paced reading experiment, they find that longer redundant VPs (ia) are read faster than shorter ones (ib) (per word). This indicates lower average surprisal on the longer VPs. Rating data show that realization of the redundant VP is more degraded in the long condition than in the short one.

- (i) a. John played football, and Bill *played football*, too.
- b. John played football in the backyard of the house, and Bill *played football in the backyard of the house*, too.



are more predictable are omitted. This is investigated with a production task in experiment 12. The production data will also show whether the messages tested in the predictable condition are indeed likely more likely than the presumably unpredictable ones.

## 5.3 Experiment 12: Script knowledge, production

### 5.3.1 Background

Experiment 11 showed that fragments are relatively more acceptable when they encode predictable messages. Even though this follows from UID, it is also predicted by a source coding account that simply assigns shorter codes to more likely messages. Experiment 12 investigates the two more fine-grained predictions of UID on the word level: First, predictable words are more likely to be omitted, because this avoids troughs in the ID profile. Second, words that reduce the surprisal of unpredictable following words are more likely to be realized, since this smooths peaks in the ID profile. Source coding predicts an overall densification strategy for predictable utterances, but only UID predicts *which* words are omitted and that channel capacity limits the densification of utterances. This upper bound will be specifically evidenced by the insertion of redundant material in order to reduce peaks in the ID profile, which is not predicted by source coding.

Investigating these predictions empirically requires (i) a data set that contains such omissions, (ii) to know which words have been omitted, and (iii) a method to estimate information in the presence of omissions. Following the approach taken by Levy & Jaeger (2007) and subsequent work, logistic regressions can then be used to test whether information-theoretic measures significantly contribute to predicting whether this word has been omitted or not in the data set. In experiment 12 I collected such a data set with a production task using the same stimuli as in experiment 11, which contains about 100 utterances for each of the script-based stories. This allows me to quantify the likelihood of utterances in the extralinguistic context modeled by the stories. The data were preprocessed so that omissions can be unambiguously reconstructed. I estimate word probabilities for omitted and realized words in this data set with a new method of surprisal estimation which is not confounded by omissions in the data.

Experiment 12 will also provide further insights on the question of whether word order variation is driven by UID. As I discussed in Section 4.2.2, it has been argued in the literature that the average surprisal of words is lower the more material precedes them in the clause. Therefore, placing unpredictable words in a sentence-final position and predictable ones sentence-initially yields a more

uniform ID profile on average. Finally, the experiment will show whether the utterances in the predictable condition of the rating study were indeed more likely than the unpredictable ones. As I observed in the discussion of experiment 11, likely events are not necessarily likely to be talked about, because the corresponding utterances are rather uninformative.

### 5.3.2 Materials

I used the context stories of experiment 11 as stimuli. These stories consist of four sentences, the first one introduces the script and the other three ones describe a sequence of three events that are likely to follow each other in DeScript. The stimulus corresponding to the pasta script is given in (12).<sup>21</sup> In the rating study, the sentence introducing the utterance was always a complete sentence, e.g. *Now Annika says to Jenny*. In the production study this sentence was replaced by a fragment of the form *<Person A> to <Person B>*. Otherwise the force of this introduction could have biased participants to produce e.g. declarative or interrogative utterances instead of the overall most likely utterance.

- (12) Annika and Jenny want to cook pasta. Annika put a pot with water on the stove. Then she turned the stove on. After a few minutes, the water started to boil.  
Annika to Jenny:

I collected data for the 24 scripts on which the materials in experiment 11 were based. Some of the materials were slightly adapted, so that there were only two characters in each script. In the rating study, some stories had three participants, for instance, the traveling couple and an airline employee.

### 5.3.3 Procedure

The experiment was completed by 198 self-reported native speakers of German, who were recruited for the experiment on the *clickworker* crowdsourcing platform. Subjects were paid €2 for participating in the study. The task consisted in reading the context story and entering the sentence the subject considered most likely into a text field. Initially, I planned to collect two data sets using the same materials, one that contains omissions and one that contains only complete sentences. The nonelliptical data set could then be used for surprisal estimation

---

<sup>21</sup>I provide English translations of the materials in this section, but the experiment was done in German.

(without surprisal being affected by omissions) whereas the data set containing omissions would have shown whether word probabilities in the nonelliptical data set predict the actual distribution of omissions. However, even though participants were instructed to produce *complete sentences*, they produced a relatively high ratio of utterances containing omissions. Therefore, I used this data set as the elliptical one and reconstructed ellipses as described in Section 5.3.4 in order to obtain a nonelliptical data set for probability estimation. In order to rule out the possibility that the term *sentence* in the instructions resulted in an artificially low rate of omissions, I collected a smaller data set ( $n = 30$  for each script) replacing the term *sentence* in the instructions by *utterance*. I then compared the ratio of elliptical utterances, which lacked at least one word that appears obligatorily in a full sentence between both data sets. In the data set collected by asking for *utterances*, the rate of elliptical utterances (25.74%) was slightly higher than in the data set collected by asking for *sentences* (23.79%). A linear mixed effects regression on the item level shows that this difference is not significant ( $\chi^2 = 0.8$ ,  $p > 0.37$ ). I conclude that the term *sentence* in the instructions did not bias subjects' behavior as compared to *utterance* and therefore used the larger data set for analysis and reconstructed the nonelliptical data set for analysis.

Subjects were also told that they would read statements like *<Person A> to <Person B>* after the story, and that these statements specified whose utterance they should produce. Subjects were assigned to one of two lists, and each list was worked on by 99 participants. Each list contained half of the 24 script-based materials and eight further stimuli. These materials had the same form as the script-based materials (a four-sentence context story), required the same task and also described potentially script-based situations. However, they were not as based on empirically measured event probabilities. The purpose of including them was to obtain a larger data base for future analyses of omissions in fragments. Materials were presented in fully randomized order. After data collection, subjects were asked to complete a questionnaire on how formal they perceived the situations described in the materials on a 5-point Likert scale (5 = very formal). These data were used as predictor for the post-hoc analysis on formality in the rating study that I reported in the discussion of experiment 5.2. There were no attention checks, but the data provided by some subjects were excluded because they entered random character strings or copied part of the stimulus into the text field.

### 5.3.4 Preprocessing

The raw data set comprised about 100 utterances for each of the 24 items. Subjects produced a diverse set of utterances even for semantically identical or very similar meanings referring to the same event. (13) shows some typical responses for the pasta script. For instance, the responses vary with respect to lexical alternatives such as the synonyms *Pasta* and *Nudeln* in (13a,b), optional adverbials like *jetzt* ‘now’ in (13c) and omissions like that of *in den Topf* ‘into the pot’ in (13d). In addition, these utterances vary in word order, which depends on the insertion of modals as *kannst* ‘can’ in (13a–c) that make the indirect speech act more polite and on sentence force, as the difference between declaratives and interrogatives in (12c,d) shows.

- (13) a. Kannst du die Pasta in den Topf tun?  
           can.2SG you the pasta in the.ACC pot do  
           ‘Can you put the pasta into the pot?’  
       b. Kannst du die Nudeln in den Topf füllen?  
           can.2SG you the noodles in the.ACC pot fill  
           ‘Can you put the pasta into the pot?’  
       c. Wir können jetzt die Nudeln rein tun.  
           we can now the pasta inside do  
           ‘Now we can put the pasta in.’  
       d. Gibst du die Nudeln rein?  
           give.2SG you the noodles inside  
           ‘Can you put the pasta in?’

The diversity in the production data results in two problems that were addressed by preprocessing the data to a standardized form. First, it complicates the resolution of ellipsis, which is necessary in order to investigate whether those words that have been omitted are really more predictable than those which are realized. If there are two synonym lexical alternatives like *Nudeln* and *Pasta*, even if it is obvious that a word referring to *the pasta* is missing, it is unclear which of the alternatives is to be inserted. Second, language models operate on word forms and treat synonyms like *Nudeln* and *Pasta* as distinct expressions. However, if a DP referring to the pasta is omitted because of its high predictability, this should be reflected in the information estimated by the language model instead of splitting its probability mass between various synonyms.<sup>22</sup>

<sup>22</sup>Higher order *n*-grams boost this problem further. For instance, only two synonym verbs *reingeben*, *reintun* (‘to put inside’) and two synonym words for the pasta split the probability mass of *pour pasta* across four bigrams.

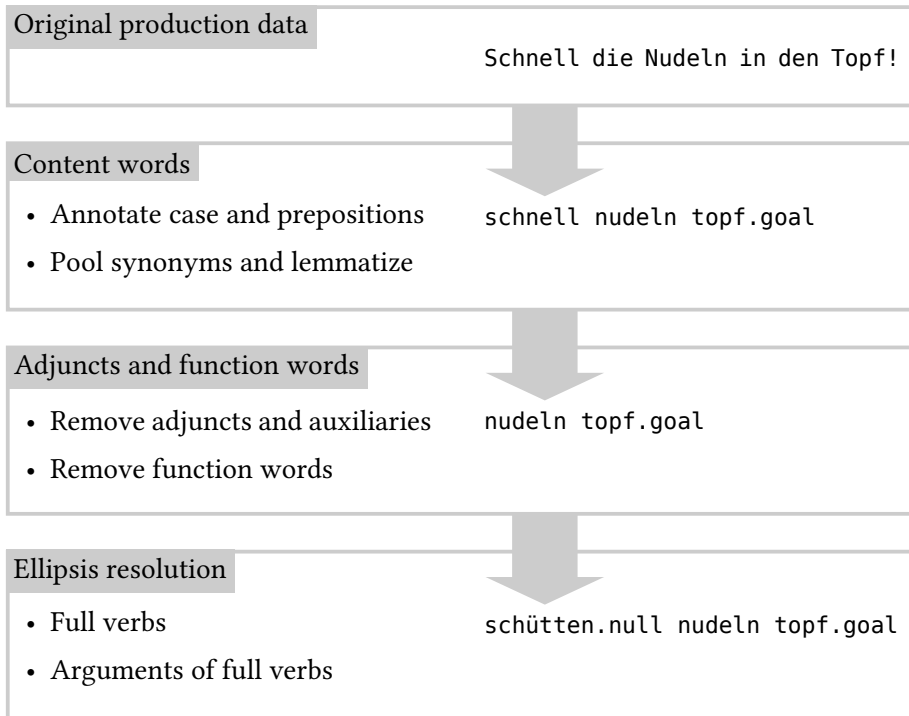


Figure 5.5: Overview of the preprocessing procedure applied to the production data. First, case and prepositions are annotated on the noun phrases, then the adverb *schnell* is removed and finally the missing verb *schütten* ('to pour') reconstructed.

Therefore, all utterances produced by the subjects were manually annotated according to a procedure, which is exemplified in Figure 5.5 and which transforms an utterance like (14a) into an abstract representation (14b). Overall, preprocessing consisted in a trade-off between the necessary standardization and the preservation of as much variation in the original data as possible. Specifically, preprocessing preserved word order and morphosyntactic properties of noun phrases, such as distinctive case inflections, which provide a cue toward the  $\vartheta$ -role of the noun in fragments. The standardization consisted in pooling synonyms and removing adjuncts, adverbials and function words. Finally, I also annotated representations of the meaning of each utterance. This annotation layer is necessary to investigate whether more predictable meanings are assigned shorter codes and to relate the production data to the rating data from experiment 11.

- (14) a. Schnell die Nudeln in den Topf!  
 quickly the pasta in the pot  
 ‘(Pour) the pasta quickly into the pot!’  
 b. schütten.null nudeln topf.goal  
 pour.NULL pasta pot.GOAL

#### 5.3.4.1 Exclusion of noisy data

Before preprocessing, data that were clearly not meaningful utterances in the situation described in the stimulus were removed from the data set. This concerns data from two subjects who provided responses in English, responses consisting in random character strings, copied and pasted parts of the stimulus material or nonsense utterances. For instance, one subject repeatedly entered ‘hello’, even in scripts where it was unnatural that one of the characters would greet the other one. Utterances that corresponded to wrong turns (e.g. by the customer in the library script, where subjects were asked to provide an utterance by the librarian) were also excluded. This resulted in a loss of 8.82% of the responses. Sometimes subjects provided two or more utterances in a single trial, such as the example in (15). In such cases, both utterances were separated and treated as individual utterances by the same subject. Separation occurred when two matrix sentences were separated by punctuation characters.

- (15) Gibst du mir bitte die Nudeln? Das Wasser kocht.  
 give.2SG you me.DAT please the noodles the water boils  
 ‘Would you give me the pasta, please? The water is boiling.’

#### 5.3.4.2 Annotation of content words

All utterances were transformed into representations that contain the matrix verb and its arguments. In the complete data from a script, i.e. item, all coreferent nouns were pooled to the same lemma, for instance *Pasta* and *Nudeln* were merged to *nudeln*. Among the synonyms, the most frequent one was chosen as a label. The same procedure was applied to synonym verbs, such as *schütten* ‘pour’, *füllen* ‘fill’, *reingeben* ‘put inside’ and *reintun* ‘put inside’ in (14), which were merged to *schütten* ‘pour’. Some verbs are ambiguous, for instance, *geben* can be used either as a part of a particle verb *reingeben* ‘put inside’, like in (14d), but it can also describe a transfer of possession ‘give’ as in (15). Therefore, only the instances of *geben* belonging to the particle verb were pooled to *schütten*. Pronouns were also resolved and pooled to the corresponding noun lemma.

For all nouns, distinctive case morphology was annotated as a suffix  $\langle \text{CASE} \rangle$  attached to the noun to ensure that the language model treated nouns with differential case marking (but only those) as different items. This is relevant because experiments 1–3 provided evidence for case connectivity: A DP fragment appears in the same case as the DP in the corresponding full sentence does. From the hearer’s perspective, case can thus be a strong cue toward a specific complete sentence. For instance, a dative DP fragment excludes all possible interpretations that require it to appear in accusative. Case was annotated only when it was morphologically distinguished, because otherwise it does not function as such a cue. Similarly, prepositions were annotated as a suffix  $\langle \text{PREPOSITION} \rangle$  attached to the noun, because the experiments on preposition omission in German showed that there is a strong tendency not to omit prepositions in PPs in German.

#### 5.3.4.3 Removal of function words, adverbials and auxiliaries

After the annotation of case and preposition information on nouns, all function words, like articles and prepositions, modal verbs and auxiliaries as well as optional words like adjuncts were removed from the data set. As for function words, the reason for this is that articles and prepositions cannot be freely omitted in German: Article omission is not grammatical in standard German but only in specific text types (Reich 2011, 2017), and my experiment 4 showed that the omission of prepositions is heavily degraded. Since the purpose of collecting this data set is to investigate omissions on the word level, treating e.g. a preposition and its complement DP as two separate units would falsely suggest that they can be omitted independently from each other. Adverbials can be inserted relatively freely at various positions of the sentence in German (Eisenberg 1999: 209), but often the information that they convey, such as time, place or manner is left implicit. Therefore, it is not reasonable to assume that e.g. (16) involves ellipsis of a temporal and a locative adverbial (Reich 2011: 1850).

(16)  $\langle \text{Yesterday evening} \rangle$  John ate a pizza  $\langle \text{at Giordano’s} \rangle$ .

Of course, the aspects of meaning contributed by adverbials might be also subject to UID: If John eats at Giordano’s frequently, the adverbial *at Giordano’s* will be less informative and therefore more likely to be omitted than if he does only rarely. However, omitting the adverbial does not yield a fragment under any theoretical account, therefore the omission of such optional expressions is not directly relevant to my research questions.

Modal verbs and auxiliaries can affect word order in German. Specifically in questions which are used as indirect speech acts, the modal verb occurs in the

sentence-initial matrix verb position and the main verb utterance-finally (17a). In declaratives, modals occupy the default left bracket position (17b). In order to keep the annotation procedure simple I omitted any occurring modals, but left the main verb in the final position where it appears in the original data. Maintaining the original word order as much as possible is necessary in order to take into account effects of word order on information. Utterance-final words can be predicted from preceding material (Levy 2008)<sup>23</sup> and will therefore be on average less informative than sentence-initial ones.

- (17) a. Kannst du die Nudeln in den Topf füllen?  
           can.2SG you the noodles in the.ACC pot fill  
           ‘Could you fill the pasta into the pot?’  
       b. Du könntest schon mal die Teller auf den Tisch stellen.  
           you can.SBJV.2SG already PRT the plates on the table put  
           ‘You could now put the plates on the table.’

#### 5.3.4.4 Ellipsis resolution

In order to estimate (and to later compare) the surprisal of realized and originally omitted words, all ellipses, i.e. omissions of words that are obligatory in full sentences, were reconstructed. This involved specifically the insertion of omitted main verbs and arguments. Note that the set of ellipses in the data encompasses not only those omissions that yield fragments according to the definition provided in Section 1.4, but also argument omissions. Testing whether omissions of individual words are in line with the predictions of UID requires not only to compare fragments and complete sentences, but to take into account also the possibility of other argument ellipses, like topic or object drop.

Ellipses could in general be unambiguously resolved, because the reconstruction procedure operated on the preprocessed simplified representations described above and there were no lexical alternatives to choose from. If there were several possibilities of enriching a fragment, the strategy that required the least amount of insertions and default subject-initial word order was pursued. All words inserted during this procedure received an additional tag *.NULL*, which indicated that they had been omitted in the utterances produced by the subjects. The purpose of the tag was to keep track of which words had been omitted and which ones had been realized in the original data set. In the statistical analysis, I used this binary *OMISSION* predictor as the dependent variable.

<sup>23</sup>But see Balling & Kizach (2017) for conflicting results on Danish.



Based on the annotated data, I created two versions of the data set: The *original* data set, where I kept track of actual omissions, and the *enriched* data set, which was used only for the purpose of surprisal estimation. This data set contains all of the originally produced words and those that had been inserted through ellipsis reconstruction. I removed the .NULL tags assigned to reconstructed words from the words in this data set, since otherwise the reconstructed and realized occurrences of the same word would be treated as different lexicon entries during surprisal calculation. Using a data set without omissions for surprisal estimation addresses a central problem of language modeling on elliptical data: If UID is correct, word probabilities estimated from a regular corpus are not proportional to the likelihood of these words. Since speakers omit predictable words more often, they are expected to be rare in the corpus, or at least not as frequent as they would be without omissions. Therefore, word probabilities estimated from a regular corpus are not proportional to the likelihood of these words. This circularity issue, which is discussed in detail below in Section 5.3.5.2, does not arise in data sets which do not contain any omissions.

#### 5.3.4.5 Annotation of messages

Finally, the meaning of each utterance was annotated as a simplified semantic representation consisting of a verb and its arguments. These representations were formally identical to those in the enriched data set, but insensitive to word order. The rationale was to assign the same single label to all meaning-equivalent expressions. The purpose of this annotation layer was to be able to quantify the likelihood of meanings and to model the mapping from signals (i.e., utterances) to messages. In this case this mapping was unidirectional, so that various signals could refer to the same message, but not *vice versa*. In principle the opposite would also be possible, but there were no ambiguous utterances in the context of the tightly constrained script-based stories.

#### 5.3.5 Surprisal estimation

I argued above that reasonably estimating the surprisal of words in fragments is possible as long as effects of extralinguistic context can be taken into account and it is known which words have been omitted. The data set whose collection and preprocessing is described in the preceding section conforms to both of these needs. The contexts used for data elicitation ensure that the probabilities of words and utterances in the data set are conditioned on these contexts. The reconstruction of ellipsis solves the circularity issue that might arise from

the frequent omission of predictable words: Surprisal can be estimated from the enriched data, which includes words that have originally been omitted and is consequently not increased by their more frequent omission.

I explore effects of three information-theoretic predictors on omissions. In order to take the script-based approximation to extralinguistic context into account, I apply all methods separately to the data for each script. The first method that I use is simply unigram surprisal. Since the language models are trained on the data for each script separately, unigram surprisal models extralinguistic context. This contrasts with unigram surprisal estimated from larger and more balanced corpora, where the frequency of a word is not conditional on a tightly controlled context. The second method that is an extension of the calculation of surprisal suggested by Hale (2001). In addition to extralinguistic context, this method also takes linguistic context into account, and unlike Hale's original method, it can deal with the possibility of arbitrary omissions in fragments. These two surprisal measures will show whether predictable words are more likely to be omitted. I interpret this as the result of a strategy to avoid troughs in the ID profile. Finally, I use a measure of surprisal reduction, which is technically similar to context-dependent surprisal, in order to investigate to what extent non-final words reduce the surprisal of the following word. This could provide evidence for a strategy to avoid peaks in the ID profile.

### 5.3.5.1 Unigram surprisal

Unigram language models with Good-Turing discount were trained on the enriched data ( $n \sim 100$ ) for each script using the SRILM toolkit (Stolcke 2002). I estimated unigram surprisal for each script independently. This method allows for the interpretation of the probability estimate  $S(w_i) = -\log_2 p(w_i)$  for a word  $w_i$  as the probability of  $w_i$  given extralinguistic context, which is set by the context story, i.e.  $S(w_i) = -\log_2 p(w_i | context_{extraling.})$ . The by-word unigram surprisal values were then extracted from the language model files and annotated with Python scripts for each word in the production data.

### 5.3.5.2 Taking linguistic context into account

In previous studies on predictability effects on linguistic encoding preferences, the probability of words in context has been estimated with  $n$ -gram language models. As I anticipated above, training  $n$ -gram models on corpora that contain elliptical data results in a circularity issue: If predictable words are more often omitted, they will appear to be rare in the corpus just because of their expected high rate of predictability-driven omission.

In the case of unigram surprisal, this issue can be addressed by training the models on the enriched data set, where all ellipses have been reconstructed. Since there are no omissions in the data, all surprisal estimates reflect the true probability of a target word. However, training higher order  $n$ -grams on the enriched data set would result in a further issue, because now words that have been originally omitted would be included in the context of a target word and consequently modulate its probability. This is of course highly implausible, because omitted words are not available to the hearer and therefore cannot contribute to this probability. For instance, if somebody uttered the fragment (18a) in the pasta scenario, the corresponding structure in the enriched data set would be (18b). A regular bigram model trained on the enriched data estimates the surprisal of the DP fragment as  $S_{pasta} = -\log_2 p(pasta | pour)$ . If the verb increases the likelihood of objects that can be poured in this context, the surprisal of pasta would be underestimated in comparison to its true value in a discourse-initial fragment like (18a).<sup>24</sup>

- (18) a. The pasta, please!  
       b. pour pasta pot.GOAL

Therefore, I estimated surprisal with a method based on the approach suggested by Hale (2001). Hale derives surprisal from a parallel parser that computes all possible parses and rejects those parses that are incompatible with the current input. The set of possible parses is correspondingly updated after processing each word. One of Hale's main insights is that the amount of work done by the parser is proportional to the total probability mass of the rejected parses: The higher this probability mass is, the more informative is a word.

This method requires knowing (i) which parses are possible in a situation and (ii) how likely each parse is. Hale (2001) obtains both of these measures from a probabilistic context-free grammar (PCFG), which comprises a set of re-write rules whose likelihood can be estimated from a corpus. The likelihood of a parse is calculated as the product of the probabilities of the individual rules that are required to generate that parse. Instead of using PCFGs, I assume that my enriched

<sup>24</sup>In the literature on grammatical function words, studies investigating effects of a function word's predictability on its own omission (avoid troughs) have used the likelihood of the construction marked by the function word, like a relative or complement clause, as an approximation to the function word's surprisal (Levy & Jaeger 2007, Jaeger 2010). In an analysis investigating whether redundant relative pronouns are inserted before unpredictable target words, Levy & Jaeger (2007) train their models on a modified version of the corpus, from which all relative pronouns have been removed. Therefore, relative pronoun omissions do not affect the estimated surprisal of the target word. Obviously both approaches cannot be applied to fragments, because omissions can target any word in my preprocessed data set.

production data set shows which complete structures are possible in that situation and how likely they are. This approach restricts the set of possible parses to those that actually have been produced and therefore excludes pragmatically odd ones, like (19). In contrast, a PCFG does not exclude odd utterances if they can be derived by its rules. If there is a rule (20a) in the PCFG, the unlikely (19) and the likely (18b) differ only in the likelihood of the two rules in (20), because the other rules required to generate both utterances are identical. Therefore, a PCFG could assign a relatively high probability to utterances that are unexpected for pragmatic reasons as compared to a human hearer, who would not expect a request to pour the salt on the table. It seems more psychologically realistic to assume that when a speaker parses the word *pour* and *salt*, the likelihood of *table.GOAL* as compared to *pot.GOAL* is not only subject to the likelihood of the re-write rules in (20) but to a particular context for which a PCFG does not account.

(19) *pour salt table.GOAL*

(20) a.  $PP \rightarrow \text{table.GOAL}$   
b.  $PP \rightarrow \text{pot.GOAL}$

Therefore, I take the set of unique produced utterances in the enriched data set to represent the range of possible structures and their frequency in the data set to reflect their likelihood in the context of the story. For instance, the German equivalent to (18b) occurs 16 times among the 115 utterances of the pasta data set, so its likelihood is 0.139. Note that this figure does not represent the likelihood of a sentence equivalent to (18b) to be actually produced, because it has been calculated based on the enriched data set and not based on the original production data that contain omissions, but the likelihood of an utterance that, if it is enriched, corresponds to this representation. Knowing the range of possible structures and their respective probabilities is necessary for the estimation of by-word information.

Given this general setup, Hale (2001) defines surprisal as the logarithm of the ratio of the prefix probability  $\alpha$ , i.e. the probability mass of the parses that are compatible with an input *before* parsing that word to the probability mass of the parses that remain active *after* processing it:

$$S(w_n) = \log_2 \left( \frac{\alpha_{n-1}}{\alpha_n} \right) \quad (5.1)$$

This measure increases the more  $w_i$  narrows the set of parses that are compatible with the input as compared to  $\alpha_{n-1}$ . If  $w_i$  is compatible with all parses that  $w_{i-1}$  is compatible with, it equals 0. In the case of utterance-initial words, sur-

praisal is thus equivalent to the negative logarithm of the cumulated probability of all parses (that is, enriched complete structures) that begin with this word. For illustration, consider the case of an utterance like (21a) in the artificial example in (21), where there are only the four parses in (21a–d).

- |      |    |                     |            |
|------|----|---------------------|------------|
| (21) | a. | pour pasta pot.GOAL | $p = 0.75$ |
|      | b. | pour salt pot.GOAL  | $p = 0.2$  |
|      | c. | set table           | $p = 0.03$ |
|      | d. | pour sauce pan.GOAL | $p = 0.02$ |

In this situation, processing *pour* excludes only the parse in (21c), so that  $\alpha_n = 0.97$ . The prefix probability before parsing *pour*,  $\alpha_{n-1} = 1$ , because before processing any input no parse is excluded. The surprisal of *pour* is thus calculated as shown in Equation 5.2.

$$S(\textit{pour}) = \log_2 \left( \frac{1}{0.97} \right) = 0.04 \quad (5.2)$$

The surprisal of *pasta* after processing *pour* is calculated as shown in Equation 5.3. Since (21c) has been previously discarded by processing *pour*,  $\alpha_{n-1} = 0.97$ . Processing *pasta* additionally excludes (21b) and (21d), so that  $\alpha_{n-1} = 0.75$ .

$$S(\textit{pasta} \mid \textit{pour}) = \log_2 \left( \frac{0.97}{0.75} \right) = 0.37 \quad (5.3)$$

### 5.3.5.3 Accounting for omissions in fragments

This method cannot be applied as-is to fragments. For instance, given the discussion on the syntax of fragments in Chapter 3, it is possible to encode (21a) with the fragment in (22). However, following the method proposed by Hale (2001), as soon as the parser processes *pasta* at all, it rejects all of the parses in (21), because none of them starts with *pasta*.

- (22) The pasta!

Therefore, I propose to extend Hale’s method by allowing for an arbitrary number of omissions to occur before and after each word. Checking whether a parse is compatible with an input now does not require for the input to exactly match the parse, starting from its onset, but that the input could have been derived by ellipsis from the parse. For instance, the fragment in (22) can only be derived

from (21a) by omitting *pour* (and *pot .GOAL*, but this does not matter for estimating the surprisal of *pasta*). Processing *pasta* thus excludes the parses in (21b–d), which do not contain this expression. The surprisal of *pasta* can then be calculated from the total probability mass of these parses by comparing the prefix probabilities as shown in Equation 5.4:

$$S(pasta) = \log_2 \left( \frac{1}{0.75} \right) = 0.42 \quad (5.4)$$

This simple modification of Hale’s approach allows in the same way to estimate the surprisal of omitted and realized words in more complex discontinuous fragments. For instance, consider the case of a fragment like (23), for which by-word surprisal is to be estimated given the probability distribution in (21). The surprisal of *pour* is calculated as described in the previous paragraph based on the exclusion of (21c) only. The surprisal of the omitted *pasta* in the context of *pour* is calculated just like described above for that of the realized word *pasta*. Note that for the purpose of surprisal estimation it does not matter whether the target word itself has been omitted. This is desirable, since a word’s surprisal must be independent from its own omission.

(23) *pour pasta .NULL pot .GOAL*

The surprisal of *pot .GOAL* given *pour* is then equivalent to the probability mass of the parses that have not been excluded after processing *pour* but are so after processing *pot .GOAL*. Since only (23d) is excluded by *pot .GOAL*, this yields a surprisal of 0.03 bits. Unlike it would in an *n*-gram model trained on the enriched data set, the omitted word *pasta* does not contribute to the surprisal of *pot .GOAL* in the example:

$$S(pot.GOAL \mid pour) = \log_2 \left( \frac{0.97}{0.95} \right) = 0.03 \quad (5.5)$$

This method avoids the circularity issues related to training language models on elliptical data discussed above. On the one hand, since the complete structures from which the probabilities are derived do not contain omissions, a target word’s surprisal is not affected by the actual frequency of its omission. On the other hand, only words that have been actually produced are included in the context of the target word and can affect its probability. This avoids the concern that *n*-gram models trained on the enriched data falsely include these words in the context. Like the approach by Hale (2001), the approach is fully incremental, because processing effort, i.e. surprisal, arises exactly at the word that triggers the rejection of a parse. By allowing for omissions between words, my approach is

similar to skip-gram models (Jans et al. 2012). However, unlike a skip-gram model, it does not skip a fixed number of words but take into account the possibility that words *could* have been omitted. A further benefit is that this approach considers as much context as possible, unlike  $n$ -gram models. The method presented here covers the complete context, that is, the contextual information about the point in the script at which the utterance occurs and all material that precedes this word within the utterance.

Although my method relies on the same reasoning as the approach in Hale (2001), the surprisal estimates that its output is mathematically not fully equivalent to Shannon information, because the probabilities of all words that could follow  $w_{i-1}$  do not sum to 1.<sup>25</sup> The reason for this is that in Hale’s approach each parse contributes only to the probability mass of a single word within the set of words  $w_i$  that might follow  $w_{i-1}$ . In order to account for the possibility that a word  $w_i$  is omitted in the actually produced string, I take also words occurring later in the parse ( $w_{i+n}, n \geq 0$ ) into account in order to calculate  $\alpha_i$ . Therefore, a single parse can contribute to the probability mass of two or more words, so the sum of the probability mass of all words in context of  $w_{i+1}$  becomes larger than 1. A possibility of dealing with this is to scale  $\alpha_i$  by dividing it by the sum of the prefix probabilities of all words  $W_{Alt}$  that could follow  $w_{i-1}$ , as shown in Equation 5.6. Scaling ensures that all word probabilities sum to 1.

$$S_{scaled} = \log_2 \left( \frac{\alpha_{i-1}}{\alpha_i \times \sum_{x \in W_{Alt}} \alpha_x} \right) \quad (5.6)$$

This equation returns a surprisal estimate based on the likelihood of encountering a word at some point after  $w_{i-1}$  in the utterance. In that sense, it is similar to unigram surprisal, the main difference being that previous linguistic context restricts the set based on which surprisal is calculated. The problem with this approach however is that it does not model the work done by the parser appropriately: In a situation where a target word is included in all parses, my approach sketched above assigns a surprisal of 0 to this word, because processing it does not result in the exclusion of any of the parses. From a theoretical perspective, this is desirable, because it models proportionality between the work done by the parser and surprisal that underlies the approach in Hale (2001). This property however is lost by scaling the ratio between the prefix probabilities before log-transformation, because if numerator and denominator in Equation 5.6 differ, the resulting surprisal estimate will never be 0 (except for the rare case that  $w_i$  is the only element in  $W_{Alt}$ ). For this reason, I do not scale the prefix probabilities and follow the approach described above instead.

<sup>25</sup>I thank an anonymous CUNY 2020 reviewer for pointing this out.

I implemented this surprisal estimation method by first calculating the probability of each of the enriched representations given the complete data set for each script. I then used these data to calculate the surprisal of each word  $w_i$ , be it omitted or realized, in the enriched data set by calculating the ratio of the total probability mass of parses compatible with  $w_i$  and the preceding  $w_{i-1}$ . Whether a parse is compatible with an input was tested using regular expressions in R. These regular expressions contained the relevant word(s) and allow for an optional arbitrary number of characters to occur before, after and between each of them, as is indicated by the gaps (...) in (24).

- (24)    `pour pot .GOAL`  
       a.    (...) `pour (...)`  
       b.    (...) `pour (...) pot .GOAL (...)`

The total probability mass of all parses that are compatible with each of these expressions yields the prefix probability  $\alpha$  of the last word within this expression. In case of (24a), it is calculated by summing up the probabilities of all parses that contain `pour`, and (24b) selects all parses that contain `pour` and `pot .GOAL` somewhere later in the parse. These prefix probabilities are then used to calculate by-word surprisal, as sketched above.

Taken together, this approach is fully incremental, applies to fragments and complete sentences alike, allows for arbitrary gaps before, between and after words in fragments, covers the complete linguistic context and still assigns a psychologically realistic surprisal estimate to each word within that fragment. However, it requires a set of utterances for which context is tightly controlled and to unambiguously resolve ellipses, therefore it cannot be straightforwardly applied to larger corpora that have not been preprocessed in the same way.

#### 5.3.5.4 Estimating the reduction of information peaks

The approximations to surprisal discussed so far are useful to predict the omission of a target word itself, that is, the tendency to omit predictable material. However, UID predicts not only that the surprisal of a word  $w_i$  determines the omission of this word. The surprisal of the next word,  $w_{i+1}$  also constrains the omission of  $w_i$ , because inserting  $w_i$  before  $w_{i+1}$  can reduce a peak in the ID profile caused by  $w_{i+1}$ , but only if this insertion makes  $w_{i+1}$  more predictable. Examples for this line of reasoning are the insertion of relative pronouns before unexpected RC onsets (Levy & Jaeger 2007) or that of articles before unpredictable nouns (Lemke et al. 2017). In previous work, this was investigated by including the surprisal of  $w_{i+1}$  as a predictor in the analysis. In order to avoid the circularity issue



discussed above, this surprisal was not estimated on the original data, but on a version of the data from which all instances of the function word in question, e.g. all articles or relative pronouns, had been previously removed. However, in the case of my data set, all words in the preprocessed data can be omitted, so that removing all potential targets of omission from the data is not an option. Furthermore, UID predicts specifically that redundancy is inserted in order to reduce peaks in the ID profile and not that any redundant expression is inserted before unpredictable words. If the insertion of a redundant  $w_i$  does not reduce  $S(w_{i+1})$ , it might actually yield a trough in the profile in addition to the following peak and thus yield an even less uniform and efficient use of the channel.

Therefore, instead of using the surprisal of  $w_{i+1}$  as a predictor for the insertion of  $w_i$ , I quantify the effect of the insertion on processing by comparing the prefix probability after processing  $w_{i+1}$  when  $w_i$  has been inserted to the prefix probability at  $w_{i+1}$  in case  $w_i$  is omitted:

$$Reduction(w_i, w_{i+1}) = \log_2 \left( \frac{\alpha_{w_{i+1}}}{\alpha_{w_i, w_{i+1}}} \right) \quad (5.7)$$

If the insertion of  $w_i$  does not change the probability distribution after processing  $w_{i+1}$ , this will result in a reduction of 0. Surprisal reduction is larger the more parses that are compatible with previous context and  $w_{i+1}$  are excluded by the insertion of  $w_i$ . To illustrate how this formula is applied, consider again the omission of *pasta* in the utterance *pour pasta.NULL pot.GOAL* in (23). If *pasta* is inserted, only the parse in (21a) is compatible with the input, whereas both (21a) and (21b) are when *pasta* is inserted. Equation 5.7 can then be applied to estimate how much inserting *pasta* reduces the surprisal of *pot.GOAL*.

$$Reduction(pasta, pot.GOAL) = \log_2 \left( \frac{0.97}{0.75} \right) = 0.34 \quad (5.8)$$

In contrast to just using the surprisal of  $w_{i+1}$  as a predictor of the omission of  $w_i$ , this method quantifies *how much* the insertion of  $w_i$  reduces the surprisal of  $w_{i+1}$ , instead of just stipulating that it always does. Therefore, this measure will only predict insertion when this results in a reduction of information on  $w_{i+1}$ , and not just because the  $w_{i+1}$  is overall unpredictable.

### 5.3.6 Results

Experiment 12 had the goal to test the prediction of UID that omissions are distributed so as to reduce peaks and troughs in the ID profile at the case of a data set collected with a production task. Given the three surprisal measures discussed in

the preceding section, this implies that a higher UNIGRAMSURPRISAL and a higher CONTEXTSURPRISAL increase the likelihood of overtly realizing a word because this avoids troughs in the ID profile. Inserting words that reduce the surprisal of the following word, that is, words with a high SURPRISALREDUCTION, reduces peaks. Furthermore, the data will allow for the investigation of word order effects predicted by UID, that is, whether predictable words tend toward being placed at the beginning and unpredictable ones at the end of the utterance. The experiment will also show whether predictable events in experiment 11 are indeed more likely to be talked about than unpredictable ones.

### 5.3.6.1 Data set

The final data set contained about 100 utterances for each of the 24 scripts. This results in 2,409 utterances and 6,816 primitive expressions that I term “words” in what follows. 1,052 words (15.43% of the total) had been omitted in the original data and were inserted during ellipsis resolution and 561 (23.29%) of the utterances contained at least one omission. Recall that this comprises both fragments as defined in section 1.4 and argument omissions.

### 5.3.6.2 Distribution of omissions across scripts

Figure 5.6 shows that scripts differ with respect to the ratio of omitted grammatically required words. In the train script, 62.3% of words were omitted, whereas there are no omissions in the cooking scrambled eggs script. This variation is not trivial to the investigation of UID effects on omissions. On the one hand, a low ratio of omissions might reflect the inappropriateness of fragments in a scripts. If the usage of fragments were blocked for independent reasons, it would be reasonable to exclude the script from further analysis, because including them could mask effects of a word’s information on its omission. On the other hand, this variation could reflect UID effects on omission that should be taken into account by the statistical analysis. Two properties of the scripts that are highly relevant in this respect are lexicon size, i.e. the number of distinct words (the primitive units resulting from annotation) in the data for each script, and the entropy in the probability distribution over these words. The lexicon size varies to a large extent between scripts, for instance, in the train script there are only 12 different words, whereas there are 64 in the driving lesson script. Everything else being equal, a larger lexicon necessarily reduces the likelihood of an average word, therefore UID predicts a lower ratio of omissions in this case.

Therefore, I first investigated whether the lexicon size has an effect on the ratio of omissions in these data. In order to account for different distributions

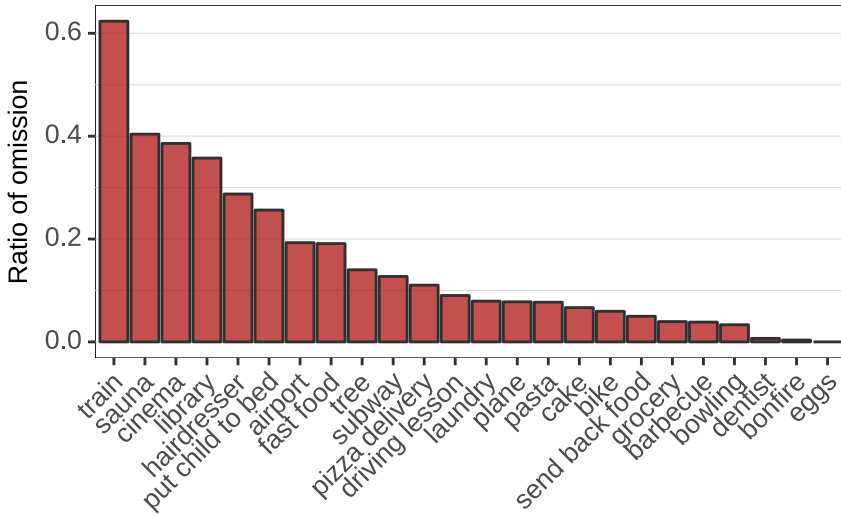


Figure 5.6: Ratio of omission across scripts.

over words I also investigated a potential effect of the *entropy* in this probability distributions. The entropy  $H$  of a random variable quantifies the degree of uncertainty about the outcome of this variable. Following Shannon (1948: 393), entropy is defined as in Equation 5.9. This measure is maximal if each outcome of the variable is equally likely and it equals 0 if there is only one possible outcome.

$$H = -K \sum_{i=1}^n p_i \log_2(p_i) \quad (5.9)$$

Entropy might be a better predictor than the lexicon size, because entropy also reflects the shape of the probability distribution over words. If this distribution is highly skewed for a script, because there are very few very probable words, a random word in this script will be on average more predictable than if the distribution is relatively flat. Since I expect that information predicts omissions, the entropy in the distribution over possible words should also predict to the ratio of omissions, possibly even better than lexicon size. Figure 5.7 illustrates the effects of lexicon size and entropy on the ratio of omissions in a script. The plots suggest that both a larger lexicon and a higher entropy result in a lower ratio of fragments. As entropy increases with lexicon size, both are highly correlated (*Pearson's*  $r^2 = 0.8, t = 6.3, p < 0.001$ ). Linear regressions predicting fragment ratio from these factors individually show that both lexicon size ( $F(1) = 6.18, p < 0.05$ ) and entropy ( $F(1) = 12.49, p < 0.01$ ) have a significant effect on fragment

ratio. A regression that tests both predictors at once reveals that raw vocabulary size has no significant effect beyond entropy ( $F(1) = 0.02, p > 0.8$ ). This suggests that the different ratios of omissions between scripts are at least in part due to properties of the data that are relevant to my research questions and not only due to independent properties of the script. For this reason, I did not exclude any scripts from the data set for further analysis.

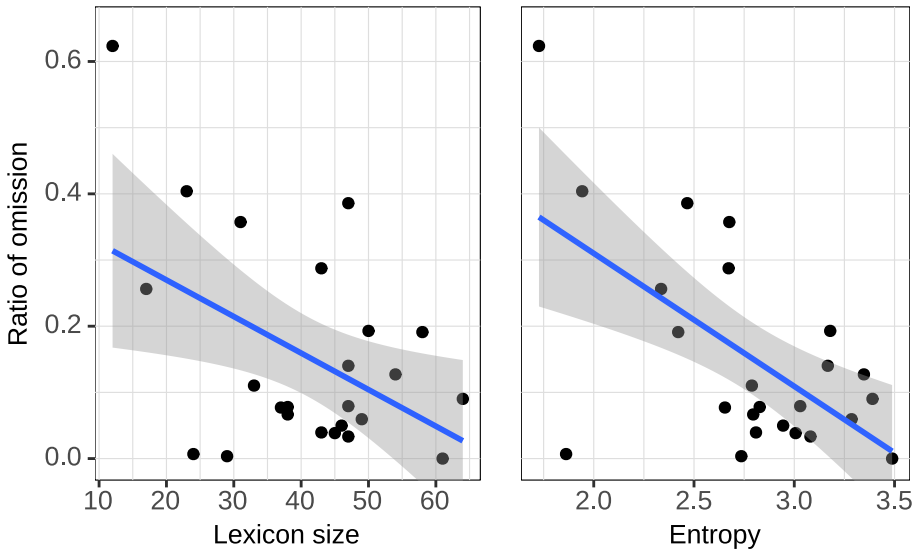


Figure 5.7: Ratio of omission as a function of lexicon size and entropy in the script.

### 5.3.6.3 Variables

The regression analyses described in what follows investigate effects of the three information-theoretic predictors described in the preceding section on the omission of words. `UNIGRAMSURPRISAL` models the likelihood of words given extralinguistic context, `CONTEXTSURPRISAL` additionally takes linguistic context into account, and `SURPRISALREDUCTION` quantifies how much inserting a word reduces the surprisal of the following one. Additionally, I annotated the position of the word in the utterance (numeric). As I discussed in Section 4.2.2, in the literature there is evidence for optimization of word order with respect to UID. On average, context reduces the information of words, therefore placing more informative words toward the end of the utterance yields a more uniform ID profile. In Section 5.3.6.5, I show that this prediction is borne out. Table 5.2 provides an

Table 5.2: Distribution of utterances by length in the production data.

| Length (words) | 1  | 2   | 3    | 4   | 5 |
|----------------|----|-----|------|-----|---|
| Count          | 74 | 680 | 1238 | 414 | 3 |

Table 5.3: Correlations between the information-theoretic predictors.

| Predictors                   | Pearson's $r^2$ | $t$   | $p$    |
|------------------------------|-----------------|-------|--------|
| Unigram, context-dependent   | 0.65            | 70.06 | <0.001 |
| Unigram, reduction           | 0.48            | 37.99 | <0.001 |
| Context-dependent, reduction | 0.62            | 54.00 | <0.001 |

overview of the distribution of utterance lengths. All preprocessed utterances had a length between one and five words. As there were only three utterances with a length of five and I analyzed utterance lengths with an ordinal model these three utterances were excluded from further analyses due to data sparseness.

#### 5.3.6.4 UID effects on omissions in fragments

The data were analyzed with logistic mixed effects regressions predicting the outcome of the binary DV OMISSION of a word in the enriched data set from the information-theoretic measures described above. The analyses were conducted with the `lme4` package (Bates et al. 2015) in R and followed the procedure described in Section 3.1.1.5: Starting with a full model containing all fixed effects and their two-way interactions as well as a maximal random effects structure, predictors that did not significantly improve model fit (as evidenced by likelihood ratio tests) were successively removed from the model. In principle, it would be desirable to include all three predictors in a single model, but as Table 5.3 shows, they are highly correlated and regression analyses require predictors to be independent. Furthermore, effects of SURPRISALREDUCTION cannot be investigated for utterance-final words, which lack a following word whose surprisal they could reduce. Therefore, I conducted three separate regression analyses. The first two analyses investigated whether UNIGRAMSURPRISAL and CONTEXTSURPRISAL predict omissions. This would provide evidence for a tendency to avoid troughs in the ID profile. In the third analysis I tested for an effect of UNIGRAMSURPRISAL and SURPRISALREDUCTION simultaneously, which could

provide evidence for the avoidance of both peaks and troughs. This analysis was conducted on a subset of the data which excluded the utterance-final words, for which SURPRISALREDUCTION cannot be estimated.

#### 5.3.6.4.1 Avoid troughs: Effects of surprisal on omissions

The density plots in Figures 5.8 and 5.9 show the distribution of omissions across the range of UNIGRAMSURPRISAL and CONTEXTSURPRISAL. For both measures, the density plots suggest that on average the surprisal of words which have originally been omitted is lower than that of realized words. The effect seems to be more pronounced for UNIGRAMSURPRISAL.

The full model for unigram surprisal contained only a main effect of UNIGRAMSURPRISAL as well as by-subject and by-item (script) random intercepts and slopes for UNIGRAMSURPRISAL. The by-subject effects account for individual preferences with respect to omission and the by-item effects for potential differences between scripts. The UNIGRAMSURPRISAL main effect was significant ( $\chi^2 = 7.39$ ,  $p < 0.01$ ): The lower the unigram surprisal of a word is, the more likely it is to be omitted (see Table 5.4). The full model for context-dependent surprisal was identical to the one for unigram surprisal except for the missing by-subject random slope for CONTEXTSURPRISAL, because the model did not converge otherwise. The effect of CONTEXTSURPRISAL was significant ( $\chi^2 = 4.86$ ,  $p < 0.05$ ) in the final model (see Table 5.5).

Table 5.4: Fixed effects in the final GLMM investigating the effect of UNIGRAMSURPRISAL on OMISSION.

| Predictor        | Estimate | SE    | $\chi^2$ | $p$   |    |
|------------------|----------|-------|----------|-------|----|
| UNIGRAMSURPRISAL | -0.337   | 0.117 | 7.39     | <0.01 | ** |

Table 5.5: Fixed effects in the final GLMM investigating the effect of CONTEXTSURPRISAL on OMISSION.

| Predictor        | Estimate | SE    | $\chi^2$ | $p$   |   |
|------------------|----------|-------|----------|-------|---|
| CONTEXTSURPRISAL | -0.28    | 0.126 | 4.86     | <0.05 | * |

Taken together, both unigram and context dependent surprisal predict omissions: Words that are more predictable are more likely to be omitted. This is in

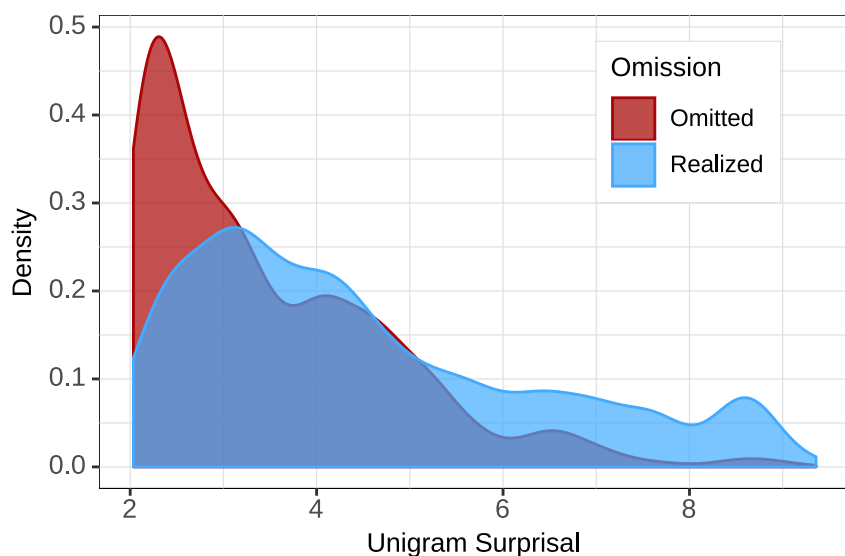


Figure 5.8: The density plot shows how the omitted and realized words are distributed across the unigram surprisal scale.

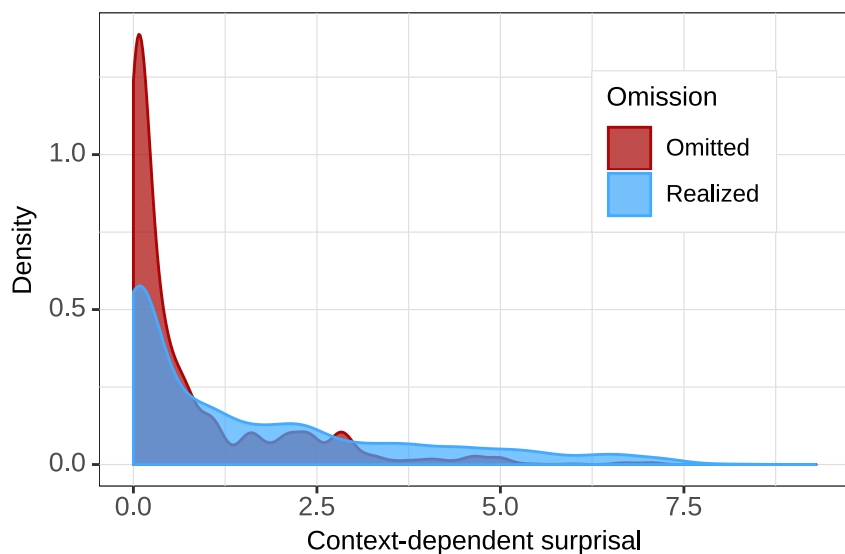


Figure 5.9: The density plot shows how the omitted and realized words are distributed across the context-dependent surprisal scale.

line with the prediction of UID that omitting predictable words in order to avoid troughs in the density profile. A somewhat unexpected finding is that UNIGRAMSURPRISAL seems to be a better predictor of omission than CONTEXTSURPRISAL. In principle the opposite would be expected, because CONTEXTSURPRISAL takes more sources of predictability into account and should therefore be a more precise measure of predictability. In part, the stronger effect of UNIGRAMSURPRISAL is probably an artifact of the data set. As the density plot in Figure 5.9 shows, the overall distribution of CONTEXTSURPRISAL is more heavily skewed due to many words having a CONTEXTSURPRISAL of 0. This results from the relatively small number of complete structures in my data: Sometimes one or two words suffice to completely disambiguate between the structures, so that all following words are fully redundant. An actual speaker however might have a larger set of possible utterances in mind, so that they are not as completely redundant as my model suggests. Therefore, I expect that CONTEXTSURPRISAL would be a better predictor of OMISSION if the same procedure is applied to a larger and more diverse data set.

#### 5.3.6.4.2 Avoid peaks: Effect of surprisal reduction on omission

In order to investigate the prediction of UID that redundancy is inserted before unpredictable words in order to smooth peaks, I conducted a third analysis that additionally considers an effect of SURPRISALREDUCTION. This analysis was conducted on a subset of the data that excluded all utterance-final words, for which SURPRISALREDUCTION cannot be estimated, and all words preceding an ellipsis. The latter were excluded because it would be unreasonable to assume that the preceding word reduced the surprisal of an expression that had been omitted in the actual data. The subset used for this analysis contained a total of 3784 words, that is 55.52% of the total data. The full model contained main effects of SURPRISALREDUCTION and UNIGRAMSURPRISAL, the interaction between both IVs and random intercepts for subjects and items. I chose UNIGRAMSURPRISAL rather than CONTEXTSURPRISAL for two reasons: First, it turned out to be a better predictor of omission than CONTEXTSURPRISAL, and second, as Table 5.3 shows, its correlation with SURPRISALREDUCTION is weaker than for CONTEXTSURPRISAL (*Pearson's*  $r^2 = 0.48$  vs. *Pearson's*  $r^2 = 0.62$ ). Table 5.6 summarizes the final model. The significant main effect of UNIGRAMSURPRISAL ( $\chi^2 = 10.39, p < 0.01$ ) replicates the finding of the previous analyses that predictable words are more likely to be omitted. The significant main effect of SURPRISALREDUCTION ( $\chi^2 = 27.03, p < 0.001$ ) shows that words that reduce the surprisal of the following word more strongly are more likely to be inserted. The



interaction between both predictors is not significant ( $\chi^2 = 0.01$ ,  $p > 0.9$ ). Taken together, this shows that optional omissions in fragments are driven by the tendency to avoid both troughs and peaks in the ID profile, just like UID predicts.

Table 5.6: Fixed effects in the final GLMM investigating effects of both UNIGRAMSURPRISAL and SURPRISALREDUCTION.

| Predictor          | Estimate | SE    | $\chi^2$ | $p$    |     |
|--------------------|----------|-------|----------|--------|-----|
| UNIGRAMSURPRISAL   | -0.151   | 0.046 | 10.39    | <0.01  | **  |
| SURPRISALREDUCTION | -0.349   | 0.07  | 27.03    | <0.001 | *** |

#### 5.3.6.5 UID effects on word order

The production data also might provide some insights into UID effects on word order. UID predicts that expressions that are relatively unpredictable in the absence of linguistic context tend toward being placed at the end of the utterance, because previous linguistic material reduces their surprisal. There are two ways of testing this prediction empirically at my data set. First, following the line of reasoning taken by Genzel & Charniak (2002), if context reduces the surprisal of expressions, words that are more informative in the absence of context should be placed at the end of the utterance because this reduces their information as compared to an initial position. Placing uninformative words at the end of the utterance would reduce their surprisal further and hence increase the risk of a trough in the ID profile. Therefore, if UID is correct, words with a higher unigram surprisal that are hence informative in the absence of linguistic context, should tend toward appearing at the end of the utterance. Second, the effect of information reduction by preceding linguistic context should be reflected in my context-dependent surprisal measure. Therefore, words with a lower context-dependent surprisal are predicted to appear later in the utterance.

These hypotheses were investigated with CLMMs predicting the outcome of an ordinal DV POSITION with four ordered levels from UNIGRAMSURPRISAL and CONTEXTSURPRISAL. The analysis was conducted on the complete data set without distinguishing between originally omitted and realized words, because, if fragments are derived from regular sentences, as my experiments in the first part of this book suggest, word order and omission are in principle independent from each other. Following the procedure applied in previous experiments, I started with a full model that contained main effects for both IVs, their interaction and a

full random effects structure. Table 5.7 summarizes the final model. The position of a word in the utterance is significantly determined by both IVs. Words with a higher UNIGRAMSURPRISAL tend to appear later ( $\chi^2 = 35.47$ ,  $p < 0.001$ ), whereas words with a higher CONTEXTSURPRISAL tend to appear earlier ( $\chi^2 = 41.78$ ,  $p < 0.001$ ). Both observations are in line with UID.

Table 5.7: Fixed effects in the final GLMM predicting POSITION from the surprisal estimates.

| Predictor        | Estimate | SE    | $\chi^2$ | $p$    |     |
|------------------|----------|-------|----------|--------|-----|
| UNIGRAMSURPRISAL | 0.475    | 0.057 | 35.47    | <0.001 | *** |
| CONTEXTSURPRISAL | -0.585   | 0.054 | 41.78    | <0.001 | *** |

### 5.3.6.6 Event likelihood vs. message likelihood

Finally, the production data can also be used to test the assumption underlying the rating study that utterances referring to predictable events are more likely than those referring to unpredictable ones. If this was not the case, the results of the rating study could not be interpreted as evidence for information-theoretic well-formedness perceptions.

I addressed this question by counting how often utterances referring to the messages used in either of the predictability conditions in the rating study were produced in the production study. There was a large degree of variation between scripts. For instance, the predictable message was produced in 96.56% of the trials in the train script, but there were three scripts where it was never produced. Averaging over the production ratios for all scripts shows that the message tested in the predictable condition was still more often produced than that in the unpredictable condition (19.1% vs. 0.4% of responses). Note that the estimate for the predictable condition is rather conservative: In three scripts where the speaker buys or orders something, only those messages that refers to the item (s)he orders in the sentential condition in the rating study were counted. Taken together, this clearly confirms the reasoning underlying the rating study that predictable messages are more likely to be talked about.

### 5.3.7 Discussion

Experiment 12 provides evidence for the hypothesis that omissions in fragments are driven by a tendency to avoid peaks and troughs in the ID profile. The main

effects of UNIGRAMSURPRISAL and CONTEXTSURPRISAL in the statistical analyses show that words that are predictable given extralinguistic and linguistic context are significantly more likely to be omitted. This reflects the tendency to avoid troughs in the ID profile, which are caused by uninformative words. The main effect of SURPRISALREDUCTION indicates that words that reduce the surprisal of the next one are more likely to be realized. This evidences a strategy of reducing peaks in the ID profile by inserting additional redundancy into the utterance.

Taken together, these findings indicate that UID constrains omissions in fragments. The acceptability rating data from experiment 11 are also compatible with a source coding account or a general tendency to omit given or redundant material, but none of these accounts predicts the effects of the *following* word's surprisal that experiment 12 reveals. UID provides a natural explanation for this observation and also accounts for the finding that predictable words themselves are more often omitted.

The observation that predictable words are more likely to be omitted also implies that the choice between a fragment and a full sentence is constrained by UID. Since the experiments show that extralinguistic context determines the predictability of individual words, the likelihood of a trough that is smoothed by the omission of predictable words is higher in predictive contexts. If the omitted word is required in a full sentence, its omission results in a fragment. This extends previous evidence for UID, where such effects were reported only for highly specific omissions of single closed-class function words to the much more diverse and semantically relevant omissions of content words in fragments.

Experiment 12 investigated only omissions of words that cannot be omitted in complete sentence, like verbs and their arguments, whose omission can result in fragments. Aspects of meaning that are conveyed by e.g. temporal or locative adjuncts, which can be implicit in full sentences, might be subject to UID too, however. Just like experiment 12 showed for words that can be omitted in fragments, adjuncts might also tend to be explicit when they convey less predictable information. Investigating omissions in adjuncts is more complicated, because they are probably more difficult to reconstruct when they are omitted, because a potentially infinite number of adjuncts can be inserted into an utterance. Therefore it is unclear which ones should be reconstructed and which ones should not. In the case of arguments, this reconstruction was more straightforward, because the absence of a syntactically required expression indicates that it must be reconstructed. However, in principle there is no reason to assume that the omission of adjuncts would not be driven by UID as well.

Experiment 12 also provides evidence for UID effects on word order. If only unigram surprisal, which is independent from linguistic context, is considered,

unpredictable words tend toward appearing late in the utterance, just like Genzel & Charniak (2002) showed for sentences within a text. This is expected under the assumption that the more linguistic context a word has, the more predictable it becomes (Genzel & Charniak 2002, Levy 2008). This assumption is in turn confirmed by the inverse effect of my context-dependent surprisal measure which shows that the words at the end of the utterance have a lower context-dependent surprisal. Taken together, the utterances in my data set are constrained by UID in two ways: First, UID constrains the omission of individual words, and second, otherwise unpredictable words are placed toward the end of the utterance, because its surprisal is reduced by preceding material.

## 5.4 The usage of fragments: Discussion

### 5.4.1 Evidence for UID effects on omissions in fragments

In this chapter I presented two experiments that investigated the predictions of UID the usage of fragments: (i) that predictable messages are preferably reduced, (ii) that predictable words are more likely to be omitted, and (iii) that words that reduce the surprisal of the next word are more likely to be realized. UID shares the first prediction with source coding, but the other two are specific to UID.

Experiment 11 supports the first of these predictions by showing that the reduction of predictable utterances is perceived as more well-formed than that of unpredictable ones. Somewhat surprisingly, sentences were on average preferred over fragments in both predictability conditions. This might be due to politeness considerations and the fact that some of the tested fragments are relatively improbable, as evidenced by the production data collected in experiment 12. Nevertheless, the experiment provides first evidence for the hypothesis that the preference for using fragments depends on the predictability of utterances. The design did not allow for the investigation of UID effects on the omission of individual words, therefore the rating study does not ultimately show whether the densification of utterances that encode likely messages is caused by the tendency to avoid peaks and troughs in the ID profile, as UID predicts.

Experiment 12 provides evidence for the more fine-grained predictions of UID that speakers avoid peaks and troughs: Words that are themselves predictable are more likely to be omitted and words that reduce the surprisal of following words are more likely to be inserted. The data set based on which these results were obtained was preprocessed to account for grammatical constraints on fragments. Since an optimization with respect to UID has been argued to occur only within the bounds defined by grammar (Jaeger 2010: 25), preprocessing ensured

that each of the primitive expressions in the data set was equivalent to a constituent whose omission does not structurally depend on surrounding material. For instance, as experiment 4 showed that prepositions cannot be freely omitted in German, I merged the noun phrase and the preposition within a PP to a single primitive unit. Experiment 12 provides clear evidence for UID effects on individual omissions in fragments. This conclusion implies that the choice between producing a fragment and producing a full sentence in a specific situation is also constrained by UID: In unpredictable contexts, the probability of troughs in the ID profile, which trigger omissions, is lower than in predictive ones. If no words that are obligatory in full sentences are omitted for this reason, the speaker will prefer to utter a full sentence rather than any of the grammatically possible fragments. This extends previous evidence for UID, where such effects were reported only for highly specific omissions of closed-class function words to the much more diverse omissions of content words in fragments.

Experiment 12 furthermore provides evidence for UID effects on word order: Words with a higher unigram surprisal, which are less predictable in the absence of context, tend to appear late in the utterance. The UID explanation for this observation is that linguistic context reduces the surprisal of words that are otherwise unpredictable. Therefore, as Fenk-Oczlon (1983) argues, placing predictable words before uninformative ones yields a more uniform ID profile. The assumption that linguistic context increases the predictability of words on average is supported by the inverse effect of context-dependent surprisal, where predictable words tend to appear *late* in the utterance. Experiment 12 thus also provides empirical evidence for a (reasonable) assumption that has only been stipulated in previous work (Fenk-Oczlon 1983, 1989, Genzel & Charniak 2002).

#### 5.4.2 UID vs. availability-based production

Following Hale (2001), I assume that the relationship between the likelihood of a word and its omission is determined by processing effort. Speakers perform audience design by adapting their message to the expected cognitive resources of the hearer, and omissions occur whenever they are beneficial to that goal. However, as I sketched in Section 4.2.4.1, predictability effects have also been explained with the effort required to retrieve a word from memory alone. Availability-based production predicts that omissions occur more often if the word following the omitted one is predictable and hence easy to retrieve, as Ferreira & Dell (2000) show for complementizer omission in English. The same effect is predicted by UID, but for a different reason: Realizing words that precede unpredictable words can reduce the surprisal of the latter and hence smooth peaks in the ID profile.

The observation that words are more likely to be inserted before unpredictable words can therefore be also interpreted as an effect of availability-based production. As Jaeger & Buz (2017) note, availability-based production and UID are not mutually exclusive and there might be independent effects of both theories, but in case both theories predict the same effect it cannot be attributed unambiguously to either of the theories.

Effects of availability-based production however are particularly expected in oral communication, because according to Ferreira & Dell (2000) the motivation for inserting optional words is to avoid disfluencies which would result from the time required to retrieve unpredictable lemmas from memory. This is the case in studies on spoken corpora (Levy & Jaeger 2007, Frank & Jaeger 2008, Jaeger 2010) or spoken production experiments (Kurumada & Jaeger 2015, Norcliffe & Jaeger 2016), but it does not concern my production study (experiment 12) in the same way: Even though I asked subjects to provide utterances that seem natural in a situation of oral communication, they provided written responses, so there is no risk of disfluencies. Therefore, even though some of the evidence for UID might be explained by production alone, this explanation is less convincing in case of the effects that I found in my production study.

Furthermore, availability-based production can only explain why the omission of a target word can be predicted from the surprisal of the following word(s), but not why it depends on the target word's own surprisal. Most of the previous studies on UID and my production study found that the predictability of the target word itself predicts its omission as well. Taken together, the predictions of UID and availability-based production partially overlap, and specifically in studies on spoken language some of the effects of UID can be explained by availability-based production as well. However, availability-based production cannot account for all of the data, and specifically in my experiments the written modality makes it implausible that omissions depend on the likelihood of disfluencies. This does not speak against the idea of availability-based production, but it provides distinctive evidence for UID.

### 5.4.3 Information theory or information structure?

The information-theoretic account that I pursue often coincides with a perspective based on information structure, which requires omitted expressions to be e-given (Merchant 2004a), not part of the focus (Reich 2007) or backgrounded (Ott & Struckmeier 2016). Words that are predictable are probably often given in a possibly implicit QuD. For instance, in the taxi example that I used to illustrate my approach above, it is reasonable to assume that a QuD like *Where should I*

*take you?* licenses ellipsis of everything but the focused word corresponding to the *wh*-phrase in the answer. From an information-structural perspective, given material *can* be omitted because it is given, and from an UID perspective it *should* be omitted, since it would probably yield a trough in the ID profile. A potential testing ground to distinguish between information structure and information theory are expressions that are given, but not predictable, or vice versa. Givenness and predictability might coincide most of the time though, so that distinguishing between these concepts requires a set of constructed materials for which the predictability of a given expression can be manipulated.

The main difference between an information-theoretic and an information-structural approach to the usage of fragments, however, is that only information theory can explain why an expression actually is (not) omitted. Information-structural concepts like givenness might license ellipsis, but it is clearly not necessary to omit each given expression. For instance, it is appropriate to answer a question with a full sentence even though most of the words contained in the answer are given. In contrast, UID provides an account of why particular words are preferably omitted and of why they might be realized. Information structure does also not explain why the surprisal of the word that follows a target word has an effect of the target word's omission. This does not neglect the role of information structure on omissions though, and it is very likely that information-theoretic concepts like givenness are reflected in and should be taken into account by more sophisticated measures of surprisal.

### 5.4.4 Script knowledge as models of extralinguistic context

Since fragments often appear discourse-initially, the predictability of words in context is determined to a large extent by extralinguistic context, which cannot be captured by standard language modeling techniques applied to speech corpora. Therefore, the context stories used in experiments 11 and 12 were based on probabilistic event chains extracted from the DeScript corpus of script knowledge (Wanzare et al. 2016), which contains crowdsourced descriptions of the stereotypical time-course of script events.

The rating study shows that utterances referring to predictable events were more acceptable, and that this holds in particular for fragments. I interpreted this as evidence for optimization of the signal with respect to information-theoretic principles because I assumed that predictable events would be more likely to be talked about. The production data collected in experiment 12 confirms that this corpus-based predictability manipulation in experiment 11 is overall in line with

subjects' expectations about upcoming utterances. The utterances in the predictable condition in the rating study had an average likelihood of 19.1%, whereas those in the unpredictable condition were produced only in 0.4% of the trials in the production task. Note that this is a conservative estimate since for instance in the pizza ordering scenarios, utterances were classified as encoding a different message depending on what the customer orders.

This manipulation did not work equally well for all scripts. In three scenarios, the presumably predictable message was never produced and some of the unpredictable fragments that were never produced received relatively high ratings. Consequently, there was no significant correlation between message frequency and the acceptability of fragment conditions. As I discuss in Section 5.5 below, the optimality of fragments might not be driven by message frequency alone, but also by the utility of the fragment to unambiguously communicate a message.

On average though, utterances referring to predictable events turned out to be more likely in the production study. This supports the procedure that I used for constructing materials for experiments 11 and 12 and shows that, even though scripts do not cover *all* aspects of extralinguistic context, script corpora provide precise estimates of the likelihood of utterances in context and therefore constitute a empirically sound approximation to extralinguistic context.

#### 5.4.5 Surprisal estimation in elliptical data

The conclusions drawn from experiment 12 rely on a novel method to estimate surprisal that is robust to a circularity issue caused by ellipses in the training data: If predictable words are omitted more often, their predictability is not proportional to their corpus frequency because they are often omitted. This would distort predictability estimates calculated with regular language models. My approach avoids this problem, because it relies on nonelliptical data for estimating surprisal. The method is similar to the approach proposed by Hale (2001), who derives surprisal from the probability mass of the parses that are disconfirmed by an input. Like Hale's method, it is fully incremental, i.e. the information that a word provides to the parser is used as soon as this word is encountered. The method is also psychologically realistic, because only those words that are available to the hearer are included in the context used for surprisal estimation. Words that are omitted in context of the target word have no effect on its surprisal.

The significant effect of context-dependent surprisal in the analysis shows that is a suitable approximation to the information of words in context. However, the effect of unigram surprisal on omission was stronger than that of context-dependent surprisal even though I expected the opposite because context-depen-



dent surprisal takes more sources of predictability into account. This might be due to the relatively small size of my data set, for which sometimes a sequence of two words completely disambiguates between parses. All following words necessarily receive a surprisal of 0. I expect stronger effects of context-dependent surprisal in larger and more diverse data sets. Linguistic context might also be more important when syntactic information, like inflectional marking on verbs, is more prominent. Other measures of context-dependent surprisal, such as word likelihoods derived from a PCFG (Hale 2001) are sensitive to hierarchical syntactic information, such as subcategorization preferences, for instance that of a preposition for a DP or specific verbs for a complementizer. Such function words had been removed from my data set during preprocessing.

The data set collected in experiment 12 was elicited with a set of carefully built context stories and required a considerable amount of manual preprocessing, which consisted in the unification of synonyms, annotation of case morphology and prepositions, removal of adverbials and the reconstruction of ellipses. Since the analysis confirmed the validity of this approach, it might be interesting to explore to which extent it can be automatized, e.g. by automatic unification of synonyms, morphosyntactic annotation and reconstruction of ellipsis. In such research, the manually preprocessed data can be used as a gold standard for the evaluation of automatized preprocessing procedures.

## 5.5 Outline of a game-theoretic model of fragment usage

This section outlines a possible game-theoretic account of fragments, which can potentially explain some aspects of the choice between a fragment and a full sentence that UID cannot. Empirically testing such an account and comparing its predictions to UID is intricate and must therefore be left to future research.

### 5.5.1 Limits of UID effects on the form of utterances

The regression approach that I took in the analysis of experiment 12 predicts the omission of individual words within a complete sentence from information-theoretic variables, such as the surprisal of the target word or how much this word's insertion reduced that of the word following it. However, as I noted in section 4.2.2, UID faces an empirical problem when a single fragment can be derived from a predictable and an unpredictable sentence. For the purpose of illustration, consider the situation where the fragment in (25a) is used to communicate the full sentences in (25b) or (25c) in the taxi scenario discussed above (the probabilities associated with each complete structure are hypothetical).

- |      |    |                                    |            |
|------|----|------------------------------------|------------|
| (25) | a. | To the university.                 |            |
|      | b. | Take me to the university.         | $p = 0.2$  |
|      | c. | Tell me the way to the university. | $p = 0.05$ |
|      | d. | [other messages, cumulated]        | $p = 0.75$ |

This issue concerns the usage of the fragment in order to communicate (25c). The ID profile Figure 4.3 suggested that omitting *tell me the way* would result in a peak in the ID profile that can be avoided by inserting these words, but I already noted there that this is a simplification. Since a hearer who perceives the fragment in (25a) does not know whether it has been derived from (25b) or (25c), processing the fragment must be equally effortful in both cases. Therefore, the fragment will either exceed channel capacity in both cases or in neither of them.

The problem with encoding the unpredictable (25c) with the fragment in (25a) therefore is not a peak in the ID profile, but that a hearer who encounters the fragment in the scenario in (25) has to guess whether the speaker intended to convey (25b), (25c) or another message. Intuitively, he will go for the more likely (25b),<sup>26</sup> because the speaker is more likely to intend to communicate this message than (25c)<sup>27</sup> and the hearer is consequently more likely to interpret the utterance as intended. UID is by definition unable to take this difference between meanings into account because it models only the encoding procedure, i.e. the choice of the most well-formed utterance to communicate a specific message given the properties of the communication system. Decodingm, the choice of a message given a received signal, is simply not covered by UID.

Before discussing a potential solution to this issue, note that this is a conceptual problem rather than an empirical one. The speaker does neither know with certainty the capacity of the channel nor the hearer's probability distribution over possible complete structures that determines surprisal. Therefore, all she can do is omit words that are *more likely* to yield a trough and to insert those that are *more likely* to smooth a peak in the ID profile. From this perspective, she will omit the predictable *take me* in (25b) and realize the less predictable *tell me the way* in (25c). The unpredictable *to the university* will be preferably realized in both situations.

---

<sup>26</sup>Only those messages that a fragment can potentially encode will be considered, i.e. those from which the fragment can be derived by ellipsis. For instance, (25a) can be derived from (25b) and (25c), whereas *Take me to the airport* is not a possible source for (25a).

<sup>27</sup>Note that this is also in line with the source coding-based prediction that more frequent messages are assigned shorter codes.

### 5.5.2 Game-theoretic pragmatics

A framework that might overcome this problem is game-theoretic pragmatics. Game-theoretic approaches provide a model of context which takes into account the knowledge and preferences of rational agents. The model allows for determining which action is the most useful one to pursue for each of the agents. Such models have been recently gaining popularity in pragmatics and been applied to phenomena like implicature (van Rooij 2004, Benz & van Rooij 2007, Franke 2009, Jäger 2012, Goodman & Stuhlmüller 2013, Gotzner & Benz 2018) and reference (Frank & Goodman 2012, Rohde et al. 2012, Sikos et al. 2019). In what follows I base my expositions on the approach taken by Franke (2009), who develops a game-theoretic account of pragmatic reasoning that models implicatures, the *Iterated Best Response* (IBR) model. He interprets communicative situations as signaling games, which are played by a speaker and a hearer and require a speaker to pick an utterance in order to get a message across.<sup>28</sup> Even though Franke investigates a different phenomenon, the problem of mapping utterances and interpretations remains the same, so in principle it can be straightforwardly applied to the interpretation of fragments.

The main ingredients of a signaling game are a set of possible messages, i.e. meanings that the speaker could intend to communicate, a set of possible interpretation actions, that is, meanings that the hearer could assign the utterance and a set of utterances that can be used for this purpose. Only the speaker knows with certainty which message she wants to convey, and she has to choose the utterance that she believes to be the best one to get a message across. The hearer has to figure out which message the speaker intended to communicate. In this setting, both the speaker and the hearer initiate a chain of recursive reasoning about each other which result in the choice of an utterance and the assignment of a meaning (a message) to this utterance. Which utterance is most optimal from the speaker perspective and which meaning the hearer assigns to it depends on a series of parameters: First, the messages often differ in their *prior probability* before any utterance is produced. In the case of the taxi example, it might be more likely that people would ask for a ride than that they would ask for directions. Second, the production of some utterances can be more effortful and hence costly. Third, a single utterance can sometimes encode more than one message, and a single message can be referred to by more than one utterance. Finally, game-theoretic

---

<sup>28</sup>The terminology used in the game-theoretic literature differs from the one that I use, which I maintain in order to keep the mapping between terms and concepts throughout this book. In the game-theoretic literature, the utterance is called the *message* and what is being communicated (I use the term *message* for this) is labeled the *state*.

models may include varying payoffs that each participant receives depending on her own and the other player's choices. In linguistic applications, where the speaker wants to get a message across and the hearer wants to figure it out correctly, the payoffs for both interlocutors are aligned, since both share the goal of successful communication (Franke 2009: 21). In what follows I present a simplified sketch of how Franke (2009: 59–61) applies this model to scalar implicature before I illustrate how this reasoning can be extended to fragments.

### 5.5.3 Game-theoretic modeling of scalar implicatures

For scalar implicatures, Franke (2009) focuses on the interpretation of the quantifier *some* as *some but not all*. Despite theoretical and empirical debates on how exactly this interpretation is generated (see e.g. Levinson 1983, 2000, Sperber & Wilson 1986, van Kuppevelt 1996, Breheny et al. 2006, Huang & Snedeker 2009, Grodner et al. 2010), the general observation is that utterances like (26a) are often interpreted as (26b), even though semantically (26a) is implied by and hence does not rule out (26c).

- (26)    a.    Some of students completed the assignment.  
           b.    Some but not all of the students completed the assignment.  
           c.    All of the students completed the assignment.

Franke (2009: 59–61) models this situation with a reference game, whose parameters are given in Table 5.8. In the game there are only two possible meanings,  $m_{\exists \rightarrow \forall}$  and  $m_{\forall}$ , two corresponding interpretation actions  $a_{\exists \rightarrow \forall}$  and  $a_{\forall}$  and two possible utterances,  $u_{some}$  and  $u_{all}$ . The two utterances correspond to (26a,c) respectively. Whereas  $u_{some}$  is true in case of both messages,  $u_{all}$  is false if  $m_{\exists \rightarrow \forall}$  is true. Therefore,  $u_{all}$  may be selected only if that the speaker wants to communicate that all students completed the assignment. Each of the messages has a prior probability  $Pr(m)$ . As I noted above, under the assumption that both agents pursue the goal of successful communication, the payoffs for speaker and hearer are matched: Both receive a payoff of 1 if the interpretation action corresponds to the intended message and no payoff if it does not.

Based on this setting, two chains of iterative reasoning of the agents about each other's behavior are initialized. One of these chains starts with a *literal* speaker and the other one with a literal hearer. Unlike higher order *pragmatic* agents, literal agents take only the general setup of the model into account but do not reason about the other agent's behavior. The literal speaker selects randomly an utterance that is true. In order to communicate  $m_{\exists \rightarrow \forall}$ , the only available utterance is  $u_{some}$ , because  $u_{all}$  is false in this situation. In contrast,  $m_{\forall}$  can be communicated

### 5.5 Outline of a game-theoretic model of fragment usage

Table 5.8: Tableau for the scalar implicature game, adapted from Franke (2009: 21). The table provides the probability  $Pr(m)$  for each message, the speaker and hearer payoffs for each combination of messages and interpretation actions and determines which utterance can be used to communicate each message.

|                            | Pr(m)   | $a_{\exists \neg \forall}$ | $a_{\forall}$ | $u_{some}$ | $u_{all}$ |
|----------------------------|---------|----------------------------|---------------|------------|-----------|
| $m_{\exists \neg \forall}$ | $1 - p$ | 1,1                        | 0,0           | ✓          | –         |
| $m_{\forall}$              | $p$     | 0,0                        | 1,1           | ✓          | ✓         |

with both utterances, because both are true for this message. This yields the following *strategies*:

$$S_0 = \left\{ \begin{array}{l} t_{\exists \neg \forall} \mapsto m_{some} \\ t_{\forall} \mapsto m_{some}, m_{all} \end{array} \right\} \quad (5.10)$$

The literal hearer calculates the posterior probability of each message to be intended by the speaker given each of the available utterances. The listener does not reason about the speaker's behavior and considers only the prior probability of each message and the truth conditions. In the first sketch of his approach, Franke (2009) uses flat priors ( $p = 0.5$ ). Table 5.9 summarizes the posterior probabilities for the scalar implicature game. If the hearer encounters  $u_{all}$ , the posterior probability of  $\mu(m_{\exists \neg \forall} | u_{all})$  equals 0, because the utterance would be false in case of the message. Therefore,  $\mu(m_{\forall} | u_{all}) = 1$ , because this is the only message for which  $u_{all}$  is true. If the hearer encounters  $u_{some}$ , both messages could be true, so he will go for the most likely one in order to maximize the likelihood of figuring out the intended meaning. Since both messages are equally likely in case of flat priors, assigning an interpretation to the utterance consists in random guessing.<sup>29</sup> This yields the interpretation strategies in Table 5.9:  $u_{some}$  will be interpreted either as  $m_{\exists \neg \forall}$  or  $m_{\forall}$ , whereas  $u_{all}$  will be always interpreted as  $m_{\forall}$ :

$$L_0 = \left\{ \begin{array}{l} u_{some} \mapsto m_{\exists \neg \forall}, m_{\forall} \\ u_{all} \mapsto m_{\forall} \end{array} \right\} \quad (5.11)$$

The pragmatic speaker takes into account these interpretation strategies of the literal hearer. Since the hearer will unambiguously interpret  $u_{all}$  as  $m_{\forall}$  but might understand  $u_{some}$  as either of the messages,  $u_{all}$  has a higher *utility* than

<sup>29</sup>However, if  $m_{\exists \neg \forall}$  is more likely than  $m_{\forall}$  (if  $p > 0.5$ ), assuming that the more likely message was intended increases the probability of success.

Table 5.9: Posterior probabilities for the literal hearer  $L_0$ .

| $\mu_0(m u)$ | $m_{\exists \rightarrow \forall}$ | $m_{\forall}$ |
|--------------|-----------------------------------|---------------|
| $u_{some}$   | $\frac{1}{2}$                     | $\frac{1}{2}$ |
| $u_{all}$    | 0                                 | 1             |

$u_{some}$  to communicate  $m_{\forall}$ . In order to communicate  $m_{\exists \rightarrow \forall}$  the utility of  $u_{all}$  is 0. Even though a literal hearer might interpret  $u_{some}$  as  $m_{\forall}$  half of the time,  $u_{some}$  is the more promising option for the speaker to get the message across. This results in the following strategies:

$$S_1 = \left\{ \begin{array}{l} t_{\exists \rightarrow \forall} \mapsto m_{some} \\ t_{\forall} \mapsto m_{all} \end{array} \right\} \quad (5.12)$$

The pragmatic hearer calculates the posterior probability for each message given each utterance, but unlike  $L_0$  he takes not only the priors but also the strategies of  $S_0$  into account. He calculates the product between the prior probability of a message  $m$  and  $S_0$  uses a  $m$  to encode this message, which is divided by the sum of this term for each message in this situation that  $u$  could potentially encode as well (Franke 2009: 27):

$$\mu(m|u) = \frac{Pr(m) \times \sigma(u|m)}{\sum_{m' \in M} Pr(m') \times \sigma(u|m')} \quad (5.13)$$

This measure increases the higher the prior probability of  $m$  is and the more likely  $u$  is to be used to encode  $m$  as compared to the alternative messages  $m'$ . In the case of the scalar implicature this yields the posterior probabilities in Table 5.10, which, applying the same reasoning as above, result in the interpretation strategies that are in line with the encoding preferences of  $S_1$  5.14:<sup>30</sup>

$$L_1 = \left\{ \begin{array}{l} u_{some} \mapsto m_{\exists \rightarrow \forall} \\ u_{all} \mapsto m_{\forall} \end{array} \right\} \quad (5.14)$$

This model is of course highly simplified because it avoids further alternative expressions and interpretations, such as other quantifiers (*many*, *most*, etc.) and the more explicit, but longer *some but not all*.<sup>31</sup> Still though, it can model scalar

<sup>30</sup>For a more detailed discussion of the formula see (Franke 2009: 26–27) and for the application to scalar implicatures (Franke 2009: 60).

<sup>31</sup>Effects of a *some but not all* message are discussed by Franke (2009: 77, 127).

implicatures both on the side of the hearer and on that of the hearer after only one recursive iteration step. The strategies of  $L_1$  or  $S_1$  can also be the input to higher order reasoning processes.

Table 5.10: Posterior probabilities for the pragmatic hearer  $L_1$ .

| $\mu_0(m u)$      | $m_{\exists \neg \forall}$ | $m_{\forall}$ |
|-------------------|----------------------------|---------------|
| $u_{\text{some}}$ | $\frac{2}{3}$              | $\frac{1}{3}$ |
| $u_{\text{all}}$  | 0                          | 1             |

#### 5.5.4 Application to fragments

The case of fragments is often more complex than the (simplified) scalar implicature example discussed above. However, the underlying situation is similar, even though the set of messages and utterances is considerably larger: There is a set of possible messages that the speaker might want to communicate to the hearer and a set of utterances that she can use for this purpose. The sets of possible messages and utterances are potentially extremely large (if not infinite, due to recursion in language), but in situations that are constrained by script knowledge (like in the data from experiment 12), the set of actually considered alternative messages and utterances is often relatively limited. Consider a simplified version of the pasta scenario which I used to illustrate the surprisal calculation methods in Section 5.3.5. In this version of the scenario there are only two messages with differing prior probabilities, which are given in (27). Again, I use the representations that resulted from preprocessing the production data collected with experiment 12. Each message  $m_i$  has is associated a (hypothetical) prior probability  $Pr(m_i)$ .

- (27) a. pour pasta pot  $Pr = 0.6$   
b. pour salt pot  $Pr = 0.4$

Under the assumption that each “word” in these utterances can be omitted independently from the surrounding words, the set of possible utterances is:

$$U = \left\{ \begin{array}{c} \text{pour pasta pot, pour salt pot,} \\ \text{pour pasta, pour pot, pour salt, pasta pot, salt pot,} \\ \text{pour, pasta, pot, salt} \end{array} \right\} \quad (5.15)$$

Even in this highly simplified scenario with only two messages that differ only in one word, there are 11 possible utterances. The combination with the two mes-

sages in (27) yields the tableau in Table 5.11, based on which the same reasoning steps like in the scalar implicature example can be performed. If more meanings were considered, the complexity of the tableau would increase, but the problem remains identical.

Table 5.11: Tableau for the simplified pasta scenario game.

| Message                     | Pr(m) | $a_{\text{pour pasta pot}}$ | $a_{\text{pour salt pot}}$ | $a_{\text{pour pasta pot}}$ | $a_{\text{pour salt pot}}$ | $a_{\text{pour pasta}}$ | $a_{\text{pour pot}}$ | $a_{\text{pour salt}}$ | $a_{\text{pasta pot}}$ | $a_{\text{salt pot}}$ | $a_{\text{pour}}$ | $a_{\text{pasta}}$ | $a_{\text{pot}}$ | $a_{\text{salt}}$ |
|-----------------------------|-------|-----------------------------|----------------------------|-----------------------------|----------------------------|-------------------------|-----------------------|------------------------|------------------------|-----------------------|-------------------|--------------------|------------------|-------------------|
| $m_{\text{pour pasta pot}}$ | 0.6   | 1,1                         | 0,0                        | ✓                           | –                          | ✓                       | ✓                     | –                      | ✓                      | –                     | ✓                 | ✓                  | ✓                | –                 |
| $m_{\text{pour salt pot}}$  | 0.4   | 0,0                         | 1,1                        | –                           | ✓                          | –                       | ✓                     | ✓                      | –                      | ✓                     | ✓                 | –                  | ✓                | ✓                 |

Applying game-theoretic reasoning to the production and interpretation of fragments requires a conceptual and a methodological modification to the approach by Franke (2009). The theoretical modification concerns the relationship between utterances and messages. In the case of Franke’s model, there are only complete utterances, and truth conditions determine whether an utterance is suitable for communicating a message. In the case of fragments however, it is not reasonable to assume that a noun phrase like *pasta* is *true* in a situation. Instead, the relationship between utterances and messages can be modeled by the suitability of an utterance to communicate a message: An utterance is classified as suitable for communicating a message if it can be derived by an arbitrary number of omissions from the message’s full form: *pour pasta* can be derived from  $m_{\text{pour pasta pot}}$  by omitting *pot*, whereas it cannot be derived from  $m_{\text{pour salt pot}}$ . Of course, this is a further simplification, because in reality each message can be encoded by a variety of nonelliptical utterances, which differ in the choice of lexical items and word order, among other properties. However, in principle the same reasoning can be applied to this situation: The set of utterances that can encode a message is restricted to those expressions that can be generated by grammatically licensed omissions from the set of full sentences that can encode this message.

The methodological problem concerns ambiguous fragments. Since speakers use fragments, so it is reasonable to assume that in some situations fragments are



more useful encodings than full sentences, so the model must allow for this outcome. However, the model so far predicts that fragments are sometimes as good as sentences, but never better than them: Complete sentences unambiguously identify the message intended by the hearer, so they are always preferred over potentially more vague fragments, just like the unambiguous  $u_{all}$  has the highest utility in the scalar implicature game by Franke (2009). Even though, at least in the example presented here, fragments like e.g. *salt* or *pasta* share this property, in this situation they are not preferred over the full sentence. The main advantage of fragments is that they perform the same function as sentences in less time and with a reduced production effort for the speaker. Therefore, it is reasonable to integrate a cost term into the model, which reduces the speaker utility of longer utterances by discounting a percentage of the utility for each word. This models the speaker's preference for producing shorter utterances, which will favor unambiguous fragments in the case of my example. In large-scale applications the cost term might even favor potentially ambiguous fragments, if the prior probability of the message as compared to competing messages and the reduction of production cost are large enough to outweigh the remaining ambiguity.

### 5.5.5 Application to natural language data

In principle it is possible to apply this model to the data set collected with experiment 12. This data set contains an annotation layer of the message underlying each utterance, so that the prior probability of each of message in the scenario can easily be calculated. Based on the preprocessed data set it is also straightforward to determine which complete structures can be used to communicate each message and hence to specify (i) the set of possible utterances and (ii) whether each of these utterances can be used to communicate a message. However, the simple example above showed that even when there are only two possible messages there are 11 possible utterances. Consequently, the about 100 utterances produced for each scenario in experiment 12 do not allow for a fine-grained investigation of the relationship between game-theoretic utility and omission. A larger and more homogeneous data set would be required to investigate the predictions of a game-theoretic account of fragment usage in detail.

### 5.5.6 Implications and comparison to UID

Even though I argued in the introduction to this section that UID and game theory make partially overlapping predictions on omissions in fragments, they differ with respect to which aspects of communication they model, the mechanism that derives their predictions, and some empirical predictions.

Game theoretic pragmatics models both the behavior of the speaker and the hearer, whereas UID focuses on the speaker and accounts only for encoding but not for decoding. The hearer is taken into account indirectly, since the speaker adapts the signal to the cognitive resources she assumes the hearer to possess. The game-theoretic approach explicitly models the actions performed by both agents, as well as their expectations and preferences. By focusing on encoding, UID compares and ranks the usefulness of various utterances to communicate a single message, and the properties and the likelihood of alternative messages are considered only insofar as they contribute to the surprisal of words. This ignores the possibility that an utterance is falsely interpreted as another message than the one intended by the speaker. In contrast, the game-theoretic account predicts a preference for more explicit forms if the intended interpretation is not the most likely one if a fragment was used.

UID and the game-theoretic approach also attribute production preferences to different mechanisms. Whereas UID is a psycholinguistic theory based on the efficient distribution of processing effort (in the interpretation that I adopt) or efficient communication in the presence of noise and is relatively indifferent to meaning, a game-theoretic approach does not take into account processing explicitly but focuses on the mapping between utterances and meanings. Even though I showed above that the predictions of UID and game theory are often aligned, the different mechanisms that trigger the choice of an encoding result in partially diverging predictions with respect to the insertion of additional redundancy to avoid troughs: From a game-theoretic perspective, a maximally informative fragment is the optimal encoding, whereas from the UID perspective it can be beneficial to insert redundant material in such an utterance in order to reduce peaks in the ID profile even though these insertions do not increase the utility of the utterance from a game-theoretic perspective. The evidence from previous research and my experiments that speakers avoid peaks in the ID profile suggests that a game-theoretic account alone cannot account for the complete empirical picture and a processing account like UID is still needed.

Since UID and the game-theoretic approach operate on different levels of analysis and model different aspects of language production and comprehension, they are not mutually exclusive. Furthermore, each of the approaches can account for empirical observations that the other one cannot: UID predicts that speakers insert additional redundancy before unpredictable words, and the game-theoretic account explains why peaks in the ID profile are not smoothed by omitting unlikely words. Therefore, in future research both approaches might be integrated in a more comprehensive model of the choice between alternative ways of encoding a message. For instance, the well-formedness with respect to UID

### *5.5 Outline of a game-theoretic model of fragment usage*

might be included in the game-theoretic model as a cost term that penalizes inefficient signals. Since the speaker has an interest in getting her message across, she will try to prevent communication failure resulting from violations of UID. Future research might spell out such an account and test its empirical predictions.



## 6 General discussion

In this book I investigated two research questions on fragments with experimental methods: First, which syntactic structure underlies fragments? And second, why do speakers use fragments at all? The results of my experiments contribute both to the research on fragments and to that on information-theoretic constraints on language production and processing.

The experiments on the syntax of fragments in the first part of this book constitute the first systematic investigation of a series of predictions derived from currently competing theories of fragments. Previously, many of these theories were founded only on partially conflicting introspective data that had not been empirically verified with experiments or corpus studies. The experiments in Chapter 3 furthermore provide relatively theory-independent insights into the form that fragments can take: Fragments can be non-constituents, they exhibit case connectivity and short answer fragments tend to match the form of their antecedent. These properties have to be taken into account by future theoretical research on fragments.

In the second part of this book, I developed an information-theoretic account of fragment usage, which explains when fragments are preferred over full sentences and which words are preferably omitted in fragments. The central predictions of this account are confirmed by two experiments. From the perspective of the research on fragments, this constitutes the first attempt to answer the almost unexplored question of why people use fragments at all that is empirically supported by actual linguistic data. From a psycholinguistic perspective, the finding that omissions in fragments are constrained by UID extends the evidence for UID in two ways. First, I find UID effects on the omission of content words, whereas previous research focused mostly on semantically relatively vacuous function words. Second, I show that not only linguistic context, but also script-based extralinguistic context modulates the predictability and consequently the reduction of words and utterances. Previous research on UID estimated surprisal almost exclusively based on local linguistic context, i.e.  $n$ -gram surprisal. In contrast to this, I developed a method to quantify effects of extralinguistic context based on script knowledge and a new approach to estimating surprisal.

## 6.1 Results on the syntax of fragments

The first goal of my research was to investigate what structure underlies fragments. Although diverse and mutually exclusive accounts of fragments have been proposed previously, most of these theories have been supported only by introspective data, but not by empirical evidence from corpus studies or experiments. The experiments in Chapter 3 constitute the first systematic investigation of a series of theoretical predictions of competing theories of fragments. These studies investigated two research questions on the syntax of fragments. First, are fragments underlyingly sentential? And, second, are fragments generated by movement and deletion?

These questions differentiate between the main generative accounts of fragments: the nonsentential account by Barton & Progovac (2005), the *in situ* deletion account by Reich (2007) and the movement and deletion account by Merchant (2004a). Since these theories make differing predictions on which fragments are grammatical, distinguishing between the theories is not only relevant from a theoretical perspective: The investigation of the usage of fragments also requires to know which utterances can be derived by grammar, because UID accounts only for variation within the limits of grammar (Jaeger 2010). In what follows I briefly summarize the main results and their implications for the theoretical analysis of fragments, before I review syntactic properties of fragments that are supported by my experiments and which must be taken into account by any theory of fragments and any empirical study on their usage.

### 6.1.1 Fragments are underlyingly sentential

#### 6.1.1.1 Case connectivity indicates unarticulated structure in fragments

Experiments 1–3 suggest that fragments contain unarticulated linguistic material, as sentential accounts of fragments assume. This is evidenced by structural case connectivity effects on DP fragments. Structural case marks the relationship between words and, unlike inherent case, it does not encode a specific  $\theta$ -role. In the case of the German accusative which I tested in my experiments it marks a DP as the direct object of the verb. If a DP fragment can appear in structural case, this suggests that there is a silent verb in such DP fragments, because otherwise the accusative cannot be checked. Relying on structural case as evidence for unarticulated structure is especially convenient because the presumable unacceptability of structural case marking is a crucial property of fragments according to the nonsentential account by Barton & Progovac (2005).

The experiments provide evidence for case connectivity and hence support a sentential account of fragments: Experiment 1 shows that accusative DP fragments are more acceptable than nominative DPs in contexts where accusative is licensed in a full sentence. Experiment 2 validates this finding with a production study that confirms that accusative is indeed more likely than nominative in the contexts tested in experiment 1. Experiment 3 rules out the possibility of a mixed account of fragments, according to which fragments can be derived both by ellipsis and as genuine nonsententials, depending on whether context provides sufficient evidence for ellipsis resolution or not. Finally, experiment 6, whose main objective was testing an alternative explanation of the P-stranding generalization, also disconfirms the prediction of the nonsentential account that prepositional case-marked DP fragments are degraded as compared to default case-marked ones.

#### 6.1.1.2 Implications for the theoretical analysis of fragments

Among the theories of fragments that I discussed, case connectivity provides evidence for a sentential account of fragments. This conclusion crucially hinges on the assumption that structural case marking is a valid diagnostic of unarticulated structure. From the perspective of Barton & Progovac (2005), this assumption could be questioned by arguing that the German accusative is inherent case too, because it often marks DPs that receive a patient  $\vartheta$ -role. Progovac et al. (2006) actually claim this for Serbian, but the tests that they adduce for this language yield the opposite result in German. Furthermore, the more instances of case are analyzed as inherent case in order to explain case connectivity under a nonsentential accounts, the less data are explained by the distinction between structural and inherent case that rules out some cases on fragments (e.g. nominative in English and Korean according to Barton & Progovac (2005)).

The conclusion that accusative case marking on DP fragments indicates unarticulated structure also relies on the generative concept of case checking. Nonsentential accounts of fragments that operate in different syntactic frameworks, like HPSG (Ginzburg & Sag 2000, Fernández & Ginzburg 2002, Schlangen 2003), explain case connectivity by linking the morphosyntactic properties of a short answer fragment to that of the *wh*-phrase in a QuD by conindexation. From this perspective, fragments do not contain any unarticulated structure and the derivation of DP fragments does not involve the PF-deletion of a verbal head, but for it to be interpreted correctly, it must match the properties of the *wh*-phrase. Such accounts make in principle very similar predictions to in situ deletion, which also relies on a (potentially implicit) QuD in order to determine which parts of the

sentence can be omitted. Empirically testing the exact predictions of the HPSG account and teasing them apart from those of in situ deletion is relatively complicated, since HPSG accounts assume that different types of fragments are based on categorically different constructions instead of a single deletion mechanism.

### 6.1.2 **Fragments are not obligatorily moved**

Experiments 4–10 investigated potential evidence for movement in fragments. As has been proposed by Merchant (2004a), I interpret restrictions on left dislocation which constrain the form of fragments in a way that cannot be explained by in situ deletion as evidence for movement. I investigated effects of three (presumable) movement restrictions: The ban on extraction of the complement of prepositions (P-stranding) in German, the impossibility to front complement clauses that lack an overt complementizer and restrictions on multiple prefield constituents in German. Taken together, these experiments do not provide evidence for movement so that, given the results on sententiality in the previous experiments, the data support an in situ deletion account of fragments. Furthermore, in the discussion of the German data I showed that the movement and deletion account suffers from serious theoretical problems concerning the placement of the E feature in this language.

#### 6.1.2.1 **Preposition omission does not evidence movement in fragments**

Experiments 4 and 5 support the data on preposition omission which are reported in Merchant (2004a) and Merchant et al. (2013) for German and English: Omitting the preposition in short answers is degraded in German but felicitous in English. However, the production study in experiment 7 provides evidence for a general tendency for the form of the answer to match that of the question. Hence the preference for realizing the preposition in German short answers can be explained by the form of the question alone, without having to assume that the generation of the answer involves P-stranding too. There are at least two ways to account for this parallelism. On the one hand, it might be due to the facilitation of language production by re-using contextually available structures (Levelt & Kelter 1982), but it might also reflect congruence between questions and answers (Reich 2002b). This is expected specifically under accounts that emphasize the relevance of QuDs to the licensing of fragments, like the in situ deletion account by Reich (2007) and the HPSG account by Ginzburg & Sag (2000). Finally, the German prepositional case-marked DP short answers were degraded in context of PP questions, but not rejected across the board, like unnatural multiple prefield configurations in experiment 10. Since Lemke (forthcoming) also



reports that some of these mismatches between the category of *wh*-phrase and answer are relatively acceptable, it is at least questionable whether the penalty for preposition omission can be attributed to the unavailability of P-stranding in German.

#### 6.1.2.2 Mismatches between left dislocation and fragments

Experiments 8 and 9 investigated effects of complementizer omission on the acceptability of topicalized complement clauses and the corresponding fragments. They show that the preference for overt complementizers in fragments is relatively weak in German and absent in English when properties of the materials which concern the acceptability of the corresponding complement clauses in complete sentences are more tightly controlled. A further surprising result of the experiments is that, unlike what has been reported in the literature for more than 40 years (Morgan 1973, Stowell 1981, Webelhuth 1992, Merchant 2004a), in none of my experiments fronting complementizer-less clauses was degraded as compared to complement clauses with overt complementizers. Since there is no evidence for the movement restriction that presumably constrains the form of fragments, it cannot explain even the subtle preference for realizing the complementizer in German fragments. Experiment 10 on German multiple prefield constituents reveals further mismatches between left dislocation and fragments, since strongly degraded prefield configurations that involve a subject and another argument DP result in acceptable fragments.

#### 6.1.2.3 Implications for the theoretical analysis of fragments

None of my experiments provides clear evidence for movement in fragments. The conclusions in Merchant et al. (2013) can either be explained by independently motivated processing constraints (in the case of P-stranding), are based on a presumed movement restriction that is not reflected in my experiments (in the case of complement clause topicalization) or not supported by the data (in the case of multiple prefield constituents). I conclude that this supports an *in situ* deletion account of fragments, which is derivationally simpler and hence to be preferred in the absence of specific evidence for movement.

#### 6.1.3 Implications for (generative) syntactic theories

Taken together, the experiments on the syntax of fragments support an *in situ* deletion account of fragments. The nonsentential account cannot explain the case marking data and the experiments on potential movement restrictions found

no clear evidence for obligatory movement in fragments. The conclusion that – within a generative framework – fragments must be analyzed as underlyingly sentential implies that syntax does not need to be modified so as to generate bare XPs as a well-formed output. Instead, fragments can be derived from regular sentences by ellipsis. The experiments on movement in fragments suggest that this ellipsis applies to regular sentences rather than obligatory left dislocations. Consequently, at least in fragments, ellipsis does not need to be triggered by the E feature proposed by Merchant (2004a). It is rather licensed by a contextually salient antecedent and it can ultimately be triggered by information-theoretic processing constraints, as experiments 11 and 12 show.

## 6.2 Results on the usage of fragments

### 6.2.1 Results on the form of fragments

The experiments on the syntax of fragments were primarily designed to test the predictions of the specific theories that I investigated, but from a theory-neutral perspective they provide evidence for several properties of fragments that must be taken into account by any theory of fragments.

- There is clear evidence for case connectivity. DP fragments appear in the same case as they do in the corresponding full sentence (or the *wh*-phrase in a preceding, potentially implicit, QuD). In discourse-initial fragments, where an implicit QuD must be retrieved, case may still vary depending on which QuD the speaker has in mind.
- Fragments tend toward matching the properties of the corresponding *wh*-phrase in the antecedent, such as a QuD. This concerns not only case connectivity, but also the omission of the preposition in PP fragments. In languages where the preposition in the antecedent cannot be stranded, like German, omitting it in the answer is degraded.
- Fragments can be non-constituents. This is indicated by the acceptability of fragment answers in experiment 10 on multiple prefield constituents.
- Fragments are not subject to specific movement restrictions. This is most clearly shown by experiment 10, where heavily degraded prefield configurations result in acceptable fragments.

The findings discussed in this section are also relevant to any empirical investigation of the usage of fragments in order to exclude ungrammatical fragments

from the set of possible utterances. As for the experiments presented here, in experiment 11, which investigated relative preferences of (un)predictable sentences and fragments, all fragments exhibited case connectivity. Similarly, the merger of PPs to a single lexical item in the preprocessing of the production data collected in the production experiment 12 accounted for the strong tendency not to omit prepositions in German PP fragments.

### 6.2.2 The usage of fragments is constrained by UID

#### 6.2.2.1 An information-theoretic account of fragment usage

In Chapter 4 I developed an information-theoretic account of the usage of fragments that explains when speakers use fragments and if they do so, which ones are preferred. In previous research this issue has been almost completely ignored. The only exception is the game-theoretic approach by Bergen & Goodman (2015), which however is based on a highly simplified example that comprises only four utterances, does not predict *which* words are omitted in fragments and has not been tested at actual linguistic data.

The UID-based account of fragment usage that I propose predicts that, taking the full sentence as a starting point, the choice between omitting and realizing words within that sentence is constrained by UID, i.e. the tendency to transmit information at a rate close to, but not exceeding channel capacity. This goal is achieved by omitting predictable words and realizing words that reduce the surprisal of following ones. Taken together, this leads to a higher ratio of fragments in predictive contexts, because predictability-driven omissions are overall more likely in such environments.

In Chapter 5 I presented two experiments that support these predictions. The acceptability rating experiment 11 shows that fragments are more acceptable when they refer to predictable messages than when they encode unpredictable ones. The production experiment 12 provides evidence for the two more fine-grained predictions of UID on the word level: Uninformative words are omitted in order to avoid inefficient troughs in the information density profile of the utterance and additional redundancy is inserted in order to reduce peaks which might otherwise exceed the hearer's processing resources.

#### 6.2.2.2 Comparison of UID to other approaches to optional omissions

The experimental results are in line with the three predictions of UID on the usage of fragments. Some of them, however, might also be explained by other accounts of optional reduction which do not share the theoretical assumptions

that UID implies, such as a parallel parser, audience design and communication through a noisy channel. In what follows I discuss to what extent these accounts (source coding, availability-based production, information structure, game theory) account for the full empirical picture: Some of them are able to explain part of the data, but UID provides a unifying account of the complete empirical picture. This does not neglect possible effects of other factors than information-theoretic optimization on omissions in fragments, but it shows that UID explains all of the predictability effects that my experiments on fragment usage reveal, whereas other frameworks account for only a part thereof.

*Source coding* is in line with UID in predicting that frequent messages will receive shorter encodings on average. UID however derives this from properties of the channel in Shannon's communication model, whereas for source coding only properties of the source, i.e. the frequency of messages, matter. Assigning shorter encodings to frequent messages reduces the average utterance length and increases the efficiency of communication. The crucial difference between source coding and UID is that only UID predicts the insertion of additional redundancy in order to reduce information peaks in the ID profile. In contrast, from a source coding perspective, maximizing encoding density is most efficient. The stronger preference for fragments to encode predictable messages that has been evidenced by experiment 11 could be interpreted as the result of source coding. However, from a source coding perspective there is no benefit in introducing additional redundancy into the signal, but the production experiment 12 suggests that speakers do so. Therefore, source coding fails to explain the complete empirical picture.

*Availability-based production* (Ferreira & Dell 2000) has been taken to explain some effects of predictability on the omission of function words by speaker-centered production preferences, without taking the processing perspective into account. The general idea is that the choice between omitting and realizing optional words is driven by the effort to retrieve words from memory and the tendency to avoid disfluencies. From this perspective, inserting optional words before unpredictable words whose retrieval causes more effort contributes to keep speech fluent, whereas from the UID perspective the trigger for such insertions is the adaptation of the signal to the channel, that is, the cognitive resource of the hearer. The result of experiment 12 that words that reduce the surprisal of following words more strongly are more often realized is in principle in line with availability-based production. However, availability-based production does not explain why predictable words are more likely to be omitted.

To my knowledge, the relationship between *information structure* and information theory has not been explicitly looked into yet, but there is probably a

close relationship between surprisal and information-theoretic concepts such as topic, focus and givenness. For instance, given expressions and specifically topics might be more likely to be talked about, and foci often mark new information, which might be on average less likely. The notion of focus is central to the in situ deletion account of fragments that my experiments in the first part of this book support. It is hence reasonable to assume that information structure has an effect on the usage and form of fragments. This might raise the concern that information-theoretic surprisal estimates actually reflect a distinction between information-structural concepts so that information structure alone is able to explain the distribution of omissions in fragments. However, information structure alone can explain when fragments are licensed, but not when they are preferred over a full sentence. Not all given words are actually omitted, as can be trivially shown by e.g. congruent sentential answers to *wh*-questions. Nevertheless, it might be a promising topic of future research to tease apart effects of information theory and information structure in experimental studies.

*Game-theoretic approaches* avoid an inherent theoretical problem to UID: They model not only how utterances are assigned to messages, but also the reverse procedure, that is, how interpretations (messages) are assigned to utterances. In Section 5.5 I sketched how the usage of fragments could be modeled in a game-theoretic framework like the Iterated Best Response model by Franke (2009). Despite its advantages, a game-theoretic account alone probably cannot explain the whole range of data in the production study, because, unlike UID, game-theoretic accounts do not predict an upper bound on densification. Therefore, just like I argued above for source coding, the evidence for the insertion of redundancy in order to smooth peaks that experiment 12 seems to contradict the predictions of the game-theoretic account. As I noted above, a careful investigation of the game-theoretic approach requires a large and correspondingly annotated data set that is currently not available.

## 6.3 Implications for predictability effects on language processing

The results of the experiments in Chapter 5 have implications for a broader range of research on script knowledge and predictability effects on language processing. From a methodological perspective, they showed that script-based probabilistic event chains can be used to manipulate and quantify effects of extralinguistic context on the predictability of utterances. The method of surprisal estimation that I applied to my production data allows for the quantification of effects of

extralinguistic context on the word level and provides a solution to a circularity issue that has previously complicated the estimation of surprisal on elliptical data. From a theoretical perspective, my experiments extend previous evidence for UID in two ways: They show that UID constraints the omission of content words and that extralinguistic context determines surprisal. This in turn has broader implications for the research on predictability effects and language processing, since it provides indirect evidence for assumptions about language production and processing that are implied by UID.

### 6.3.1 Script-based event chains as a model of context

In experiments 11 and 12 I based my materials on probabilistic event chains extracted from the DeScript corpus of script knowledge (Wanzare et al. 2016) in order to determine the likelihood of target utterances. Previous studies that investigated effects of script knowledge relied on stimuli constructed according to researcher intuitions and/or based on norming studies, which are specific to a particular experiment (see the references in Section 5.1).

Script corpora like DeScript (Wanzare et al. 2016) are a valuable resource for constructing empirically founded models of script knowledge. Due to the large number of contributors to such corpora, they provide a reasonable approximation to the representation of scripts in the memory of a larger population. This is crucial to experimental studies that investigate effects of script knowledge, because script knowledge manipulations presuppose that subjects possess the relevant script knowledge. The use of probabilistic event chains rather than deterministic ones takes into account both the uncertainty about the next event and differing script representations between subjects. This is desirable, since interlocutors in actual conversations must consider the possibility that their script representations differ to some extent.

Experiment 11 confirms the validity of the script-based manipulation of predictability. In the rating experiment 11, utterances referring to predictable events were perceived as more natural. This was specifically the case when subjects possessed the relevant script knowledge, which was assessed with a questionnaire following the main experiment. Event chains extracted from script corpora thus are a promising method for constructing materials in research on script knowledge: They are psychologically realistic representations of an average speaker's script knowledge and reduce the amount of cloze and norming studies required for stimulus generation.

### 6.3.2 Surprisal estimation in elliptical data

For the analysis of the production data from experiment 12 I developed a method of surprisal estimation that is specifically suitable for elliptical data. The approach is based on the insight by Hale (2001) that the surprisal of a word is proportional to the cumulated probability mass of the parses that it disconfirms. Unlike Hale (2001), however, it allows for omissions to occur before and after each word in the actually produced utterance.

This approach avoids a circularity issue that affects *n*-gram surprisal estimated from corpora that contain elliptical data: Since I estimate surprisal based on the probability of complete structures, the omission of a word does not affect its own surprisal. Furthermore, the method is psychologically realistic, because omitted words preceding a target word in the complete structures have no effect on the target word's predictability: Only the realized words that are available to the hearer modulate surprisal.

This method requires to know which nonelliptical utterances are possible in a specific situation and how likely they are. In experiment 12 I collected a data set that constrained by extralinguistic context stories based on which the likelihood of utterances can be estimated, and which contains the relevant omissions. In order to estimate the surprisal of both omitted and realized words, a procedure to reconstruct all omissions within this data set is also necessary. In the case of my data set, omissions were reconstructed manually. This required a large extent of annotation work. Future research might extend this approach to larger data sets in case the preprocessing procedure can be at least in part automatized.

### 6.3.3 UID constrains the omission of content words

Previous evidence for UID focused mostly the omission of semantically relatively vacuous function words. Investigating closed-class function words like relative pronouns (Levy & Jaeger 2007) and complementizers (Jaeger 2010) has several methodological advantages over focusing on content words: Both realized function words and instances of omissions are easy to find in corpora and in case of omissions reconstructing the missing expression is relatively straightforward. The surprisal of the target word itself can be equated with that of the syntactic construction that it encodes (e.g. a relative or complement clause) and that of the surrounding words can be estimated with *n*-gram models (Levy & Jaeger 2007).

Content words in contrast require a sophisticated preprocessing approach, a strategy for the reconstruction of omissions and a different method for surprisal estimation that is psychologically realistic and not affected by omissions in the

actual data. This surprisal estimation method in turn requires a particular data set, which contains a sufficiently large number of utterances produces in the same context to calculate reasonable surprisal estimates. I proposed solutions to these issues and was able to show that the central predictions hold not only for the omission of function words, but also for that of content words.

### 6.3.4 Effects of extralinguistic context on predictability

In principle, information-theoretic approaches to language predict that the likelihood of a word depends on a variety of sources, which comprise both linguistic and extralinguistic context. Previous research in the field however estimated the likelihood of utterances and words only based on linguistic, and specifically very local intrasentential, context. At most, context comprised some utterances preceding a target word (see e.g. Tily & Piantadosi 2009, Kravtchenko 2014), who used guessing experiments (Shannon 1951) to quantify this context's effect on predictability. Most of the time however, surprisal is estimated with *n*-gram models, take only a few words preceding the target word into account.

Experiments 11 and 12 constitute to my knowledge the first investigation of effects of extralinguistic context on the predictability and the omission of words. Experiment 11 indicates that utterances that refer to events which are predictable in a script-based extralinguistic context are more likely to be reduced. Experiment 12 shows that even unigram surprisal calculated on the utterances for a single scenario only is a significant predictor of omission: Words that are more likely to appear in an utterance in that scenario are more often omitted. This shows that not only local linguistic context, but also script-based extralinguistic context determines the likelihood of words and of their omission.

### 6.3.5 Psycholinguistic implications of UID

Extending the available evidence for UID to omissions of content words and effects of extralinguistic context indirectly supports more general assumptions about language production and processing that UID presupposes.

#### 6.3.5.1 Predictability is related to processing effort

The assumption that processing predictable words requires less effort is crucial to the interpretation of UID that I take, which interprets channel capacity as an upper bound to the processing resources of the hearer. From an empirical perspective, it is relatively uncontroversial that predictable words are easier to process, which is indicated for instance by faster reading times (Demberg & Keller



2008, Levy 2008, Smith & Levy 2013, Brothers & Kuperberg 2019) and a reduced N400 in ERP studies (Frank et al. 2015, Delogu et al. 2017). In my experiments I did not explicitly measure processing effort, but since it is indexed by surprisal, the optimization of utterances with respect to UID ensures a uniform distribution of processing effort.

#### **6.3.5.2 The human parser is parallel**

The relationship between predictability and processing effort is theoretically explained by the derivation of surprisal from the work done by the human parser (Hale 2001, Levy 2008), which consists in the rejection of structures that a word disconfirms. Processing effort is proportional to the probability mass of the rejected parses, hence unlikely words are harder to process. This assumption presupposes that the human parser is fully parallel. Serial or bounded parallel parsers, do not keep the complete set of possible parses, so it is impossible to calculate how large the share of the probability mass of the rejected parses is. Levy (2008) however notes that the assumption of a fully parallel parser might not be psychologically realistic, since it involves the calculation of an extremely large amount of low-probability structures which never become likely throughout the parsing process. Still though, Levy (2008: 1135–1136) argues that even if it is assumed that the human parser is not fully parallel, computational surprisal estimates still provide a reasonable approximation to processing effort, as long as only those parses that are assigned very low probabilities during the parsing process are ignored.

#### **6.3.5.3 Speakers perform audience design**

Since optimization with respect to UID consists in the adaptation of the signal to the channel, whose capacity I interpret as an upper bound to the hearer's cognitive resources, UID presupposes audience design. The result of this optimization will differ depending on properties of the hearer and the situation. For instance, if the hearer's processing resources are reduced by an interfering effortful task, the speaker will choose a less dense encoding. Predictability effects do not necessarily evidence audience design: For instance, source coding predicts that more likely expressions are more often reduced even though it takes only statistical properties of the source into account. Similarly, availability-based production explains the distribution of optional omissions only with production effort. As I argued above in the comparison of the predictions of UID to other approaches to optional deletion, unlike UID, neither source coding nor availability-based production are able to explain the full empirical picture.

## 6.4 Implications for other reduction phenomena

The question of why speakers sometimes prefer a reduced form over a syntactically complete one is not only relevant to fragments, but also to other omission and reduction phenomena. This holds for antecedent-based ellipses like verb phrase ellipsis (Sag 1976, Williams 1977), sluicing (Ross 1969) or gapping (Ross 1970), but also for the omission of topics, subjects and objects and pronominalization. For all of these phenomena, UID predicts that the more predictable an expression is, the more likely it will be omitted, provided that the omission is permitted by grammar, and that omission is dispreferred when it reduces information density peaks on following material.

For some of these phenomena there is evidence for predictability effects on the choice of encodings that are in line with UID. Tily & Piantadosi (2009) find that more predictable referents are more often pronominalized, and Kravtchenko (2014) observes that predictable subjects are more often omitted in Russian. More recently, Schäfer et al. (2021) show that verb phrase ellipsis is more strongly preferred the longer, i.e. the more redundant, a VP is.

A further question that predictability effects on omissions raise is whether UID explains only whether a specific ellipsis occurs provided that it is licensed, or whether some licensing conditions on ellipsis actually reflect UID effects. For instance, Chung (2006) proposes a syntactic identity condition on sluicing that requires that all words that are omitted in the sluice must be given in the antecedent. This intends to account for the acceptability of preposition omission under sluicing (1a), but not under sprouting (1b).

- (1) a. John danced with somebody, but I don't know who ~~John danced with~~.
- b. John danced, but I don't know who ~~John danced with~~.

Two recent studies suggest that this can be explained by predictability: Poppels & Kehler (2019) show that the acceptability of sluices that violate Chung's constraint increases the more accessible the QuD that contains the antecedent to the sluice is. This suggests that predictable words can be omitted whereas unpredictable ones cannot. In a self-paced reading experiment on similar full forms to (1) in German, Lemke et al. (forthcoming) find structurally mismatching sluices are read faster in the sluicing condition than under sprouting which also indicates that the redundant *John danced with* is simply more predictable in the sluicing condition than under sprouting. Future research will show whether the result that UID constrains omissions in fragments can be extended to other instances of ellipsis and whether it can explain away specific identity conditions that have been postulated for some ellipsis phenomena.

# Appendix: Models

## Experiment 1 (CLMM)

### Full model

```
Rating ~ (Case + XPlease + Position)2 + (1 + Case * XPlease | Subject)
+ (1 + Case | Item)
```

### Final model

```
Rating ~ Case + XPlease + (1 + Case * XPlease | Subject) + (1 + Case
| Item)
```

## Experiment 2 (Logistic mixed effects regression)

### Full model

```
Answer ~ 1 + (1 | Subject) + (1 | Item)
```

## Experiment 3 (CLMM)

### Full model

```
Rating ~ (Case + Predictability + Position)2 + (1 + Case * Predictabil-
ity | Subject) + (1 + Case * Predictability | Item)
```

### Final model

```
Rating ~ Predictability + (1 + Case * Predictability | Subject) + (1
+ Case * Predictability | Item)
```

## Experiment 4 (CLMM)

### Full model

Rating ~ (Preposition + CaseDative + CaseGenitive + Position)<sup>2</sup> + (1 + Preposition \* CaseDative + Preposition \* CaseGenitive | Subject) + (1 + Preposition | Item)

### Final model

Rating ~ Preposition + CaseDative + CaseGenitive + Position + (1 + Preposition \* CaseDative + Preposition \* CaseGenitive | Subject) + (1 + Preposition | Item)

## Experiment 5 (CLMM)

### Full model

Rating ~ (Preposition + Sententiality + Position)<sup>2</sup> + (1 + Preposition | Subject) + (1 + Preposition \* Sententiality | Item)

### Final model

Rating ~ Preposition \* Sententiality + (1 + Preposition | Subject) + (1 + Preposition \* Sententiality | Item)

## Experiment 6 (CLMM)

### PP v Prepositional case DP

#### Full model

Rating ~ FragmentType \* Position + (1 + FragmentType + Position | Subject) + (1 + FragmentType + Position | Item)

#### Final model

Rating ~ FragmentType + Position + (1 + FragmentType + Position | Subject) + (1 + FragmentType + Position | Item)

## Default case DP v Prepositional case DP

### Full model

Rating ~ FragmentType \* Position + (1 + FragmentType + Position | Subject) + (1 + FragmentType + Position | Item)

### Final model

Rating ~ FragmentType + (1 + FragmentType + Position | Subject) + (1 + FragmentType + Position | Item)

## Experiment 7 (Logistic mixed effects regression)

### Locative v subcategorized PPs

#### Full model

Answer ~ 1 + (Question + Function + Index)<sup>2</sup> + (1 | Subject) + (1 | Item)

#### Final model

Answer ~ 1 + (1 | Subject) + (1 | Item)

### Complete data set, after pooling

#### Full model

Answer ~ 1 + Question + Function + Index + Question:Function + Question:Index + Function:Index + (1 | Subject) + (1 | Item)

#### Final model

Answer ~ 1 + Question + Function + (1 | Subject) + (1 | Item)

## Experiment 8 (CLMM)

### Analysis of the verb-second conditions only

#### Full model

Rating ~ (Sententiality + CCType + Position + MatrixVerb)<sup>2</sup> + (1 + CCType \* MatrixVerb | Subject) + (1 + CCType \* Sententiality | Item)

#### Final model

Rating ~ Sententiality + Position + Sententiality:Position + (1 + CCType + MatrixVerb + CCType:MatrixVerb | Subject) + (1 + CCType + Sententiality + CCType:Sententiality | Item)

### Main analysis after pooling

#### Full model

Rating ~ (Sententiality + CCType + Position + MatrixVerb)<sup>2</sup> + (1 + CCType \* MatrixVerb | Subject) + (1 + CCType \* Sententiality | Item)

#### Final model

Rating ~ Sententiality + CCType + Position + Sententiality:CCType + Sententiality:Position (1 + CCType \* MatrixVerb | Subject) + (1 + CCType \* Sententiality | Item)

## Experiment 8, follow-up (CLMM)

### 6.4.0.1 Full model for all analyses

Rating ~ (Sententiality + CCType + MatrixVerb)<sup>2</sup> + (1 + CCType \* MatrixVerb | Subject) + (1 + CCType \* Sententiality | Item)

### 6.4.0.2 Final model for all analyses

Rating ~ Sententiality \* CCType + (1 + CCType \* MatrixVerb | Subject) + (1 + CCType \* Sententiality | Item)

## Experiment 9 (CLMM)

### Main analysis of the complete data set

#### Full model

Rating ~ (Sententiality + CCType + Position + MatrixVerbBelieve + MatrixVerbMean + MatrixVerbSay)<sup>2</sup> + (1 + CCType + MatrixVerbBelieve + MatrixVerbMean + MatrixVerbSay + CCType | Subject) + (1 + CCType \* Sententiality | Item)

#### Final model

Rating ~ Sententiality \* CCType + Position + MatrixVerbBelieve + (1 + CCType + MatrixVerbBelieve + MatrixVerbMean + MatrixVerbSay + CCType | Subject) + (1 + CCType \* Sententiality | Item)

### Analysis of the sentential conditions only

#### 6.4.0.1 Full model

Rating ~ (CCType + Position + MatrixVerbBelieve + MatrixVerbMean + MatrixVerbSay)<sup>2</sup> + (1 + CCType | Subject) + (1 + CCType | Item)

#### 6.4.0.2 Final model

Rating ~ CCType + Position + (1 + CCType | Subject) + (1 + CCType | Item)

## Experiment 10 (CLMM)

### Full model for all pairwise comparisons

Rating ~ Prefield \* Sententiality + (1 + Sententiality | Subject) + (1 + Prefield | Item)

### Final model for all pairwise comparisons without a significant PREFIELD:SENTENTIALITY interaction<sup>1</sup>

Rating ~ Prefield + Sententiality + (1 + Sententiality | Subject) + (1 + Prefield | Item)

---

<sup>1</sup>In case of a significant interaction, the final model was identical to the full model.

## Experiment 11 (CLMM)

### Full model

Rating ~ (Sententiality + Predictability + Position + ScriptType + ScriptKnowledge)<sup>2</sup> + (1 + (Sententiality + Predictability + ScriptKnowledge)<sup>2</sup> | SubjectID) + (1 + Sententiality + Predictability + SNZscore | Item)

### Final model

Rating ~ Sententiality + Predictability + Position + Sententiality: Predictability + Predictability:ScriptKnowledge + (1 + (Sententiality + Predictability + ScriptKnowledge)<sup>2</sup> | SubjectID) + (1 + Sententiality + Predictability + SNZscore | Item)

## Experiment 12

### Entropy and lexicon size analysis (Linear regression)

#### Lexicon size

Full and final model

FragmentRatio ~ 1 + LexiconSize

#### Entropy

Full and final model

FragmentRatio ~ 1 + Entropy

#### Entropy v lexicon size

Full model

FragmentRatio ~ 1 + Entropy \* LexiconSize

Final model

FragmentRatio ~ 1 + Entropy



## Surprisal analyses (Logistic mixed effects regression)

### Unigram surprisal analysis, complete data set

Full and final model

$$Omission \sim 1 + \text{UnigramSurprisal} + (1 + \text{UnigramSurprisal} \mid \text{Subject}) + (1 + \text{UnigramSurprisal} \mid \text{Item})$$

Final model

### Context-dependent surprisal analysis, complete data set

Full and final model

$$Omission \sim 1 + \text{ContextSurprisal} + (1 + \text{ContextSurprisal} \mid \text{Subject}) + (1 + \text{ContextSurprisal} \mid \text{Item})$$

### Both surprisal predictors, non-final words only

Full model

$$Omission \sim 1 + \text{UnigramSurprisal} * \text{ContextSurprisal} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

Final model

$$Omission \sim 1 + \text{UnigramSurprisal} + \text{ContextSurprisal} + (1 \mid \text{Subject}) + (1 \mid \text{Item})$$

## Word order analysis (CLMM)

Full model

$$\text{Position} \sim \text{UnigramSurprisal} * \text{ContextSurprisal} + (1 + \text{UnigramSurprisal} * \text{ContextSurprisal} \mid \text{Subject}) + (1 + \text{UnigramSurprisal} * \text{ContextSurprisal} \mid \text{Item})$$

Final model

$$\text{Position} \sim \text{UnigramSurprisal} + \text{ContextSurprisal} + (1 + \text{UnigramSurprisal} * \text{ContextSurprisal} \mid \text{Subject}) + (1 + \text{UnigramSurprisal} * \text{ContextSurprisal} \mid \text{Item})$$



# References

- Abbott, Valerie, John B. Black & Edward E. Smith. 1985. The representation of scripts in memory. *Journal of Memory and Language* 24(2). 179–199. DOI: 10.1016/0749-596X(85)90023-3.
- Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding*. University of Connecticut. (PhD thesis). DOI: 10.1515/9783110284225.
- Abels, Klaus. 2012. *Phases: An essay on cyclicity in syntax*. Berlin & Boston: Walter de Gruyter. DOI: 10.1515/9783110284225.
- Aelbrecht, Loebke. 2009. *You have the right to remain silent: The syntactic licensing of ellipsis*. Brussels: Hogeschool-Universiteit Brussel. (PhD thesis).
- Arregi, Karlos. 2003. Clitic left dislocation is contrastive topicalization. *U. Penn Working Papers in Linguistics* 9(1). 31–44.
- Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246. DOI: 10.1016/j.jml.2006.02.005.
- Asr, Fatameh 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, UK: Association for Computational Linguistics.
- Attneave, Fred. 1959. *Applications of information theory to psychology*. Oxford: Holt-Dryden.
- 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.
- Balling, Laura Winther & Johannes Kizach. 2017. Effects of surprisal and locality on Danish sentence processing: An eye-tracking investigation. *Journal of Psycholinguistic Research* 46(5). 1119–1136. DOI: 10.1007/s10936-017-9482-2.
- Bar-Hillel, Yehoshua & Rudolf Carnap. 1953. Semantic information. *The British Journal for the Philosophy of Science* 4(14). 147–157.

## References

- 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. 1990. *Nonsentential constituents: A theory of grammatical structure and pragmatic interpretation*. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/pbns.2.
- Barton, Ellen. 1998. The grammar of telegraphic structures: Sentential and non-sentential derivation. *Journal of English Linguistics* 26(1). 37–67. DOI: 10.1177/007542429802600103.
- Barton, Ellen. 2006. Toward a nonsentential analysis in generative grammar. In Ljiljana Progovac, Kate Paesani, Eugenia Casielles & Ellen Barton (eds.), *Linguistik Aktuell/Linguistics Today*, vol. 93, 11–31. Amsterdam: John Benjamins. DOI: 10.1075/la.93.03bar.
- Barton, Ellen & Ljiljana Progovac. 2005. Nonsententials in Minimalism. In Reinaldo Elugardo & Robert J. Stainton (eds.), *Ellipsis and Nonsentential Speech*, 71–93. Dordrecht: Springer Netherlands. DOI: 10.1007/1-4020-2301-4\_4.
- 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.
- Baumann, Stefan, Johannes Becker, Martine Grice & Doris Mücke. 2007. Tonal and articulatory marking of focus in German. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1029–1032. Saarbrücken.
- Baumann, Stefan, Martine Grice & Susanne Steindamm. 2006. Prosodic marking of focus domains: Categorical or gradient? In *Proceedings of Speech Prosody 2006*, 301–304.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. DOI: 10.1016/j.jml.2008.06.003.
- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2). 1001–1024. DOI: 10.1121/1.1534836.
- Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204. DOI: 10.1017/S004740450001037X.
- Benincà, Paola & Cecilia Poletto. 2004. Topic, focus and V2: Defining the CP sublayers. In Luigi Rizzi (ed.), *The structure of CP and IP: The cartography of syntactic structures*, vol. 2, 52–75. Oxford: Oxford University Press.

- Benz, Anton & Robert van Rooij. 2007. Optimal assertions, and what they implicate: A uniform game theoretic approach. *Topoi* 26(1). 63–78. DOI: 10.1007/s11245-006-9007-3.
- Bergen, Leon & Noah D. Goodman. 2015. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science* 7(2). 336–350. DOI: 10.1111/tops.12144.
- Bicknell, Klinton, Jeffrey L. Elman, Mary Hare, Ken McRae & Marta Kutas. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language* 63(4). 489–505. DOI: 10.1016/j.jml.2010.08.004.
- Bildhauer, Felix. 2011. Mehrfache Vorfelddbesetzung und Informationsstruktur: Eine Bestandsaufnahme. *Deutsche Sprache* 4(11). 362–379.
- Bock, Kathryn. 1987. An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language* 26(2). 119–137. DOI: 10.1016/0749-596X(87)90120-3.
- Bogal-Allbritten, Elizabeth. 2013. Modification of DPs by epistemic modal adverbs. In Maria Aloni, Michael Franke & Floris Roelofsen (eds.), *Proceedings of the 19th Amsterdam Colloquium 2013*, 51–58.
- Bösković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59. 1–45. DOI: 10.1111/j.1467-9582.2005.00118.x.
- Botvinick, Matthew & David C. Plaut. 2004. Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review* 111(2). 395–429. DOI: 10.1037/0033-295X.111.2.395.
- Bower, Gordon H., John B. Black & Terrence J. Turner. 1979. Scripts in memory for text. *Cognitive Psychology* 11(2). 177–220. DOI: 10.1016/0010-0285(79)90009-4.
- Brandner, Ellen. 2004. Head-movement in minimalism, and V2 as FORCE-marking. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*. Berlin & Boston: de Gruyter. DOI: 10.1515/9783110912111.97.
- Brandt, Erika, Frank Zimmerer, Bistra Andreeva & Bernd Möbius. 2017. Mel-cepstral distortion of German vowels in different information density contexts. In *Proceedings of Interspeech 2017*, 2993–2997. ISCA. DOI: 10.21437/Interspeech.2017-838.
- Brandt, Erika, Frank Zimmerer, Bistra Andreeva & Bernd Möbius. 2018. Impact of prosodic structure and information density on dynamic formant trajectories in German. In *9th International Conference on Speech Prosody 2018*, 119–123. ISCA. DOI: 10.21437/SpeechProsody.2018-24.
- Breheny, Richard, Napoleon Katsos & John Williams. 2006. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100(3). 434–463. DOI: 10.1016/j.cognition.2005.07.003.

## References

- Brothers, Trevor & Gina Kuperberg. 2019. *Linear (not logarithmic) effects of lexical predictability*. Poster presented at CUNY 2019. Boulder, CO.
- Büring, Daniel. 1997. *The meaning of topic and focus: The 59th Street Bridge accent*. London: Routledge.
- Burnard, Lou. 2000. Reference guide for the British National Corpus (World Edition). Oxford University Computing Services.
- Camblin, C. Christine, Peter C. Gordon & Tamara Y. Swaab. 2007. The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language* 56(1). 103–128. DOI: 10.1016/j.jml.2006.07.005.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and topic*. New York: Academic Press.
- Chambers, Nathanael & Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL '08: HLT*, 789–797. Columbus, OH.
- Chambers, Nathanael & Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *ACL '09: Proceedings of the Joint Conference of the 47th annual meeting of the ACL and the 4th International Joint Conference on Natural Language Processing*, vol. 2, 602–610. Suntec, Singapore: Association for Computational Linguistics. DOI: 10.3115/1690219.1690231.
- Chevallier, Coralie, Ira Noveck, Tatjana Nazir, Lewis Bott, Valentina Lanzetti & Dan Sperber. 2008. Making disjunctions exclusive. *The Quarterly Journal of Experimental Psychology* 61(11). 1741–1760. DOI: 10.1080/17470210701712960.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1973. Conditions on transformations. In Steven Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart and Winston.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.
- Christensen, Rune Haubo B. 2015. *ordinal: Regression models for ordinal data*. R package version 2015.6-28. <http://cran.r-project.org/package=ordinal/>.
- Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. In *Annual meeting of the Berkeley Linguistics Society*, vol. 31, 73–91. DOI: 10.3765/bls.v31i1.896.
- Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1994. On the evidence for partial N-movement in the Romance DP. In Guglielmo Cinque, Jan Koster, Jean-Yves Pollock, Luigi Rizzi & Raffaella Zanuttini (eds.), *Paths towards universal grammar: Studies in honor of Richard S. Kayne*, 85–110. Washington, D.C.: Georgetown University Press.

- Citko, Barbara. 2011. Small clauses. *Language and Linguistics Compass* 5/10. 748–763. DOI: 10.1111/j.1749-818x.2011.00312.x.
- Cooper, Richard & Tim Shallice. 2000. Contention scheduling and the control of routine activities. *Cognitive Neuropsychology* 17(4). 297–338. DOI: 10.1080/026432900380427.
- Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. New York: Oxford University Press. DOI: 10.1093/acprof:oso.
- Delogu, Francesca, Matthew W. Crocker & Heiner Drenhaus. 2017. Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs. *Cognition* 161. 46–59. DOI: 10.1016/j.cognition.2016.12.017.
- Delogu, Francesca, Heiner Drenhaus & Matthew W. Crocker. 2018. On the predictability of event boundaries in discourse: An ERP investigation. *Memory & Cognition* 46(2). 315–325. DOI: 10.3758/s13421-017-0766-4.
- Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109. 193–210.
- Demberg, Vera, Asad B. Sayeed, Philip J. Gorinski & Nikolaos Engonopoulos. 2012. Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of EMNLP-CoNLL 2012*.
- den Besten, Hans. 1989. On the interaction of root transformations and lexical deletive rules. *Studies in West Germanic Syntax* 20. 14–100.
- den Dikken, Marcel. 2013. *On the syntax of locative and directional adpositional phrases*. Manuscript, CUNY.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2019. *BERT: Pre-training of deep bidirectional transformers for language understanding*. <http://arxiv.org/abs/1810.04805>.
- Di Meola, Claudio. 2003. Grammaticalization of postpositions in German. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Current issues in linguistic theory*, vol. 243, 203–222. Amsterdam: John Benjamins. DOI: 10.1075/cilt.243.14dim.
- Dixon, Robert M. W. 1977. Where have all the adjectives gone? *Studies in Language* 1(1). 19–80.
- Döring, Sandra. 2016. *A morphosyntactic analysis of German außer 'except'*. Rencontres d'Automne de Linguistique formelle : Langage, Langues et Cognition.
- Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt am Main: Diesterweg.
- Ehrlich, Susan F. & Keith Rayner. 1981. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior* 20(6). 641–655. DOI: 10.1016/S0022-5371(81)90220-6.

## References

- Eisenberg, Peter. 1999. *Grundriß der deutschen Grammatik*. Vol. 2: Der Satz. Stuttgart: Metzler.
- Eldridge, R C. 1911. *Six thousand common English words*. Buffalo: The Clement Press.
- Engonopoulos, Nikolaos, Asad Sayeed & Vera Demberg. 2013. Language and cognitive load in a dual task environment. In M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth (eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society*, 2249–2254. Berlin: Cognitive Science Society.
- Fanselow, Gisbert. 1993. Die Rückkehr der Basisgenerierer. *Groninger Arbeiten zur Germanistischen Linguistik* 36. 1–74.
- Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3). DOI: 10.1515/TL.2007.020.
- 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. 1983. Ist die SVO-Wortfolge die natürlichste? *Papiere zur Linguistik* 29. 23–32.
- Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27. 517–556.
- Fenk-Oczlon, Gertraud. 1990. Ökonomieprinzipien in Kognition und Kommunikation. In N Boretzky, W Enninger & T Stolz (eds.), *Spielarten der Natürlichkeit – Spielarten der Ökonomie*, 37–50. Bochum: Brockmayer.
- Fernández, Raquel & Jonathan Ginzburg. 2002. Non-sentential utterances: A corpus study. *TAL* 43(2). 13–42.
- 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.
- Féry, Caroline & Shinichiro Ishihara. 2009. How focus and givenness shape prosody. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 36–63. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199570959.003.0003.
- Fortin, Catherine. 2007. Some (not all) nonsententials are only a phase. *Lingua* 117(1). 67–94. DOI: 10.1016/j.lingua.2005.09.004.
- Fox, Danny & Howard Lasnik. 2003. Successive-cyclic movement and island repair: The difference between sluicing and VP-Ellipsis. *Linguistic Inquiry* 34(1). 143–154. DOI: 10.1162/002438903763255959.
- Frank, Austin F & T. Florian Jaeger. 2008. Speaking rationally: Uniform Information Density as an optimal strategy for language production. *Proceedings of the annual meeting of the Cognitive Science Society* 30. 939–944.



- Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998. DOI: 10.1126/science.1218633.
- 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.
- Franke, Michael. 2009. *Signal to act: Game theory in pragmatics*. Universiteit van Amsterdam. (PhD thesis).
- Frazier, Lyn & Charles Clifton. 2001. Parsing coordinates and ellipsis: Copy alpha. *Syntax* 4. 1–22. DOI: 10.1111/1467-9612.00034.
- 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. Göteborg: Acta Universitatis Gothoburgensis.
- Gengel, Kirsten. 2007. *Focus and ellipsis: A generative analysis of pseudogapping and other elliptical structures*. Universität Stuttgart. (PhD thesis).
- Genzel, Dmitriy & Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL)*, 199–206. Philadelphia.
- 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*, vol. 10, 65–72. Sapporo: Association for Computational Linguistics. DOI: 10.3115/1119355.1119364.
- Gibson, Edward, Steve Piantadosi & Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5(8). 509–524. DOI: 10.1111/j.1749-818X.2011.00295.x.
- Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations: The form, meaning, and use of English interrogatives*. Stanford, California: CSLI Publications.
- Goodman, Noah D. & Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5(1). 173–184. DOI: 10.1111/tops.12007.
- Gotzner, Nicole & Anton Benz. 2018. The Best Response Paradigm: A new approach to test implicatures of complex sentences. *Frontiers in Communication* 2. DOI: 10.3389/fcomm.2017.00021.
- Grave, Edouard, Moustapha Cisse & Armand Joulin. 2017. Unbounded cache model for online language modeling with open vocabulary. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (eds.),

## References

- Advances in neural information processing systems* 30, 6042–6052. Red Hook, NY: Curran Associates.
- Grice, H. Paul. 1975. Logic and conversation. In P. Cole & J.L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.
- Griffiths, James. 2019. A Q-based approach to clausal ellipsis: Deriving the preposition stranding and island sensitivity generalisations without movement. *Glossa* 4(1). 12. DOI: 10.5334/gjgl.653.
- Griffiths, James, Güliz Güneş & Anikó Lipták. 2018. *A minimalist approach to reprise fragments*. Presentation at DGfS 2018. Stuttgart.
- Grimshaw, Jane. 2000. Locality and extended projection. In Peter Coopmans, Martin B. H. Everaert & Jane Grimshaw (eds.), *Lexical specification and insertion* (Current Issues in Linguistic Theory 197), 115–133. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/cilt.197.07gri.
- Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary & Michael K. Tanenhaus. 2010. “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116(1). 42–55. DOI: 10.1016/j.cognition.2010.03.014.
- Gussenhoven, Carlos. 1983. *A semantic analysis of the nuclear tones of English*. Bloomington, IN: Indiana University Linguistics Club.
- Haider, Hubert. 2000. Adverb placement: Convergence of structure and licensing. *Theoretical Linguistics* 26. 95–134. DOI: 10.1515/thli.2000.26.1-2.95.
- Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL (Vol. 2)*, 159–166. DOI: 10.3115/1073336.1073357.
- Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30(4). 643–672. DOI: 10.1207/s15516709cog0000\_64.
- Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass* 10(9). 397–412. DOI: 10.1111/lnc3.12196.
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In K. L. Hale & S. J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT.
- Hare, Mary, Michael Jones, Caroline Thomson, Sarah Kelly & Ken McRae. 2009. Activating event knowledge. *Cognition* 111(2). 151–167. DOI: 10.1016/j.cognition.2009.01.009.
- Häuser, Katja I., Vera Demberg & Jutta Kray. 2019. Effects of aging and dual-task demands on the comprehension of less expected sentence continuations: Evidence from pupillometry. *Frontiers in Psychology* 10. DOI: 10.3389/fpsyg.2019.00709.
- Hawkins, John A. 2004. *A performance theory of order and constituency*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511554285.

- Heck, Fabian. 2008. *On pied-piping: Wh-movement and beyond*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110211467.
- Hegarty, Michael Vincent. 1992. *Adjunct extraction and chain configuration*. Cambridge, MA: MIT. (PhD thesis).
- Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology* 58(3). 376–415. DOI: 10.1016/j.cogpsych.2008.09.001.
- Iyer, Rukmini M. & Mari Ostendorf. 1996. Modeling long distance dependence in language: Topic mixtures vs. dynamic cache models. In *Proceedings of the fourth International Conference on Spoken Language Processing, ICSLP '96*, vol. 1, 236–239. Philadelphia, PA, USA: IEEE. DOI: 10.1109/ICSLP.1996.607085.
- Jacobs, Joachim. 1986. The syntax of focus and adverbials in German. In Werner Abraham & Sjaak de Meij (eds.), *Topic, focus, and configurationality: Papers from the 6th Groningen Grammar Talks, Groningen, 1984*, 103–127. Amsterdam & Philadelphia: John Benjamins.
- 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.
- Jaeger, T. Florian & Esteban Buz. 2017. Signal reduction and linguistic encoding. In Eva M. Fernández & Helen Smith Cairns (eds.), *The Handbook of Psycholinguistics*, 38–81. Hoboken, NJ, USA: John Wiley & Sons. DOI: 10.1002/9781118829516.ch3.
- Jäger, Gerhard. 2012. Game theory in semantics and pragmatics. In Claudia Maienborn, Paul Portner & Klaus von Stechow (eds.), *Semantics: An international handbook of natural language meaning*, vol. 3, 2487–2516. Berlin: de Gruyter. DOI: 10.1515/9783110253382.2487.
- Jans, Bram, Steven Bethard, Ivan Vulic & Marie Francine Moens. 2012. Skip n-grams and ranking functions for predicting script events. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 336–344. Avignon: Association for Computational Linguistics.
- Junqua, Jean-Claude. 1994. A duration study of speech vowels produced in noise. In *Proc. 3rd International Conference on Spoken Language Processing (ICSLP 1994)*, 419–422. Yokohama.
- Junqua, Jean-Claude. 1996. The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication* 20(1-2). 13–22. DOI: 10.1016/S0167-6393(96)00041-6.
- Kaeding, Friedrich Wilhelm. 1897. *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz bei Berlin: Author's edition.
- Kayne, Richard S. 1981. ECP extensions. *Linguistic Inquiry* 12(1). 93–133.

## References

- Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Stanford, California: CSLI.
- Khandelwal, Urvashi, He He, Peng Qi & Dan Jurafsky. 2018. Sharp nearby, fuzzy far away: How neural language models use context. In *Proceedings of the 56th annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 284–294. Melbourne, Australia: Association for Computational Linguistics. DOI: 10.18653/v1/P18-1027.
- Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in Linguistics*, 143–173. Den Haag: Mouton.
- Klein, Dan & Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st meeting of the Association for Computational Linguistics*, 423–430. Sapporo: Association for Computational Linguistics. DOI: 10.3115/1075096.1075150.
- Klein, Wolfgang. 1993. Ellipse. In Jochen Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Venneman (eds.), *Syntax: An international handbook of contemporary research*, 763–799. Berlin & New York: de Gruyter.
- Konieczny, Lars & Philipp Döring. 2003. Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In Peter Slezak (ed.), *Proceedings of the 4th International Conference on Cognitive Science*, 330–335. University of New South Wales, Sydney, Australia.
- Kravtchenko, Ekaterina. 2014. Predictability and syntactic production: Evidence from subject omission in Russian. In *Proceedings of the annual meeting of the Cognitive Science Society*, vol. 36, 785–790.
- Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25. 209–257.
- Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *Interdisciplinary studies on information structure*, vol. 6, 13–56. Potsdam: Universitätsverlag Potsdam.
- Kuperberg, Gina, Trevor Brothers & Edward W. Wlotko. 2020. A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience* 32(1). 12–35. DOI: 10.1162/jocn\_a\_01465.
- Kuperman, Victor & Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66(4). 588–611. DOI: 10.1016/j.jml.2012.04.003.
- 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.

- Lasnik, Howard. 2006. *On ellipsis: Is material that is phonetically absent but semantically present present or absent syntactically?* 22nd Scandinavian Conference of Linguistics. Aalborg, Denmark.
- Lasnik, Howard. 2015. *Islands and ellipsis: A reexamination*. Grasping ellipsis: Its syntax, semantics, acquisition and processing. University of Campinas, Brasil.
- Lemke, Robin. 2017. Sentential or not?: An experimental study on the syntax of fragments. In Sam Featherston, Robin Hörnig, Reinhild Steinberg, Birgit Umbreit & Jenni Wallis (eds.), *Proceedings of Linguistic Evidence 2016*. University of Tübingen, online publication system.
- Lemke, Robin. 2020. *Experimental investigations on the syntax and usage of fragments*. Saarbrücken: Saarland University. (PhD thesis).
- Lemke, Robin. Forthcoming. Satzäquivalente: Syntax oder Pragmatik? In Rita Finkbeiner & Robert Külpmann (eds.), *Selbständigkeit von Sätzen und Satzwertigkeit von Äußerungen* (Linguistische Berichte Sonderheft). Hamburg: Buske.
- Lemke, Robin, Eva Horch & Ingo Reich. 2017. Optimal encoding!: Information theory constrains article omission in newspaper headlines. In *Proceedings of the 15th Conference of the european Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 131–135.
- Lemke, Robin, Ingo Reich, Lisa Schäfer & Heiner Drenhaus. 2021. Predictable words are more likely to be omitted in fragments: Evidence from production data. *Frontiers in Psychology* 12. DOI: doi:10.3389/fpsyg.2021.662125.
- Lemke, Robin, Lisa Schäfer, Heiner Drenhaus & Ingo Reich. 2020. Script knowledge constrains ellipses in fragments: Evidence from production data and language modeling. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2020*, vol. 3, 441–444. New Orleans, LA. DOI: 10.7275/mpby-zr74.
- Lemke, Robin, Lisa Schäfer & Ingo Reich. 2021. Modeling the predictive potential of extralinguistic context with script knowledge: The case of fragments. *PLOS ONE* 16. 1–20. DOI: 10.1371/journal.pone.0246255.
- Lemke, Robin, Lisa Schäfer & Ingo Reich. Forthcoming. Can identity conditions on ellipsis be explained by processing principles? In Sam Featherston, Robin Hörnig, Andreas Konietzko & Sophie von Wietersheim (eds.), *Proceedings of Linguistic Evidence 2020*. University of Tübingen, online publication system.
- Levelt, Willem J. M. & Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14(1). 78–106. DOI: 10.1016/0010-0285(82)90005-6.
- Levinson, Stephen. 1983. *Pragmatics*. London: Cambridge University Press.
- Levinson, Stephen. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge: MIT Press. DOI: 10.7551/mitpress/5526.001.0001.

## References

- Levy, Roger. 2005. *Probabilistic models of word order and syntactic discontinuity*. Stanford: Stanford University. (PhD thesis).
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.
- Levy, Roger. 2018. *Using R formulae to test for main effects in the presence of higher-order interactions*. <https://arxiv.org/pdf/1405.2094.pdf>.
- Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hoffman (eds.), *Advances in neural information processing*, vol. 19, 849–856. Cambridge, MA: MIT Press.
- Libben, Gary & Jürgen Tesak. 1994. *English corpus of telegraphese*. European Telegram Project.
- LimeSurvey GmbH. 2012. LimeSurvey: An open source survey tool. Hamburg.
- Lombard, Étienne. 1911. Le signe de l'élévation de la voix. *Annales des maladies de l'oreille, du larynx, du nez et du pharynx* 37. 101–119.
- Loper, Edward & Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *ETMTNLP '02 Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, vol. 1, 63–70. Philadelphia, PA: Association for Computational Linguistics.
- Mahowald, Kyle, Isabelle Dautriche, Edward Gibson & Steven T. Piantadosi. 2018. Word forms are structured for efficient use. *Cognitive Science*. 1–19. DOI: 10.1111/cogs.12689.
- 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.
- Malisz, Zofia, Erika Brandt, Bernd Möbius, Yoon Mi Oh & Bistra Andreeva. 2018. Dimensions of segmental variability: Interaction of prosody and surprisal in six languages. *Frontiers in Communication* 3. 1–18. DOI: 10.3389/fcomm.2018.00025.
- Manshadi, Mehdi, Reid Swanson & Andrew S. Gordon. 2008. Learning a probabilistic model of event sequences from Internet weblog stories. In *Proceedings of the Twenty-First International FLAIRS Conference*, 159–164.
- Martin, James E. 1969. Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior* 8(6). 679–704.
- Matsuki, Kazunaga, Tracy Chow, Mary Hare, Jeffrey L. Elman, Christoph Scheepers & Ken McRae. 2011. Event-based plausibility immediately influences online language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(4). 913–934. DOI: 10.1037/a0022964.



- McDonald, Scott A. & Richard C. Shillcock. 2003. Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14(6). 648–652. DOI: 10.1046/j.0956-7976.2003.psci\_1480.x.
- McKoon, Gail & Roger Ratcliff. 1986. Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12(1). 82–91.
- Merchant, Jason. 2001. *The syntax of silence. Sluicing, islands, and the theory of ellipsis*. New York: Oxford University Press.
- Merchant, Jason. 2004a. Fragments and ellipsis. *Linguistics and Philosophy* 27(6). 661–738. DOI: 10.1007/s10988-005-7378-3.
- Merchant, Jason. 2004b. Resumptivity and non-movement. *Studies in Greek Linguistics* 24. 471–481.
- Merchant, Jason. 2008. Variable island repair under ellipsis. In Kyle Johnson (ed.), *Topics in ellipsis*, 132–153. New York: Cambridge University Press. DOI: 10.1017/CBO9780511487033.006.
- Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1). 77–108. DOI: 10.1162/LING\_a\_00120.
- Merchant, Jason, Lyn Frazier, Thomas Weskott & Charles Clifton. 2013. Fragment answers to questions. A case of inaudible syntax. In Laurence Goldstein (ed.), *Brevity*, 21–35. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199664986.003.0002.
- Metusalem, Ross, Marta Kutas, Thomas P. Urbach, Mary Hare, Ken McRae & Jeffrey L. Elman. 2012. Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language* 66(4). 545–567. DOI: 10.1016/j.jml.2012.01.001.
- Minsky, Marvin. 1974. *A framework for representing knowledge*. MIT-AI Laboratory Memo 306, June, 1974.
- 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.
- Morgan, Jerry. 1989. Sentence fragments revisited. In B. Music, R. Graczyk & C. Wiltshire (eds.), *CLS Parasession on language in context*, 228–241. Chicago, IL: Chicago Linguistic Society.
- Müller, Gereon. 2004. Verb-second as vP-first. *Journal of Comparative Germanic Linguistics* 7. 179–234.
- Müller, Gereon. 2011. *Constraints on displacement: A phase-based approach*. Amsterdam: John Benjamins. DOI: 10.1075/lfab.7.

## References

- Müller, Stefan. 2002. Syntax or morphology: German particle verbs revisited. In Nicole Dehé, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.), *Verb-particle explorations*, 119–139. Berlin & Boston: de Gruyter. DOI: 10.1515/9783110902341.119.
- Müller, Stefan. 2003. Mehrfache Vorfeldbesetzung. *Deutsche Sprache* 31(1). 29–62.
- Müller, Stefan. 2005. Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung. *Linguistische Berichte* 203. 297–330.
- Murphy, Andrew. 2018. Pronominal inflection and NP ellipsis in German. *The Journal of Comparative Germanic Linguistics* 21(3). 327–379. DOI: 10.1007/s10828-018-9102-z.
- Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79(4). 682–707. DOI: 10.1353/lan.2003.0260.
- Norcliffe, Elisabeth & T. Florian Jaeger. 2016. Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition* 8(2). 167–205. DOI: 10.1017/langcog.2014.39.
- Nuthmann, Antje & Elke van der Meer. 2005. Time's arrow and pupillary response. *Psychophysiology* 42(3). 306–317. DOI: 10.1111/j.1469-8986.2005.00291.x.
- Nykiel, Joanna. 2013. Clefts and preposition omission under sluicing. *Lingua* 123. 74–117. DOI: 10.1016/j.lingua.2012.10.011.
- Nykiel, Joanna. 2014. *The ellipsis alternation: Remnants with and without prepositions*. Manuscript.
- Nykiel, Joanna. 2017. Preposition stranding and ellipsis alternation. *English Language and Linguistics* 21(1). 27–45. DOI: 10.1017/S1360674315000477.
- Ott, Dennis & Volker Struckmeier. 2016. Deletion in clausal ellipsis: Remnants in the middle field. *UPenn Working Papers in Linguistics* 22(1). 225–234. <https://repository.upenn.edu/pwpl/vol22/iss1/25>.
- Otten, Marte & Jos J.A. Van Berkum. 2007. What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research* 1153. 166–177. DOI: 10.1016/j.brainres.2007.03.058.
- Oualil, Youssef, Clayton Greenberg, Mittul Singh & Dietrich Klakow. 2017. *Sequential recurrent neural networks for language modeling*. <http://arxiv.org/abs/1703.08068>.
- Oualil, Youssef, Mittul Singh, Clayton Greenberg & Dietrich Klakow. 2016. Long-short range context neural networks for language modeling. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1473–1481. Austin, TX: Association for Computational Linguistics. DOI: 10.18653/v1/D16-1154.



- Pate, John K & Sharon Goldwater. 2015. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language* 78. 1–17. DOI: 10.1016/j.jml.2014.10.003.
- Philippova, Tatiana. 2014. P-omission under sluicing, [P clitic] and the nature of P-stranding. In *Proceedings of ConSOLE XXII*, 133–155.
- Piantadosi, Steven, Harry J. Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9). 3526–3529. DOI: 10.1073/pnas.1012551108.
- Pichotta, Karl & Raymond Mooney. 2014. Statistical script learning with multi-argument events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 220–229. Gothenburg, Sweden: Association for Computational Linguistics. DOI: 10.3115/v1/E14-1024.
- Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry L. Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press. DOI: 10.7916/D8KD24FP.
- Poppels, Till & Andrew Kehler. 2019. *Ellipsis and the QuD: Sluicing with nominal antecedents*. Poster presented at CUNY 2019. Boulder, CO.
- Progovac, Ljiljana. 2006. The syntax of nonsententials: Small clauses and phrases at the root. In Ljiljana Progovac, Kate Paesani, Eugenia Casielles & Ellen Barton (eds.), *Linguistik Aktuell/Linguistics Today*, vol. 93, 33–71. Amsterdam: John Benjamins. DOI: 10.1075/la.93.04pro.
- Progovac, Ljiljana, Kate Paesani, Eugenia Casielles & Ellen Barton. 2006. Epilogue: Wherefrom and whereto? In Ljiljana Progovac, Kate Paesani, Eugenia Casielles & Ellen Barton (eds.), *The syntax of nonsententials: Multidisciplinary perspectives*, 323–354. Amsterdam: John Benjamins. DOI: 10.1075/la.93.15pro.
- Pullum, Geoffrey K. & Rodney Huddleston. 2002. Prepositions and preposition phrases. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 1st edn., 597–662. Cambridge: Cambridge University Press. DOI: 10.1017/9781316423530.008.
- Pullum, Geoffrey K. & Arnold M. Zwicky. 1986. Phonological resolution of syntactic feature conflict. *Language* 64. 751–773.
- R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria. R Foundation for Statistical Computing.
- Rabs, Elisabeth, Heiner Drenhaus, Francesca Delogu & Matthew W. Crocker. 2017. *The influence of script knowledge on language processing: Evidence from ERPs*. AMLaP 2017. Lancaster (UK).
- Regneri, Micaela. 2013. *Event structures in knowledge, pictures and text*. Saarland University. (PhD thesis).

## References

- Reich, Ingo. 2002a. Pied piping and the syntax and semantics of complex *wh*-phrases. In Simon Mauck & Jenny Mittelstaedt (eds.), *Georgetown University Working Papers in Theoretical Linguistics*, vol. 2, 263–286.
- Reich, Ingo. 2002b. Question/answer congruence and the semantics of *wh*-phrases. *Theoretical Linguistics* 28(1). 73–94. DOI: 10.1515/thli.2002.28.1.73.
- Reich, Ingo. 2007. Toward a uniform analysis of short answers and gapping. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 467–484. Amsterdam: John Benjamins. DOI: 10.1075/la.100.25rei.
- Reich, Ingo. 2011. Ellipsis. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 1849–1874. Berlin & New York: Mouton de Gruyter. DOI: 10.1515/9783110255072.1849.
- 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.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-011-5420-8\_7.
- Roark, Brian. 2001. Probabilistic top-down parsing and language modeling. *Computational Linguistics* 27(2). 249–276. DOI: 10.1162/089120101750300526.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. Hak Yoon & A. Kathol (eds.), *Ohio State University Working Papers in Linguistics*, vol. 49. Ohio State University.
- Rodman, Robert. 1974. On left dislocation. *Paper in Linguistics* 7(3–4). 437–466. DOI: 10.1080/08351817409370381.
- Rodrigues, Cilene, Andrew Ira Nevins & Luis Vicente. 2009. Cleaving the interactions between sluicing and P-stranding. In Danièle Torck & W. Leo Wetzels (eds.), *Current issues in linguistic theory*, vol. 303, 175–198. Amsterdam: John Benjamins. DOI: 10.1075/cilt.303.11rod.
- Rohde, Hannah, Scott Seyfarth, Brady Clark, Gerhard Jaeger & Stefan Kaufmann. 2012. Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue*, 107–116.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. DOI: 10.1007/BF02342617.
- Ross, John R. 1967. *Constraints on variables in syntax*. MIT. (PhD thesis).

- Ross, John R. 1969. Guess who? In Robert Binnick, Alice Davidson & Georgia Green (eds.), *Papers from the 5th regional meeting of the Chicago Linguistic Society*, 252–286. Chicago, IL: Chicago Linguistic Society.
- Ross, John R. 1970. Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics*, 249–259. The Hague: Mouton.
- Saab, Andrés & Anikó Lipták. 2016. Movement and deletion after syntax: Licensing by inflection reconsidered. *Studia Linguistica* 70(1). 66–108. DOI: 10.1111/stul.12039.
- Sag, Ivan A. 1976. *Deletion and logical form*. Cambridge, MA: MIT. (PhD thesis).
- Sato, Yosuke. 2011. P-stranding under sluicing and repair by ellipsis: Why is Indonesian (not) special? *Journal of East Asian Linguistics* 20(4). 339–382. DOI: 10.1007/s10831-011-9082-3.
- 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. 1672. DOI: 10.3389/fpsyg.2021.661087.
- Schank, Roger & Robert Abelson. 1977. *Scripts, plans, goals, and understanding: An enquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Schank, Roger C. 1975. *Conceptual information processing*. Amsterdam: North Holland.
- Schlangen, David. 2003. *A coherence-based approach to the interpretation of non-sentential utterances in dialogue*. University of Edinburgh. (PhD thesis).
- Schütze, Carson T. 2001. On the nature of default case. *Syntax* 4(3). 205–238. DOI: 10.1111/1467-9612.00044.
- Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141–177. DOI: 10.1023/A:1008370902407.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. MIT Press.
- Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155. DOI: 10.1016/j.cognition.2014.06.013.
- Shannon, Claude E. 1948. A mathematical theory of communications. *Bell Systems Technical Journal* 27(4). 623–656.
- Shannon, Claude E. 1951. Prediction and entropy of printed English. *Bell Systems Technical Journal* 3. 50–64.

## References

- Sikos, Les, Clayton Greenberg, Heiner Drenhaus & Matthew W. Crocker. 2017. Information density of encodings: The role of syntactic variation in comprehension. In *Proceedings of the 39th annual conference of the Cognitive Science Society (CogSci 2017)*, 3168–3173.
- Sikos, Les, Noortje J. Venhuizen, Heiner Drenhaus & Matthew W. Crocker. 2019. *Reevaluating pragmatic reasoning in web-based language games*. Rational Approaches in Language Science (RAILS) 2019, Saarbrücken.
- Singh, Mittul, Clayton Greenberg & Dietrich Klakow. 2016. The custom decay language model for long range dependencies. In Petr Sojka, Aleš Horák, Ivan Kopeček & Karel Pala (eds.), *Text, speech, and dialogue*, 343–351. Cham: Springer. DOI: 10.1007/978-3-319-45510-5\_39.
- Smith, Nathaniel J. & Roger Levy. 2008. Optimal processing times in reading: A formal model and empirical investigation. In *The 30th annual meeting of the Cognitive Science Society*, 595–600. Washington, D.C.
- Smith, Nathaniel J. & Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3). 302–319. DOI: 10.1016/j.cognition.2013.02.013.
- Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and cognition*. 2nd edn. Oxford: Blackwell.
- Speyer, Augustin & Robin Lemke. 2017. Information density as a factor for variation in the embedding of relative clauses. <http://arxiv.org/abs/1705.06457>.
- Stainton, Robert. 2006. *Words and thoughts: Subsentences, ellipsis and the philosophy of language*. Oxford, New York: Oxford University Press. DOI: 10.1093/mind/fzn105.
- Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25. 701–721. DOI: 10.1023/A:1020867916902.
- Stigliano, Laura. 2018. *Syntactic identity in clausal ellipsis: Evidence from P-stranding effects in Spanish*. Poster presented at NELS 49. Ithaca, NY.
- Stigliano, Laura. 2019. *Revisiting the interaction between P-stranding and clausal ellipsis in Spanish*. Sluicing +50. Chicago, IL.
- Stolcke, Andreas. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics* 21(2). 38.
- Stolcke, Andreas. 2002. SRILM: An extensible language modeling toolkit. In *Proc. Intl. Conf. Spoken Language Processing*. Denver, Colorado.
- Stowell, Tim. 1981. *Origins of phrase structure*. MIT. (PhD thesis).

- Summers, W. Van, David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow & Michael A. Stokes. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America* 84(3). 917–928. DOI: 10.1121/1.396660.
- Szczegielniak, Adam. 2008. Islands in sluicing in Polish. In Natasha Abner & Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 404–412. Somerville, MA: Cascadilla Proceedings Project.
- Tancredi, Christopher. 1992. *Deletion, deaccenting and presupposition*. Cambridge, MA: MIT. (PhD thesis).
- Telljohann, Heike, Erhard Hinrichs & Sandra Kübler. 2004. The Tüba-D/Z treebank: Annotating German with a context-free backbone. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon, Portugal, 2229–2232. Paris: European Language Resources Association (ELRA).
- Tily, Harry, Susanne Gahl, Inbal Arnon, Neal Snider, Anubha Kothari & Joan Bresnan. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1(2). 147–165. DOI: 10.1515/LANGCOG.2009.008.
- Tily, Harry & Steven Piantadosi. 2009. Refer efficiently: Use less informative expressions for more predictable meanings. In Kees van Deemter, Albert Gatt, Roger van Gompel & Emiel Krahmer (eds.), *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*. Amsterdam.
- Tokizaki, Hisao. 2010. Stress location and the acquisition of morphosyntactic parameters. In *Proceedings of WCCFL 28*.
- Trissler, Susanne. 2000. *Syntaktische Bedingungen für w-Merkmale: Zur Bildung interrogativer w-Phrasen im Deutschen*. Universität Tübingen. (PhD thesis).
- Trissler, Susanne & Uli Lutz. 1992. Einige Überlegungen zu syntaktischen +w-Merkmalen, Interrogativsätzen und w-Phrasen im Deutschen. In *Arbeitspapiere des Sonderforschungsbereichs 340*, vol. 7, 105–178. Tübingen, Stuttgart.
- van den Broek, Paul. 1994. Comprehension and memory of narrative texts. In Morton Ann Gernsbacher (ed.), *Handbook of psycholinguistics*, 539–588. San Diego, CA: Academic Press.
- van der Meer, Elke, Reinhard Beyer, Bertram Heinze & Isolde Badel. 2002. Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(4). 770–779. DOI: 10.1037//0278-7393.28.4.770.

## References

- van Craenenbroeck, Jeroen. 2010. Invisible last resort. *Lingua* 120(7). 1714–1726. DOI: 10.1016/j.lingua.2010.01.002.
- van Kuppevelt, Jan. 1996. Inferring from topics. *Linguistics and philosophy* 19. 393–443.
- van Riemsdijk, Henk. 1978. *A case study in syntactic markedness: The binding nature of prppositional phrases*. Lisse: Peter de Ridder Press.
- van Rooy, Robert. 2004. Utility, informativity and protocols. *Journal of Philosophical Logic* 33(4). 389–419. DOI: 10.1023/B:LOGI.0000036830.62877.ee.
- Vat, Jan. 1981. Left dislocation, connectedness and reconstruction. In Elena Anagnostopoulou, Henk van Riemsdijk & Frans Zwarts (eds.), *Materials on left dislocation*, 67–92. Amsterdam: John Benjamins.
- von Stechow, Arnim. 1981. Topic, focus and local relevance. In Wolfgang Klein & Willem J. M. Levelt (eds.), *Crossing the boundaries in linguistics* (Synthese Language Library 13), 95–130. Dordrecht: Springer.
- Wanzare, Lilian D. A., Alessandra Zarcone, Stefan Thater & Manfred Pinkal. 2016. DeScript: A crowdsourced corpus for the acquisition of high-quality script knowledge. In *Proceedings of LREC 2016*, 3494–3501. Portoroz, Slovenia.
- Webelhuth, Gerd. 1992. *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press.
- Weir, Andrew. 2014a. Fragment answers and the Question under Discussion. In Jyoti Iyer & Leland Kusmer (eds.), *Proceedings of NELS 44*, 255–266.
- Weir, Andrew. 2014b. *Fragments and clausal ellipsis*. University of Massachusetts Amherst. (PhD thesis).
- Weir, Andrew. 2015. Fragment answers and exceptional movement under ellipsis: A PF-movement account. In Thuy Bui & Deniz Özyıldız (eds.), *Proceedings of NELS 45*, 175–188. University of Ghent.
- Weir, Andrew. 2018. Cointensional questions, fragment answers, and structured meanings. In Robert Trueswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde (eds.), *Proceedings of Sinn und Bedeutung 21*, 1289–1306.
- Williams, Edwin. 1975. Small clauses in English. In John Kimball (ed.), *Syntax and semantics*, 249–273. Orlando, FL: Academic Press.
- Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8. 101–139.
- Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37(1). 111–130. DOI: 10.1162/002438906775321175.
- Wulff, Stefanie. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8(2). 245–282. DOI: 10.1075/ijcl.8.2.04wul.

- Yoon, Jeong-Me. 2001. Feature percolation, movement and cross-linguistic variation in pied-piping. In *Language, Information and Computation: Proceedings of The 15th Pacific Asia Conference*, 283–292. Hong Kong.
- Zhan, Meilin, Roger Levy & Edward Gibson. 2017. *Rational inference and sentence interpretation in Mandarin Chinese*. Poster presented at CUNY 2017. Cambridge, MA.
- Zimmermann, Malte & Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121(11). 1651–1670. DOI: 10.1016/j.lingua.2011.06.002.
- Zipf, George Kingsley. 1935. *The psycho-biology of language*. Oxford: Houghton, Mifflin.
- Zondervan, Albert Jan. 2010. *Scalar implicatures or focus: An experimental approach*. Utrecht: LOT/Netherlands Graduate School of Linguistics. (PhD thesis).
- Zwarts, Joost. 2005. *The case of prepositions: Government and compositionality in German PPs*. Haifa. 21st annual meeting of the Israel Association for Theoretical Linguistics.





# Name index

- Abbott, Valerie, 180  
Abels, Klaus, 79–81, 90  
Abelson, Robert, 25, 175, 177–183, 188, 190, 191  
Aelbrecht, Loebke, 71  
Arregi, Karlos, 38  
Arregui, Ana, 24  
Asr, Fatameh Torabi, 165  
Attneave, Fred, 156  
Aylett, Matthew, 162, 166, 169, 170  
  
Balling, Laura Winther, 206  
Bar-Hillel, Yehoshua, 155  
Barr, Dale J., 57  
Barton, Ellen, 1, 2, 7–12, 26, 28–33, 35, 40, 46, 49–52, 60, 64, 70, 71, 81, 91–93, 97, 99, 112, 146, 148, 150, 244, 245  
Bates, Douglas, 57, 63, 108, 155, 197, 219  
Baumann, Stefan, 41  
Bell, Alan, 166, 170  
Bell, Allan, 64  
Benincà, Paola, 37, 38  
Benz, Anton, 233  
Bergen, Leon, 3, 7, 8, 26, 27, 33, 46, 47, 60, 70, 90, 98, 153, 249  
Bicknell, Klinton, 181  
Bildhauer, Felix, 140  
Bird, Steven, 186  
Bock, Kathryn, 172  
Bogal-Allbritten, Elizabeth, 138  
  
Bösković, Željko, 45  
Botvinick, Matthew, 180  
Bower, Gordon H., 177, 178, 180, 181, 188  
Brandner, Ellen, 136  
Brandt, Erika, 166, 170  
Breheny, Richard, 64, 234  
Bresnan, Joan, 166  
Brothers, Trevor, 155, 170, 255  
Büring, Daniel, 38, 140  
Burnard, Lou, 3  
Buz, Esteban, 172, 173, 228  
  
Camblin, C. Christine, 181  
Carnap, Rudolf, 155  
Chafe, Wallace L, 167  
Chambers, Nathanael, 184  
Charniak, Eugene, 162, 176, 223, 226, 227  
Chevallier, Coralie, 41, 64  
Chomsky, Noam, 1, 8, 10, 13, 19, 35, 41, 78, 105, 114  
Christensen, Rune Haubo B., 56, 57  
Chung, Sandra, 256  
Cinque, Guglielmo, 19, 46, 168  
Citko, Barbara, 34  
Clifton, Charles, 24  
Cooper, Richard, 180  
Culicover, Peter, 2, 70, 151  
  
Dell, Gary S., 172, 227, 228, 250  
Delogu, Francesca, 181, 255

## *Name index*

- Demberg, Vera, 165, 166, 170, 254  
den Besten, Hans, 38, 136  
den Dikken, Marcel, 92  
Devlin, Jacob, 177  
Di Meola, Claudio, 168  
Dixon, Robert M. W., 168  
Döring, Philipp, 167  
Döring, Sandra, 71  
Drach, Erich, 135
- Ehrlich, Susan F., 170  
Eisenberg, Peter, 51, 205  
Eldridge, R C, 160  
Engonopoulos, Nikolaos, 171
- Fanselow, Gisbert, 141  
Featherston, Sam, 116  
Fenk, August, 162, 169  
Fenk, Gertraud, *see* Fenk-Oczlon, Gertraud  
Fenk-Oczlon, Gertraud, 162, 167, 169, 176, 227  
Fernández, Raquel, 1–3, 7, 70, 151, 245  
Ferreira, Victor S., 172, 227, 228, 250  
Féry, Caroline, 173  
Fortin, Catherine, 7  
Fox, Danny, 42  
Frank, Austin F, 162, 164, 165, 228  
Frank, Michael C., 233  
Frank, Stefan L., 255  
Franke, Michael, 233–236, 238, 239, 251  
Frazier, Lyn, 24  
Frey, Werner, 38
- Gengel, Kirsten, 19, 21  
Genzel, Dmitriy, 162, 176, 223, 226, 227  
Gibson, Edward, 57
- Ginzburg, Jonathan, 1–3, 7, 70, 151, 245, 246  
Goldwater, Sharon, 159, 166  
Goodman, Noah D., 3, 7, 8, 26, 27, 33, 46, 47, 60, 70, 90, 98, 153, 233, 249  
Gotzner, Nicole, 233  
Grave, Edouard, 177  
Grice, H. Paul, 23  
Griffiths, James, 71, 100, 112  
Grimshaw, Jane, 78  
Grodner, Daniel J., 234  
Gussenhoven, Carlos, 41
- Haider, Hubert, 36, 138  
Hale, John, 154, 156, 169–171, 176, 177, 208–213, 227, 230, 231, 253, 255  
Halle, Morris, 10  
Hare, Mary, 181  
Häuser, Katja I., 170, 171  
Hawkins, John A, 121  
Heck, Fabian, 80, 90  
Hegarty, Michael Vincent, 117, 120  
Hirschberg, Julia, 41  
Huang, Yi Ting, 234  
Huddleston, Rodney, 45, 79, 87, 105
- Ishihara, Shinichiro, 173  
Iyer, Rukmini M., 177
- Jackendoff, Ray, 2, 70, 151  
Jacobs, Joachim, 140  
Jaeger, T. Florian, 3, 117, 154, 159, 162–165, 170–173, 176, 199, 209, 214, 226, 228, 244, 253  
Jäger, Gerhard, 233  
Jans, Bram, 184, 213  
Junqua, Jean-Claude, 170

- Jurafsky, Dan, 184
- Kaeding, Friedrich Wilhelm, 160
- Kayne, Richard S., 114
- Kehler, Andrew, 24, 256
- Keller, Frank, 170, 254
- Kelter, Stephanie, 81, 100, 102–104, 112, 113, 149, 246
- Khandelwal, Urvashi, 177
- Kiparsky, Carol, 117, 120
- Kiparsky, Paul, 117, 120
- Kizach, Johannes, 206
- Klein, Dan, 186
- Klein, Wolfgang, 5, 6, 53
- Konieczny, Lars, 167
- Kravtchenko, Ekaterina, 165, 254, 256
- Krifka, Manfred, 1, 21, 38, 82, 168
- Kuperberg, Gina, 155, 168, 170, 255
- Kuperman, Victor, 166
- Kurumada, Chigusa, 162, 165, 228
- Lasnik, Howard, 42, 78
- Lemke, Robin, v, 26, 49, 75, 76, 111, 165, 168, 173, 175, 214, 246, 256
- Levelt, Willem J. M., 81, 100, 102–104, 112, 113, 149, 246
- Levinson, Stephen, 234
- Levy, Roger, 3, 123, 154, 155, 159, 162–165, 167, 169–171, 176, 177, 199, 206, 209, 214, 226, 228, 253, 255
- Libben, Gary, 9
- Lipták, Anikó, 71
- Lombard, Étienne, 170
- Loper, Edward, 186
- Lutz, Uli, 78
- Mahowald, Kyle, 160
- Malisz, Zofia, 166, 170
- Manning, Christopher D., 186
- Manshadi, Mehdi, 184
- Marantz, Alec, 10
- Martin, James E., 168
- Matsuki, Kazunaga, 181
- McDonald, Scott A., 170
- McKoon, Gail, 181
- Merchant, Jason, 2, 6, 7, 10, 13, 15–21, 24–26, 28, 29, 31, 32, 35–47, 49–51, 54, 71–74, 76–78, 82–87, 90–92, 96, 99, 103, 106, 111, 113–120, 124, 125, 128–130, 132–138, 142, 145, 146, 149–151, 173, 228, 244, 246–248
- Metusalem, Ross, 181
- Minsky, Marvin, 178
- Mooney, Raymond, 184
- Morgan, Jerry, 1, 2, 5, 7, 13, 41, 64, 114, 117, 247
- Müller, Gereon, 36, 38, 42, 136
- Müller, Stefan, 36, 138–141, 144, 148
- Murphy, Andrew, 71
- Newmeyer, Frederick J., 153
- Norcliffe, Elisabeth, 162, 165, 228
- Nuthmann, Antje, 181
- Nykiel, Joanna, 75, 81, 86, 100, 102–105, 109, 110, 113, 149
- Onea, Edgar, 46
- Ostendorf, Mari, 177
- Ott, Dennis, 2, 15, 71, 138, 228
- Otten, Marte, 181
- Oualil, Youssef, 177
- Pate, John K, 159, 166
- Philippova, Tatiana, 81

*Name index*

- Piantadosi, Steven, 160, 166, 254, 256  
Pichotta, Karl, 184  
Pierrehumbert, Janet, 41  
Plaut, David C., 180  
Poletto, Cecilia, 37, 38  
Poppels, Till, 256  
Progovac, Ljiljana, 1, 2, 7–12, 26, 28–35, 46, 49–52, 60, 64, 70, 71, 81, 91–95, 97, 99, 112, 146–148, 150, 244, 245  
Pullum, Geoffrey K., 24, 45, 79, 87, 105  
  
Rabs, Elisabeth, 190  
Ratcliff, Roger, 181  
Rayner, Keith, 170  
Regneri, Micaela, 187  
Reich, Ingo, 2, 3, 5–7, 13–17, 22–24, 32–35, 37, 38, 46, 53, 71, 81, 100, 101, 112, 136, 146, 148–151, 153, 173, 175, 205, 228, 244, 246  
Reinhart, Tanya, 168  
Rizzi, Luigi, 1, 12, 19, 37, 46  
Roark, Brian, 177  
Roberts, Craig, 13  
Rodman, Robert, 39  
Rodrigues, Cilene, 32, 74, 75, 112  
Rohde, Hannah, 233  
Rooth, Mats, 14, 174  
Ross, John R., 45, 256  
  
Saab, Andrés, 71  
Sag, Ivan A., 2, 7, 70, 151, 245, 246, 256  
Sato, Yosuke, 71, 75, 78, 81  
Schäfer, Lisa, 175, 198, 256  
Schank, Roger, 25, 175, 177–183, 188, 190, 191  
Schank, Roger C, 178  
Schlangen, David, 2, 7, 245  
Schütze, Carson T., 10, 11  
Schwarzschild, Roger, 18  
Selkirk, Elisabeth, 15  
Seyfarth, Scott, 166  
Shallice, Tim, 180  
Shannon, Claude E., 153–161, 166, 169–171, 217, 250, 254  
Shillcock, Richard C., 170  
Sikos, Les, 168, 233  
Singh, Mittul, 177  
Smith, Nathaniel J., 155, 170, 171, 255  
Snedeker, Jesse, 234  
Sperber, Dan, 23, 25, 64, 153, 234  
Speyer, Augustin, 168  
Stainton, Robert, 1, 2, 8, 12, 40, 151  
Stalnaker, Robert, 154  
Stigliano, Laura, 75  
Stolcke, Andreas, 177, 188, 208  
Stowell, Tim, 33, 34, 41, 114, 116, 117, 247  
Struckmeier, Volker, 2, 15, 71, 138, 228  
Stuhlmüller, Andreas, 233  
Summers, W. Van, 170  
Szczegielniak, Adam, 75, 76, 112  
  
Tancredi, Christopher, 173  
Telljohann, Heike, 120  
Tesak, Jürgen, 9  
Tily, Harry, 166, 254, 256  
Tokizaki, Hisao, 81  
Trissler, Susanne, 78  
Turk, Alice, 162, 166, 169, 170  
  
Van Berkum, Jos J.A., 181  
van Craenenbroeck, Jeroen, 32, 75  
van den Broek, Paul, 181

van der Meer, Elke, 181  
van Kuppevelt, Jan, 234  
van Riemsdijk, Henk, 73, 79, 105, 111  
van Rooij, Robert, 233  
Vat, Jan, 39  
von Stechow, Arnim, 100, 101  
  
Wanzare, Lilian D. A., 65, 175–178,  
183, 195, 229, 252  
Webelhuth, Gerd, 41, 113–115, 118,  
124, 247  
Weir, Andrew, 2, 19–22, 24, 37, 38,  
40–42, 44–47, 71, 74, 77, 80,  
90, 101, 102, 111, 112, 115, 137,  
146, 150, 173  
Williams, Edwin, 33, 256  
Wilson, Deirdre, 23, 25, 64, 153, 234  
Woods, Rebecca, 133  
Woolford, Ellen, 51  
Wulff, Stefanie, 168  
  
Yoon, Jeong-Me, 78  
  
Zhan, Meilin, 110  
Zimmermann, Malte, 46  
Zipf, George Kingsley, 159, 160  
Zondervan, Albert Jan, 41  
Zwarts, Joost, 62, 93, 99  
Zwicky, Arnold M., 24



# Language index

Arabic, 72<sup>15</sup>

Basque, 72<sup>15</sup>

Bosnian/Croatian/Serbian, 11, 12, 29,  
30, 31<sup>18</sup>, 60, 70, 72<sup>15</sup>, 91, 92,  
93<sup>31</sup>, 94, 147, 245

Bulgarian, 73<sup>17</sup>

Chinese, 160

Czech, 72<sup>15</sup>, 73<sup>17</sup>

Danish, 72<sup>15</sup>, 73<sup>17</sup>, 206<sup>23</sup>

Dutch, 11, 31, 73<sup>16</sup>, 102, 103

English, 5, 9–11, 11<sup>3</sup>, 17<sup>9</sup>, 19<sup>11</sup>, 21, 21<sup>13</sup>,  
28, 30–32, 32<sup>19</sup>, 34, 35, 39,  
41, 42, 44, 45, 50, 51, 70, 72,  
72<sup>15</sup>, 73, 73<sup>17</sup>, 74, 75, 77–  
81, 81<sup>23</sup>, 82, 83, 86, 87, 87<sup>26</sup>,  
87<sup>27</sup>, 89–92, 98, 99, 99<sup>33</sup>,  
100, 102–106, 110–113, 116–  
118, 125, 129, 130, 132–134,  
137, 149, 150, 160, 165, 172,  
183, 186, 187, 200<sup>21</sup>, 204, 227,  
245–247

French, 31, 32

Frisian, 72<sup>15</sup>

German, 6, 11, 11<sup>3</sup>, 15<sup>8</sup>, 17<sup>9</sup>, 19<sup>11</sup>, 26,  
28–30, 30<sup>17</sup>, 31, 31<sup>18</sup>, 34–36,  
38, 39, 41, 47, 50–52, 58–  
60, 63, 67, 69, 70, 72, 72<sup>15</sup>,

73, 73<sup>16</sup>, 73<sup>17</sup>, 74–76, 76<sup>19</sup>,  
76<sup>20</sup>, 77–80, 80<sup>22</sup>, 81, 81<sup>23</sup>,  
82, 82<sup>24</sup>, 83<sup>25</sup>, 84–87, 89–  
92, 93<sup>31</sup>, 94, 97–99, 99<sup>33</sup>,  
100, 102, 103, 111–113, 117–  
121, 124, 125, 129, 130<sup>50</sup>, 132–  
136, 136<sup>51</sup>, 137–139, 142, 144,  
146–150, 167, 168, 192, 193,  
200, 200<sup>21</sup>, 205, 210, 227,  
244–249, 256

Greek, 31, 72<sup>15</sup>, 73<sup>17</sup>

Hebrew, 72<sup>15</sup>, 73<sup>17</sup>, 92<sup>29</sup>

Icelandic, 72<sup>15</sup>, 73<sup>17</sup>, 92<sup>29</sup>

Irish, 11, 72<sup>15</sup>

Korean, 64, 245

Norwegian, 11, 72<sup>15</sup>, 73<sup>17</sup>

Polish, 72<sup>15</sup>, 75, 75<sup>18</sup>, 76<sup>19</sup>

Portuguese, 32<sup>19</sup>

Russian, 11, 72<sup>15</sup>, 73<sup>17</sup>, 165, 256

Slovene, 72<sup>15</sup>

Spanish, 32<sup>19</sup>, 38<sup>27</sup>, 74, 75<sup>18</sup>

Swedish, 72<sup>15</sup>, 73<sup>17</sup>

Yiddish, 72<sup>15</sup>, 73<sup>17</sup>





# Subject index

- Acceptability rating task, 2, 52, 53,  
55–59, 61, 62, 65, 67, 69,  
76<sup>20</sup>, 77, 82–89, 91, 94–98,  
112, 115, 116, 119, 121–124,  
126, 126<sup>49</sup>, 130, 132–134, 140,  
142, 143, 145, 148, 153, 176,  
189, 192, 195, 196<sup>17</sup>, 197, 200,  
201, 203, 224–226, 229, 230,  
244, 247–249, 256
- Accusative case, 9–11, 11<sup>3</sup>, 28–30,  
30<sup>17</sup>, 31, 31<sup>18</sup>, 32<sup>19</sup>, 34, 51–55,  
56<sup>5</sup>, 57–62, 62<sup>9</sup>, 63, 65, 65<sup>12</sup>,  
66, 69–71, 75, 84, 93, 93<sup>31</sup>,  
94, 95, 139<sup>55</sup>, 147, 148, 205,  
244, 245
- Adjunct, 15, 20, 37, 40, 79<sup>21</sup>, 87<sup>27</sup>, 105,  
106, 108–111, 137<sup>52</sup>, 138, 140,  
146, 203, 205, 225
- Adverbial, 36, 135, 138, 138<sup>54</sup>, 140, 141,  
145, 146, 202, 203, 205, 231
- Ambiguity, 120, 238, 239
- Antecedent-based ellipsis, 5, 22, 24,  
256
- Article omission, 9, 26, 75, 165, 205,  
214
- Audience design, 4, 64<sup>10</sup>, 110<sup>40</sup>, 166,  
185, 192, 227, 250, 255
- Availability-based production, 172–  
173, 227–228, 250, 255
- Bigram language model, 176, 188,  
190, 202<sup>22</sup>, 209
- Case connectivity, 7, 12, 13, 16, 32<sup>19</sup>,  
27–33, 39, 46, 50, 52, 53,  
58–61, 63–65, 69–71, 98, 147,  
205, 243–245, 248, 249
- Case feature, 9, 10, 10<sup>2</sup>, 11, 12, 29, 34,  
50, 69, 90–93, 98, 99, 99<sup>33</sup>,  
112, 151, 245
- Channel capacity, 157, 158, 161–  
163, 164<sup>11</sup>, 165, 166, 169–171,  
171<sup>17</sup>, 199, 232, 249, 254, 255
- Channel coding, 159, 161–163
- Cleft, 32<sup>19</sup>, 75, 76, 76<sup>19</sup>, 76<sup>20</sup>, 91, 95,  
97–99, 112
- Clitic left dislocation, 19<sup>10</sup>, 31, 38,  
38<sup>27</sup>, 39
- Complement clause, 5, 41, 42, 49,  
50, 72, 113, 114, 114<sup>43</sup>, 115–  
117, 117<sup>45</sup>, 118, 118<sup>46</sup>, 119–  
135, 138, 147, 149, 150, 176<sup>2</sup>,  
209<sup>24</sup>, 246, 247, 253
- Complementizer omission, 47, 114,  
114<sup>41</sup>, 115, 117, 117<sup>45</sup>, 118,  
118<sup>46</sup>, 119, 120, 123–125, 129,  
131–134, 137–138, 147, 227,  
247
- Complementizer phrase, 1, 15, 17–20,  
37–40, 43, 78, 114, 136, 136<sup>51</sup>,  
137, 137<sup>52</sup>, 138, 145, 146
- Construction grammar, 70
- Context, extralinguistic, 4, 12, 22, 23,  
25, 52–53, 65, 160–161, 163–  
164, 175–177, 177<sup>4</sup>, 180, 199,

## *Subject index*

- 207, 208, 218, 225, 227, 229,  
230, 233, 243, 251–254
- Context, linguistic, 5, 6, 12, 16, 22, 60,  
147, 156, 167–168, 208, 209,  
213, 214, 218, 222, 223, 225–  
227, 230, 231, 243, 254
- Context-free grammar, 209, 210, 231
- Corpus, 3, 3<sup>3</sup>, 9, 30, 34, 36, 65–67,  
86, 103, 104, 109, 110, 113,  
117, 120, 136<sup>51</sup>, 138, 149, 155<sup>3</sup>,  
164<sup>11</sup>, 166, 171<sup>16</sup>, 175, 176,  
176<sup>2</sup>, 177, 178, 180–184, 186,  
187<sup>9</sup>, 188<sup>11</sup>, 190<sup>14</sup>, 195, 198<sup>19</sup>,  
207–209, 209<sup>24</sup>, 214, 228–  
230, 243, 244, 252, 253
- Dative case, 12, 28, 29, 50, 51, 60, 62,  
62<sup>9</sup>, 75, 84, 90, 92, 93, 93<sup>31</sup>,  
94, 98, 205
- Default case, 6, 10, 11, 11<sup>3</sup>, 12, 28–30,  
32–34, 46, 49–52, 60, 64, 65,  
69, 70, 91, 92, 94, 98, 112, 135,  
148, 245
- Determiner phrase, 2, 3, 5, 9, 10, 10<sup>2</sup>,  
11, 11<sup>3</sup>, 12, 19<sup>11</sup>, 20, 25, 28–  
30, 31<sup>18</sup>, 32, 33, 35, 39, 41, 42,  
45, 49–51, 51<sup>2</sup>, 52–55, 58–  
64, 65<sup>12</sup>, 66, 69–71, 74–76,  
76<sup>19</sup>, 76<sup>20</sup>, 77, 78, 80<sup>22</sup>, 81<sup>23</sup>,  
82<sup>24</sup>, 83–86, 89–92, 92<sup>29</sup>,  
93–99, 99<sup>33</sup>, 102–106, 111–  
113, 136<sup>51</sup>, 137, 138, 138<sup>54</sup>,  
140, 145–151, 168, 168<sup>15</sup>, 176,  
189, 191, 197, 202, 205, 209,  
231, 244–248
- Distributed morphology, 10<sup>2</sup>
- E feature, 16, 17, 17<sup>9</sup>, 18–20, 35, 37,  
38<sup>26</sup>, 39, 41, 43, 44, 136<sup>51</sup>,  
137, 137<sup>52</sup>, 146, 150, 246, 248
- E-givenness, 15<sup>7</sup>, 18, 25, 173, 174, 228
- Entropy, 162, 217, 218
- ERP, 255
- Event chain, 4, 65, 66, 175, 177, 178,  
181–182, 184–188, 190, 195,  
200, 229, 251, 252
- Event sequence description, 183–186
- Exceptional movement account, 20,  
21, 38, 38<sup>26</sup>, 40–42, 44, 45,  
47, 74, 77, 80, 90, 112, 115,  
137<sup>52</sup>, 146, 150, 173
- Eye tracking, 170<sup>16</sup>
- F-marking, 15, 16, 38, 40
- Factivity, 41, 114–118, 118<sup>46</sup>, 120, 132–  
134
- Feature percolation, 78–81, 99<sup>33</sup>
- Focus, contrastive, 32, 82, 85, 140
- Fragment, discourse-initial, 6, 14, 18,  
21–25, 53, 60, 69, 175, 177,  
209, 229, 248
- Fragment, short answer, 6, 9, 13<sup>6</sup>, 14,  
15, 18, 20–22, 24, 28, 31, 31<sup>18</sup>,  
32, 53, 60, 72–74, 76<sup>20</sup>, 77,  
79<sup>21</sup>, 82, 82<sup>24</sup>, 83, 83<sup>25</sup>, 84–  
87, 89, 91, 92<sup>29</sup>, 94, 95, 97–  
103, 103<sup>36</sup>, 104, 107, 109, 111,  
113, 115, 125, 132, 134, 142,  
148, 149, 243, 245, 246
- Game theory, 3, 26, 47, 231, 233<sup>28</sup>,  
233–241, 249–251
- Gapping, 4, 5, 13, 13<sup>6</sup>, 14–16, 22–24,  
256
- Generative grammar, 1–3, 7, 8, 27, 77,  
136, 153, 244, 245, 248
- Genitive case, 12, 29, 50, 51, 75, 84,  
93<sup>31</sup>, 94
- HPSG, 2, 7, 36, 70, 245, 246

- In situ deletion account, 3, 13–16, 32, 33, 42, 45, 47, 53, 60, 71, 72, 78, 79, 81, 90, 91, 99, 99<sup>33</sup>, 100<sup>34</sup>, 102, 112, 113, 115, 116, 120, 133, 134, 146, 147, 149–151, 153, 173, 244–247, 251
- Information density, 162–165, 165<sup>12</sup>, 166, 168, 169, 171<sup>17</sup>, 172, 199, 200, 208, 214–216, 218, 219, 223–227, 229, 232, 240, 249, 250
- Information structure, 2, 14, 20, 27, 37–41, 51, 140, 172–174, 228–229, 250–251
- Information theory, 3, 4, 6, 26, 91, 151, 153–160, 162, 170, 172, 173, 173<sup>20</sup>, 174, 176, 193–195, 198, 199, 208, 218, 219, 224, 228–229, 231, 243, 248–251, 254
- Inherent case, 10, 12, 29, 30, 31<sup>18</sup>, 32, 50, 51, 51<sup>3</sup>, 60, 70, 92, 93, 93<sup>31</sup>, 94, 244, 245, 251
- Island, 32, 39, 42–44, 79, 137, 145, 146
- Iterated Best Response model, 233–237, 251
- Middle field, 135
- Minimalist program, 8–10, 28, 39<sup>29</sup>, 69, 80<sup>22</sup>, 77–81, 92<sup>30</sup>
- Mixed account, 24, 64<sup>11</sup>, 63–65, 68–70, 148, 245
- Movement and deletion account, 2, 3, 7, 13, 19<sup>11</sup>, 16–21, 32, 33, 35–37, 37<sup>25</sup>, 41–43, 45–47, 50, 53, 60, 71, 72, 74, 75, 77–79, 81, 82, 82<sup>24</sup>, 83, 85–87, 90, 91, 97–99, 100<sup>34</sup>, 104, 106, 111–115, 119, 120, 129, 133–135, 136<sup>51</sup>, 137–139, 144–150, 173, 244, 246–248
- Movement restriction, 3, 12, 20, 26, 27, 41–45, 47, 49, 50, 71, 72, 77, 79, 81, 82, 91, 99, 106, 113–115, 118–120, 123–125, 129, 132–135, 144, 147, 149–151, 246–248
- N-gram language model, 176, 177, 184, 188, 202<sup>22</sup>, 208, 209, 212, 213, 243, 253, 254
- Noisy channel, 26, 154, 156–159, 161, 166, 169–170, 172, 250
- Nominative case, 9–11, 11<sup>3</sup>, 12, 28–30, 31<sup>18</sup>, 34, 51, 51<sup>2</sup>, 52–54, 56<sup>5</sup>, 57–65, 65<sup>12</sup>, 69, 70, 75, 76, 76<sup>20</sup>, 77, 85, 91, 92, 94–99, 112, 148, 245
- Nonsentential account, 2, 7–12, 21, 25, 27–30, 32, 32<sup>20</sup>, 33–37, 40, 46, 50, 52–54, 59, 60, 64, 65, 69–71, 72<sup>14</sup>, 91–94, 97–99, 112, 113, 146–148, 150, 151, 244, 245, 247
- Noun phrase, 5, 92, 187, 203, 227, 238
- Object drop, 206, 256
- P-Stranding Generalization, 73<sup>16</sup>, 72–82, 86, 89–92, 93<sup>32</sup>, 94, 99, 100<sup>34</sup>, 100<sup>35</sup>, 103, 111–113, 149, 151
- Parser, human, 4, 117, 158, 167, 171, 209–211, 213, 230, 250, 255
- Parser, parallel, 171, 171<sup>18</sup>, 172, 209–211, 230, 250, 253, 255
- Parser, serial, 255

## *Subject index*

- Pied-piping, 72, 77–81, 86, 87, 87<sup>27</sup>, 89, 90, 100, 100<sup>34</sup>, 101–113, 149
- Prefield, 36, 37<sup>25</sup>, 38, 38<sup>28</sup>, 39, 47, 49, 50, 72, 118<sup>46</sup>, 135, 136, 136<sup>51</sup>, 137–139, 139<sup>55</sup>, 140–147, 149, 150, 246–248
- Preposition omission, 32<sup>19</sup>, 47, 72–75, 76<sup>20</sup>, 79, 79<sup>21</sup>, 81–83, 85, 86, 88–93, 93<sup>31</sup>, 99<sup>33</sup>, 100<sup>34</sup>, 97–104, 107–113, 134, 135, 147–149, 205, 227, 246, 246, 247, 247, 248, 256
- Preposition phrase, 5, 12, 36, 72, 73<sup>16</sup>, 74, 75, 76<sup>20</sup>, 77–79, 79<sup>21</sup>, 80, 80<sup>22</sup>, 81, 81<sup>23</sup>, 82<sup>24</sup>, 83, 83<sup>25</sup>, 84–86, 87<sup>27</sup>, 89–91, 92<sup>29</sup>, 92<sup>30</sup>, 94, 96–98, 99<sup>33</sup>, 100, 100<sup>34</sup>, 102–113, 116, 140, 148, 149, 168, 197, 205, 210, 227, 246, 248, 249
- Preposition stranding, 41, 45, 50, 72, 73, 73<sup>16</sup>, 74, 76–80, 80<sup>22</sup>, 81, 81<sup>23</sup>, 86, 87, 87<sup>27</sup>, 89–92, 92<sup>29</sup>, 94, 97–99, 99<sup>33</sup>, 100, 100<sup>34</sup>, 101–113, 149, 150, 247
- Prepositional case, 62, 62<sup>9</sup>, 75, 76<sup>19</sup>, 77, 81, 83–85, 90–92, 92<sup>30</sup>, 93, 93<sup>31</sup>, 94, 96–99, 112, 148, 245, 246
- Processing effort, 3, 24, 64, 81, 91, 98, 102, 104, 110, 111, 113, 120, 151, 154, 156<sup>4</sup>, 162, 170<sup>16</sup>, 169–171, 171<sup>16</sup>, 171<sup>17</sup>, 172, 181, 212, 227, 232, 240, 249, 250, 254–255
- Production task, 2, 52, 61, 63, 69, 76<sup>20</sup>, 100, 104, 110<sup>40</sup>, 111, 113, 125, 142, 153, 167<sup>13</sup>, 176, 189, 196, 199, 200, 202, 203, 210, 215, 223, 224, 226, 228–230, 237, 245, 246, 249, 251, 253
- Question under Discussion, 13, 14, 16, 22–25, 52, 59, 153, 174, 228, 245, 246, 248, 256
- Relevance theory, 25, 153
- Right node raising, 22
- Script knowledge, 4, 23, 25, 65, 175–177, 177<sup>4</sup>, 178<sup>5</sup>, 180<sup>6</sup>, 178–181, 181<sup>7</sup>, 182–184, 184<sup>8</sup>, 185, 186, 188, 188<sup>11</sup>, 189, 190, 190<sup>14</sup>, 190<sup>15</sup>, 191–193, 193<sup>16</sup>, 194–197, 198<sup>19</sup>, 199–202, 204, 207, 208, 213, 214, 216–218, 220, 224, 229, 230, 237, 243, 251–252, 254
- Shannon information, 154–156, 156<sup>4</sup>, 157, 161–164, 164<sup>11</sup>, 166–170, 170<sup>16</sup>, 171, 171<sup>16</sup>, 172, 173, 173<sup>20</sup>, 174–176, 176<sup>2</sup>, 177, 177<sup>4</sup>, 197, 198, 198<sup>20</sup>, 199–202, 205–209, 209<sup>24</sup>, 210–220, 222–232, 240, 243, 249–255
- Simpler syntax, 70
- Situation-based ellipsis, 22
- Skip-gram language model, 184<sup>8</sup>, 213
- Sluicing, 4, 5, 13, 16, 17<sup>9</sup>, 18–20, 43, 44, 72–75, 76<sup>20</sup>, 93<sup>32</sup>, 173<sup>20</sup>, 256
- Small clause, 33–36, 146, 148
- Source coding, 159–162, 169, 175, 198, 199, 225, 226, 232<sup>27</sup>, 250–251, 255
- Structural case, 10, 12, 14, 18, 20, 27–30, 31<sup>18</sup>, 33, 34<sup>21</sup>, 37, 46, 59,

- 69, 71, 92–94, 112, 147, 148,  
151, 244, 245
- Subject drop, 165, 186, 256
- Surprisal, *see* Shannon information
- Tense phrase, 5, 9, 10, 17, 19, 20, 38<sup>26</sup>,  
43, 137
- That is-ellipsis, 76<sup>20</sup>
- Topic drop, 206, 256
- Topic, contrastive, 38, 38<sup>27</sup>, 88, 140
- Topic, hanging, 11, 31, 32, 39
- Topological field model, 135
- Trace (movement), 39, 43–45, 137,  
145, 146
- Trigram language model, 177
- Ungrammaticality of fragments, 7, 8,  
26–27, 33, 53, 60, 70, 90, 98
- Uniform Information Density, 3, 4,  
154, 159, 161–163, 165–170,  
171<sup>17</sup>, 172, 172<sup>19</sup>, 173–176,  
185, 189, 192, 194, 196–199,  
205–207, 214–216, 218, 222–  
229, 231, 232, 239–241, 243,  
244, 249–256
- Unigram language model, 176, 188,  
208, 209, 213, 216, 219, 220,  
222, 223, 225, 227, 230, 254
- Verb phrase, 5, 9, 24, 35, 36<sup>23</sup>, 36<sup>24</sup>,  
38<sup>28</sup>, 138, 198<sup>20</sup>, 256
- Verb phrase ellipsis, 4, 5, 16, 17<sup>9</sup>, 22,  
24, 43, 198<sup>20</sup>, 256
- Word order, 30, 66, 67, 105, 140, 146,  
167–168, 199, 202, 203, 205–  
207, 216, 218, 223–225, 227
- Zipf's Law, 159–160



# Experimental investigations on the syntax and usage of fragments

This book investigates the syntax and usage of fragments (Morgan 1973), apparently sub-sentential utterances like “A coffee, please!” which fulfill the same communicative function as the corresponding full sentence “I’d like to have a coffee, please!”. Even though such utterances are frequently used, they challenge the central role that has been attributed to the notion of sentence in linguistic theory.

The first part of the book is dedicated to the syntactic analysis of fragments, which is investigated with experimental methods. Currently, there are competing theoretical analyses of fragments, which relied almost only on introspective judgements. The experiments presented in this book constitute a first systematic evaluation of their predictions and, taken together, support an *in situ* ellipsis account of fragments, as has been suggested by Reich (2007).

The second part of the book addresses the questions of why fragments are used at all, and under which circumstances they are preferred over complete sentences. Syntactic accounts impose licensing conditions on fragments, but they do not explain, why fragments are sometimes (dis)preferred provided that their usage is licensed. This book proposes an information-theoretic account of fragments, which is supported by two experiments: In order to distribute processing effort uniformly across the utterance, predictable words are more likely to be omitted and additional redundancy is inserted before unpredictable words.

ISBN 978-3-96110-331-7



9 783961 103317