

Rethinking principle C reconstruction in ATB movement and parasitic gaps

Data driven perspectives on the too many tools problem

Dissertation
zur Erlangung des Doktorgrades der Philosophie (Dr. phil.)
vorgelegt an der Humanwissenschaftlichen Fakultät der Universität Potsdam

von
Timea Szarvas

Potsdam, 2025

Acknowledgements

I have no idea how to begin. Do you know that feeling of anticipating something for a long time and having a clear expectation of what it will feel like, and then when it is finally here, it feels completely different, making you wonder why you thought so much about what it might make you feel in the first place? That is me in a nutshell. During these last four and a half years, I felt supported, encouraged, accepted and loved by many people. I owe them my gratitude for accompanying me not only on this academic pursuit, but also the journey of personal growth that allows me to pause and be proud of my achievement.

First and foremost, I owe my utmost gratitude to my supervisor Doreen Georgi. In your class on Differential Argument Coding in 2018, you asked us about our interests, and I said I liked syntax and experimental linguistics. I remember feeling like I had to pick one eventually and abandon the other. Thank you for encouraging me not to. Thank you for your support and your contagious enthusiasm. Thank you for trusting me with decisions I would not have trusted myself with, for following me wherever my curiosity led me, and for nudging me back to my path whenever I risked going astray.

I thank Malte Zimmermann for co-supervising me, and I apologize for taking this project as far from semantics as I possibly could have. Still, you have been an academic advisor to me and made sure to let me know that I was exactly where I belonged. Thank you for that.

Martin Salzmann deserves special thanks for naturally taking on the role of an advisor the day he arrived in Potsdam, no questions asked. Thank you for your mentorship, guidance, and fun collaborations.

None of this work would exist in this form without Johannes Rothert. Thank you for introducing me to the practicalities of designing and running experiments through countless discussions, and for projects that started out as assignments in class and ended up paving the way for us to become researchers. By extension, I thank the late Gisbert Fanselow, not only for teaching Johannes what he passed on to me, but for building a legacy of experimental syntax in Potsdam that I continue to benefit from (and hope to have contributed something meaningful to).

I thank various past and present residents of Haus 14 for their feedback,

academic collaborations, earnest advice and silly chats; for sharing beers and the unique experience of Golm: Nina Adam, Giuliano Armenante, Nadine Bade, Carla Bombi, Aron Finholt, Johannes Hein, Andreas Hölzl, Rebecca Jarvis, Jeanne Lecavelier, Andrew Murphy, Andreas Pregla, Masha Privizentseva, Anna Struck, Maximilian Wiesner, and Stan Zompì. Max Wiesner I also have to thank along with Chiara Meisch for their support as former student assistants, for joining me on the dreadful quest to create items for some of the experiments that made it into this dissertation (and for some that did not).

I am grateful that this research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 317633480, SFB 1287. The SFB's graduate school provided me with financial means, but more importantly, with an infrastructure other PhD students in Europe only dare to dream of. Most of all, I thank Raúl Bendeزú Araujo for coordinating our cohort, for checking in with me even when I seemed fine and especially for doing so when I didn't, and for holding our SFB together by providing advice and having our backs, always, no matter what the issue was. He is also the person I owe the suggestion of the *too many tools* metaphor to.

I thank the people at UMass who welcomed me for a short visit in spring 2025, particularly Rajesh Bhatt, Brian Dillon, Kyle Johnson and Shota Momma for discussing my work and where it was going. An additional thank you to Shota for agreeing to read and review this thesis. Thanks to Lulu Guo, Polina Kasyanova, Andrea Matticchio, Satoru Ozaki, Jonathan Pesetsky and Carla Spellerberg for their company.

Despite feeling a bit isolated in the niche I created for myself sometimes, I never felt alone in being a PhD student in linguistics, mostly because of the great people I met at conferences, workshops, summer schools and other visits throughout the years. In particular, I thank Iva Kovač and Leonardo Russo Cardona for lunches spent in the botanical garden of the University of Göttingen. I also thank everyone whom I may not be in touch with otherwise, but who reliably save me from awkwardly standing around by myself at conferences. I thank the cohort of CreteLing 2023 for a ridiculously fun and inspiring two weeks.

I thank Magdalena Lohninger for proving that having a drink with your peers is equally as important for your career as connecting over mutual research interests. Thank you for a friendship that is rooted in navigating academia and reaches beyond.¹

Jennifer Sander deserves my thanks for much needed check-ins. For reminding me why I chose this path despite all of its difficulties. Thank you also for cat sitting, not as in sitting my cat or me sitting yours, but for your cats sitting on me.

¹You guys better get ready for Syntoxicated 2039.

I thank Marie Hansen for sitting next to me in our first stats class in 2018 and for being by my side ever since. Thank you for supporting me, and thank you, especially, for constantly challenging the way I think. Thank you for being both a colleague and like a sister to me, und dass du mir als meine Anwältin immer wieder rätst, die Dinge nicht so schrecklich ernst zu nehmen.

There are many lovely people in my life that deserve my gratitude for being equally capable of distracting me from this project and engaging with my thoughts related to it, despite not knowing what it is exactly that I do and only vaguely understanding the framework of academia. I thank Michi for convincing me to get out of my head and dare asking for what I want. Thank you for being the reason why I ended up in Potsdam. Anne, Birte, Carl, Clara, Florin, Gina, Sarah, Sebastian, Vi, thank you for time spent in kitchens, bars and on dance floors. For cooking and devouring elaborate dinners, for staying up way too late, for all the walking and talking. Thank you for reassurance, advice, support, and most of all, lightheartedness.

Lolle, by a landslide, your company is what I miss the most about my time in Tübingen. I'm glad we got to tackle this formative decade of our lives together despite the distance. Thank you for surviving kindergarden with me, for phone calls, for sharing your thoughts, and for always caring.

I thank Laura for an incredibly deep and completely unhinged friendship. Thank you for having an extremely fine-tuned and simultaneously nonsensical humor that makes us laugh until we can't articulate ourselves anymore. Thank you for always meeting me no matter where I am at, for telling me the things I need to hear even if I don't want to hear them, and for believing in me when I don't believe in myself.

Fabi, how can I thank you for all the things I can barely put into words? Thank you for being my partner in sourdough, crime and life. Thank you for making me giggle and blush. Thank you for making everything feel easy. Thank you for making everything make sense.

Mókinak, Mayának, Mamónak, Gabinak és Zsoltinak köszönöm a szeretetet és hogy bármikor számíthatok rájuk. Anyunak köszönöm az önfeláldozását és főleg a türelmét, ha csak ritkán halljuk egymást. Köszönöm, hogy a végtelenségig bízol bennem és abban, hogy a végén majd minden lesz valahogy, még ha halvány fogalmunk sincs róla, hogyan. Köszönöm, hogy soha az életben nem jutott eszedbe németül beszélni velem és hogy szerbül káromkodtál (főleg ezt köszönöm). Ki tudja, mi lett volna belőlem.

The best is yet to come.

Abstract

Multi-gap dependencies, such as ATB movement and parasitic gaps, constitute a puzzle for syntactic theory because they violate the strict one-to-one mapping between a filler and a gap. Over the last few decades, multiple ‘tools’ have been proposed to model this unusual one-to-many relation, some of them unifying ATB and parasitic gaps under a single approach, others treating them as unrelated. These tools either aim to re-shape how we think about movement via multidominance or sideward movement, or apply independently established mechanisms, like, for example, ellipsis or operator movement. Despite their fundamental differences, nearly all tools remain to coexist, and it appears that the field prioritizes proposing new theories over identifying which of them has the best empirical coverage. The lack of clarity regarding the empirical facts themselves might be the main cause of this. On the one hand, different researchers focus on different types of diagnostics, ultimately yielding theories that empirically over- or undergenerate. On the other hand, there are certain diagnostics whose outcomes researchers outright disagree about. After illustrating these circumstances and motivating the need for more, but also more structured means of data collection in chapter 1, the aim of this thesis is to take a step towards ‘decluttering the toolkit’ by adopting a more data driven, experimental approach to the study of multi-gap dependencies.

Instead of presenting different types of evidence in support of a single theory, this thesis deals with one particularly controversial piece of evidence and its implications for the plethora of existing theories. The core of the thesis investigates the contested observation that principle C reconstruction of nominal PP modifiers reveals an asymmetry between the initial and non-initial gap in ATB and parasitic gap constructions in extraposed adjunct clauses. In chapter 2, two experiments on principle C reconstruction of PP modifiers in German demonstrate that the effect is detectable in simple wh-dependencies, but not universally among speakers, items or even experimental setups. Then, in chapter 3, the limited cross-linguistic comparability of ATB movement and ‘English-type’ parasitic gaps is motivated by an acceptability judgment experiment of German parasitic gaps in extraposed adjunct clauses, revealing that the majority of participants judge parasitic gaps to be marginal or entirely unacceptable. The investigation of principle C reconstruction

in German multi-gap dependencies thus focuses exclusively on ATB movement, presenting three experiments that severely challenge the impact of principle C reconstruction on coreference resolution.

The methodological insights from chapters 2 and 3 are then applied to the study of principle C reconstruction in English multi-gap dependencies in chapter 4. A multi-part experiment shows the limited but non-negligible relevance of the basic principle C reconstruction effect, the vanishing of its impact under ATB movement, and the strong resemblance of the coreference pattern observed in ATB movement compared to parasitic gaps. A follow-up experiment studying the effect of structural vs. linear distance on the coreference pattern in parasitic gaps in purpose clauses delivers inconclusive results, and I discuss methodological complications supported by post-hoc statistical analyses.

Based on the findings from chapters 2-4, I argue throughout the thesis that the apparent asymmetry between the gaps with respect to principle C does not need to be accounted for in the syntactic derivation of ATB movement and parasitic gaps. What instead needs to be accounted for is the lack of robustness in principle C reconstruction not only with increased distance, but particularly with increased dependency complexity. In turn, the previously reported principle C reconstruction data do not constitute evidence for the movement path of the filler in multi-gap dependencies, and do not support asymmetrical derivations in which only the initial gap is created via movement.

Chapter 5 recapitulates the findings presented in this thesis and what the orthogonality of principle C reconstruction implies for existing approaches to multi-gap dependencies. Finally, I discuss how we could approach reconstruction in multi-gap dependencies in future work, with a special focus on which types of reconstruction could lead to more firm conclusions about the derivation of ATB and parasitic gap dependencies.

Preface

The contents of this dissertation were presented at various international linguistics conferences and as invited talks from 2021 to 2025. The publications that it is partially based on, as well as a repository of supplementary materials available at the Open Science framework, are listed in the following.

- Szarvas, Timea (2024a). “On reconstruction in German ATB-movement and the optimization of experimental designs”. In: *Proceedings of the 32nd Conference of the Student Organization of Linguistics in Europe*. Ed. by Federica Longo, Leonardo Russo Cardona, Tommaso Sgrizzi, Atefeh Shahbazi, and Hadis Tamleh. Leiden: Leiden University Centre for Linguistics, 2024, pp. 195–212. URL: <https://ling.auf.net/lingbuzz/008687>.
- (2024b). “Unscrambling German parasitic gaps”. In: *Gisbert Fanselow’s Contributions to Syntactic Theory*. Ed. by Artemis Alexiadou, Doreen Georgi, Fabian Heck, Gereon Müller, and Florian Schäfer. Universität Leipzig: Institut für Linguistik, Leipzig, 2024, pp. 133–144. URL: <https://ling.auf.net/lingbuzz/008679>.
- (2025a). “An asymmetry without theoretical consequences: Revisiting principle C reconstruction in ATB movement experimentally”. In: *Proceedings of the 55th Annual Meeting of the North East Linguistic Society*. Ed. by Duygu Demiray, Roger Cheng-yen Liu, and Nir Segal. Vol. 2. University of Massachusetts, Amherst: GLSA, 2025, pp. 239–248. URL: <https://lingbuzz.net/lingbuzz/008992>.
- (2025b). *Supplementary materials for “Rethinking principle C reconstruction in ATB movement and parasitic gaps”*. Open Science Framework. 2025. URL: <https://doi.org/10.17605/OSF.IO/HVM8U>.
- (2025c). “Why we’re still debating principle C reconstruction. A comparative experimental study”. In: *Glossa: A journal of general linguistics* 10.1, pp. 1–42. URL: <https://doi.org/10.16995/glossa.18707>.

Contents

1	Introduction	1
1.1	An overview of multi-gap dependencies	6
1.2	Existing approaches to multi-gap dependencies	10
1.2.1	Symmetrical approaches	10
1.2.2	Asymmetrical approaches	15
1.3	Evidence	26
1.3.1	Case matching, repair by syncretism and beyond	26
1.3.2	Reconstruction	40
1.4	Too many tools for multi-gap dependencies	51
1.4.1	The problem	51
1.4.2	The way out	53
2	Why we're still debating principle C reconstruction	56
2.1	Background	57
2.1.1	Principle C reconstruction	57
2.1.2	Previous experimental investigations	59
2.1.3	Evaluation of previous experiments	62
2.2	Experiments	68
2.2.1	Baseline	68
2.2.2	Experiment 1: simplification of the task	72
2.2.3	Experiment 2: omission of the matrix referent	80
2.3	General discussion	88
2.3.1	Methodological comparison	88
2.3.2	Inter-speaker variability	89
2.3.3	Optionality of reconstruction	91
2.3.4	Limitations	93
2.4	Conclusion	94
3	Experiments on German multi-gap dependencies	96
3.1	On the acceptability of German parasitic gaps	97
3.1.1	Existing reports	97

3.1.2	Empirical complications	104
3.1.3	Experiment 3: German parasitic gaps	106
3.1.4	Discussion	117
3.1.5	Conclusion	123
3.2	Principle C reconstruction in German ATB movement	124
3.2.1	Testing principle C reconstruction in ATB movement experi- mentally	125
3.2.2	Experiment 4: Two forced choice tasks	127
3.2.3	Experiment 5: One forced choice about possibility of embedded referent	137
3.2.4	Experiment 6: One forced choice between embedded and unnamed referent	144
3.2.5	General discussion	150
3.2.6	Conclusion	155
3.3	Interim summary	156
4	Experiments on English multi-gap dependencies	160
4.1	Foundations	161
4.1.1	Experiment 7: Principle C reconstruction in English	161
4.1.2	Experiment 8: The effect of finiteness on the acceptability of parasitic gaps in English	167
4.1.3	Interim discussion	173
4.2	Principle C reconstruction in English multi-gap dependencies . . .	174
4.2.1	Experiment 9: Difference between ATB and parasitic gap constructions	177
4.2.2	Experiment 10: Principle C reconstruction in ATB dependen- cies	183
4.3	Follow-up investigations	189
4.3.1	Experiment 11: The effect of distance on coreference	190
4.3.2	Correlation of tasks	195
4.4	General discussion	200
4.5	Beyond PP modifiers	204
5	Synthesis and outlook	207
5.1	Empirical neutrality	209
5.2	Alternative diagnostics	210
5.3	Final conclusion	216

List of Figures

2.1	Coreference rates from Salzmann, Wierzba, and Georgi (2023) . . .	70
2.2	Coreference rates in experiment 1	76
2.3	Participants' slopes in experiment 1	76
2.4	Coreference rates in experiment 2	83
2.5	Participants' slopes in experiment 2	84
2.6	Participants' slopes from Salzmann, Wierzba, and Georgi (2023) . .	90
3.1	Mean acceptability ratings from experiment 3	111
3.2	Response distribution in experiment 3	112
3.3	Heatmap of experiment 3	113
3.4	Ratings by region in experiment 3	115
3.5	Effect of trial number in experiment 3	116
3.6	Coreference rates in experiment 4	132
3.7	Participants' slopes in experiment 4	134
3.8	Coreference rates in experiment 5	139
3.9	Participants' slopes in experiment 5	141
3.10	Participant migration in experiment 5	142
3.11	Coreference rates in experiment 6	146
3.12	Participants' slopes in experiment 6	148
3.13	Participant migration in experiment 6	149
4.1	Coreference rates in experiment 7	164
4.2	Participants' slopes in experiment 7	166
4.3	Acceptability ratings in experiment 8	170
4.4	Response distribution in experiment 8	171
4.5	Heatmap of experiment 8	172
4.6	Coreference rates in experiment 9	179
4.7	Participant migration in experiment 9	181
4.8	Acceptability ratings in experiment 9	183
4.9	Coreference rates in experiment 10	185
4.10	Participants' slopes in experiment 10	187
4.11	Participant migration in experiment 10	188

4.12 Coreference rates in experiment 11	193
4.13 Correlation of tasks in experiments 9 and 10	198

List of Tables

1.1	Approaches to multi-gap dependencies	3
1.2	Data availability for reconstruction in multi-gap dependencies . . .	40
1.3	Reported patterns of reconstruction in multi-gap dependencies . .	50
2.1	Results of Salzmann, Wierzba, and Georgi’s experiment	70
2.2	Model estimates for experiment 1	77
2.3	Log-transformed estimate summary for experiment 1	78
2.4	Model estimates for experiment 2	85
2.5	Log-transformed estimate summary for experiment 2	86
2.6	Coreference rates across experiments 1-2	87
3.1	Design of experiment 3	106
3.2	Model estimates of experiment 3	110
3.3	Mean ratings by condition.	111
3.4	Model estimates for experiment 4	133
3.5	Model estimates for experiment 5	140
3.6	Model estimates for experiment 6	147
3.7	Coreference rates across experiments 4-6	150
3.8	Number of observations across experiments 4-6	151
4.1	Model estimates for experiment 7	165
4.2	Mean ratings in experiment 8	169
4.3	Model estimates for experiment 8	170
4.4	Optimal experimental design for experiments 9 and 10	175
4.5	Model estimates for experiment 9	180
4.6	Model estimates for experiment 10	186
4.7	Model estimates for experiment 11	194
4.8	Model estimates for acc. \sim ref. in experiments 9 and 10	197
4.9	Model estimates for ref. \sim acc. in experiments 9 and 10	197
4.10	Model estimates for acc. \sim ref. in experiment 8	199
4.11	Model estimates for ref. \sim acc. in experiment 8	199

Chapter 1

Introduction

Multi-gap dependencies constitute a notable exception to the one-to-one mapping found across syntactic movement dependencies. In standard filler-gap dependencies, a filler occupies a single position on the surface, i.e. the position at which it is pronounced, originating from a single base position where it leaves a gap. By this definition, every gap is related to exactly one filler. In multi-gap dependencies, on the other hand, a single filler appears to be related to more than one gap. This thesis focuses on Across-the-Board (ATB) movement, where a single \bar{A} -moved filler is related to a gap in each conjunct of a coordinate structure (1a), and parasitic gap (PG) dependencies, where such a filler is related to a main gap and another (optional) gap that must be licensed by the first (1b).

- (1) a. [Which paper] did John file ___ and Mary read ___ ?
 b. [Which paper] did John file ___ before reading ___ pg ?

The one-to-many relation between the filler and the gaps gave rise to numerous ‘tools’ that are used to derive underlying syntax of multi-gap dependencies. For the purposes of this thesis, the extant ones can be classified based on two criteria: (i) whether they cover both ATB and parasitic gaps, treating them on a par to varying extents, and (ii) which position(s) the moving filler originates from, bearing on the nature of the relationship between the filler and each gap. Regarding the first criterion, authors who have proposed a unitary treatment of ATB movement and parasitic gaps primarily cite superficial resemblance and various empirical similarities (Altshuler and Truswell 2022; Bruening and Al Khalaf 2017; Franks 1993; Haik 1985; Hornstein and Nunes 2002; Huybregts and van Riemsdijk 1985; Munn 1993; Nunes 2001; Pesetsky 1982; Pollard and Sag 1994; Seguin and Thoms 2025; Williams 1987). However, there are at least as many pieces of evidence pointing to fundamental differences between the two constructions (see Postal 1993 for a summary). Some authors therefore straightforwardly argue against the unitary treatment of ATB and parasitic gaps in their proposals (Nissenbaum 2000;

Salzmann 2012), while other accounts do not address the potential connection between the two constructions at all (An 2007; Citko 2005; Levine 2001; Zhang 2010).

Coming to the second criterion, theories vary with respect to the relation between the filler and each gap. One strand of approaches maintains that there is an instance of the filler moving from each gap, having to explain why only one filler is pronounced. Qualifying explanations for ATB involve the deletion of the additional instance at Phonological Form (Biskup 2018; Wilder 1994); fusion of the two instances over the course of the derivation (Hein and Murphy 2020; Williams 1978), or that the structure is multidominant, i.e. derived by an operation combining properties of internal and external Merge, called Parallel Merge (Bachrach and Katzir 2009; Citko 2005, see also Goodall 1987; Moltmann 1992; Williams 1978). Approaches involving parallel extraction from both gaps have also been proposed for parasitic gaps in Dutch (Huybregts and van Riemsdijk 1985), English (Haik 1985; Williams 1987), and Italian (Seguin and Thoms 2025). These approaches will be called symmetric in the following because they assume that all existing gaps are created symmetrically via movement. It is also worth mentioning that ATB and parasitic gap structures played a crucial role in the early stages of Head-driven Phrase Structure Grammar (Gazdar 1981). The approach taken therein is likewise symmetric in the sense that slash feature percolation proceeds from both conjuncts of ATB and both clauses of a parasitic gap construction (Pollard and Sag 1994).

The other strand of approaches argues that there is only one instance of the filler in the derivation that is extracted to the left periphery from one of the gaps, having to explain how the other gap arises. The non-initial gap has been proposed to arise via operator movement (Bošković and Franks 2000; Franks 1995 for ATB, Chomsky 1986; Nissenbaum 2000 for PGs, Munn 1993, et seq. for both) or to be occupied by a null pronominal (Zhang 2010 for ATB, Cinque 1990 for PGs). For ATB, ellipsis approaches have been proposed as well, crucially differing with respect to whether ellipsis targets the initial (Ha 2008) or non-initial conjunct (Salzmann 2012). These approaches will be called asymmetric, for they postulate that material is extracted asymmetrically from only one of the conjuncts.

There is a third type of approach that is neither symmetric nor asymmetric, strictly speaking. A special type of movement is proposed where the two domains (conjuncts under ATB and clauses in parasitic gaps) are built in two independent work spaces. The extracted constituent is inserted from the numeration only in the non-initial one, being copied into the initial domain from there—i.e., the constituent *sideward* moves from one independent syntactic object to another rather than moving *upward* within a single syntactic domain. The core idea of the approach is that Move (as in internal Merge) is not a primitive operation of the

computational system, instead deriving movement by combining Copy, Merge, Form Chain and Chain Reduction. This operation has been proposed to be the source of ATB and parasitic gaps alike (Altshuler and Truswell 2022; Hornstein and Nunes 2002; Nunes 2001, 2004). The attempt to squeeze the sideward movement approach into a box based on a traditional, internal Merge view of movement admittedly does not do it justice. However, for the sake of generating testable hypotheses, it will be treated as an asymmetric approach since the filler moving to the left periphery is base generated in only one of the domains and copied into the other. Furthermore, based on parallels to successive-cyclic movement (Nissenbaum 2000, for example, equates the two for the most part), we will assume that the filler is base generated in the non-initial conjunct, even if the notions of initial and non-initial only gain relevance after copying applies.

		Unificational	Non-unificational	
			ATB	PGs
Symmetric		Parallel extraction: Haik (1985), Huybregts and van Riemsdijk (1985), Seguin and Thoms (2025), and Williams (1987); Multidominance: Citko (2013); HPSG: Pollard and Sag (1994)	Parallel extraction: Biskup (2018), Blümel (2014, 2017), Hein and Murphy (2020), Wilder (1994), and Williams (1978); Multidominance: Citko (2005)	
Asymmetric	Initial	Operator movement: Munn (1993, 1994, 1999, 2001)	Ellipsis: Salzmann (2012)	Operator movement: Chomsky (1986) and Nissenbaum (2000); Fission & Fusion: Assmann (2012)
	Non-initial	Sideward movement: Altshuler and Truswell (2022), Hornstein and Nunes (2002), and Nunes (1995, 2001)	Ellipsis: Ha (2008)	

Table 1.1: Approaches to multi-gap dependencies categorized by the gap(s) from which the filler is extracted to the left periphery (symmetric or asymmetric; if asymmetric, initial or non-initial gap) and whether they pursue a unificational account of ATB and parasitic gaps.

A ‘taxonomy’ of approaches is given in Table 1.1.¹ Notably, there exists no symmetric approach to parasitic gap constructions that does not propose a unification with ATB movement. Likewise, there exists no non-unificational parasitic gap approach assuming that the moved filler only originates in the non-initial gap, i.e. the parasitic gap site itself. For ATB movement, we see that all combinatorial possibilities are exhausted by the existing literature. Although some favorites have emerged over the past years—multidominance is the most frequently cited explanation for ATB; operator movement is the strongest contender for parasitic gaps; and for those who treat both on a par, sideward movement is the most popular—the different approaches more or less remain to co-exist. The field has yet to identify the proposal that is empirically superior, and whether each construction requires its own derivation or if both can be covered by a single approach. The fact that we have not succeeded so far is quite puzzling considering that the approaches seem to make clear and distinct predictions. It is moreover especially desirable to evaluate the empirical accuracy of these tools because some of them fundamentally challenge state-of-the-art assumptions about the scope and nature of core syntactic operations (multidominance and sideward movement), while others are more conservative, using independently established mechanisms (such as ellipsis or empty operator movement, and to some extent, parallel extraction).

This thesis approaches the problem of having too many syntactic tools² at our disposal for the derivation of multi-gap dependencies by taking a deep dive into principle C reconstruction and its implications for the underlying syntactic derivation of ATB and parasitic gaps. Theories of multi-gap dependencies diverge regarding the base position(s) from which movement launches and how many elements are moving. A few examples: Under multidominance, a single element occupies both positions and thus also moves from both of them *simultaneously* (Citko 2005); under ellipsis, one of the elements moves from either position while the other one is elided (with the possibility of vehicle change, as argued by Salzmann 2012 based on Fiengo and May 1994); under sideward movement, a single element moves from the non-initial position *through* the initial

¹The collection of approaches listed in Table 1.1 is by no means complete. Notably absent are null pronominal approaches to either construction due to crucial arguments for movement at least within the domain hosting the non-initial gaps (see Levine, Hukari, and Calcagno 2001 against Cinque 1990 and de Vries 2017 against Zhang 2010). Furthermore, certainly more authors have suggested that a specific treatment of multi-gap dependencies is to be favored given individual empirical observations, and some new way to model these constructions is suggested, it seems, at every major syntactic conference every single year. I thus limit myself to approaches that are specifically discussed in this thesis and invite the reader to evaluate any missing proposal of interest to them against the evidence presented in the following chapters.

²The titular term is based on the eponymous workshop tackling this problem in semantics and pragmatics at *Sinn und Bedeutung* 29.

gap (Hornstein and Nunes 2002; Nunes 2001, 2004); whereas under operator movement, depending on the implementation, the operator in the non-initial gap site is either morphosyntactically ‘empty’, meaning it does not bear the syntactic features of the extracted element from the initial gap (Chomsky 1986; Nissenbaum 2000), or it bears the properties of a trace of the extracted element (Munn 1994). Principle C has been argued to reconstruct to base positions only (Nissenbaum 2000; Sportiche 2017) and as such is considered a particularly informative diagnostic in this regard, as the absence of a principle C reconstruction effect in either gap would indicate that (i) no movement is launched from the respective position, and (ii) no syntactic representation bearing the extractee’s features is present in the respective gap site.

The second reason for investigating principle C reconstruction is that despite its ubiquitous use, the empirical facts on it are severely unclear. This thesis is not the first piece of work to report on principle C reconstruction in ATB movement and parasitic gaps (Citko 2005; Nissenbaum 2000; Salzmann 2012), and not even the very first experimental attempt to investigate it (Bruening and Al Khalaf 2017). It is the first piece of work, however, to ultimately eliminate the uncertainty that has become associated with principle C reconstruction of nominal PP modifiers, its experimental investigation, and how reliably the associated pattern maps to the underlying syntactic configuration. As we shall see in section 1.3, we crucially do not know what the reconstruction pattern looks like in ATB or parasitic gaps due to conflicting reports in the literature—that is, our current empirical practices keep us from making theoretical progress. Because of the uncertainty associated with the data, we are facing the titular *too many tools problem*. We have numerous ways to derive multi-gap dependencies that remain to exist in parallel, and as such, making progress towards a unified, general theory of syntax is less likely. To overcome this state, we need to rethink our empirical practices, and this thesis takes a step in that direction.

The remainder of chapter 1 gives an overview of ATB and parasitic gaps and why they still remain a puzzling subject half a century after their first description. First, we look at their core properties in section 1.1, followed by a sketch of the most widely accepted theoretical approaches to them in section 1.2. Then, two different types of empirical evidence often cited to evaluate the different approaches will be summarized in section 1.3: case related effects in section 1.3.1 and reconstruction effects in section 1.3.2. Chapter 1 ends with an assessment of the current theoretical and empirical situation in section 1.4. After identifying the problem in section 1.4.1 and motivating the need to fix it, the data driven solution to be pursued in this thesis is outlined in section 1.4.2.

In chapter 2, we shall see that a more consistent approach to data collection

is the first step to declutter the syntactic toolkit. In German, principle C reconstruction simply does not follow the clear-cut pattern predicted by binding theory due to the overall complexity of coreference resolution. This conclusion can be drawn based on comparative experimental work that leaves behind threshold-based designs and the reliance on null results. Instead, we see that the detectability of the effect depends on the experimental setup, whether an alternative referent is present or not, and on the individual speaker.³

In chapter 3, we will then see how these considerations map to principle C reconstruction in German ATB movement. Again, based on three experiments that vary the setup and task without manipulating the lexical content of the items, I conclude that the slight asymmetry observed between the initial and non-initial gap in German ATB movement is a confound of distance and not the result of asymmetrical principle C reconstruction. The study of parasitic gaps in German is set aside due to the limited acceptability of ‘English-type’ parasitic gaps in adjunct clauses, demonstrated by an acceptability judgment experiment.⁴

Chapter 4 will then take the methodological insights from chapters 2 and 3 and present a comparison between ATB and parasitic gaps in English, showing that the same explanation provided for the German data applies here too—crucially, for ATB as well as for parasitic gaps. I argue that the similarity between the two constructions is real, but it does not justify conclusions about syntactic relatedness. The chapter ends with a follow-up investigation of the observed effect of distance and whether it stems from the linear or structural distance between the pronoun and the R-expression.

Chapter 5 wraps up the thesis by synthesizing the results and suggesting possible directions for future research.

1.1 An overview of multi-gap dependencies

The first mention of multi-gap dependencies can be traced back to Ross (1967) who observed them alongside a series of general syntactic constraints in relation to movement. Ross’s observations are based on Right Node Raising (RNR), reporting the data in (2).

- (2)
- a. Tom picked these grapes, and I washed these grapes, and Suzie will prepare these grapes.
 - b. Tom picked, and I washed, and Suzie will prepare, these grapes.
 - c. Tom picked these grapes, and I washed some turnips, and Suzie will

³This chapter appeared as Szarvas (2025c).

⁴The contents of this chapter appeared individually as Szarvas (2024a,b, 2025a).

1.1. An overview of multi-gap dependencies

prepare these grapes.

- d. *Tom picked, and I washed some turnips, and Suzie will prepare, these grapes. (Ross 1967, pp. 176–177)

In (2a), the (identical) direct object of all conjuncts is spelled out in situ. In (2b), the identical direct object is shared across conjuncts by being moved to the right: this is what we know as RNR today. (2c) shows that as long as the direct objects remain in situ, they are free to differ. However, as shown in (2d), it is impossible to spell out one of the objects in situ and share the other one between a subset of the conjuncts. Although this is the first discussion of constituent sharing in coordinate structures, Ross did not discuss wh-extractions from coordinate structures—the first formal assessment of what we consider to be ATB movement today stems from Williams (1978). However, he did note that parasitic gaps, though not labeled as such, might be related to the construction in (2). Example (3) shows the licensing of a parasitic gap via relativization:

- (3) a. I want to peruse that contract before filing it away.
b. I suspect that the contract which I wanted to peruse before filing away may have some loopholes. (Ross 1967, p. 191)

The term *parasitic gap* has been coined a few years later by Taraldsen (1981) and Engdahl (1983) independently, who introduced the construction to the formal literature based on data from Swedish and English.

Let us examine the two relevant multi-gap dependencies by having another look at (1), repeated below as (4). In (4a) showing ATB and (4b) showing a PG dependency, *which paper* simultaneously serves as the direct object of both *file* and *read(ing)*, a relation that cannot be created by a simple wh-movement step due to violating the one-to-one mapping between filler and gap.

- (4) a. [Which paper] did John file ___ and Mary read ___ ?
b. [Which paper] did John file ___ before reading ___ pg ?

The two constructions have in common that movement from them is subject to certain restrictions. In coordinate structures such as (4a), movement out of either conjunct is severely ungrammatical, see (5a) and (5b). This led Ross (1967) to the formulation of the Coordinate Structure Constraint (CSC) in (6):

- (5) a. *Which paper did John file the mail and Mary read ___ ?
b. *Which paper did John file ___ and Mary read the mail?
(cf. Ross 1967, p. 158)

(6) The Coordinate Structure Constraint (Ross 1967, p. 161)

In a coordinate structure, no conjunct may be moved, nor may any element contained in a conjunct be moved out of that conjunct.

The two subparts of the CSC operate independently of one another (Grosu 1973): According to the Conjunct Constraint, no conjunct may be moved, and according to the Element Constraint, no element of a conjunct may be moved out of the coordinate structure (for an overview of the literature, see de Vries 2017). The syntactic and pragmatic dimensions of the CSC were recently discussed by Altshuler and Truswell (2022).

The crucial discovery made by Ross (1967) is that conjunct reduction in coordinate structures is nonetheless possible, however ‘[it] must work “across the board”’.⁵ In other words, the dependency is only grammatical if the filler is related to a gap in each conjunct (7).

(7) [Which paper] did John file ___ and Mary read ___ ?

Turning now to parasitic gaps, we find that a single gap in the adjunct clause (8) is likewise ungrammatical without a gap in the main clause, like under ATB in (5a):

(8) *Which paper did John file the mail before reading ___ ?

(cf. Engdahl 1983, p. 14)

At first glance, the scenario in (8) bears striking similarity to (5a): In (9), extracting from the main clause makes the ungrammaticality disappear, just like ATB-rule application did in (7):

(9) [Which paper] did John file ___ before reading ___ pg ?

However, there is one crucial difference between ATB structures such as (7) and parasitic gaps such as (9): In the former, it does not matter which conjunct is asymmetrically targeted by subextraction, the outcome is always bad, see (5). In (8), the root of the ungrammaticality is not that we asymmetrically extract from only one of the domains—extracting only from the main clause instead of the adjunct clause like in (10b) produces a well-formed utterance, and this well-formedness does not hinge on the presence of a gap in the adjunct clause. The ungrammaticality in (8) is caused by the gap in the adjunct clause, which is only allowed in the presence

⁵Ross (1967) has observed that this rule can be violated if the two conjuncts are semantically asymmetric. Lakoff (1986) defines the possible cases as denoting a sequence of events that are either (i) in line with conventionalized expectations (type A), (ii) run against them (type B) or (iii) are causatively linked (type C). See Neeleman and Tanaka (2024) for a series of experiments showing that the empirical profile of type A coordination is virtually the same as that of rationale clauses, supporting the claim that the non-initial conjunct in type A coordinate structures is an adjunct.

of a gap in the main clause, see (10a) vs. (10c).

- (10) a. [Which paper] did John file ___ before reading ___ pg ?
b. [Which paper] did John file ___ before reading the mail/it?
c. #[Which paper] did John file the mail/it before reading ___ pg ?
(cf. Engdahl 1983, p. 14)

In other words, while any gap appearing in only one conjunct of a coordinate structure is unacceptable, the licensing gap in parasitic gap constructions, also called the ‘host’, is perfectly acceptable without its ‘parasite’, i.e. the gap in the adjunct clause. Only the parasitic gap itself cannot ‘survive’ without its host, as shown in (10c). This is because parasitic gaps crucially appear in domains that are syntactic islands. The defining property of parasitic gaps is that the island violation caused by their presence is ameliorated by a suitable host in the main clause, viz. an independent gap created by a licit \bar{A} -movement dependency. Without the licensing \bar{A} -gap, such gaps are simply island violations as in (10c).⁶ A plausible antecedent (11a) or a wh-phrase alone (11b) do not suffice, and neither does A-movement.⁷

- (11) a. John filed a bunch of articles without reading * ___ pg /them.
b. John filed which articles without reading * ___ pg /them?
(Engdahl 1983, pp. 12, 14)

The fact that the parasitic gap depends on the main gap whereas the gaps in ATB are co-dependent seems to challenge the idea that PG constructions are syntactically related to ATB movement. Naturally, research was quick to start debating whether the similarities or differences should bear greater significance for the syntactic derivation of the two. Before we look at the different proposals, it has to be noted that the strict symmetry between the two gaps of ATB constructions does not remain unchallenged. Zhang (2010) reports that asymmetrical extraction from an island embedded in the non-initial conjunct of ATB structures (12b) is more acceptable than from an island embedded in the initial conjunct (12a). This suggests that the status of the two gaps may be different in ATB, making them in some sense similar to the main and parasitic gap in PG constructions. This observation has been disputed (de Vries 2017).

- (12) a. *Who did [Bill lose business [because he hired ___]] and [Mary praise

⁶Islands differ with respect to their strength: adjuncts, for example, are considered weak or selective islands, meaning that they permit the subextraction of arguments but not adjuncts. See Phillips (2006) for an overview of existing approaches to islandhood, ranging from purely syntactic or pragmatic to purely processing-based accounts.

⁷For this reason, the licensing of parasitic gaps is often used as a diagnostic tool itself to assess A/ \bar{A} -properties of dependencies. See section 3.1.1 of chapter 3 for a discussion of how this practice has affected the study of parasitic gaps in German.

- ___ a lot]?
- b. ?Who did [Mary praise a lot ___] and [Bill lose business because he hired ___ ?] (Zhang 2010, p. 226)

Recall that Ross (1967) made his observations about the CSC based on RNR, and we have only discussed ATB and parasitic gaps so far. The reason is that RNR does not involve any (overtly recognizable) movement—a property that ATB and PG constructions have in common, and a property that gives rise to reconstruction effects. Since this thesis deals with reconstruction observed under \bar{A} - and specifically *wh*-movement, we will set the study of RNR aside. The interested reader is referred to Barros and Vicente (2011), Belk, Neeleman, and Philip (2024), Georgi et al. (forthcoming), and Hein (2025) for some recent publications on RNR.

1.2 Existing approaches to multi-gap dependencies

Existing approaches are divided based on whether they assume a mutual derivation for ATB and parasitic gaps, and based on whether the filler is related to all or only one of the gaps via movement. Based on the latter criterion, I will distinguish between symmetrical and asymmetrical approaches in this thesis. Because the focus is principle C reconstruction which is claimed to reconstruct “all the way down” (Nissenbaum 2000; Sportiche 2017), I focus on whether an element is extracted from its base position in each gap—that is, I make the somewhat controversial choice to group sideward movement with the asymmetrical approaches, for reasons that are explained in the respective section.

1.2.1 Symmetrical approaches

1.2.1.1 Parallel extraction

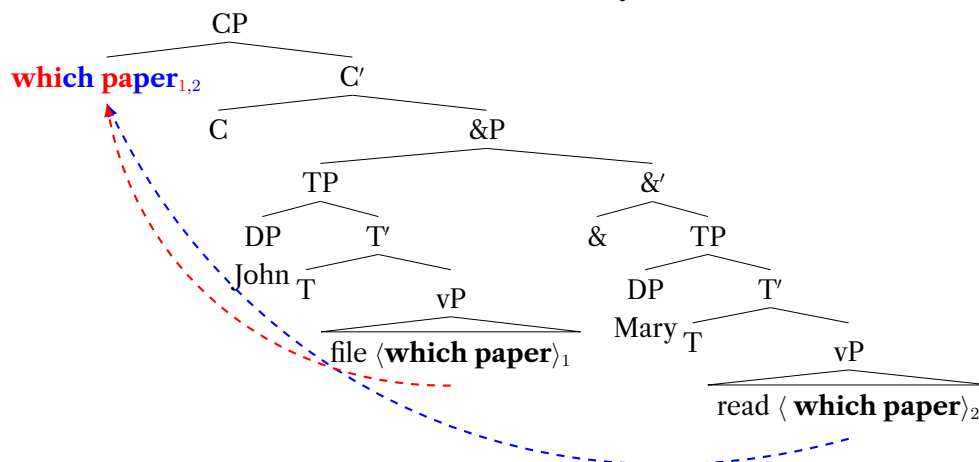
Parallel extraction, i.e. movement of an element from each gap site, is the earliest of all proposals to multi-gap dependencies (Ross 1967; Williams 1978). In most proposals, the two elements are extracted from each conjunct at the same time. There have been a number of accounts, some fairly recent, on how exactly a single phonological realization of the filler can be derived despite the movement of two distinct elements.⁸ As illustrated in (13), one approach is to postulate that the two elements fuse together at some point over the course of the derivation before

⁸In the following trees, red is used to mark the element originating in the initial conjunct/clause whereas blue is used for the element originating in the non-initial one.

1.2. Existing approaches to multi-gap dependencies

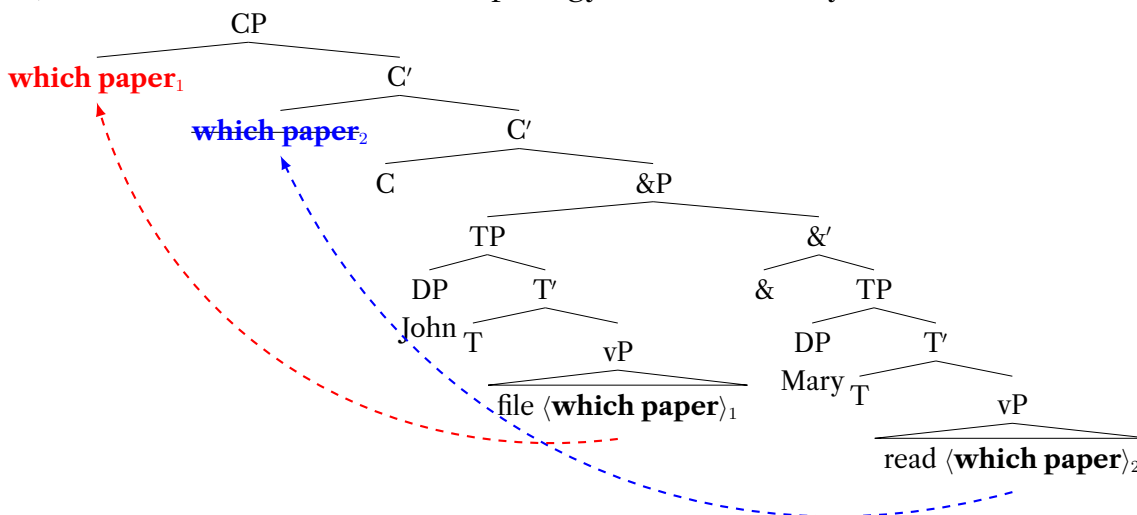
ultimately reaching SpecCP (Blümel 2014, 2017; Hein and Murphy 2020; Williams 1978).

(13) Parallel extraction and fusion analysis of ATB



On the other hand, it has been proposed that an instance of the filler moves from each conjunct to the left periphery independently with the one in the specifier of C being deleted—this reduction mechanism is called haplology reduction, depicted in (14) (Biskup 2018).

(14) Parallel extraction and haplology reduction analysis of ATB

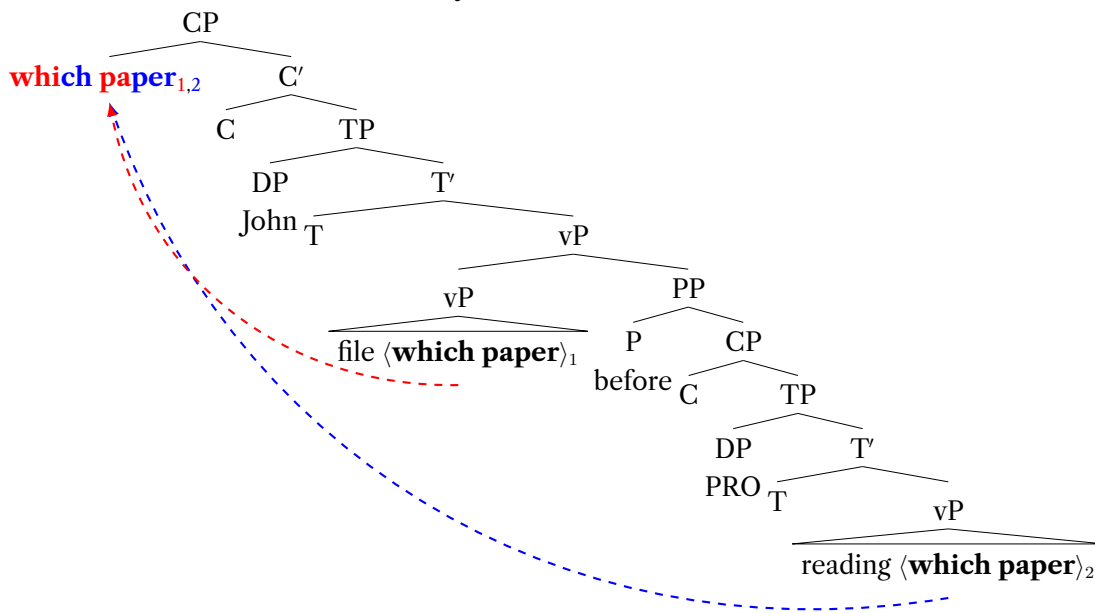


A very detailed and recent implementation comes from Hein and Murphy (2020). Complex specifiers are assembled in a space that is distinct from the one where the overall derivation is taking place (Nunes and Uriagereka 2000, p. 22) and thus, movement to a derived position as in ATB always requires an external syntactic workspace. The derivation starts with movement of an instance of the filler out of each conjunct into this external workspace. Here, the two elements are merged into one via feature set intersection. That is, the syntactic and semantic features

of the two copies are evaluated, and the elements are successfully merged if all of their features intersect. The features are decomposed into binary sub-features, and the most suitable form is then selected based on the subset principle. If no matching form can be found, the derivation crashes.

Parallel extraction approaches to parasitic gaps postulate that the underlying structure is coordinate and thus subject to the CSC. The details of the derivations diverge based on the languages they have been proposed for: Huybregts and van Riemsdijk (1985) only derive Dutch parasitic gaps via parallel extraction, and only when the subordinator is construed as a coordinator (given that the distribution of PGs in Dutch is different from English). For Haik (1985), English parasitic gap constructions always involve a coordinate structure and are always derived via parallel extraction, but crucially only at LF. Williams (1987), similarly to Huybregts and van Riemsdijk (1985), argues that subordinators can be construed as coordinators underlyingly, giving rise to a parasitic gap due to ATB-rule application. The parallel extraction approach to parasitic gaps is sketched in (15). Presumably, the P head should act as a coordinate head in this structure, although the exact syntactic implementation is left open.

(15) Parallel extraction analysis of PGs



Although Williams (1987) does not provide any empirical arguments in favor of the underlying coordinate construal, Fanselow (2001) argues that some adverbials introducing adjunct clauses in German are similar to coordinators.⁹

⁹The reason I do not include Fanselow (2001) among the listed proposals is that the paper does not actually make any claims about the derivation of parasitic gaps. See section 3.1 of chapter 3 for a discussion of German parasitic gaps as well as Szarvas (2024b) for a discussion of the problems that come with this ‘coordinate construal’.

A very recent proposal based on Williams (1987) comes from Seguin and Thoms (2025). The assumption is that ATB constructions are derived via movement in the shape of a forking chain. Parasitic gap constructions have an added twist under this analysis—they feature an operator constructions in the adjunct clause, such as under operator movement analyses presented in section 1.2.2. However, there is also a full DP within this operator construction, which is smuggled to the edge of the adjunct clause alongside with it (Hicks 2009 based on Collins 2005). This smuggling to the left edge then allows the DP to move out of the adjunct clause via parallel extraction with the DP in the main clause. As such, the operator simply acts as a vessel to get the constituent to the left edge.

Early parallel extraction approaches, regardless of whether they are aimed at ATB only or also parasitic gaps, predict that the two gaps in a multi-gap dependency should equally reflect the morphosyntactic and semantic properties of the filler. Morphological mismatches between the gaps are not predicted by earlier approaches, while Hein and Murphy (2020) specifically develop their ATB analysis so that it allows for case mismatches between the gaps under syncretism. Under Seguin and Thoms’s analysis, the DP is smuggled to the edge of the adjunct clause and also undergoes ATB movement before being assigned case, therefore allowing for case mismatches between the main gap and the parasitic gap.¹⁰ Both gaps are predicted to show reconstruction effects under all existing accounts, equally for scope, variable binding and binding principles.

1.2.1.2 Multidominance

Based on Chomsky (2001), there are two distinct types of Merge: External and Internal Merge. External Merge takes two independent syntactic objects and joins them into one. Internal Merge, on the other hand, takes two subparts of a single syntactic element and joins them, thus yielding the effect of Move (see section 1.2.2.3 for Nunes’ criticism of the ambiguity of Merge in Chomsky 1995). Citko (2005) argues that the existence of Internal and External Merge predicts a third type of operation combining the properties of the two, called ‘Parallel Merge’ (16).

(16) Parallel Merge (Citko 2005, p. 476)

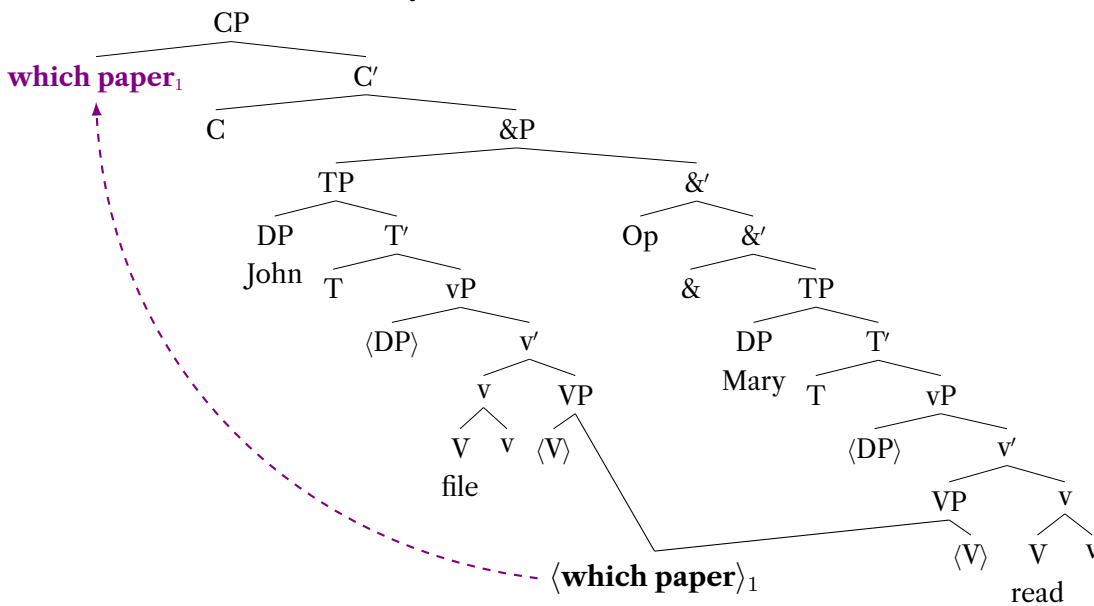


Like External Merge, it takes two distinct syntactic objects, XP and Y, but like

¹⁰The varying tolerability of case mismatches and syncretism repair effects are acknowledged, and the authors hypothesize that this type of Hicks-style operator movement is allowed only in certain syntactic environments and possibly certain languages. That is, languages that allow for mismatches do so because they allow for smuggling in an operator construction, whereas languages that do not tolerate mismatches do so because they lack the mechanism.

Internal Merge, it merges Y with the subpart of XP rather than XP's root. Thus, the element Y is additionally dominated by the node XP, while maintaining its original position under domination of node ZP. Under this operation, movement is redefined as involving no elimination of the element in its old position, but rather as merging the element into a new position in addition to its old one.¹¹ Citko specifically proposes ATB movement to derive in this fashion since Parallel Merge straightforwardly establishes the one-to-many relation absent from standard syntactic assumptions so far: The single ATB moved element can be dominated by multiple nodes simultaneously. The constituent is Parallel Merged with each VP in the two conjuncts and subsequently moved to SpecCP (17).¹²

(17) Multidominant analysis of ATB



With respect to the classification of approaches given in the introduction, there is only one wh-phrase present in the derivation, yet, by virtue of being dominated by two nodes, it simultaneously occupies both gap sites and thus is likewise technically extracted from both. As such, multidominance is considered a symmetric approach. Aside from multi-gap dependencies, wh-questions with conjoined wh-phrases, serial verb constructions, as well as free relatives have been proposed to originate from multidominant structures. Thus, despite making use of the special operation Parallel Merge, multidominance has a fairly wide empirical scope and has been proposed to cover a range of phenomena beyond the currently investigated ones. Nevertheless, the special nature of multidominant structures poses a potential problem: Parallel Merge runs into linearization issues due to the Linear

¹¹It is acknowledged that this theory needs to be supplemented with a proposal on where the targeted element is pronounced and interpreted, see Gračanin-Yüksek (2013) and Johnson (2018).

¹²I am using purple in this particular tree to illustrate that there are no unique initial and non-initial positions but rather a single position that is dominated by a node in each conjunct.

Correspondence Axiom by Kayne (1994). According to this axiom, the linear order of the terminal nodes in a syntactic structure is determined based on asymmetric c-command relations. Because the shared element in several of the configurations remains in situ, i.e. it is dominated by two distinct nodes at the same time in the final syntactic structure, it is unclear in which position exactly it should be spelled out. Solutions to this problem fall into two categories: authors have proposed to either (i) modify the linearization algorithm (Goodall 1987; Moltmann 1992), or (ii) alter the definition of c-command exclusively in multidominant structures (Bachrach and Katzir 2009; Wilder 1999). Overall, there have been several proposals how to tackle the linearization problem, although at the cost of proposing construction-specific mechanisms that are not applicable elsewhere in grammar. For our current purposes, no linearization problem arises under ATB movement since the *wh*-phrase ultimately moves from its multiply dominated position to SpecCP.

Multidominant analyses of parasitic gaps exist as well, but to my knowledge, have not received a lot of attention (Citko 2013; Kasai 2010). For ATB, the analysis predicts case matching, with Citko (2005) arguing for the possibility of repair by syncretism (see section 1.3.1.1 for details). Citko (2013) additionally argues that the size of the shared constituent derives why parasitic gaps exhibit case proximity effects, while ATB requires case matching or syncretism (see section 1.3.1.1 for a discussion of the data). Reconstruction effects should appear in both gaps since the element has a trace in the scope of binders in both conjuncts. Kasai (2010) pursues an explanation of anti-reconstruction effects in the parasitic gap site based on split lexical insertion and Wholesale Late Merger (Agbayani and Ochi 2007 and Takahashi and Hulsey 2009, respectively). The predictions of the approach get a bit more complicated when considering variable binding reconstruction, as will be briefly discussed in section 5.2 of chapter 5.

1.2.2 Asymmetrical approaches

1.2.2.1 Ellipsis

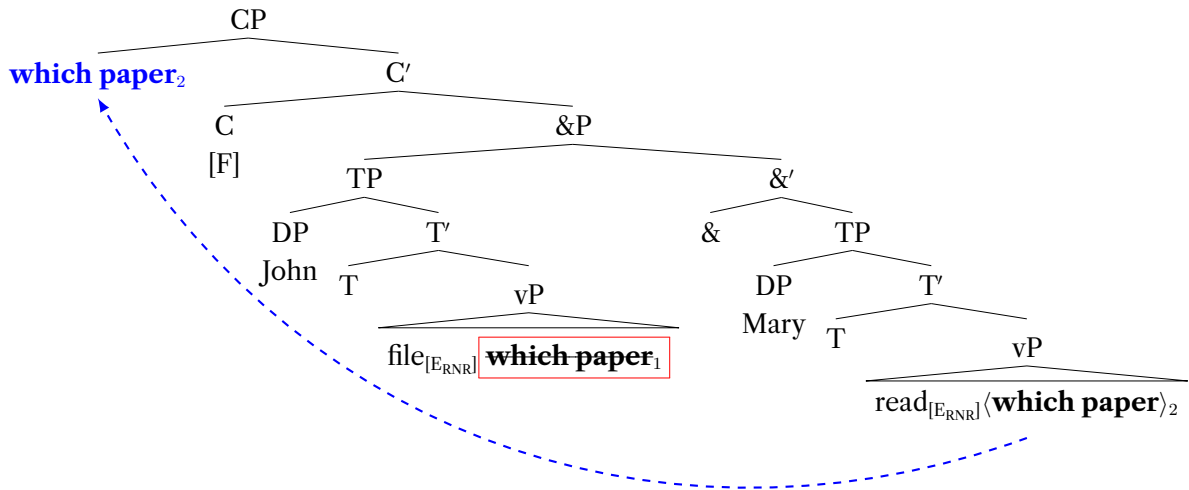
Ellipsis approaches have been proposed for ATB movement only. Under such approaches, the shared constituent is base generated in each conjunct. In the backward ellipsis approach proposed by Ha (2008), it is the constituent in the initial conjunct that is elided, with the constituent in the non-initial conjunct undergoing successive-cyclic movement to SpecCP.¹³ The operation is licensed by an E_{RNR}

¹³Williams (1978) mentions such an analysis but discards it because only ATB can target subjects:

- (i) a. Who does John like and we hope will win?
- b. *John likes, and we hope will win, Peter.

feature in each conjunct on the contrastively focused constituent preceding the material to be elided in the initial conjunct and its antecedent in the non-initial conjunct which will ultimately undergo movement—that is, the ellipsis feature is born by the verbs as sketched in (18) if the operation targets the direct object. Semantic identity is required between the elided constituent and its antecedent based on an eGIVENness feature (Merchant 2001).

(18) Backward ellipsis analysis of ATB

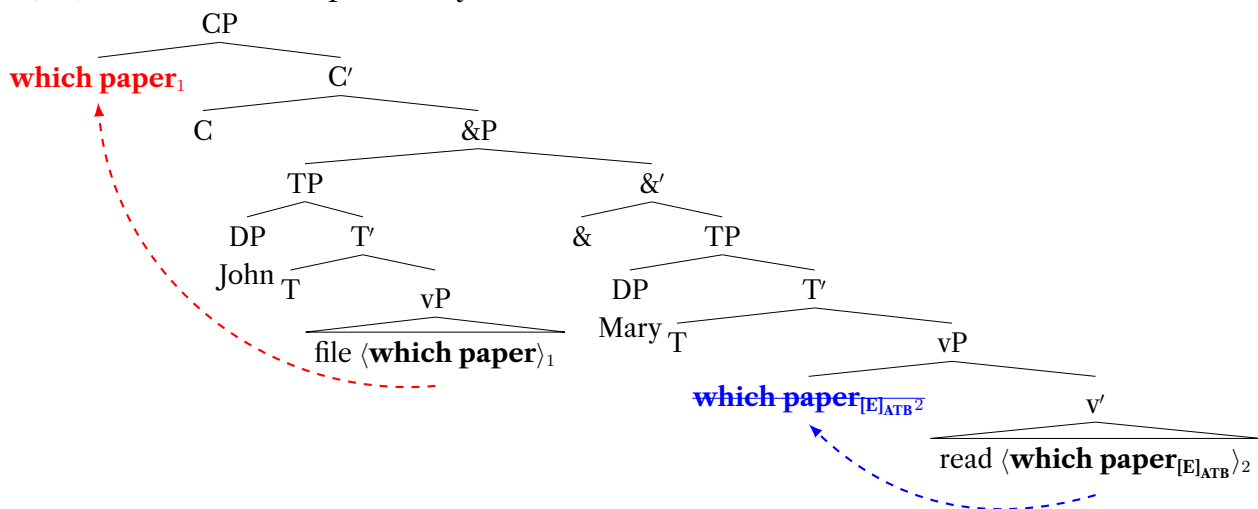


Due to constituents in both conjuncts bearing an E_{RNR} feature, locality determines which of them is elided: A head outside of the coordinate structure (in this case C) bears an F feature and probes for a goal with an E_{RNR} feature. Agreement is established between C and the constituent in the initial conjunct since it is the closest goal. As a result, the constituent in the initial conjunct is elided, and the one in the non-initial conjunct moves to SpecCP via successive-cyclic movement. The probe will never target the constituent in the non-initial conjunct because the constituent in the initial conjunct is always the closer goal. Asymmetrical extraction from coordinate structures is argued not to be an issue since the CSC is treated as an interpretive, not a syntactic constraint (see also Ruys 1992 and Fox 2000).

Under the forward ellipsis approach to ATB proposed by Salzmann (2012), it is the constituent in the non-initial conjunct that is elided, with the constituent in the initial conjunct moving to SpecCP. Here, the Agree relation licensing ellipsis is established between the &-head and an E_{ATB} feature on the targeted constituent in the non-initial conjunct. The two conjuncts are first built in two separate work spaces with the wh-phrase moving to the left edge within each conjunct. Then, the &-head is merged with the non-initial conjunct and agrees with the E_{ATB} feature on the wh-phrase, triggering ellipsis. The initial conjunct is then merged in the

specifier position of &. This is followed by movement of the wh-phrase in the initial conjunct to SpecCP. The derivation is sketched in (19).

(19) Forward ellipsis analysis of ATB



At PF, the chain in the initial conjunct is reduced via deletion of non-top copies. Therefore, the chain in the non-initial conjunct remains unpronounced and only the copy in SpecCP is realized. At LF, the Preference Principle applies: Intermediate copies are deleted, the operator binds both its own lower copy in the initial conjunct as well as the copy in the non-initial conjunct by virtue of them bearing the same index (cf. Chomsky 1995, p. 209). In this derivation, too, the CSC is treated as an LF constraint: Since bound variables are of the same syntactic type, no violation occurs. Salzmann (2012) further argues that, although PF deletion also gives rise to the correct surface structure and LF, ellipsis is in fact necessary to ensure a well-formed syntax. Since the sentence has a single question reading, there must be only one C, the C whose specifier the constituent from the initial conjunct moves to in order to get its μ F feature checked. If the structure is shipped off to the interfaces, the derivation will crash due to the unchecked μ F feature on the wh-phrase in the non-initial conjunct. If it is elided instead, the unchecked feature will be elided with it and the derivation will not crash.

VP-ellipsis is well known to allow for morphological mismatches between the antecedent and the elided constituent. Thus, approaches based on ellipsis predict case mismatches to be possible with essentially no restrictions. The approaches furthermore predict that both conjuncts should exhibit reconstruction effects since the constituent is present in both conjuncts following the recoverability requirement on ellipsis. Crucially, Salzmann (2012) argues that ellipsis predicts the possibility of circumventing reconstruction effects in the ellipsis site, i.e. in the initial conjunct in the case of backward and the non-initial conjunct in the case of forward ellipsis due to vehicle change (Fiengo and May 1994). Under vehicle

change, R-expressions and anaphors can correspond to pronouns in the ellipsis site. Reconstruction effects for principles A and C should therefore only occur in the gap site the element is extracted from, but crucially not the ellipsis site.

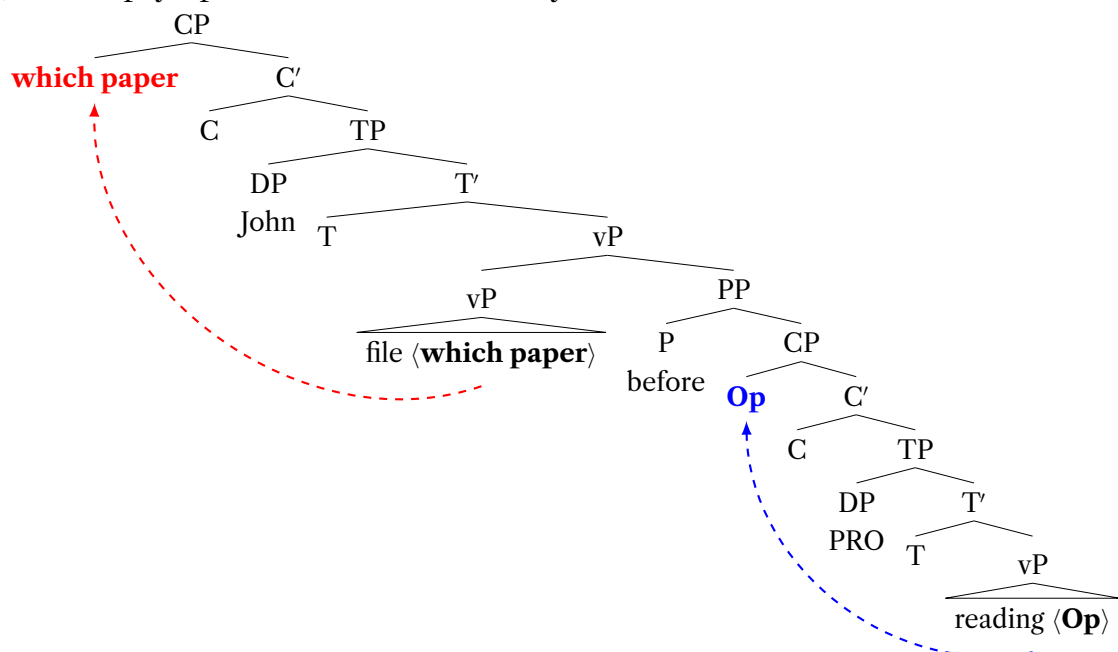
1.2.2.2 Empty operator movement

An approach involving movement of a phonologically empty operator to create the non-initial gap was first proposed for parasitic gaps by Chomsky (1986). The idea goes back to an observation by Kayne (1983) that parasitic gaps are unacceptable when they are embedded in islands, which is indicative of movement. This is shown in (20), where in (20b) the parasitic gap is embedded in a relative clause within the adjunct clause:

- (20) a. the article [which we filed ____ [without reading _____{pg}]]
 b. *the article [which we filed ____ [without meeting the person [who wrote _____{pg}]]] (Culicover 2001, p. 27)

Chomsky (1986) proposes that there is an empty operator moving within to the vP-adjoined adjunct clause to its left edge, but crucially not beyond. This derives that (20a) is fine since the operator can move freely within the adjunct clause, whereas (20b) is out because the operator has to cross the boundary of the relative clause in order to get to the left edge of the adjunct clause, violating the relative clause island. The filler spelled out on the surface originates in the gap site of the main clause and undergoes movement to SpecCP. That is, the main gap and the parasitic gap are created independently of one another.

- (21) Empty operator movement analysis of PGs



1.2. Existing approaches to multi-gap dependencies

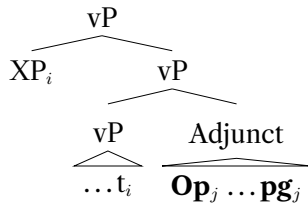
The interpretability of the parasitic gap is ensured via Chain Composition, see (22), a process applying at S-structure. The conditions under which this operation is licensed are ultimately left open.

(22) Chain Composition (Chomsky 1986, p. 56):

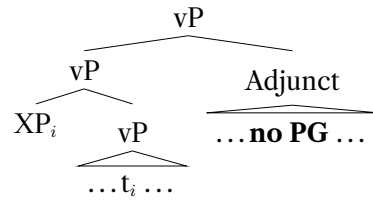
If $C = (\alpha_1, \dots, \alpha_n)$ is the chain of the real gap, and $C' = (\beta_1, \dots, \beta_n)$ is the chain of the parasitic gap, then the composed chain is $(C, C') = (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n)$.

Nissenbaum (2000) argues that parasitic gaps are obligatory whenever the syntactic structure in (23) arises: A vP has an outer specifier formed by successive-cyclic movement of the filler leaving behind the main gap, and an inner adjunct in which empty operator movement gives rise to a parasitic gap (following Chomsky 1986). The optionality of parasitic gaps is derived by the alternative structure in (24) where the scope of the specifier and the adjunct is reversed:

(23) Outer Spec, inner adjunct



(24) Inner Spec, outer adjunct



The generalization is thus that a parasitic gap must appear in the adjunct whenever a DP in the specifier of vP scopes over it. Evidence for this claim comes from Heavy NP Shift (HNPS), extraposition and stacked vP adjuncts. In the case of the latter, a parasitic gap can be present in both (25a), neither (25b) or only the innermost of the adjunct clauses (25c), but crucially not only in the outermost of the adjunct clauses (25d):

- (25) a. Who did you praise ___ to the sky [after criticizing ____{pg}] [in order to surprise ____{pg}]?
 b. Who did you praise ___ to the sky [after criticizing him] [in order to surprise the poor man]?
 c. Who did you praise ___ to the sky [after criticizing ____{pg}] [in order to surprise him]?
 d. *Who did you praise ___ to the sky [after criticizing him] [in order to ____{pg}]?
 (Nissenbaum 2000, p. 547)

Nissenbaum argues that this pattern follows from how the possible configurations in (23) and (24) can be extended. The specifier could be above both adjuncts, i.e. if there is another inner adjunct merged into (23), and the result would be (25a). If the

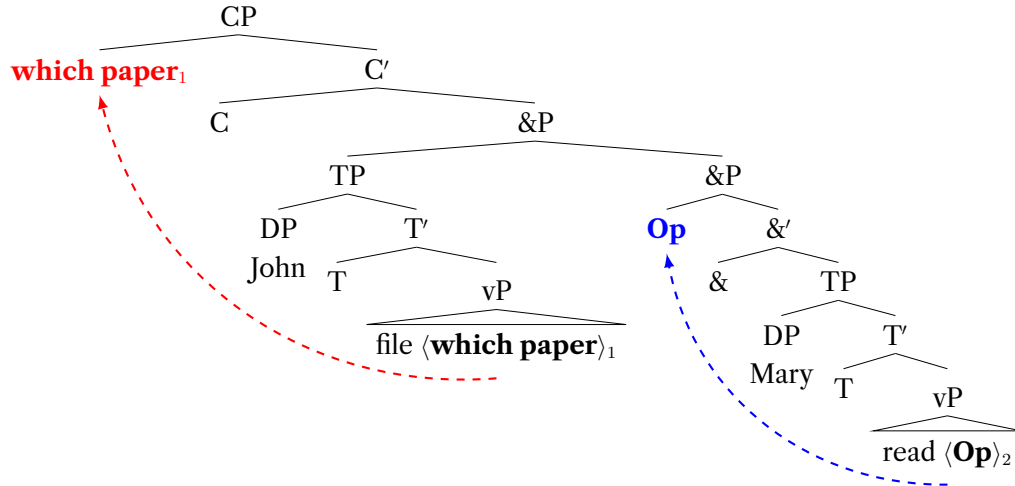
specifier is below both adjuncts, i.e. in a configuration like (24) with an additional adjunct on top, the consequence would be (25b). In the third scenario, the specifier is above the inner but not the outer adjunct, i.e. if there is another adjunct stacked on top of the configuration in (23), and the result is (25c). Crucially, there is no configuration in which the specifier can take scope over the outer adjunct without also taking scope over the inner adjunct. According to Nissenbaum (2000), this derives the observation that (25d) is ruled out: All adjuncts in the scope of SpecvP must contain a parasitic gap, while all adjuncts that outscope the specifier must not.

It is further argued that independently needed semantic mechanisms suffice to account for the parasitic gap's semantic interpretation, thereby alleviating the need for an additional semantic rule like (22). The assumption that an empty operator binds the parasitic gap and scopes over the island containing it entails that the adjunct is interpreted as a one-place predicate rather than a clause. If clausal adjuncts without parasitic gaps normally modify clausal vPs, there must be an alternative way in which a predicate (i.e. the null operator structure containing a parasitic gap) can also compose with the vP. This semantic composition, like the syntactic licensing of a parasitic gap, is supposedly enabled by movement of the filler to the outer specifier position of vP. The consequence is that the lower segment of the vP will be turned into a derived predicate and interpreted as a lambda abstract that binds a variable in the position of the gap. The two predicates—the adjunct and the vP—compose via predicate modification, and the result applies to the raised XP by function application. Nissenbaum's work has far reaching syntactic consequences: He further presents an explanation for Engdahl's generalization as well as its exceptions, i.e. that in certain cases, *wh*-in-situ licenses parasitic gaps as well, and argues based on these exceptions that overt syntactic operations must precede covert ones.

We have seen in section 1.2.1.1 that several authors had proposed parasitic gaps to be an instance of ATB-rule application, postulating the adjunction structure to be coordinate in some sense. Notably, the opposite has likewise been proposed, namely that ATB movement involves a phonologically null operator like parasitic gaps have been argued to (Munn 1992, 1993). A sketch of the derivation in a coordinate structure is provided in (26). Munn argues for the unification of parasitic gaps and ATB under a mechanism that is independently required in grammar. He argues that this avoids two problems: First, that of postulating parasitic gaps to involve coordination, and second, the idiosyncratic formalism that is ATB-rule application. Here as well, the CSC is argued to impose semantic identity on the conjuncts and not to be a constraint on movement. The approach is further refined in Munn (1994) where it is argued that null operators and their traces are copies

of the elements that license them, which is one of the necessary requirements to account for the complex reconstruction facts (in addition to postulating that principles A and C apply at different levels of representation, see section 1.3).

(26) Empty operator movement analysis of ATB



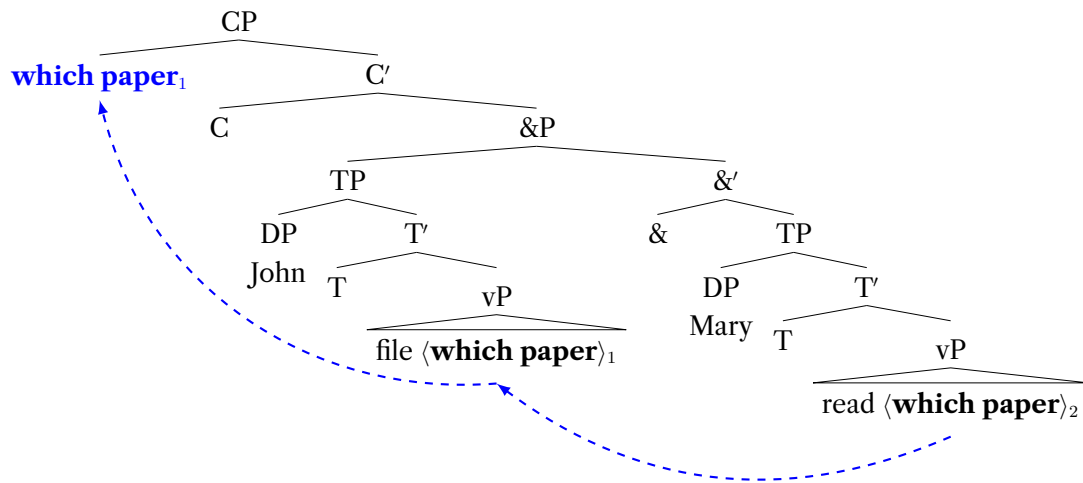
The predictions of empty operator approaches diverge based on the properties of the operator. Under the approaches proposed to account for parasitic gaps only (Chomsky 1986; Nissenbaum 2000), the operator and the parasitic gap it gives rise to do not share the morphosyntactic properties of the filler that they are bound by. That is, these approaches predict morphological mismatches to be possible and no reconstruction to the parasitic gap site, whereas the licensing gap in the main clause is a regular trace of \bar{A} -movement. Under the account given by Munn (1994), the empty operator behaves like a trace of the element that licenses it, i.e. the operator as well as its trace share the properties of the main gap. Under Munn's approach, both gaps in ATB and parasitic gaps should show reconstruction effects, and case matching is likewise expected.

1.2.2.3 Sideward movement

We have seen in section 1.2.1.2 that solving the puzzle of multi-gap dependencies can give rise to innovative proposals, such as Parallel Merge by Citko (2005). Arguably, it was Nunes (1995) who paved the way to rethinking syntactic movement within a Minimalist theory of grammar. Nunes proposes a modified theory of syntactic movement called 'Copy + Merge' (Nunes 2001, 2004). According to the Minimalist Program, Merge is a concatenative binary transformation of two disconnected syntactic objects. Move, on the other hand, is an operation on two subparts K and α of a single syntactic object Σ , first targeting K, then raising α , and then merging it to K, resulting in Σ' (Chomsky 1995, p. 250). Nunes argues that Move has several conceptual problems: One of them is the lack of motivation for

the phonological deletion of lower copies. The other one is the fact that Merge is taken to exceptionally operate on *two subparts of a single syntactic object* within the definition of Move instead of *two independent syntactic objects* like it does elsewhere in the computational system. Nunes proposes to reframe Move as the interaction between the independent operations Copy, Merge, Form Chain and Chain Reduction instead of a primitive of operation in its own right. Getting rid of the dedicated definition of Move eliminates the ambiguity of Merge, and effects that movement necessarily targets two independent syntactic objects rather than two subparts of a single syntactic object.

(27) Sideward movement analysis of ATB



This combination of operations results in what is called sideward movement: A constituent α of a syntactic object K is copied (into the workspace) and then merged with a syntactic object L, which is assembled independently and bears no connection to the syntactic object K. Instead of a constituent moving *upward* to a higher position within a tree, it moves *sideward* to another tree. Crucially, according to Nunes (2001), this rephrasing of Move straightforwardly allows for the derivation of ATB movement (27) as well as parasitic gaps (32). Both derivations follow the same logic. For ATB, the syntactic objects K (28b) and L (28c) are assembled with elements from the numeration N (28a). We will focus on examples (28c)-(28f) which track the assembly of the initial conjunct. The wh-phrase, in an instance of sideward movement, is copied and merged with L (28c), resulting in M (28d). Through an application of Select and Merge, the subject and the v-head are introduced in (28e) from the numeration. T is present only once in the numeration, therefore it is copied and merged from K (28b) to O (28e) through another sideward movement step, resulting in P (28f).¹⁴

¹⁴The superscripts indicate the order in which the copies are created: 1 indicates the original object from the numeration, 2 and 3 indicate the copies created through sideward movement.

1.2. Existing approaches to multi-gap dependencies

- (28)
- a. $N = \{\text{which, paper, did, } Q, \text{John, } v_2, \text{file, and, Mary, read}\}$
 - b. $K = [_{TP} \text{did}^1 [_{VP} \text{Mary } v [_{VP} \text{read} [\text{which paper}]^1]]]$
 - c. $L = \text{file}$
 - d. $M = [_{VP} \text{file} [\text{which paper}]^2]$
 - e. $O = [_{VP} \text{John } v [_{VP} \text{file} [\text{which paper}]^2]]$
 - f. $P = [_{TP} \text{did}^2 [_{VP} \text{John } v [_{VP} \text{file} [\text{which paper}]^2]]]$ (Nunes 2001, p. 337)

The subjects of K and P move to the SpecTP positions in their structure and the $\&$ -head is merged with one of them in the following step, resulting in R (29a) and S (29b). R and S merge to create U (28c), then, the final object from the numeration, the interrogative complementizer Q , merges with the resulting object U to form W (29d), followed by another instance of sideward movement of *did* and *which paper*, respectively, to check Q 's features (29e).

- (29)
- a. $R = [_{\&P} \text{and} [_{TP} \text{Mary did}^1 \text{read} [\text{which paper}]^1]]]$
 - b. $S = [_{TP} \text{John did}^2 \text{file} [\text{which paper}]^2]$
 - c. $U = [_{\&P} [_{TP} \text{John did}^2 \text{file} [\text{which paper}]^2] [_{\&} \text{and} [_{TP} \text{Mary did}^1 \text{read} [\text{which paper}]^1]]]$
 - d. $W = [_{CP} Q [_{\&P} [_{TP} \text{John did}^2 \text{file} [\text{which paper}]^2] [_{\&} \text{and} [_{TP} \text{Mary did}^1 \text{read} [\text{which paper}]^1]]]]]$
 - e. $[_{CP} [\text{which paper}]^3 \text{did}^3 + Q [_{\&P} [_{TP} \text{John did}^2 \text{file} [\text{which paper}]^2] [_{\&} \text{and} [_{TP} \text{Mary did}^1 \text{read} [\text{which paper}]^1]]]]]$ (Nunes 2001, p. 337)

The notions of initial and non-initial conjunct only gain meaning once the $\&$ -head is merged. Which object the head merges with does not appear to be governed by any specific criteria. Finally, the chains in (30) are formed such that the highest copies of *which paper* and *did* are linked to the lower copies in individual chains. Since neither of the lower copies c-commands the other, neither of them is closer to the topmost copy, allowing it to form a chain with each of them.

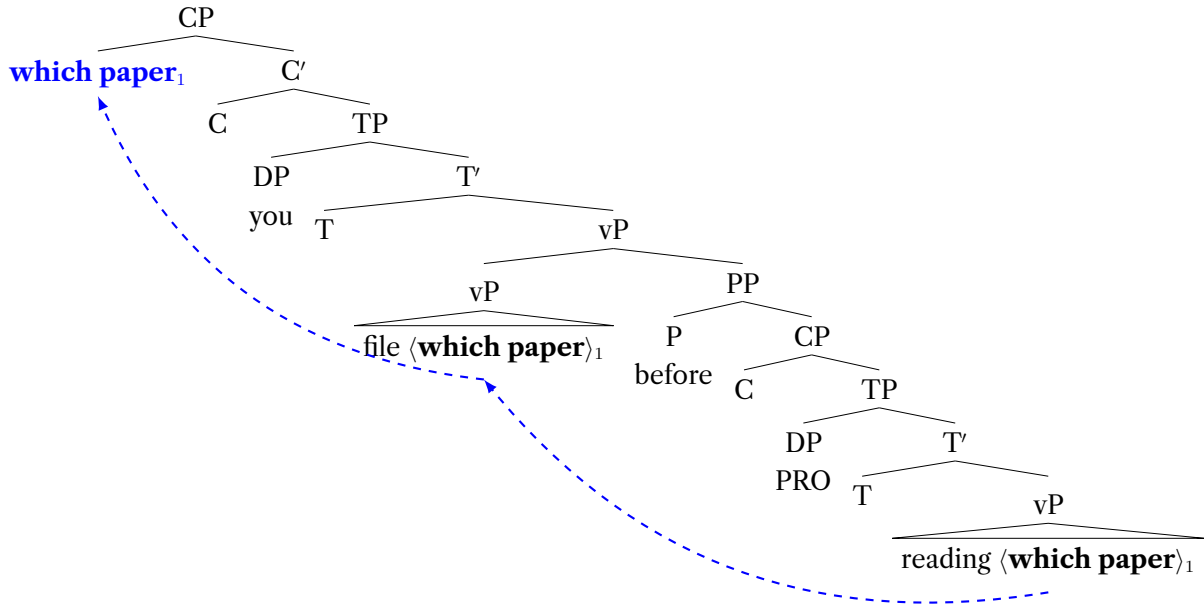
- (30)
- a. $CH_1 = ([\text{which paper}]^3, [\text{which paper}]^1)$
 - b. $CH_2 = ([\text{which paper}]^3, [\text{which paper}]^2)$
 - c. $CH_3 = (\text{did}^3, \text{did}^1)$
 - d. $CH_4 = (\text{did}^3, \text{did}^2)$

The chains are linearized in accordance with asymmetry and irreflexivity (Kayne 1994). Only the topmost copies of the chain are pronounced (31) due to considerations of economy. This principle of economy is based on the number of checking relations a copy is associated with. Lower copies bear unchecked features which can be eliminated at PF by Formal Feature Elimination (Nunes 1995, p. 291). Subsequently, retaining a copy with checked features will avoid this

operation to apply, thus making the derivation less costly.

(31) $[_{CP}[\text{which paper}]^3 \text{ did}^3 + Q [_{\&P}[_{TP} \text{ John did}^2 \text{ file } [\text{which paper}]]^2 [_{\&'} \text{ and } [_{TP} \text{ Mary did}^1 \text{ read } [\text{which paper}]^1]]]]]$

(32) Sideward movement analysis of PGs



The derivation of parasitic gaps follows the exact same logic, with a sketch of the tree provided in (32). First, the adjunct clause is assembled as a syntactic object K (33b) along with the independent syntactic object L (33c) containing only the matrix verb from the numeration N (33a). In (33d), we see the application of sideward movement of *which paper* from (33b). The P-head is merged with K (33b), resulting in O (33e). This is followed by merging the subject of the main clause with M (33d), resulting in R (33f). Finally, the subject of the main clause and the T-head of O (33e) move, and the interrogative complementizer Q from the numeration is merged with the result, yielding (33g). The strong wh-feature on Q triggers copying of *which paper* to SpecCP, yielding (33h).

- (33) a. $N = \{\text{which}_1, \text{paper}_1, Q_1, \text{you}_1, \text{did}_1, v_2, \text{file}_1, \text{before}_1, C_1, \text{PRO}_1, T_1, \text{reading}_1\}$
 b. $K = [_{CP} C [_{TP} \text{PRO}_j [_{T'} T [_{vP} t_j [_{v'} v [_{VP} \text{reading } [\text{which paper}]^i]]]]]]]$
 c. $L = \text{file}$
 d. $M = [_{VP} \text{file } [\text{which paper}]^i]$
 e. $O = [_{PP} \text{before } [_{CP} C [_{TP} \text{PRO}_j [_{T'} T [_{vP} t_j [_{v'} v [_{VP} \text{reading } [\text{which paper}]^i]]]]]]]]]$
 f. $R = [_{vP} \text{you } [_{v'} v [_{VP} \text{file } [\text{which paper}]^i]]]$
 g. $S = [_{CP} \text{did} + Q [_{TP} \text{you } [_{vP} [_{vP} \text{file } [\text{which paper}]^i] [_{PP} \text{before PRO reading } [\text{which paper}]^i]]]]]$

- h. $[_{CP} [\text{which paper}]^i \text{ did} + Q [_{TP} \text{ you } [_{VP} [_{VP} \text{ file } [\text{which paper}]^i]]_{PP} \text{ without PRO reading } [\text{which paper}]^i]]]$ (Nunes 2001, p. 325)

Here, too, the highest copy of *which paper* forms a chain with each lower copy, and deleting the lower copies at PF is the most economical option. The existence and binding of parasitic gaps is thus a natural consequence of the chain reduction mechanism employed here: A chain containing a parasitic gap as one of its links must undergo chain reduction like any other chain in order to be linearized. Parasitic gaps are therefore treated to be the same as a trace of any other movement chain—their creation via PF-deletion of the lowest copy of the respective chain is more economical than the deletion of another link of the chain, i.e. the topmost copy. If the parasitic gap does not form a chain with a higher copy, it will not be deleted and violate the asymmetry and irreflexivity conditions on linear order, and the derivation will crash.

Nunes discusses one of the most obvious questions that this analysis raises, namely how the lowest copy of the *wh*-phrase can originate in the parasitic gap site without causing an island violation. Crucially, the adjunct clause in (33b) is not yet an adjunct (and therefore not yet an island) when the *wh*-phrase is sideward moved into the object position of *file* in (33c). Note that the tree sketches provided in (27) and (32) are thus only partially correct in that the initial movement step from the non-initial to the initial gap site takes place before the conjuncts or clauses are joined as one syntactic object.

Given that the sideward movement approach precisely aims to rephrase the notion of syntactic movement, deriving predictions from it for syntactic phenomena that are rooted in the understanding of movement as internal Merge is tricky without disfiguring the core idea of the approach. One attempt to is to treat sideward movement as successive-cyclic movement as suggested by Nissenbaum (2000). There is a single *wh*-phrase base generated in what ends up as the non-initial gap site, from which it moves to the initial gap site and ultimately to SpecCP. Under this treatment, the predictions of Nunes' account are fairly straightforward: If the approach is on the right track, reconstruction phenomena targeting the two gaps in multi-gap dependencies should exhibit the same pattern as reconstruction phenomena in long \bar{A} -dependencies. Sideward movement predicts symmetrical effects for scope, idioms and variables, but things get a bit more tricky with respect to binding principles. According to Sportiche (2017), an anaphor can reconstruct into any of its previous positions, meaning that under sideward movement, a suitable binder *c*-commanding either the initial or non-initial gap in ATB and parasitic gap constructions should satisfy principle A with no difference between the two configurations. For principle C, Sportiche

observes that it reconstructs all the way down, meaning that it may be violated in intermediate positions, but crucially not the base position of the constituent. In other words, sideward movement predicts an asymmetrical pattern yielding a violation of principle C only if the non-initial gap is c-commanded by a matching pronoun. Regarding morphological mismatches, since there is only a single element which is sideward moved, there should be strict case matching effects.

This concludes the summary of the most popular approaches to multi-gap dependencies. The theories were classified based on the relationship they postulate between the filler and each gap. Within this division, we have seen approaches maintaining that the same mechanism gives rise to ATB as well as parasitic gaps, but also ones that treat the two constructions to be entirely unrelated. In the next section, we discuss some of the evidence most frequently cited in support of the theories.

1.3 Evidence

The previous sections gave an overview of the most popular approaches to ATB movement and parasitic gaps. We shall now turn to how they hold up against the evidence reported in the literature so far. The focus is on two different types of evidence: The case marker borne by the filler as well as reconstruction effects. While the former type of evidence targets the question about the morphosyntactic complexity of the underlying representations in each gap, the latter is informative about the number and type of movement steps involved in the derivation. This is precisely where the predictions of the approaches vary. The aim of this section is to illustrate that the evidence paints an inconsistent picture in multiple dimensions: (i) a test is reported to yield distinct outcomes across languages, (ii) a test is reported to yield distinct outcomes within a language depending on the author, and (iii) different tests are reported to yield outcomes that are inconsistent with either single approach within the same language. As such, this section motivates the need for more controlled means of data collection.

1.3.1 Case matching, repair by syncretism and beyond

1.3.1.1 Case under ATB movement

Based on data from Polish, it was noted that ATB movement appears to be subject to a case matching effect as shown in (34) (Bondaruk 2003; Citko 2005; Dylą 1984; Franks 1993).

1.3. Evidence

- (34) a. Kogo_{ACC} Jan lubi_{ACC} ____ a Maria podziwia_{ACC} ____ ?
 who.ACC Jan likes and Maria admires
 ‘Who does Jan like and Maria admire?’
 b. *Kogo_{ACC}/Komu_{DAT} Jan lubi_{ACC} ____ a Maria ufa_{DAT} ____ ?
 who.ACC/DAT Jan likes and Maria trusts
 ‘Who does Jan like and Maria trust?’ (Citko 2005, p. 485)

In (34a), the verbs *lubi* (‘likes’) and *podziwia* (‘admire’) both assign accusative case to their direct object. As a result, the filler bears accusative case—so far so good. In (34b), *lubi* is replaced by the verb *ufa* (‘trusts’) which assigns dative case. The filler is thus assigned conflicting cases in each conjunct. This mismatch is unresolvable as neither an overt accusative nor an overt dative case marker can rescue the construction. However, the mismatch is repairable by syncretism as shown in (35a). The phonological realization of *kto* (‘who’) in the accusative assigned by *lubi* and the genitive assigned by *nienawidzi* (‘hates’) is the same, i.e. *kogo*. (35b) illustrates that the acceptability of (35a) is not due to the morphological feature specification of the accusative and the genitive. If we use a *wh*-phrase with two distinct phonological realizations of the accusative and genitive, such as *co* (‘what’), the mismatch becomes unacceptable, just like the mismatch between the dative and the accusative in (34b):

- (35) a. Kogo_{GEN/ACC} Jan nienawidzi_{GEN} ____ a Maria lubi_{ACC} ____ ?
 who.GEN/ACC Jan hates and Maria likes
 ‘Who does Jan hate and Maria like?’
 b. *Czego_{GEN}/Co_{ACC} Jan nienawidzi_{GEN} ____ a Maria lubi_{ACC} ____ ?
 what.GEN/ACC Jan hates and Maria likes
 ‘What does Jan hate and Maria like?’ (Citko 2005, p. 487)

Others have argued that Polish shows a case proximity effect, permitting non-syncretic case mismatches as well. Rothert (2022) reports experimental data showing that case mismatches are in fact acceptable as long as the filler bears the case assigned by the linearly closest verb, see (36).¹⁵

- (36) a. Słyszalem tylko że [tej nowej sasiadki]_{GEN} Agata nie
 heard only that this.GEN new.GEN neighbor.GEN Agata NEG
 przytulila_{GEN} ____ ale Julia powitala_{ACC} ____ .
 hugged but Julia welcomed
 ‘I only heard that the new neighbor, Agata didn’t hug but Julia welcomed.’
 b. *Słyszalem tylko że [te nowa sasiadke]_{ACC} Agata nie
 heard only that this.ACC new.ACC neighbor.ACC Agata NEG

¹⁵Judgments are based on the experimental findings reported in Rothert (2022).

przytulila_{GEN} ____ ale Julia powitala_{ACC} ____ .
 hugged but Julia welcomed

‘I only heard that the new neighbor, Agata didn’t hug but Julia welcomed.’ (Rothert 2022, p. 60)

In (36), the negated verb in the initial conjunct (*nie przytulila*, ‘didn’t hug’) assigns genitive of negation, while the non-negated verb in the non-initial conjunct (*powitala*, ‘welcomed’) assigns accusative. The topicalized filler can only bear genitive in this case. If the accusative is assigned in the initial conjunct and the genitive in the non-initial, the filler must bear accusative:

- (37) Ktora ksiazke obejrzal ____ ACC a nie zabral ____ GEN?
 which.ACC book.ACC looked.through but NEG took
 ‘Which book did he look through but not take?’ (Bondaruk 2003, p. 231)

Russian and Czech also show syncretism effects in ATB (Biskup 2018; Franks 1995). The case proximity effect is yet to be investigated beyond Polish.

Turning to non-Slavic languages, they seem to require case matching between the gaps in ATB. Greek is reported to be sensitive even to surface-identical but structurally distinct genitives (Georgi, Paparounas, and Salzmann 2025), as shown in (38).

- (38) a. pjon dhen eci kanenas endopisi_{ACC} ____ ke katigori_{ACC} ____ ?
 who.ACC NEG has nobody located_{ACC} and accused
 ‘Who has nobody located and accused?’
 b. *pjon/*pjanu dhen eci kanenas telefonisi_{GEN} ____ ke
 who.ACC/who.GEN NEG has nobody telephoned and
 katigori_{ACC} ____ ?
 accused
 ‘Who has nobody telephoned and accused?’
 c. ??pjanu_{GEN} dhen eci kanenas epitethi_{GEN2} ____ ke iperiscisi_{GEN3}
 who.GEN NEG has nobody attacked and prevailed.over
 ____ ?
 ‘Who has nobody attacked and prevailed over?’

In (38a), both verbs assign accusative. In (38b), the initial gap is assigned genitive and the non-initial gap accusative, rendering both a genitive and an accusative filler unacceptable due to a case matching requirement. Notice that the proximity of the filler to the verb assigning the respective case does not rescue the example. In (38c), we see that distinct classes of genitives cannot be combined either. Georgi, Paparounas, and Salzmann (2025) interpret this as showing a structural difference between the two genitives, one of them being introduced as the PP complement of V (class 3), the other classes as a DP occupying SpecAppl (class 1 and 2).

Hungarian appears to exhibit a strict case matching requirement, too.¹⁶ In (39a), both verbs assign accusative. In (39b) and (39c), the filler bears the case assigned to the closest gap, i.e. accusative or nominative, while the case assigned by the verb in the non-initial conjunct is mismatched—the outcome is ungrammatical.

- (39) a. Hány lány-t_{ACC} hív-ott_{ACC} meg Petra ____ és
how.many girl-ACC invite-PST.3SG PV Petra and
köszönt-ött_{ACC} Kata ____ ?
greet-PST.3SG Kata
‘How many girls did Petra invite and Kata greet?’
- b. *Hány lány-t_{ACC} köszönt-ött_{ACC} Kata ____ és ____
how.many girl-ACC greet-PST.3SG Kata and
jö-tt_{NOM} el?
come-PST.3SG PV
‘How many girls did Kata greet and (how many girls) came?’
- c. *Hány lány_{NOM} ____ jö-tt_{NOM} el és köszönt-ött_{ACC} Kata
how.many girl come-PST.3SG PV and greet-PST.3SG Kata
____ ?
‘How many girls came and did Kata greet?’

The effect of syncretism is unknown since there are no syncretic case markers in the language. However, there is one intriguing case marking pattern that might tell us something about the case matching requirement observed in (39). Long \bar{A} -moved subjects can appear with accusative case in the matrix clause (É. Kiss 1985). The different agreement possibilities on both the matrix and the embedded verb make for an intricate and highly complex pattern, the derivation of which is not fully settled yet (den Dikken 2009). Nevertheless, speakers have clear judgments about the constructions, though they fall into distinct groups (Gervain 2003). The two patterns relevant for our purposes are shown in (40).

- (40) a. Hány lány_{NOM} gondol-t-ad hogy el-jö-tt_{NOM}?
how.many girl think-PST-DEF.2SG that PV-come-PST.3SG
‘How many girls did you think came?’
- b. Hány lány-t_{ACC} gondol-t-ál hogy el-jö-tt_{NOM}?
how.many girl-ACC think-PST-INDEF.2SG that PV-come-PST.3SG
‘How many girls did you think came?’

(40a) shows the standard, expected pattern where the extracted subject has no special case marking and the matrix verb shows definite object agreement,

¹⁶The data provided in this section are modeled after the parasitic gap examples in É. Kiss (1985) discussed in section 1.3.1.2; and modified according to the considerations in den Dikken (2009) whereby agreement with ‘how many’ phrases distinguishes long \bar{A} -movement from prolepsis. The judgments are my own and need to be verified.

i.e. agreeing with the embedded CP.¹⁷ The pattern in (40b) is the exceptional pattern where the extracted subject is accusative marked and the matrix verb shows indefinite object agreement, i.e. as if the extracted subject were its object. Crucially, however, the agreement on the embedded verb rules out a proleptic parse, since the subject *hány lány* ('how many girls') is formally singular but referentially plural. The proleptic structure would thus require plural agreement on the embedded verb, contrary to what we see in (40b) (den Dikken 2009). Importantly, the nominative-accusative mismatch rendering examples (39b) and (39c) ungrammatical becomes possible under long extraction in (41) if and only if the filler is marked accusative and the matrix verb shows indefinite agreement (cf. É. Kiss 1985):

- (41) a. *Hány lány-t_{ACC} gondol-t-ál hogy ____*
 how.many girl-ACC think-PST-INDEF.2SG that
el-jö-tt_{NOM} és köszönt-ött_{ACC} Kata ____?
 PV-come-PST.3SG and greet-PST.3SG Kata
 'How many girls did you think came and did Kata greet?'
- b. *Hány lány-t_{ACC} köszönt-ött_{ACC} Kata ____ és*
 how.many girl-ACC greet-PST.3SG Kata and think-PST-INDEF.2SG
gondol-t-ál hogy ____ el-jö-tt_{NOM}?
 that PV-come-PST.3SG
 'How many girls did Kata greet and you think came?'
- c. **Hány lány_{NOM} gondol-t-ad hogy ____*
 how.many girl think-PST-DEF.2SG that PV-come-PST.3SG
el-jö-tt_{NOM} és köszönt-ött_{ACC} Kata ____?
 and greet-PST.3SG Kata
 'How many girls did you think came and Kata greeted?'
- d. **Hány lány_{NOM} köszönt-ött_{ACC} Kata ____ és*
 how.many girl greet-PST.3SG Kata and
gondol-t-ad hogy ____ el-jö-tt_{NOM}?
 think-PST-DEF.2SG that PV-come-PST.3SG
 'How many girls did Kata greet and you think that came?'

That is, while the extracted phrase is presumably assigned conflicting cases in its base position(s), the mismatch becomes acceptable if the conjunct that the nominative originates from is embedded. Although the exact syntax of 'case switching' is not agreed upon yet, the observations from ATB seem to suggest that the filler can be assigned conflicting cases as long as the mismatch is resolved over the course of the derivation. If movement out of the embedded conjunct rather than to the left edge of it is a necessary requirement, this phenomenon could serve as evidence for a parallel extraction analysis of ATB.

¹⁷The terms definite and indefinite or subjective and objective agreement are most frequently used to describe the pattern, although they are not quite fitting, see Bárány (2015).

1.3. Evidence

For German, Hartmann, Konietzko, and Salzmann (2016) report experimental data showing that syncretic case forms do not increase the acceptability of case mismatches, rendering ATB with a syncretic nominative/accusative filler such as the feminine DP in (42a) as unacceptable as cases with non-syncretic mismatched forms such as the masculine DP in (42b).¹⁸

- (42) a. *Diese Komödiantin hat der Hans besucht_{ACC} ___ und
this.NOM/ACC comedian has the.NOM Hans visited and
___ den Peter amüsiert_{NOM}.
the.ACC Peter amused
'This (female) comedian, Hans visited and (she) amused Peter.
- b. *Diesen Komödianten hat der Hans besucht_{ACC} ___ und ___
this.ACC comedian has the.NOM Hans visited and
den Peter amüsiert_{NOM}.
the.ACC Peter amused
'This (male) comedian, Hans visited and (he) amused Peter.

(Hartmann, Konietzko, and Salzmann 2016, pp. 68, 75)

The authors maintain that several examples with case mismatches which were not tested in the experiments (provided by a reviewer) are perfectly acceptable based on informal judgments:

- (43) a. Was_{NOM/ACC} ___ hat in der Zeitung gestanden_{NOM} und wohl auch
what has in the newspaper stood and PRT also
jeder ___ gelesen_{ACC}?
everyone read
'What has been reported in the newspaper and (has) everyone read?'
- b. Was_{ACC/DAT} hat mir ___ gefallen_{NOM} aber der Karl ___ sich
what has me.DAT pleased but the Karl REFL
nicht angesehen_{ACC}?
not looked.at
'What pleased me but Karl didn't look at?'
- c. Was für Frauen_{ACC/DAT} hat der Hans ___ getroffen_{ACC} und der
what for women has the Hans met and the
Peter ___ geholfen_{DAT}?
Peter helped
'What kind of women did Hans meet and Peter help?'

(Hartmann, Konietzko, and Salzmann 2016, p. 81)

Judgments seem to vary among native speakers, however. To me, (43a) appears quite acceptable while the other two are increasingly degraded. Exploring potential

¹⁸(42) are sample experimental items from Hartmann, Konietzko, and Salzmann (2016) and the indicated judgments are based on the mean ratings the respective conditions scored in their experiments.

effects of the filler's animacy and D-linking is up to future research.

Taken at face value, these findings imply cross-linguistic variability with respect to the syntax of ATB movement. It appears as though the Polish pattern could be best explained in terms of forward ellipsis. If Russian and Czech would likewise turn out to exhibit a case proximity effect, the forward ellipsis approach may receive further support from other Slavic languages as well. Greek, Hungarian and German appear to require strict case matching. The case switch effect in Hungarian additionally lends support to approaches postulating extraction from both conjuncts, assuming that a nominative constituent can only 'switch' to accusative under \bar{A} -movement. German appears to resist repair by syncretism based on experimental evidence, although the empirical situation may be more complex due to item variability. The next section turns to case effects in parasitic gaps cross-linguistically.

1.3.1.2 Case in parasitic gap constructions

The case marking of the two gaps has also been studied in parasitic gap constructions. This section aims to systematically go through the languages presented in the previous section and compare the properties of their parasitic gaps to their ATB constructions. In Polish, parasitic gaps seem to pattern with ATB in that the case assigned to each gap has to match (44a), be syncretic (44b) or match the case assigned by the closest verb (44c)-(44d).¹⁹

- (44) a. To jest dziewczyna ktora_{ACC} Jan tolerowal_{ACC} ____ zanim
 this is girl which Jan tolerated before
 polubil_{ACC} ____ pg .
 liked
 'This is the girl Jan tolerated before he grew to like.'
- b. To jest dziewczyna ktora_{ACC/INS} Jan ignorowal_{ACC} ____ zanim sie
 this is girl which Jan ignored before REFL
 zainteresowal_{INS} ____ pg .
 got.interested.in
 'This is the girl Jan ignored before he became interested in.'
- c. To jest dziewczyna ktora_{ACC}/*ktorej_{DAT} Jan lubil_{ACC} ____ zanim
 this is girl which Jan liked before
 zaczal pomagac_{DAT} ____ pg .
 started help
 'This is the girl Jan liked before he started to help.'
- d. To jest dziewczyna ktorej_{DAT}/*ktora_{ACC} Jan ufal_{DAT} ____ zanim
 this is girl which Jan trusted before

¹⁹Note that there is disagreement between Citko (2005) and Rothert (2022) regarding the case proximity effect in ATB, while for parasitic gaps, the existence of the effect is unanimously reported by Bondaruk (2003), Citko (2005), and Himmelreich (2017).

polubil_{ACC} — pg ·
liked

‘This is the girl Jan trusted before he grew to like.’

(from Himmelreich 2017, p. 16 based on Citko 2013)

(44a) shows the basic scenario in which the filler bears the accusative assigned by both verbs. The syncretism effect is shown in (44b). In (44c) and (44d) we see the case proximity effect where the mismatch in case assigned to the real and parasitic gap is tolerated as long as the filler matches the case of the main gap. Note that evaluating the empirical status of these examples is difficult since as Himmelreich (2017) points out, parasitic gaps in Polish are perceived as marginal.

We have seen that Greek requires strict case matching in ATB constructions. Himmelreich (2017) reports that this is the case for parasitic gap constructions as well (45):

- (45) a. pion andra_{ACC} pandreftike_{ACC} horis na agapa_{ACC}?
which man.ACC married without to love
‘Which man did she marry without loving?’
b. *pion andra_{ACC} voithise_{ACC} horis na dosi_{GEN} hrimata?
which man.ACC helped without to give money
‘Which man did she help without giving money to?’

(Himmelreich 2017, p. 86)²⁰

In (45a), both verbs assign accusative. In (45b), the verb in the main clause assigns accusative and the verb in the adjunct clause assigns genitive. Yet, contrary to what we observe in (44c)-(44d) for Polish, Greek does not tolerate a case mismatch between the filler and the parasitic gap, even if case proximity is obeyed.

Similarly, we have seen that Hungarian requires the case assigned to the two gaps in ATB to match. The same pattern can be observed in parasitic gaps: The two gaps are subject to a case matching requirement (46).²¹

- (46) a. Hány lány-t hív-ott meg Petra mielőtt köszönt-ött
how.many girl-ACC invite-PST.3SG PV Petra before greet-PST.3SG
Kata?
Kata
‘How many girls did Petra invite before Kata greeted (them)?’
b. *Hány lány-t köszönt-ött Kata mielőtt el-jö-tt?
how.many girl-ACC greet-PST.3SG Kata before PV-come-PST.3SG
‘How many girls did Kata greet before (they) came?’

²⁰Based on Iatridou 1995 for (45a) and A. Alexiadou, p.c. for (45b).

²¹Subject parasitic gaps such as (46b) should be ruled out on independent grounds due to the anti-c-command condition, see É. Kiss (1985, 2008) and Horvath (1992) for a discussion of the matter specifically in Hungarian. The more general ban on subject parasitic gaps in monoclausal adjuncts can be explained by means of anti-locality, see Davis (2025).

- c. *Hány lány jö-tt el mielőtt köszönt-ött Kata?
 how.many girl came-PST.3SG PV before greet-PST.3SG Kata
 ‘How many girls came before Kata greeted (them)?’

Crucially, under ATB, this requirement is satisfied even if the two gaps start out with distinct cases and the mismatch is ameliorated over the course of the derivation, i.e. if a long extracted nominative is assigned accusative. Parasitic gaps exhibit the same effect (47). Again, this implies that the two gaps can start out with conflicting case values as long as the mismatch is ameliorated over the course of the derivation.

- (47) a. Hány lány-t gondol-t-ál hogy el-jö-tt
 how.many girl-ACC think-PST-INDEF.2SG that PV-come-PST.3SG
 mielőtt köszönt-ött Kata?
 before greet-PST.3SG Kata
 ‘How many girls did you think came before Kata greeted?’
 b. Hány lány-t köszönt-ött Kata mielőtt gondol-t-ál hogy
 how.many girl-ACC greet-PST.3SG Kata before think-PST-2SG that
 el-jö-tt?
 PV-come-PST.3SG
 ‘How many girls did Kata greet before you thought that came?’
 c. *Hány lány gondol-t-ad hogy el-jö-tt mielőtt
 how.many girl think-PST-DEF.2SG that PV-come-PST.3SG before
 köszönt-ött Kata?
 greet-PST.3SG Kata
 ‘How many girls did you think came before Kata greeted?’
 d. *Hány lány köszönt-ött Kata mielőtt gondol-t-ad hogy
 how.many girl greet-PST.3SG Kata before think-PST-DEF.2SG that
 el-jö-tt?
 PV-come-PST.3SG
 ‘How many girls did Kata greet before you think that came?’

Finally, turning to German, the case matching requirements appear to be slightly more lenient with parasitic gaps than in ATB. Aside from matching case (48a), syncretic mismatches are also tolerated (48d).²² (48b) and (48c) show that non-syncretic case mismatches are ruled out regardless of proximity:²³

²² Notice that Hartmann, Konietzko, and Salzmann tested nominative/accusative syncretism in their study. Re-examining ATB using the genitive/dative syncretism from (48) appears to render syncretism perfectly acceptable under ATB as well:

- (i) [Welcher Verstorbenen]_{GEN/DAT} hat Hans in einer Grabrede ____ gedacht_{GEN}
 which deceased has Hans in a eulogy commemorated and
 und Peter ____ ein Gedicht gewidmet_{DAT}?
 Peter a poem dedicated
 ‘Which deceased has Hans commemorated in a eulogy and Peter dedicated a poem to?’

²³In the data provided by Himmelreich originally discussed among others by Felix (1985) and Kathol (2001), the parasitic gap is licensed via scrambling. There are several conceptual and

- (48) a. Welche_{ACC} Frau hat Hans — begrüßt_{ACC} ohne — pg zu
 which.ACC woman has Hans greeted without pg to
 umarmen_{ACC}?
 hug
 ‘Which woman has Hans greeted without hugging?’
- b. *Welche_{ACC}/*welcher_{DAT} Frau hat Hans — begrüßt_{ACC} ohne
 which.ACC woman has Hans greeted without
 — pg zu helfen_{DAT}?
 to help
 ‘Which woman did Hans greet without helping?’
- c. *Welcher_{DAT}/*welche_{ACC} Frau hat Hans — geholfen_{DAT} ohne
 which-DAT woman has Hans helped without
 — pg zu begrüßen_{ACC}?
 to greet
 ‘Which woman did Hans greet without helping?’
- d. [Welcher Verstorbenen]_{GEN/DAT} hat Hans in einer Grabrede
 which.GEN/DAT deceased has Hans in a euology
 — gedacht_{GEN} anstatt — pg ein Gedicht zu widmen_{DAT}?
 commemorate instead a poem to dedicate
 ‘Which deceased did Hans commemorated in a euology instead of
 dedicating a poem to?’ (cf. Himmelreich 2017, p. 14)

Several disclaimers are in order before we move on. First of all, parasitic gaps are reported to be marginal in most languages, such as in German (Szarvas 2024b) and Polish. Himmelreich (2017) notes that while three of her Polish informants found parasitic gaps acceptable enough to intuitively form judgments about them, one of them outright rejected all items. Second, in an ideal scenario, the cross-linguistic status of case mismatches would be determined in the absence of differences in item structure, particularly with respect to D-linking and animacy which might skew the picture. However, it must be acknowledged that the array of syncretic case forms and verbs assigning them make it difficult if not impossible to yield a perfectly uniform, balanced, and large enough sample. As such, the reported differences between informal judgments and experiments in Polish and German may be additionally enhanced by structural and lexical biases. Third, recall also that Hartmann, Konietzko, and Salzmann (2016) emphasize that individual counterexamples illustrating the acceptability of case syncretism in German ATB movement do exist. The type of syncretism investigated might have an effect, too (as noted in footnote 22). Syncretic mismatches between more oblique cases, such as genitive and dative, could be more acceptable than less oblique ones, such as nominative and accusative, which is precisely the type of syncretism that did not

empirical complications to treating these gaps on a par with typical parasitic gaps licensed via wh-movement, which is why the following German examples are modified versions of the original data (cf. Szarvas 2024b). See also section 3.1.2 in chapter 3 for a discussion.

repair mismatches in the experiments by Hartmann, Konietzko, and Salzmann (2016).

The next section assesses how theories of ATB movement and parasitic gaps hold up against these empirical reports.

1.3.1.3 Theoretical implications

We now turn to the discussion of how effects related to case are derived by theories of multi-gap dependencies. Case matching effects are the least problematic and can be derived by all approaches. In fact, the sideward movement approach as well as the empty operator approach by Munn (1993) predict a strict case matching requirement and cannot account for alternative patterns without additional assumptions.

Syncretism effects straightforwardly follow from ellipsis approaches as they do not impose restrictions on the morphological properties of the gaps at all (Fiengo and May 1994; Merchant 2013). Other approaches need to make special assumptions, though not all of them do. Citko (2005) spells out a solution for the syncretism effect in the terms of Distributed Morphology (Halle and Marantz 1993): When a node bears several value sets for case, only a vocabulary item which is compatible with all value sets can be inserted. *Kogo* ('who') in (35a) is assumed to be underspecified such that it is compatible both with accusative and genitive case features. Seguin and Thoms (2025) adopt the repair by syncretism proposal put forth by Citko (2005). Note that this is different from the standard view where the most specific vocabulary item is inserted, predicting that either just one of the feature sets should be realized or both (if the feature sets are equally specific). Under the parallel extraction account of Hein and Murphy (2020), repair by syncretism follows from the derivation of ATB movement involving feature set intersection. As such, ATB movement is only predicted to be possible if the extracted items are either fully identical, deriving the case matching effect, or overlap in at least one relevant feature, leading to a syncretism effect.

The effect of 'case switching' that we observe in Hungarian suggests, on the one hand, that underlying mismatches between the gap sites are tolerated as long as the mismatch is repaired over the course of the derivation. On the other hand, the explanation that the nominative subject is assigned accusative case in an intermediate position it \bar{A} -moves to is in conflict with standard theoretical assumptions about case assignment and its interactions with \bar{A} -movement (or lack thereof; see also Abramovitz 2021 for discussion of a similar pattern in Koryak). Given that long \bar{A} -movement appears to be a genuine requirement for case switching, these data are irreconcilable with asymmetrical approaches where movement only targets one of the gaps. Indeed, É. Kiss (1985) interprets the

parasitic gap data in favor of a parallel extraction approach.

The case proximity effect, on the other hand, is where things become puzzling for symmetrical approaches. Recall that the case born by the filler is determined by the position of the verbs, implying that the feature composition of the competing vocabulary items does not matter at all. As such, Citko's multidominance approach, regardless of the extra assumptions about syncretism, cannot account for the effect. Likewise, Hein and Murphy's predictions are too narrow since mismatches are only predicted under the guise of syncretism. The parallel extraction account by Seguin and Thoms (2025) could derive the effect, but crucially predicts it only to be available in parasitic gaps—the reason being that the smuggled DP can be mismatched in case, but no smuggling is involved in the derivation of ATB movement. Rothert (2022) concludes that the case proximity effect in Polish can only be accounted for by an ellipsis approach to ATB. Given that morphological mismatches are predicted under ellipsis, the forward ellipsis approach in particular neatly derives why only the case assigned in the initial conjunct is acceptable on the filler since the filler is directly extracted from the initial conjunct, thus bearing only the case assigned by the closest verb. The backward ellipsis approach to ATB predicts the opposite, i.e. that (36b) is acceptable.

In the previous sections, we saw that neither of the construction types forms an empirically uniform class across languages. We now summarize the findings from each individual language to determine whether ATB and parasitic gaps pattern together within languages, or if there is variability along other parameters.

1.3.1.4 Cross-linguistic variability

Polish parasitic gaps pattern with Polish ATB movement, both showing a case proximity effect (on the basis of experimental evidence provided by Rothert 2022 for ATB). This raises the question whether a parallel pattern observed in ATB and parasitic gaps should be taken as evidence for a mutual syntactic derivation. Recall that Rothert (2022) convincingly argues in favor of the forward ellipsis approach to ATB for Polish. Salzmann (2012), who proposes the forward ellipsis approach to ATB, cites several arguments against a uniform treatment of the two constructions—among them the cross-linguistically limited availability of parasitic gaps—and therefore does not entertain the idea of an ellipsis approach to parasitic gaps. Likewise, the empty operator approach to parasitic gaps proposed by Chomsky (1986) and Nissenbaum (2000) does not draw any parallels between ATB and parasitic gaps, postulating that the empty operator giving rise to the parasitic gap does not bear the morphosyntactic properties of the real gap and the filler moving from it. Nissenbaum (2000) explicitly argues against the uniform treatment of the two constructions, citing diverging reconstruction phenomena

(see section 1.3.2). Seguin and Thoms (2025) argue in favor of a unificational account acknowledging the existence of case mismatches in Polish. Because the DP is smuggled to the edge of the adjunct clause from the parasitic gap site before it is assigned case, mismatches are predicted to be possible in principle. However, because ATB movement does not involve smuggling, it is predicted to be more conservative with respect to mismatches and require case matching or at least syncretism. Thus, the proponents of the unitary analysis of ATB and parasitic gaps still predict their case requirements to be different, while the proponents of the two approaches that could derive the case proximity effect in Polish multi-gap dependencies rule out their unitary treatment on independent grounds. Of the existing proposals, the empty operator approach to parasitic gaps and the ellipsis approach to ATB are the most promising candidates for Polish.

Turning to Greek, the strict case matching observed in both ATB and parasitic gap constructions appears to be best compatible with the sideward movement approach, the multidominance approach (under omission of the special assumptions made to allow for repair by syncretism) or the empty operator approach proposed by Munn (1993). An interesting piece of data missing from the picture is whether parasitic gaps in Greek likewise show the strict categorial and structural matching requirements between the different genitive classes demonstrated by Georgi, Paparounas, and Salzmann (2025) for ATB. If so, this would likely lend further support to approaches where a single filler is responsible for creating both gaps, such as sideward movement or multidominance.

Hungarian likewise imposes a case matching requirement on both ATB and parasitic gaps, with the added twist that underlying mismatches can be repaired via ‘case switching’. If we assume that the nominative constituent is assigned accusative in an intermediate position, the ATB and parasitic gap data could only be derived if material is extracted from both gaps, whether that is in parallel, or because there is only a single constituent such as under sideward movement or multidominance. Note, however, that due to the conflict between case assignment and \bar{A} -movement, it has been proposed that the relevant construction in Hungarian does not feature successive-cyclic movement at all, but merely \bar{A} -movement within the embedded clause (den Dikken 2009). Given that ‘case switching’ is poorly understood, it is somewhat unclear what exactly it reveals about the derivation of multi-gap dependencies.

In German, parasitic gaps and ATB movement come apart based on the reports by Himmelreich (2017) and the experimental data by Hartmann, Konietzko, and Salzmann (2016). While parasitic gaps do show syncretism effects, case mismatches under ATB movement are irreparable via syncretism. That is, a joint approach to both constructions via sideward movement or empty operator movement can be

ruled out since it predicts strict case matching in both constructions. One could, however, still consider the approaches for ATB only, or a modification of the empty operator approach where the operator in ATB is morphosyntactically more complex than in parasitic gap constructions. A multidominant approach without the extra assumptions made by Citko (2005) is likewise promising for ATB, and so is parallel extraction. Finally, the ellipsis approaches both overgenerate, predicting repair by syncretism as well as a case proximity effect in ATB.

The pattern found in German parasitic gaps is a bit tricky. Case does not seem to be irrelevant, i.e. non-syncretic mismatches are not allowed, suggesting that an empty operator with default morphosyntactic features does not suffice. At the same time, the sideward movement and empty operator approach by Munn (1993) are too strict as they predict case matching. Considering that multidominant approaches to parasitic gaps do exist, they appear promising (Citko 2013; Kasai 2010). Note, however, that Citko (2013) argues for case proximity effects in parasitic gaps based on Polish, and that this extension does not seem to apply to German. It appears that further assumptions are necessary to correctly rule out the case proximity effect while ruling in the syncretism effect for German parasitic gaps, and simultaneously ruling out both for ATB. A solution along the lines of Seguin and Thoms (2025) could be a good fit given that parallel extraction could derive the case matching pattern in ATB. The involvement of smuggling in the derivation of parasitic gaps additionally opens up the possibility of (syncretic) mismatches.

While all theories in section 1.2 implicitly set out to apply cross-linguistically, the evidence from case-related phenomena quite clearly speaks against this. This mismatch in theoretical assumptions and empirical facts is unsurprising given that the majority of approaches we have discussed only feature data from English, a language in which case marking plays an extremely limited role. On the other hand, the most prominent case-related data come from Polish, which appears to be on the more liberal end of the spectrum compared to Greek, Hungarian and German.

To summarize the evidence related to case, we see ample cross-linguistic differences with respect to repair by syncretism as well as case proximity effects. While there seem to exist languages with strict case matching requirements in both constructions like Greek, others like German allow for repair by syncretism only in parasitic gaps. Based on Hungarian, we saw that special instances of case marking under long \bar{A} -movement have the potential to reveal the nature of multi-gap dependencies, although we must better understand the phenomenon itself before applying it as a diagnostic tool. Polish appears to be particularly lenient, allowing for case mismatches as long as the case on the filler matches the case assigned by the closest verb in ATB as well as parasitic gaps. Thus, the cross-linguistic insights

presented in this section reveal a much more complex empirical landscape than most theoretical proposals seem to acknowledge. Languages differ with respect to whether they show the same type of effect across multi-gap dependencies. Consequently, ATB movement and parasitic gaps can be, but are not necessarily subject to the same restrictions. The evidence does not converge in favor of a single superior proposal, not even within a construction type. Future research should investigate these cross-linguistic differences more carefully to settle whether they can be accounted for by independent properties of the respective languages, or if there is genuine cross-linguistic variability regarding the syntax of multi-gap dependencies.

1.3.2 Reconstruction

In the previous section, we have seen the complexity of the cross-linguistic empirical landscape regarding case-related effects in multi-gap dependencies. We shall now move on to a group of phenomena for which the reported data comes exclusively from English, i.e. reconstruction effects. Reconstruction for \bar{A} -movement is, in principle, possible into any position previously occupied by a constituent, including intermediate positions (Sportiche 2017, p. 15). As such, assessing reconstruction effects in multi-gap dependencies is a particularly straightforward way to test the accuracy of existing theoretical proposals. Recall that this is exactly where approaches differ—some of them postulate that only one of the gaps contains material that is morphosyntactically related to the filler, others postulate that both do independently of one another, and yet others postulate that a single constituent moves through or (simultaneously) from both positions.

Diagnostic	ATB movement	Parasitic gaps
Variable binding	(!)	(✓)
Idioms	✓	✗
Scope	✓	✗
SCO	✓	✓
WCO	✓	!
Principle A	!	✓
Principle C	!	✓

Table 1.2: Data availability for reconstruction effects in multi-gap dependencies based on diagnostic and construction type.

Instead of examining multiple languages, this section focuses on different types of reconstruction which supposedly all diagnose the path of movement of the filler(s). While in the previous section, the results of a single diagnostic differed within and across languages depending on the construction that was tested, in

this section, the results of multiple diagnostics imply each construction to have conflicting properties within a single language. Table 1.2 gives an overview of the diagnostics that this section covers, marking data availability by construction type. Check marks indicate that there are data from the literature, cross marks indicate that there are no data available, and exclamation points indicate conflicting data reports. Parentheses around the symbols indicate that the data are confounded and likely inconclusive. Due to the complexity of the empirical facts, this section goes through the diagnostics one by one without trying to group them by outcome or construction type.

1.3.2.1 Variable binding

Variable binding reconstruction involves the reconstruction of a phrase containing a variable, i.e. a pronoun, into the scope of a suitable binder, usually a QP. In (49a), the QP *no politician* binds the variable *his* from its base position. In (49b), the variable is part of an \bar{A} -moved wh-phrase whose base position is c-commanded by the QP. The variable reconstructs to its base position and is therefore bound by the QP. In (49c), the base position of the \bar{A} -moved phrase is not c-commanded by the QP. Therefore, the bound reading is unavailable.

- (49) a. No politician_i ignores many of his_i collaborators.
 b. Which of his_i collaborators does no politician_i ignore ____i?
 c. *Which of his collaborators_i ___ ignore no politician_i?
 (Sportiche 2017, p. 9)

Turning to multi-gap dependencies, the data in (50) are taken to indicate that reconstruction symmetrically targets each gap under ATB, but the exact implications are not discussed any further by Citko (2005).

- (50) a. Which picture of his mother did every Italian like ___ and every Frenchman dislike ___?
 b. #Which picture of his mother did every Italian like ___ and Mary dislike ___?
 c. #Which picture of his mother did Mary dislike ___ and every Italian like ___?
 (Citko 2005, p. 492)

It is unclear whether (50b) and (50c) are ruled out entirely or if it is merely the bound reading that is ruled out unless there is a suitable binder in each conjunct like in (50a). In the latter case, the variable is presumably interpreted as being bound by *every Italian* in the first conjunct and *every Frenchman* in the second. However, it is unclear what these data exactly reveal about the derivation of ATB

movement. Similar data were first reported by Nissenbaum (2000) for both ATB and parasitic gaps. However, he comments that the contrast between (51a) and (51b) might be irrelevant due to the possibility of QR:

- (51) a. Which picture of his mother did you give to every Italian ___ after buying from Mary ____{pg} ?
 b. #Which picture of his mother did you give to Mary ___ after buying from every Frenchman ____{pg} ?
 c. #Which picture of his mother did you give to every Italian ____{pg} after buying from every Frenchman ____{pg} ? (Nissenbaum 2000, p. 34)

Furthermore, there are doubts regarding the plausibility of the examples in (51). For an exploration of reconstruction for variable binding, the different patterns it hypothetically exhibits and what they imply for theories of ATB movement, see section 5.2 of chapter 5. For now, I conclude that the status of variable binding in multi-gap dependencies is inconclusive.

1.3.2.2 Idioms

Citko (2005) also provides data showing that idioms reconstruct symmetrically in ATB constructions (52).

- (52) a. Which picture did John take ___ and Bill pose for ___ ?
 b. Which picture did John pose for ___ and Bill take ___ ?
 (Citko 2005, p. 492)

The filler *which picture* must be interpreted in the scope of the verb *take* in order to derive the idiomatic interpretation. Given that this interpretation is acceptable regardless of the position of the verb, the filler seems to have previously occupied both the initial and the non-initial gap site. Similar data have not yet been reported for parasitic gaps.

1.3.2.3 Scope

Moltmann (1992) notes that *how many* ATB-questions are multiply ambiguous (53). Again, parallel data is not available for parasitic gaps.

- (53) a. How many books did every student like and every professor dislike?
 b. Seven books. (*how many* > \mathfrak{E} > *every*)
 c. Student A liked seven books and Prof. B disliked two books; Student C liked nine books and Prof. D disliked four books. (\mathfrak{E} > *every* > *how many*)

- d. Every student liked seven books and every professor disliked three books. ($\&$ > *how many* > *every*) (Moltmann 1992, pp. 137–138)

How many books can have wide scope, prompting an answer like (53b). If it takes narrow scope, the appropriate answer is a list of pairs (53c). It can also have narrow scope with respect to *and* but wide scope with respect to *every*, deriving the reading in (53d). Citko (2005) shows that *how many books* cannot have wide scope in one of the conjuncts but narrow scope in the other, like *almost every professor* necessarily would require, as in (54).

- (54) a. How many books did every student like and almost every professor dislike?
 b. #Student A liked seven books, and Student B liked three books, and Student C liked nine books, and almost every professor disliked six books. (Citko 2005, p. 493)

As such, she argues that scope reconstruction affects both conjuncts simultaneously.

1.3.2.4 Strong crossover

The strong crossover (SCO) effect occurs when a phrase has ‘crossed over’ a pronoun via movement, manifesting in disjoint reference between the two (Postal 1971; Wasow 1972). In parasitic gaps, it was first observed by Williams (1987), who provides (55a) to illustrate that the parasitic gap site does show reconstruction effects and as such should be treated on a par with the non-initial gap of ATB movement. The judgments for (55a) are reproduced by Munn (1993) who supplies the relevant data for ATB in (55b), which are then reproduced by Citko (2005) and Salzmann (2012).

- (55) a. *Whose_i mother did we talk to ____ after he_i saw _____{pg} ?
 b. *Whose_i mother did we talk to ____ and he_i never visit ____ ?
 (Munn 1993, p. 53)

The conclusion drawn from these data is that the phrase *whose mother* must originate in a position from which it crosses over the pronoun *he* when it is moved. If no extraction would take place from the non-initial gap, the effect should not occur with a pronoun in the subject position of the non-initial conjunct, contrary to the judgments in (55).

1.3.2.5 Weak crossover

Turning now to weak crossover (WCO), it is likewise an effect that occurs when a phrase crosses over a pronoun via movement, yet contrary to SCO, the respective pronoun is not in an argument position. According to Munn (2001), a WCO effect only occurs in the initial gap in ATB as well as in parasitic gap constructions:

- (56) a. Which man_i did you hire ____ and his_i boss fire ____ ?
 b. *Which man_i did his_i boss fire ____ and you hire ____ ?
 c. Which man_i did you visit ____ just before his_i boss fired _____{pg} ?
 d. *Which man_i did his_i boss fire ____ just after you visited _____{pg} ?

(Munn 2001, p. 5)

The data for ATB are reproduced by Citko (2005) and Salzmann (2012). Bruening and Al Khalaf (2017) ran an experiment on WCO in parasitic gap constructions. The experiment's aim was to clarify whether the parasitism of the non-initial gap or its mere linear position accounted for the effect (based on existing linear accounts of WCO such as Chomsky 1976; Higginbotham 1980). The authors chose parasitic gaps in subjects because they precede the licensing gap, testing the hypothesis that the parasitic gap site is subject to a WCO effect, and only seems not to be when it is in an adjunct clauses because it linearly follows the licensing gap. The experiment had a 2×2 design crossing the factors LINEAR ORDER (whether the pronoun *precedes* or *follows* the gap) and GAPS (*single* or *multiple*, i.e. whether the dependency featured a parasitic gap or a simple wh-dependency).

- (57) a. *Precedes, multiple*
 Sally is wondering which girl her teacher's shouting at _____{pg} really upset ____ .
 b. *Follows, multiple*
 Sally is wondering which girl the teacher's shouting at _____{pg} made her mother take out of school ____ .
 c. *Precedes, single*
 Sally is wondering which girl her mother shouted at ____ .
 d. *Follows, Single*
 Sally is wondering which girl ____ shouted at her mother.

(Bruening and Al Khalaf 2017, p. 13)

Participants were asked about the pronoun's referent (e.g. for (57a), *Whose teacher shouted at a girl?*) and given a forced choiced task between *Sally* or *the other girl's*. Bruening and Al Khalaf (2017) argue that the evidence is inconclusive due to neither of the tested factors turning out significant, citing high by-item variability

to be among the causes. The authors also hypothesize that the acceptability of the sentences they tested may be compromised. Nevertheless, they reach the conclusion that “the facts do not support treating parasitic gaps as different in kind from regular gaps, nor do they support treating non-initial gaps in ATB movement differently” (Bruening and Al Khalaf 2017, p. 16). However, because the evidence is uninterpretable, it does not provide clear support for the alternative pattern based on linearity that they propose either. Rather, it seems that the design or the nature of the items was inadequate to assess the research questions: The authors make claims about WCO in both ATB and parasitic gaps without having tested items involving ATB movement at all, and likewise, a control condition akin to Munn’s examples with parasitic gaps in adjunct clauses would be crucial to detect any contrasts. Thus, it is unclear what is left to support Bruening and Al Khalaf’s claims at all. At the end of this thesis, in section 5.2 of chapter 5, I discuss how one could modify the experiment presented by the authors to account for these shortcomings.

1.3.2.6 Principle A

Principle A reconstruction, like WCO, is likewise maintained to asymmetrically affect only the initial gap in ATB (Citko 2005; Salzmann 2012). According to principle A, an anaphor must be locally bound.

- (58) a. Which pictures of himself_i did John_i buy ___ and Mary paint ___ ?
b. *Which pictures of herself_i did John buy ___ and Mary_i paint ___ ?
(Citko 2005, p. 493)

Given that *Mary* is a suitable binder of *herself*, the condition would be met in (58b) if the filler reconstructed to the non-initial gap site, which does not seem to be the case. Together with the observation that (58a) is acceptable, the data are interpreted such that the filler reconstructs only to the main gap. Likewise, the same asymmetrical pattern is reported for parasitic gap constructions (after Kearney 1983 by Chomsky 1986; Munn 1994; Nissenbaum 2000).

- (59) a. Which pictures of himself did John sell ___ before Mary had a chance to look at ___ pg ?
b. *Which pictures of himself did Mary sell ___ before John had a chance to look at ___ pg ?
(Nissenbaum 2000, p. 30)

Note, however, that the principle A data are controversial. Nissenbaum (2000) maintains that while reconstruction is asymmetrical in parasitic gap constructions (59), targeting only the main gap, it is symmetrical in ATB (60), contradicting the

data in (58). Munn (1993) and Salzmann (2012) seem to disagree with Nissenbaum, citing the judgments in (58) from Citko (2005).

- (60) a. *Which pictures of himself did John sell ____ and Mary buy ____?
 b. *Which pictures of himself did Mary sell ____ and John buy ____?

(Nissenbaum 2000, p. 31)

Bruening and Al Khalaf (2017) note that all examples of principle A reconstruction reported in the literature feature ‘picture NPs’, which have been shown to be exempt from principle A due to logophoricity (Pollard and Sag 1992; Reinhart and Reuland 1993). To avoid this confound, they present examples with parasitic gaps licensed by topicalization. They claim that the contrast between (61a) and (61b) indicates that reconstruction is symmetric because null operators are expected to behave like pronouns (Browning 1987):

- (61) a. *Himself_i, John_i admires ____ without Mary liking _____{pg} in the least.
 b. Himself_i, John_i admires ____ without Mary liking him_i in the least.

(Bruening and Al Khalaf 2017, p. 9)

Note that Chomsky (1986) and Nissenbaum (2000) do not state explicitly the properties of the null operator that they assume. Munn (1993, et seq.), on the other hand, whom Bruening and Al Khalaf (2017) are arguing against most extensively, states that null operators are copies of the traces that license them, not pronouns. Therefore, the account by Munn does not in fact predict the data in (61) to pattern alike in the first place, making the comparison orthogonal to the question about reconstruction (a-)symmetries.²⁴

The authors go on to report the results of an acceptability judgment experiment comparing principle A reconstruction in ATB and parasitic gaps using cleft constructions. In both construction types, three different patterns were tested: One where all gaps were c-commanded by a matching binder (62a), (62d); one where only the initial gap was (62b), (62e), and one where only the non-initial gap was (62c), (62f):

- (62) a. It’s himself that John took a hard look at ____ and became more aware of ____.
 b. It’s himself that John took a hard look at ____ and Melinda became

²⁴Bruening and Al Khalaf (2017) pursue a unified account of ATB and parasitic gaps in a ‘left-to-right’ model of syntax. As such, the data in (61) are crucial for distinguishing the predictions of their account from those made by Munn, yet, as argued in the main text, Munn explicitly states that null operators behave like traces of the licenser rather than pronouns, rendering the contrast irrelevant. Notably, there exists a proposal that parasitic gaps do not involve any movement and are instead represented as null pronominals (Cinque 1990), though the evidence seems to quite clearly speak against this (Levine, Hukari, and Calcagno 2001).

- more aware of ____ .
- c. It's himself that Melinda took a hard look at ____ and John became more aware of ____ .
 - d. It's himself that John taking a hard look at _____{pg} could make him more aware of ____ .
 - e. It's himself that John taking a hard look at _____{pg} could make Melinda more aware of ____ .
 - f. It's himself that Melinda taking a hard look at _____{pg} could make John more aware of ____ .
- (Bruening and Al Khalaf 2017, p. 9)

The results are not straightforwardly interpretable as participants gave all experimental items poor ratings. Ungrammatical controls elicited a mean rating of 2.3. As a comparison, only the condition illustrated by (62a) elicited a mean rating above the ungrammatical controls (2.98), with all other conditions being rated lower, and particularly all parasitic gap conditions, i.e. (62d)-(62f), scoring mean ratings between 1.58 and 1.68. The authors nevertheless maintain that the results most closely match the scenario under which reconstruction symmetrically targets both gaps in both constructions.

1.3.2.7 Principle C

Principle C reconstruction is a similarly controversial subject in multi-gap dependencies. According to principle C, an R-expression must be free, i.e. it may not be bound by a pronoun that c-commands it. In practical terms, this means that the pronoun will necessarily be in disjoint reference with the R-expression it c-commands. According to Nissenbaum (2000), the reading where *John* and the pronoun corefer is available if the pronoun c-commands the parasitic gap (63b) but not if it c-commands the main gap (63b). That is, the constituent does not seem to reconstruct to the parasitic gap site whereas it does to the main gap. Such a contrast does not exist between the initial and the non-initial gap of ATB constructions based on (63c) and (63d). Regardless of the position of the pronoun, it cannot corefer with the R-expression in the filler.

- (63)
- a. *Which picture of John_i did he_i buy ____ without letting Mary look at _____{pg} ?
 - b. Which picture of John_i did Mary buy ____ without letting him_i look at _____{pg} ?
 - c. *Which picture of John_i did he_i buy ____ and not let Mary look at ____ ?
 - d. *Which picture of John_i did Mary buy ____ and not let him_i look at ____ ?
- (Nissenbaum 2000, p. 33)

As with principle A reconstruction, the reports diverge. Citko (2005) maintains that the pattern under ATB movement is asymmetrical as well, i.e. that the pronoun and the R-expression can corefer under ATB if the pronoun c-commands the non-initial conjunct, as in (64a) compared to (64b).

- (64) a. *Which picture of John_i did he_i like ___ and Mary dislike ___ ?
 b. Which picture of John_i did Mary like ___ and he_i dislike ___ ?

(Citko 2005, p. 494)

Bruening and Al Khalaf (2017) also ran an experiment testing principle C reconstruction. The authors go on to argue that reconstruction for principle C does not exist (followed up by a dedicated experimental investigation of principle C reconstruction in simple wh-dependencies in Bruening and Al Khalaf 2019, see section 2.1.2 of chapter 2 for a discussion). They claim that reconstruction in both ATB and parasitic gap constructions symmetrically targets both gaps. Any asymmetries that arise are due to linear order. In their experiment, the authors only test ATB items, providing no explanation for the lack of data for parasitic gaps. RNR is taken as the control condition to ATB. In the RNR conditions, the pronoun precedes the R-expression, while in the ATB conditions, the order is reversed:

- (65) a. Bill explained [which picture of John in renaissance costume] he liked ___ but Mary disliked ___ .
 b. Bill explained [which picture of John in renaissance costume] Mary disliked ___ but he liked ___ .
 c. Bill explained why he liked ___ , but Mary disliked, [that picture of John in renaissance costume].
 d. Bill explained why Mary disliked ___ , but he liked, [that picture of John in renaissance costume]. (Bruening and Al Khalaf 2017, p. 5)

The authors report that when given a forced choice between *Bill* and *John* as the referent of the pronoun, *John*, i.e. the principle C violation, was chosen in 25.33% of trials in the condition represented by (65a). In the condition represented by (65b), *John* was chosen in 30.67% of trials. This difference is not statistically significant. On the other hand, *John* was only chosen between 4-5.34% of trials testing the conditions represented by the RNR structures in (65c) and (65d) were. The authors interpret this to indicate that the judgments reported by Citko (2005) shown in (64) are incorrect and that there is in fact no difference between the two gaps under ATB movement (in line with the judgments supplied by Nissenbaum 2000 shown in (63)). Despite not having tested parasitic gap constructions, the authors suggest that principle C reconstruction therein follows the same pattern, concluding that “most importantly, there is no asymmetry in multiple-gap constructions” (Bruening

and Al Khalaf 2017, p. 7).

As with the previously discussed experiments by Bruening and Al Khalaf (2017), there are some decisive drawbacks to this experiment and the conclusions drawn from it. First of all, these results do not clearly indicate the absence of reconstruction effects. This is due to the comparison of RNR constructions to ATB movement—reversing the superficial c-command relation between the pronoun and the R-expression coincides with switching their linear order. For all we know, the significantly lower coreference rates in the RNR conditions could indicate that coreference is further depressed when pronominal reference is cataphoric in addition to a principle C effect. Second, the authors adopt a threshold-based reasoning according to which a principle C reconstruction effect should rule out coreference entirely, and any coreference rate that is significantly higher than zero is interpreted as a lack of reconstruction. This threshold-based reasoning could be avoided by isolating c-command as an experimental factor, allowing for a more straightforward interpretation of the results based on the presence of an effect of the pronoun c-commanding the R-expression at any point in the derivation rather than a null result (see section 2.1.3.4 of chapter 2 for a detailed discussion). Finally, no data is provided for parasitic gaps regarding principle C reconstruction at all: The conclusions from the RNR vs. ATB comparison are suggested to carry over to parasitic gaps as well. As such, this experiment suffers a very similar fate to the WCO experiment mentioned in 1.3.2.5 in the sense that only half the data necessary to make substantial claims are actually collected. This is particularly unfortunate given the authors' very important observation that the judgments from the literature are not necessarily reproducible under controlled experimental conditions with naive participants. I show in chapter 4 that experimenting with parasitic gaps is indeed a tricky endeavor, but nonetheless one that is worth undertaking. In sum, the claims put forth by Bruening and Al Khalaf (2017) are objectionable, though the paper highlights relevant concerns about the previously reported reconstruction data in multi-gap dependencies.

1.3.2.8 Theoretical implications

We now turn to the overview of the different reconstruction diagnostics and what they reveal about multi-gap dependencies. Overall, four out of seven reconstruction diagnostics suggest that the filler is interpreted in the non-initial as well as the initial gap site under ATB movement, while only one out of these four shows the same for the parasitic gap. On the other hand, three out of seven diagnostics are reported to show an asymmetrical pattern under both ATB and parasitic gaps. Especially in this latter group, the status of the evidence is contested. Table 1.3 summarizes the empirical reports on reconstruction phenomena in

multi-gap dependencies.

Diagnostic	ATB movement	Parasitic gaps
Variable binding	(controversial)	(asymmetric)
Idioms	symmetric	NA
Scope	symmetric	NA
SCO	symmetric	symmetric
WCO	asymmetric	controversial
Principle A	controversial	(controversial)
Principle C	controversial	(controversial)

Table 1.3: Reported patterns of reconstruction effects in multi-gap dependencies based on diagnostic and construction type.

Aside from SCO, none of the diagnostics’ outcomes remain unchallenged. What is more, it is nearly impossible to evaluate which sources are more reliable. Most of the reports rely on introspective judgments of individual data points, and the only set of quantifiable (experimental) data are largely uninterpretable. The experimental data are either (i) incomplete, i.e. only testing one of the two constructions despite drawing universal conclusions for both, or (ii) confounded, either because the relevant hierarchical factors are not distinguished from linear ones or because the constructions were not rated to be acceptable at all by participants, or (iii) reliant on null results rather than contrasts between conditions. As such, the predictions made by the individual theories cannot be evaluated based on these data.²⁵ It is also evident how multiple theories that are so distinct could exist in parallel: Given not only the disagreement but also the lack of information surrounding the majority of judgments and their sources, it is impossible to estimate the reliability and confoundedness of the data and the theories based on them. The following section describes this problem in more detail and the direction taken by this thesis to take a step towards solving it.

²⁵Make no mistake, it is quite likely that the case assignment data discussed in section 1.3.1 are similarly controversial at least in some languages. Recall that the experimental investigations conducted by Rothert (2022) for Polish and Hartmann, Konietzko, and Salzmann (2016) for German contradict the introspective data previously reported. The picture may just seem a bit clearer because there are at most two sources reporting the same diagnostic within a single language, and because case assignment, being a core morphosyntactic phenomenon, is not subject to pragmatic biases the way reconstruction is.

1.4 Too many tools for multi-gap dependencies

1.4.1 The problem

Through the characterization of the state of affairs in the previous sections, I hope to have convinced the reader that proposing yet another theory of multi-gap dependencies is not the logical next step. There is absolutely no shortage of theories dealing with multi-gap dependencies, and this scenario gives rise to what I shall call the *too many tools problem*. It is fair to assume that neither of the proponents of the aforementioned theories envisioned all of them to exist in parallel, at least not within a single language, and that the shared goal of the field is to identify which approach has the best empirical coverage. As we have seen, this appears to be quite difficult: While each theory comes with a set of supportive data points, the full array of empirical evidence, once combined, paints an incoherent picture. Within each type of multi-gap dependency, certain phenomena suggest a symmetry between the two gaps, while others seem to suggest that they behave differently. Subsequently, the evidence becomes even less clear when we try to compare the properties of ATB and parasitic gaps, as some phenomena suggest a mutual derivation but others highlight crucial differences.

In section 1.3, I presented part of the empirical basis on which these approaches are built, showcasing what can essentially only be described as an empirical crisis. One dissertation alone is not going to solve this problem. Instead, my aim in this thesis is to try and disentangle at least one of these diagnostics and why its results vary from source to source. For the remainder of this thesis, we shall focus on principle C reconstruction because of the diagnostic power assigned to it, and because despite this power, the empirical reports do not match the clarity of the predictions in the slightest. Recall that reconstruction can technically target any position the constituent has previously occupied. For principle C, it was observed that constituents must reconstruct ‘all the way down’. That is, the possibility of reconstructing to an intermediate position such as in (66b) cannot rescue the construction from a violation in the lowest position.

- (66) a. *Whose characterization of the typical male viewer_i does he_i object to ___?
 b. *Whose characterization of the typical male viewer_i do you think ___
 he_i objects to ___? (Sportiche 2017, p. 31)

As such, principle C appears to be a particularly promising diagnostic in multi-gap dependencies: Instead of merely revealing any position a constituent has previously occupied, it should reveal the position in which a constituent was base generated.

This is relevant especially for the study at hand because approaches differ precisely in this respect. Symmetrical extraction approaches as well as the multidominance approach predict reconstruction into both gaps since the filler is base generated in all of them. As for the asymmetrical approaches, the backward ellipsis approach predicts a principle C violation in the non-initial gap only, assuming that vehicle change obviates principle C effects involving elided R-expressions. Following Nissenbaum's logic, the sideward movement approach to multi-gap dependencies would predict this pattern, too, since movement from the non-initial to the initial domain can be treated as successive-cyclic movement with the base position of the *wh*-phrase being in the non-initial domain. However, things might get a bit more complicated as copying from the non-initial to the initial domain might precede indexing, and the notions of initial and non-initial domains only gain relevance when the prepositional or coordinate head is merged, i.e. after sideward movement has already applied. Also, as shown in 1.2.2.3, each lower copy of the *wh*-phrase forms its own chain with the higher copy, meaning that sideward movement could likely predict a symmetrical pattern as well. Yet a third pattern is predicted by the forward ellipsis approach as well as null operator approaches where the filler is only extracted from the initial gap: Reconstruction should only target the initial gap. Again, the forward ellipsis approach predicts vehicle change in, and thus no principle C reconstruction to, the non-initial conjunct of ATB movement (Salzmann 2012).

To reiterate what the evidence suggests, Nissenbaum (2000), who primarily deals with parasitic gaps, maintains that reconstruction for principle C is symmetric in ATB and asymmetric in parasitic gaps. He therefore concludes that parasitic gaps and ATB are fundamentally different, providing an account of parasitic gaps based on empty operator movement. Citko (2005), on the other hand, takes the position that reconstruction phenomena form an inconsistent class, in that principles A and C as well as WCO only target the initial conjunct, while idioms, scope effects, variable binding and SCO target both conjuncts. She therefore disagrees with the data cited by Nissenbaum (2000), pursuing an analysis of ATB via multidominance, suggesting an extension of the theory to parasitic gaps as well given some modifications (Citko 2013).

A completely new perspective is taken by Bruening and Al Khalaf (2017): Reconstruction in ATB and parasitic gap constructions is always symmetric. If it appears not to be, this is due to the fact that there is no reconstruction at all, instead resulting from purely linear effects. As argued throughout this thesis, this characterization will turn out to be the most accurate out of the four, even if claims of symmetry are unsolicited if merely based on a lack of asymmetry. There are a number of design related issues that make the conclusions reached on the basis of

the experiments premature, as discussed in 1.3.2.7.

Thus, we have three different data reports: Principle C reconstructs asymmetrically to the initial gap in ATB (Citko 2005); asymmetrically to the licensing gap in parasitic gaps but symmetrically to both gaps in ATB (Nissenbaum 2000); and lastly, there is no principle C reconstruction, therefore the asymmetrical pattern is only apparent (Bruening and Al Khalaf 2017). That the empirical reports about principle C are as complicated as they could be is further illustrated by the fact that the first publication to address principle C reconstruction under ATB-movement and parasitic gaps was actually Munn (1994). Note that this report was not addressed in section 1.3.2.7. The reason is that although Munn refers to principle C reconstruction in text and even works out a theory as to why the principle A and C diagnostics yield distinct results, he appears to equate it to SCO in multi-gap dependencies (see especially Munn 1994, pp. 404–405). In essence, SCO can be seen as a type of principle C violation—the pronoun is in an argument position and as such c-commands the base position of the wh-phrase that crosses over it. The crucial difference is that principle C reconstruction is most often tested with R-expressions embedded in PP modifiers of wh-phrases, whereas in SCO, the wh-phrase itself acts as the R-expression involved in the violation.²⁶ Precisely the status of these PP modifiers is subject to a long-standing debate as it is unclear whether they reconstruct along with the wh-phrases they modify. This issue is addressed in chapter 2. Before we get there, the next section outlines the solution this thesis offers to this problem more broadly.

1.4.2 The way out

At this point, we simply do not know what the empirical facts are. A conceptual debate might help us identify which theory is the most elegant, but it will not determine which of them is the most accurate. It is evident that something needs to change about our data collection standards when our most widespread method, i.e. the collection of informal judgments, leads us in all possible directions. Such a change is necessary at least to get us out of crises like the one at hand, but more generally, I invite the reader to consider matching the complexity of our theories with an adequate degree of empirical rigor when it comes to the evidence supporting them. This does not mean that informal judgments are useless—they did get us pretty far considering experimental methods specifically in formal linguistics have only caught on in the past three decades (Schütze 1996), and that many areas of linguistics make progress even without experimental work,

²⁶I return to this observation in section 5.2 of chapter 5.

although often drawing their data from fieldwork instead.²⁷ It simply appears that multi-gap dependencies really push us to the edge of what is and what is not an acceptable utterance, especially when paired with complicated interpretive phenomena like reconstruction effects. Given the reported data, our ability to tell a genuine syntactic effect from a pragmatic or linear bias playing tricks on us has apparently reached its limit.

One may be tempted to interpret the findings provided by Bruening and Al Khalaf (2017) to imply that even experiments cannot supply the necessary evidence to settle our current empirical issues. Not only do I believe this to be an extremely pessimistic conclusion, it is also a conclusion that will ultimately turn out wrong. As Bruening and Al Khalaf (2017) propose, we need to be open to the possibility that some of the effects are in fact not tied to syntactic mechanisms. We also have to accept that some of the previously reported effects may not even exist at all. In order to make progress in the area of multi-gap dependencies, we have to focus on the data rather than the theories we are trying to motivate with them. That being said, simply collecting *more* data and declaring all existing theories to be false if the acceptability rates do not straightforwardly align with the expected grammaticality pattern is not the solution. Instead, we need to be more cautious about *how* we collect data. If data are collected solely with the intention to support or challenge one specific theory, there is a high risk of sacrificing the completeness and conciseness of the study for the sake of getting to theoretical implications faster, even if they might be unwarranted. More importantly, if the relevant experimental factors are not isolated from superficial influences, the experimental data will simply reproduce the conflicting judgments reported in the literature over the past decades on a larger scale. The point of controlled experiments is that they allow us to tease apart the aforementioned non-syntactic factors from the syntactic ones by employing an adequate experimental design, and allowing us to quantify the effect that each of these factors has on the outcome. Crucially, they also allow us to quantify how variable individual speakers are with respect to their responses, and how much the outcome depends on which experimental item speakers are confronted with. That is, experiments have the potential to reveal (i) why linguists have been arguing over the acceptability of individual data points for decades, (ii) whether the data points are relevant for theorizing at all, and if so, (iii) which of the existing theories, if any, come closest to accurately modeling the effects we observe.

Part of the reason why the reconstruction data lead to conflicting conclusions

²⁷Introspective judgments and the risks they do (not) carry have been debated by Fanselow (2007) vs. Featherston (2007) on a conceptual level and Sprouse, Schütze, and Almeida (2013) vs. Gibson and Fedorenko (2013) on an empirical level. Phillips (2009) provides a well-rounded discussion of the topic.

may be, as Citko (2005) hypothesizes, that the different diagnostics do not form a coherent class. This was also addressed by Munn (1994), who aims to establish that principles A and C apply at different levels of representation, deriving why they yield distinct outcomes (but recall his use of SCO when meaning principle C reconstruction). The next chapter attempts to establish the status of principle C reconstruction in simple *wh*-dependencies. Recently, there has been a surge in experimental investigations addressing this matter, yet they did not settle the debate about its existence or strength. Before tackling the challenge of principle C reconstruction in multi-gap dependencies, we thus need to explore the nature of principle C reconstruction itself. More precisely, the next chapter aims to settle the question whether principle C reconstruction of PP modifiers exists and to quantify the size of the effect, taking into consideration contextual and methodological factors. This foundation will then anchor the subsequent study of principle C reconstruction in multi-gap dependencies pursued in chapters 3 and 4.

Chapter 2

Why we're still debating principle C reconstruction

The existence and robustness of reconstruction for principle C has been subject to a longstanding debate. Nevertheless, it remains a popular diagnostic in the formal syntactician's toolkit—whether the aim is to tease apart A- from \bar{A} -movement (Lohninger 2025; van Urk 2015) or to assess the base position of an element in a complex dependency (Citko 2005; Nissenbaum 2000), principle C reconstruction is prominently featured in the study of syntactic structures and assumed to hold universally across the languages of the world. The empirical picture, however, is rather blurry to say the least. Up until recently, claims were exclusively built on conflicting introspective judgments, and later experimental work has failed to settle the status of the phenomenon. For German, it has been argued that there is reconstruction, but that principle C is a violable constraint (Salzmann, Wierzba, and Georgi 2023). Experiments on English have been interpreted both in favor of and against reconstruction (Adger et al. 2017; Bruening and Al Khalaf 2019; Stockwell, Meltzer-Asscher, and Sportiche 2021, 2022). This chapter aims to address the discrepancies that have emerged from these recent experiments. For the very first time, the impact of methodological and experimental manipulations on participants' responses to principle C reconstruction are examined, focusing on German.

The chapter is structured as follows: Section 2.1 provides background on principle C reconstruction, including a detailed assessment of the experimental studies so far and how they differ. Section 2.2 outlines the study on German principle C reconstruction by Salzmann, Wierzba, and Georgi (2023) which served as a baseline for the two novel experiments presented thereafter. Section 2.3 provides a general discussion of the experiments, highlighting the methodological comparison, inter-speaker variability, the implications of the study as well as its

limitations. Section 2.4 concludes.¹

2.1 Background

2.1.1 Principle C reconstruction

This section gives a brief background on principle C reconstruction. Binding principle C states that referring expressions (R-expressions) must not be bound, resulting in disjoint reference between *Poirot* and *he* in (1):

- (1) *He_i says that Poirot_i is leaving. (Haegeman 1994, p. 226)

Although coreference and binding can seem almost identical in many cases, only binding requires that the binder c-command the bindee, while coreference is specified in a discourse model with no syntactic requirements (Reinhart 1983a,b). One environment where binding and coreference yield distinct interpretations is ellipsis (Heim and Kratzer 1998; Reinhart 1983b; Sag 1976).

- (2) Gina called her mother. The teacher did, too.
 a. *sloppy reading (binding)*: ‘The teacher called the teacher’s mother.’
 b. *strict reading (coreference)*: ‘The teacher called Gina’s mother.’

In cases with only one potential binder, binding and coreference are indistinguishable. Nevertheless, all else being the same, binding is preferred over coreference (Grodzinsky and Reinhart 1993; Heim 1998). Reinhart (1983b) proposes that only proper binding should be regulated by Binding Theory, a view that has been widely adopted (Büring 2005), with the noteworthy exceptions of Heim (2007) and more recently Bruening (2021). Most of the discussion in the literature, including the claims and data informing the current investigation, relies on a c-command-based definition of binding Bruening (though, for this, too, there exist alternative proposals, such as 2014).

It is maintained that \bar{A} -movement does not create new binding options (Büring 2005). The effect is shown in (3), where *he* cannot be bound by *a successful athlete* in (3a) but can be bound by *no one* in (3b). This is taken to indicate that the entire extracted phrase reconstructs to the internal argument position of the verb.

- (3) a. *I wonder [whose picture of a successful athlete_i] he_i reminded Bill that you saw ____ .
 b. I wonder [whose comments about him_i] no one_i reported ____ .

¹All supplementary materials from this thesis, including experimental items, raw data and analysis scripts can be accessed on OSF at <https://osf.io/hvm8u/> (Szarvas 2025b).

(Sportiche 2017, p. 16)

On the premise that this is correct, binding can be used to establish A/ \bar{A} -distinctions and to probe into the underlying structure of complex dependencies. Notice that in these cases of wh-extraction, it is not the head of the NP itself that causes the violation, but rather a nominal contained in the PP modifier of the wh-phrase. These data are particularly controversial. Some researchers maintain that argumental PP modifiers reconstruct, such as *of a successful athlete* in (3a) and *about him* in (3b). Crucially, PPs functioning as adjuncts do not reconstruct under this view. Those in favor of this distinction cite data such as (4) (Barss 1988; Freidin 1986; Lebeaux 1988; Sauerland 1998; Takahashi and Hulsey 2009; van Riemsdijk and Williams 1981).

- (4) a. *Which investigation of Nixon_i did he_i resent ____ ?
 b. Which investigation near Nixon_i's house did he_i resent ____ ?

(Safir 1999, p. 589)

The widely adopted explanation for the lack of a reconstruction effect in (4b) is that the adjunct is merged countercyclically after the NP has moved (Freidin 1986; Lebeaux 1988). The adjunct cannot be interpreted in the deeper syntactic position along with the head of the phrase because it has never occupied that position to begin with.² Those who disagree with the data in (4), and therefore also (3), argue that nominals cannot take syntactic arguments, meaning that the distinction is merely semantic—syntactically, all PP modifiers are adjuncts, therefore neither of them reconstructs and no principle C violation can be triggered. The reconstructing material, under this view, is subject to deletion up to interpretability at LF (Bianchi 1995; Fox 1999; Henderson 2007; Kuno 2004; Lasnik 1998; Safir 1999).

Recall that the validity of the principle C reconstruction test hinges on the assumption that disjoint reference must result from c-command. On the flipside, coreference is taken to be indicative of the absence of a principle C violation. However, we know that several factors can render a principle C violation quite acceptable, such as affectedness (Temme and Verhoeven 2017), interpretive economy, antilogophoricity, processing complexity (Varaschin, Culicover, and Winkler 2023), plausibility, salience of the antecedent and at-issueness (Gor 2020). More generally, topichood and contrastive focus (Cowles, Walenski, and Kluender 2007; Kaiser 2011), subjecthood (Kaiser 2011), first mention (Järvikivi et al. 2005) and linear proximity (Cunnings, Patterson, and Felser 2014) make referents more likely antecedents for a pronoun, showing that anaphora resolution is a highly

²Note that proponents of the argument-adjunct asymmetry do not unanimously agree with the implementation of Late Merger, criticizing it as being unconstrained and therefore an undesirable solution to the problem (Sportiche 2016, 2019).

complex issue going well beyond binding theory. Despite all of these confounds and the conflicting introspective reports, principle C reconstruction remains a popular test to establish A vs. \bar{A} -properties of movement types and c-command relations in complex dependencies. The following section reviews the existing experiments on principle C reconstruction and discusses why they failed to clear up the empirical controversies.

2.1.2 Previous experimental investigations

We now turn to the experiments conducted so far addressing coreference possibilities under principle C reconstruction. First, I give an overview of the studies and their findings, then we take a look at the specifics in which they differ. Understanding these details is vital to appreciate the effect that they have on the experimental outcomes, which is precisely what the novel experiments in section 2.2 are designed to address. The existing studies exemplify that multiple roads lead to coreference judgments. Examples (5)-(9) schematically illustrate the different setups.

Adger et al. (2017) employed a forced choice task, collecting data from 91 participants. With positive responses to the alleged principle C violation being at 30%, further increasing with distance, the authors conclude that PP modifiers of nouns do not reconstruct.

- (5) Item structure by Adger et al. (2017, p. 25)
- a. (short) Which side of Elizabeth does she prefer ____?
 - b. (long) Which side of Elizabeth does she say Philip prefers ____?
 - c. (longer) Which side of Elizabeth did Philip say she prefers ____?

Bruening and Al Khalaf (2019) included two matching referents in the sentence, forcing a choice between them. They compared adjunct vs. argument PP modifiers in surface violations of principle C as well as underlying ones, yielding a 2×2 design. Analyzing data from 75 participants, they report no significant difference between responses for the referent contained in the PP (*the countess*) in the ‘wh, adjunct’ and ‘wh, argument’ condition, likewise concluding that PP modifiers do not reconstruct.

- (6) Item structure by Bruening and Al Khalaf (2019, pp. 254–255)
- a. (wh, argument) The chambermaid told me which portrait of the countess she considered to be the most valuable.
 - b. (wh, adjunct) The chambermaid told me which portrait in the countess’s collection she considered to be the most valuable.
 - c. (no wh, argument) The chambermaid told me that she considered one

particular portrait of the countess to be the most valuable.

- d. (no wh, adjunct) The chambermaid told me that she considered one particular portrait in the countess's collection to be the most valuable.

Stockwell, Meltzer-Asscher, and Sportiche (2021), while using items similar to Adger et al. (2017), varied underlying c-command instead of movement by comparing causatives to transitives. They additionally manipulated dependency length, but did not include a comparison between adjuncts and arguments. Participants were asked what the sentence was about, judging the naturalness of two readings on a sliding Likert scale from 0 to 7. One reading indicated coreference (*A picture that Harry framed*) and the other one disjoint reference (*A picture that someone else framed*). Based on the data from 223 participants, the authors conclude that there is reconstruction due to a mean rating of 5.67 for the unnamed referent and 1.95 for *Harry* in the condition 'violation, short'. The significance of the effect vanished in 'violation, long' as well as in 'no violation, short'.

- (7) Item structure by Stockwell, Meltzer-Asscher, and Sportiche (2021, p. 206)
 - a. (violation, short) Which picture of Harry did he frame ____ ?
 - b. (violation, long) Which picture of Harry did Meghan say he framed ____ ?
 - c. (no violation, short) Which picture of Harry ____ made him laugh?
 - d. (no violation, long) Which picture of Harry ____ made Meghan say he has good taste?

In a follow-up study closely matching the previous one, Stockwell, Meltzer-Asscher, and Sportiche (2022) included argument vs. adjunct PPs instead of varying the distance of movement. Based on data from 275 participants, they report a mean rating of 2.19 for the coreferent reading for arguments, and a rating of 3.24 for adjuncts in the presence of a violation. This contrast is statistically significant, taken by the authors to indicate that there is an argument-adjunct asymmetry.

- (8) Item structure by Stockwell, Meltzer-Asscher, and Sportiche (2022, p. 147)
 - a. (violation, argument) Which picture of Harry did he frame ____ ?
 - b. (violation, adjunct) Which picture arranged by Harry did he frame ____ ?
 - c. (no violation, argument) Which picture of Harry ____ made him laugh?
 - d. (no violation, adjunct) Which picture arranged by Harry ____ made him laugh?

Salzmann, Wierzba, and Georgi (2023) investigated the phenomenon in German with yet another kind of experimental task. As in the experiment by Bruening

and Al Khalaf (2019), there were two matching referents in the sentence, but also two yes-no questions per trial. One of them asked whether the sentence could be understood such that the referent in the PP modifier (*embedded referent*, i.e. *Hanna*), in the case of the item presented here, overheard a comment, while the other task asked the same about the referent in the matrix clause (*matrix referent*, i.e. *Lisa*). The relevant measure is the proportion of *yes*-responses to the question about the embedded referent. The authors contrasted the grammatical function of the NP, whether it was moved or in situ, and whether the PP modifier was an argument or adjunct.³ Based on data from 32 participants, the authors report that coreference with the embedded referent elicited positive responses in 35.9% of the observations for both the condition *object, moved, argument* and *object, moved, adjunct*. The proportion of *yes*-responses to the condition *subject, moved, argument* and *subject, moved, adjunct* were at 50.8% and 51.6%, respectively. The authors report a significant effect of GRAMFUNC, but no significant effect of the argument-adjunct distinction, concluding that both arguments and adjuncts reconstruct in German. Relatively high coreference rates led to the conclusion that principle C is violable in German.

(9) Item structure by Salzmann, Wierzba, and Georgi (2023, pp. 602–603)

a. Object, moved, argument

Lisa erzählt, welche Geschichte über Hanna
 Lisa-NOM recount-3SG which story-ACC about Hanna-ACC
 sie — ärgerlich fand.
 she-NOM — upsetting find.PST.3SG

‘Lisa recounts which story about Hanna she found upsetting.’

b. Object, moved, adjunct

Lisa erzählt welche Geschichte im Buch über
 Lisa-NOM recount-3SG which story-ACC in.DAT book-DAT about
 Hanna sie — ärgerlich fand.
 Hanna-ACC she-NOM — upsetting find.PST.3SG

‘Lisa recounts which story in the book about Hanna she found upsetting.’

c. Subject, moved, argument

Lisa erzählt, welche Geschichte über Hanna —
 Lisa-NOM recount-3SG which story-NOM about Hanna-ACC
 sie verärgert hat.
 she-ACC upset have.PST.3SG

‘Lisa recounts which story about Hanna has upset her.’

d. Subject, moved, adjunct

³Due to space restrictions, the in situ conditions are omitted from (9). They featured embedded *that*-clauses.

Lisa erzählt, welche Geschichte im Buch über
 Lisa-NOM recount-3SG which story-NOM in.DAT book-DAT about
 Hanna — sie verärgert hat.
 Hanna-ACC she-ACC upset have.PST.3SG

‘Lisa recounts which story in the book about Hanna has upset her.’

To summarize, two groups argue that PP modifiers never reconstruct in English, one group argues that argumental PP modifiers do, and the group studying German argues that all types of PP modifiers reconstruct, concluding that it is principle C rather than reconstruction that is compromised. Despite aiming to test the same phenomenon, the experimental data are not straightforwardly comparable due to the use of different methods. Further confounding factors include the experimental design, item complexity and structure. It is also reasonable to assume that researchers interpret the data differently. While syntactic theory makes categorical predictions, the results obtained do not depict these extremes. Experiments show that participants respond positively to a supposed principle C violation in 20-50% of the cases. Speakers therefore seem to have tendencies at most, which is expected based on reports of non-syntactic factors making principle C violations acceptable (Gor 2020; Temme and Verhoeven 2017; Varaschin, Culicover, and Winkler 2023), and work arguing for the importance of linear order in binding (Bruening 2014). The subsequent section critically reviews the limits of what we can conclude based on the data so far.

2.1.3 Evaluation of previous experiments

2.1.3.1 Methodological differences

We now delve into the differences between the previous studies with respect to the methods they have employed. Adger et al. (2017) collected yes-no responses to a question asking about coreference with the only matching referent in the sentence. They did so by highlighting the R-expression and the pronoun, and asking if participants could use the sentence when the two referred to the same individual. Sentences were presented without context. Only those participants who passed a test where they had to indicate the impossibility of coreference with surface-level principle C violations were included in the sample.

Bruening and Al Khalaf (2019) collected a forced choice response between the two matching referents in the sentence, measuring which referent was preferred among participants. Although we cannot tell whether participants truly disallow a referent, the authors argue that the preference-based task is fairly simple and natural. If a grammatical constraint rules out either referent, responses should be close to zero, while chance level performance would indicate no such constraint.

2.1. Background

It is acknowledged that other factors, such as word order and the alternation between cataphoric and anaphoric reference may distort the picture further. The authors mention that rates should be ‘significantly different from zero’ if there is no reconstruction, though it is unclear what exactly counts as significantly different. Sentences were presented without context in this experiment, too.

Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) used a sliding Likert scale for each intended reading ranging from 0 to 7, arguably the most complex method. In the relevant condition, participants had to judge the naturalness of the reading violating principle C and of the reading where the pronoun referred to an unknown referent. There were thus two independent tasks per trial. In this experiment, too, sentences were presented without specific contexts, but participants were informed before the experiment that they should imagine being at a party and joining an ongoing conversation. Although the aim of this complex task was likely to allow for more nuanced responses, the authors report that the mean ratings of the two readings in each condition add up roughly to 8.0. This indicates that an increased rating for one reading led to a decreased rating for the other reading. Rather than improving the precision of the measurement, it appears that the complexity encouraged people to engage in simplification strategies of their own.

In the experiment by Salzmann, Wierzba, and Georgi (2023), participants had to complete two forced choice tasks. The two questions were whether the sentence could be interpreted such that (i) the referent in the matrix clause could corefer with the pronoun or (ii) the referent in the PP could corefer with the pronoun, respectively. The intention was to collect explicit judgments on both interpretive options. In addition, the authors informed participants that the sentences may have multiple interpretations, even if one of them is more readily available than the other, and that they should consider each of the options carefully. In a training sequence before the experiment, participants were shown an ambiguous sentence to illustrate this point (*Maria hat Anna besucht, weil sie nett ist* ‘Mary visited Anna because she is nice’). Though other groups do not report about their instructions in as much detail, it appears that Salzmann, Wierzba, and Georgi (2023) explicitly communicated the intention of the task to their participants, which on the one hand, ensures that they understood it, but on the other hand, may also encourage participants to spend more time thinking about the items and being less spontaneous with their judgments. In general, there is nothing inherently wrong about instructing participants more explicitly or inviting them to think about their responses. However, in the case of coreferent judgments, which are known to be highly susceptible to pragmatic factors, participants may end up accepting the coreferent reading regardless of a syntactic violation if they take enough time to

contemplate.

While there is some level of uncertainty associated with giving people only one forced choice task, providing them with more complex tasks and instructing them to consider each option carefully may induce a bias to be more accepting and potentially consider non-syntactic cues that may facilitate the acceptability of a violation. Given the involvement of non-syntactic factors in coreference resolution, these factors may have additionally influenced the outcomes.

2.1.3.2 Item structure

Let us now turn to the differences between the experimental items themselves across studies. Authors vary with respect to whether they presented participants with interrogative or declarative sentences embedding the *wh*-dependency. This coincides with the presence of an alternative referent that the pronoun could saliently refer to.

- (10) a. Adger et al. (2017)
Which side of Elizabeth does she prefer?
- b. Bruening and Al Khalaf (2019)
The chambermaid told me which portrait of the countess she considered to be the most valuable.
- c. Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022)
Which picture of Harry did he frame?
- d. Salzmann, Wierzba, and Georgi (2023)
Lisa erzählt, welche Geschichte über Hanna
Lisa-NOM recount-3SG which story-ACC about Hanna-ACC
sie ärgerlich fand.
she-NOM upsetting find.PST.3SG
'Lisa recounts which story about Hanna she found upsetting.'

Notice the three-way opposition in previous experiments regarding alternative referents: in the case of Adger et al. (2017), the possibility of an alternative referent was not mentioned at all, Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) gave participants scales to assess coreference with the referent in the sentence and *someone else*, while Bruening and Al Khalaf (2019) and Salzmann, Wierzba, and Georgi (2023) included a named alternative referent in their experimental items. If not given an alternative at all or if the alternative is not particularly salient, participants may be biased to resolve pronominal reference with whatever referent is available (Gordon and Hendrick 1998). It is therefore unclear if, in the case of Adger et al. (2017), fairly high coreference rates were obtained because there was no reconstruction, or if people simply chose the referent present in the sentence

because it is less costly than postulating another referent in the discourse model without any prompt to do so. On the other hand, the properties of the alternative referent included by Bruening and Al Khalaf (2019) and Salzmann, Wierzba, and Georgi (2023) may have the opposite effect. In both experiments, the referent was the subject of the matrix clause, making it a strong contender for coreference and a potential distraction from the embedded referent. Participants may have responded less favorably to the referent in the PP simply because the alternative referent was extremely prominent.

There are also some potential confounds resulting from the choice of verbs. In particular, Salzmann, Wierzba, and Georgi (2023) used psych verbs to maintain minimal pairs across conditions with subject and object extraction, allowing for pairs such as *which story about Hanna delighted her* and *which story about Hanna she found delightful*. Temme and Verhoeven (2017) argue that affectedness can increase the acceptability of principle C violations particularly in German, meaning that higher coreference rates should be obtained with psych verbs than transitive verbs. However, in that case one may find that varying verb type across different conditions of the same item plays a role, and arguably, this would lead to greater variability across items than consistently using verbs with potential confounds. The next section discusses the experimental designs, in particular the control construction(s) that the supposed underlying principle C violation has been compared to in the existing studies.

2.1.3.3 Experimental designs

We now examine the experimental designs employed by previous researchers in more detail. Adger et al. (2017) did not collect data for PP modifiers functioning as adjuncts, only arguments, given that reports in the literature converge about the former but not the latter. All authors collected responses to surface violations of principle C, intended as a baseline as to what response a violation should elicit.⁴ Salzmann, Wierzba, and Georgi (2023) and Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) additionally used items where only underlying c-command was manipulated without changing the linear order of the R-expression and the pronoun. Salzmann, Wierzba, and Georgi (2023) varied the grammatical function of the extracted phrase, while Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) used causative constructions such as *which picture of Harry made him laugh*, comparing them to regular transitives to achieve this distinction.

The divide between the groups arguing for and those arguing against the

⁴In the case of Adger et al. (2017) and Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022), movement was not included as an experimental factor, supplying merely descriptive statistics for these items.

reconstruction of PP modifiers corresponds to the use of c-command as an experimental factor. Recall that Salzmann, Wierzba, and Georgi (2023) and Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) argue that there is reconstruction while Adger et al. (2017) and Bruening and Al Khalaf (2019) argue that there is not. In the former groups' experiments, items without an underlying principle C violation allow for coreference more readily than ones where reconstruction leads to a violation. The experiments by Adger et al. (2017) and Bruening and Al Khalaf (2019) base their arguments on the observation that supposed underlying violations are indicated to be possible much more frequently than surface violations of principle C, which, however, is confounded with the linear order of the pronoun and the R-expression Adger et al. (via controls in the case of 2017). The evidence provided by Salzmann, Wierzba, and Georgi (2023) and Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022), on the other hand, provides straightforward information on the relevance of c-command. The review of experimental designs in this section highlights that previous claims have been made on quite disparate grounds, making the experiments difficult to evaluate side by side. This is why interpreting the findings and placing them into context is particularly important. In what follows, we take a closer look at the different ways in which previous authors chose to do this.

2.1.3.4 Interpreting the results

To contextualize previous authors' conclusions, it is crucial to understand how they interpret their results. Groups measuring proportions of responses unanimously report them to be below chance for the condition in which a principle C violation is predicted to occur, yet interpret this below-chance performance in different ways. The groups arguing against reconstruction, i.e. Adger et al. (2017) and Bruening and Al Khalaf (2019), rely on the observation that responses indicating the possibility of or the preference for coreference between the pronoun and the R-expression are well above zero. They argue that if there was a grammatical constraint, these kinds of results would be unexpected, especially because they found that responses were below 10% in items where a principle C violation without movement is involved. The authors therefore conclude that PP modifiers do not reconstruct based on a contrast between items with and without movement. Salzmann, Wierzba, and Georgi (2023) also find coreference rates well above zero but below chance in conditions with PP modifiers functioning as arguments (35.9%). Crucially, however, they reach a different conclusion due to separating linear precedence from underlying c-command by comparing object and subject extraction (instead of surface vs. underlying violations of principle C, where a null result would indicate full reconstruction). Because coreference rates differ by

roughly 15% between the conditions with and without an underlying principle C violation, and this effect is statistically significant, they conclude that there is reconstruction. They further argue against an argument-adjunct asymmetry in German because the exact same coreference rate was obtained with argument and adjunct PPs. We see, therefore, that similar coreference rates are interpreted in different ways depending on what comparisons research groups have included. Designs relying on a null effect, such as Adger et al. (2017) and Bruening and Al Khalaf (2019), are difficult to evaluate given that one has to rely on an arbitrary threshold-based reasoning (for a thorough discussion, see Salzmann, Wierzba, and Georgi 2023).

Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) collected judgments on a scale for each available reading rather than a binary coreference judgment. In their first experiment where the effect of distance was tested with PP arguments, they report that the coreferent reading scored a mean rating of 1.95, while the reading where the pronoun referred to an unnamed referent scored a mean rating of 5.67. They argue in favor of reconstruction because the rating for the unnamed referent increases if a violation is present. In their second experiment, the mean rating for the coreferent reading with PP arguments was 2.19 and 3.24 for PP adjuncts. Because this contrast is statistically significant, the authors argue that there is an argument-adjunct asymmetry in English. Recall that this group also isolated c-command from linear order by comparing causative and transitive structures. The experiment revealed a higher mean rating for the coreferent reading in the absence of a principle C violation, leading the authors to conclude that PP modifiers do reconstruct.

Recall that Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) hypothesize that participants may have responded with preferences rather than possibilities. This is supported by the observation that the mean ratings of the two readings in each condition add up to roughly 8.0, indicating that an increase of one reading's rating led to the decrease of the other's rating. Notice also that the authors interpret an increased rating for the alternative referent as an indicator for a principle c violation rather than focusing on whether the rating for the embedded referent decreased. The definition of a principle C effect is the unavailability of coreference with the R-expression that is c-commanded by the pronoun, not the prominence of alternatives. The results of Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) thus may be quite similar to those obtained by Bruening and Al Khalaf (2019), but the groups crucially differ in how they interpret them. For Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022), determining whether there is a principle C reconstruction effect is based on two components: the decrease of ratings for the named referent as well as the increase of ratings for the unnamed

referent. That is, while mean ratings for the coreferent reading may look similar in the argument and adjunct conditions (reflected in the lack of a significant main effect of `RELATION` reported by Stockwell, Meltzer-Asscher, and Sportiche 2022), it is crucially the larger increase of the mean rating for the unnamed referent that leads the authors to the conclusion that there is an argument-adjunct asymmetry (the significant interaction of `RESPONSE` and `RELATION`). For Bruening and Al Khalaf (2019), and arguably also Adger et al. (2017), the possibility of coreference with the R-expression in the PP being well above zero is taken to indicate that there cannot be a syntactic violation. These differences in what researchers decide to focus on reveal a fundamental problem—syntactic theory makes categorical predictions, yet the data do not reflect them. It is up to the researcher to tease apart the confounds and identify the confounding factors that may lead to the observed results. The fact that in most studies, a threshold-based reasoning needs to be adopted to interpret the data creates problems due to the lack of objective parameters.

Running experiments is not only a welcome addition to introspective data, it is necessary if the introspective judgments diverge. However, if supposedly objective, experimental evidence seems to further fuel the debate rather than settle it, we need to investigate whether the experimental design and methods are causing this. The existing experiments, when compared, reveal a lot about what to consider when studying principle C reconstruction experimentally. However, they do not provide a definitive answer to whether principle C reconstruction exists. I came up with two experiments following up on the work by Salzmann, Wierzba, and Georgi (2023), focusing on the experimental task and the presence of an alternative referent in the target item, while keeping the lexical content of the items consistent. By doing so, the novel data allow us to verify how much of the discrepancy between the results stems from actual properties of principle C reconstruction, and how much is indebted to methodological and item-related confounds.

2.2 Experiments

2.2.1 Baseline

Before presenting the novel experiments on principle C reconstruction in German, I discuss the baseline study by Salzmann, Wierzba, and Georgi (2023) in more detail. Although the following investigation is on German, given that the manipulations are rather methodological and not language-specific, the findings should hold cross-linguistically. To reiterate briefly, one of the experiments by Salzmann, Wierzba, and Georgi (2023) investigated whether there was an argument-adjunct

2.2. Experiments

asymmetry in the reconstruction of PP modifiers, which is the experiment discussed herein. The authors manipulated the factors MOVEMENT (moved vs. in situ), GRAMFUNC (subject vs. object) and STATUS (argument vs. adjunct). The relevant conditions for the current matter are ‘moved, subject, argument’ and ‘moved, object, argument’—focusing on the methods, the argument-adjunct asymmetry will not be included in the experiments. Each target item contains a wh-extracted NP with a PP argument modifying it. The PP contains an R-expression matching the features of the pronoun that linearly follows it in the sentence. This extraction dependency is embedded into a matrix clause containing another matching R-expression. The example in (11) repeats the item structure of the two conditions discussed hereafter:

(11) Item structure by Salzmann, Wierzba, and Georgi (2023)

a. Object, moved, argument

Lisa erzählt, welche Geschichte über Hanna
Lisa-NOM recount-3SG which story-ACC about Hanna-ACC
sie — ärgerlich fand.
she-NOM — upsetting found

‘Lisa recounts which story about Hanna she found upsetting.’

Can the sentence be understood such that...

Lisa found a story upsetting? ☐ yes ☐ no

Hanna found a story upsetting? ☐ yes ☐ no

b. Subject, moved, argument

Lisa erzählt, welche Geschichte über Hanna —
Lisa-NOM recount-3SG which story-NOM about Hanna-ACC —
sie verärgert hat.
she-ACC upset has

‘Lisa recounts which story about Hanna has upset her.’

Can the sentence be understood such that...

A story has upset Lisa? ☐ yes ☐ no

A story has upset Hanna? ☐ yes ☐ no

Using a Latin square design, the experiment tested 32 items in eight conditions in addition to 44 distractors. Items were not accompanied by a context sentence. 32 native speakers of German were recruited over Prolific. Participants were instructed to read each item carefully. They were also explicitly instructed to consider both of the interpretations offered, noting that in some cases, both may be possible. Items were presented in a randomized order. Two forced choice

questions were shown per sentence, one assessing coreference with the matrix referent and the other with the embedded referent. The order of the two questions was likewise shown in a randomized order. Table 2.1 shows the results.

Phrase	Arg/adj	Movement	Q matrix	Q embedded
object	argument	in situ	97.8%	7.0%
object	argument	moved	94.5%	35.9%
object	adjunct	in situ	96.9%	7.8%
object	adjunct	moved	96.9%	35.9%
subject	argument	in situ	93.0%	65.6%
subject	argument	moved	95.3%	50.8%
subject	adjunct	in situ	96.1%	55.5%
subject	adjunct	moved	96.1%	51.6%

Table 2.1: Results of experiment 2 reported in Salzmann, Wierzba, and Georgi (2023, p. 38), conditions tested in experiments reported herein are highlighted.

The authors report a significant simple effect of GRAMFUNC within the levels moved and argument. Coreference rates with the referent in the PP illustrated by Figure 2.1, meaning the proportion of *yes*-responses to the respective question, were 35.9% in the object, argument, moved (11a) and 50.8% in the subject, argument, moved condition (11b). The proportion of positive responses to the question about coreference with the matrix referent was above 90% across all conditions, indicating that participants understood the task. The authors take the findings to indicate that PPs functioning as arguments reconstruct, acknowledging that surface position matters as well based on an effect of MOVEMENT.

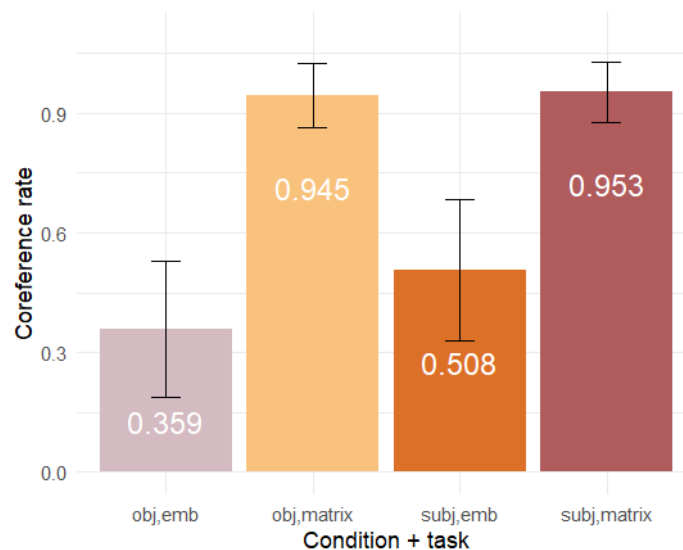


Figure 2.1: Coreference rates across conditions in experiment 2 reported in Salzmann, Wierzba, and Georgi (2023). Error bars indicate standard error.

The authors include a discussion of why coreference rates are neither close

to 0% nor to 100%, just like in the experiments on English. They speculate that non-syntactic factors may be at play, addressing two prominent accounts in follow-up experiments. One of them assessed whether parallel function of the R-expression and the pronoun had an effect, the other whether subjects were particularly prominent referents. Neither of the experiments yielded a significant outcome. The authors therefore conclude (i) that there is reconstruction for PP modifiers, and (ii) that principle C informs coreference possibilities, although it is a violable constraint in German.

There are some remaining issues, however, that receive no straightforward explanation. The contrast between the conditions *object, argument, moved* and *subject, argument, moved* is arguably low at 14.9%. If the source of the contrast is a principle C violation under reconstruction, one would expect a contrast similar to the one found between *object, argument, in situ* and *subject, argument, in situ*, which is at 58.6%. The authors note that participants may vary with respect to how much they prefer anaphoric over cataphoric reference, since the pronoun precedes the referent in in situ conditions, but follows it in moved conditions. However, this is in conflict with the conclusion that reconstruction is robust and that it is principle C that is violable in German—the effect of the principle C violation in in situ conditions is much more pronounced than in moved conditions.⁵ Further, no difference in coreference rate is obtained between arguments and adjuncts in the reconstructing conditions. This is in line with the view that there is no argument-adjunct asymmetry. However, note that those arguing against the asymmetry maintain that neither adjuncts nor arguments should show reconstruction effects, while the authors conclude the opposite, arguing that both types of PPs reconstruct in German. This entails cross-linguistic variability. Note also that the coreference rates in the *subject, in situ* conditions are still fairly low at 65.6% and 55.5%, respectively, despite the absence of a syntactic violation. They crucially also differ quite substantially despite no difference in linear and structural precedence as well as c-command. Syntactic theory predicts equal availability of coreference in both of these conditions. This demonstrates that cataphoric reference is dispreferred regardless of a violation.

Salzmann, Wierzba, and Georgi (2023) designed the instructions, task and item structure with a lot of attention to detail. While the aim was to eliminate ambiguity both for the participants as well as the interpretability of the data, the level of precision made the experiment particularly demanding. Note that participants were presented 78 items with two questions each. The results may be skewed if participants became inattentive in later trials. Furthermore, the high saliency

⁵Nonetheless, linear order and c-command coincide in the in situ conditions, making it difficult to determine which of the two produces the contrast.

of the matrix referent suggests that it may have distracted participants from the embedded referent, introducing another unwanted bias in the experiment. To investigate these issues in particular, two experiments were designed constituting simplified versions of the study by Salzmann, Wierzba, and Georgi (2023). The first experiment, which we turn to in the following section, departs from the original by including only one forced choice task per trial.

2.2.2 Experiment 1: simplification of the task

2.2.2.1 Materials

In the first experiment, the 32 items from experiment 2 by Salzmann, Wierzba, and Georgi (2023) were used, including only the conditions *object, argument, moved* and *subject, argument, moved*.

(12) Context:

Bei Lisas und Hannas Familienfest wurden peinliche
 at Lisa's and Hanna's family.party AUX.PST.PASS embarrassing-PL
 Geschichten aus ihrer Kindheit erzählt.
 story-PL from 3PL.POSS.DAT childhood tell.PTCP

'At Lisa's and Hanna's family reunion, embarrassing stories from their childhood were told.'

a. Object, matrix

Lisa erzählt, welche Geschichte über Hanna
 Lisa-NOM recount-3SG which story-ACC about Hanna-ACC
 sie — ärgerlich fand.
 she-NOM upsetting found

'Lisa recounts which story about Hanna she found upsetting.'

Lisa found a story upsetting. ☐ yes ☐ no

b. Subject, matrix

Lisa erzählt, welche Geschichte über Hanna —
 Lisa-NOM recount-3SG which story-NOM about Hanna-ACC
 sie verärgert hat.
 she-ACC upset has

'Lisa recounts which story about Hanna has upset her.'

A story has upset Lisa. ☐ yes ☐ no

c. Object, embedded

Lisa erzählt, welche Geschichte über Hanna
 Lisa-NOM recount-3SG which story-ACC about Hanna-ACC
 sie — ärgerlich fand.
 she-NOM — upsetting found

‘Lisa recounts which story about Hanna she found upsetting.’

Hanna found a story upsetting.

☐ yes ☐ no

d. Subject, embedded

Lisa erzählt, welche Geschichte über Hanna —
 Lisa-NOM recount-3SG which story-NOM about Hanna-ACC —
 sie verärgert hat.
 she-ACC upset has

‘Lisa recounts which story about Hanna has upset her.’

A story has upset Hanna.

☐ yes ☐ no

Additionally, 24 target items from another experiment exploring a different research question were presented. Eight unambiguous distractors were also included. Before testing started, participants were guided through three training trials, including feedback explaining the logic of the task. Targets items were distributed in a 2×2 Latin square design.

2.2.2.2 Method

Participants were presented a single coreference judgment task per trial. Participants only had to decide whether the pronoun could refer to either of the referents, varying which referent the task was about across trials. Metalinguistic terms were avoided by repeating the sentence with the intended reading (cf. Salzmann, Wierzba, and Georgi 2023). Each sentence was accompanied by a neutral context introducing both referents, aiming to mitigate the prominence of the matrix referent by having both referents present in the discourse before encountering them in the target item. Whether the wh-phrase was an object or a subject was encoded as the factor GRAMFUNC. This factor was fully crossed with another factor, REFERENT, encoding whether the task inquired about the matrix or the embedded referent, with the matrix conditions serving as a sanity check. In subject conditions, both the surface and the base position of the extracted phrase c-commands the pronoun, i.e. there should be no principle C violation regardless of reconstruction, and coreference between the embedded referent and the pronoun should always be possible. In object conditions, however, the surface position of the displaced phrase c-commands the pronoun, but under

reconstruction, this relation is reversed, i.e. the pronoun c-commands the base position of the displaced phrase, yielding a principle C violation.

2.2.2.3 Participants

A total of 175 participants were recruited over the platform Prolific, with mean age 32.09, $n = 87$ identifying as male, $n = 84$ as female, and $n = 4$ as non-binary. Monetary compensation was provided for everyone who completed the experiment successfully. Participants were native speakers of German living and raised in Germany (76.3%), Austria (19.08%) or Switzerland (4.62%). The experimental conditions *object, matrix* and *subject, matrix* served as attention checks, with failure to indicate coreference with the matrix referent in more than 20% of cases leading to exclusion. This threshold was also implemented to assess whether participants understood the task and considered coreference possibilities rather than preferences. Based on this criterion, 25 participants were excluded.

2.2.2.4 Procedure

The experiment was set up using the platform L-Rex (Starschenko and Wierzba 2024). Participants were instructed to read the items carefully, but to decide based on their first impression. They were also encouraged to use the comment function in case they felt that the forced choice task alone did not fully express their judgment. The sentences were displayed simultaneously with the context, the latter in italics. The question was shown below the sentence with the answer options *yes*, indicating coreference between the respective referent and the pronoun, and *no*, indicating disjoint reference. The first block included the training items in non-randomized order, first showing an example of coreference, then disjoint reference, and an ambiguous sentence as the third example. Participants were explicitly guided through this training sequence and were given feedback indicating that *yes* should be chosen in cases of ambiguity. This was done to ensure that the experiment assessed coreference *possibilities*, not *preferences*. The second block contained target items, pseudodistractors and distractors in pseudo-randomized order, such that two items from the same set of materials were never shown consecutively.

2.2.2.5 Hypotheses

Assuming that the mediocre proportions of coreference reported by Salzmann, Wierzba, and Georgi (2023) are due to the complexity of the experimental task, and that their conclusions about reconstruction are correct, the majority of responses in the condition *object, embedded* should be *no*. Responses indicating

coreference with the referent in the sentence are expected to occur in 25% or less of the observations in *object, embedded*. Simultaneously, the majority of responses in the condition *subject, embedded* should be *yes*, given that there is no principle C violation. Responses indicating the possibility of coreference are expected to occur in around 75% of the observations, indicating above chance performance. Neither floor nor ceiling effects were expected in either condition due to non-syntactic influences on coreference.⁶ Overall, simplifying the experimental task is expected to yield clearer results than found in previous work. This should manifest in a significant main effect of GRAMFUNC if principle C reconstruction is a robust syntactic constraint based on c-command. On the other hand, if PP modifiers do not reconstruct, we expect no pronounced difference between the conditions *object, embedded* and *subject, embedded*, meaning no significant effect of GRAMFUNC.

2.2.2.6 Results

Coreference rates across conditions are illustrated in Figure 2.2. Participants chose the answer *yes*, indicating the possibility of coreference, in 19.4% of the observations in the condition *object, embedded*, and in only 26.4% of the observations in the condition *subject, embedded*. At the same time, coreference rates in conditions inquiring about the matrix referent are between 96.1% and 98.8%.

Figure 2.3 visualizes the crucial information regarding how individual participants' responses change from one condition to the other. Only 42.7% of participants show an effect in the predicted direction, while 57.3% show the opposite or no effect at all. Even among the 42.7%, the increase from *object* to *subject* condition is nowhere near the predictions of a syntactic constraint. There are barely cases moving from below chance to above chance, let alone indicating the clear possibility of the coreferent reading in the *subject* condition and the opposite in the *object* condition. The directionality in the data along with the visual cues is in line with the statistical models, or rather, the problem of fitting any. The data were analyzed in R (R Core Team 2024), modeled by a generalized linear mixed effects model using the *glmer* function with the family 'binomial' (logit link) and the optimizer 'bobyqa' (D. Bates et al. 2015). The model only included the conditions asking about the embedded referent, since the conditions inquiring about the matrix referent solely served as a sanity check.

⁶The justification for the cut-off points at 25 and 75% are based on the previous study by Salzmann, Wierzba, and Georgi (2023), who found coreference rates around 35-50%. Since the aim of the study was to amplify the contrast between the conditions, the expectation was that coreference rates would be pushed further toward the extremes depending on the presence or absence of a principle C violation under reconstruction.

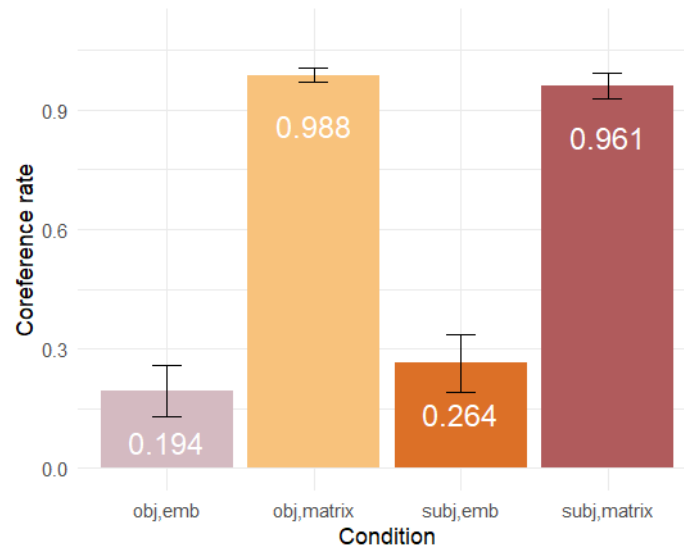


Figure 2.2: Coreference rates across conditions in experiment 1. Error bars indicate standard error.

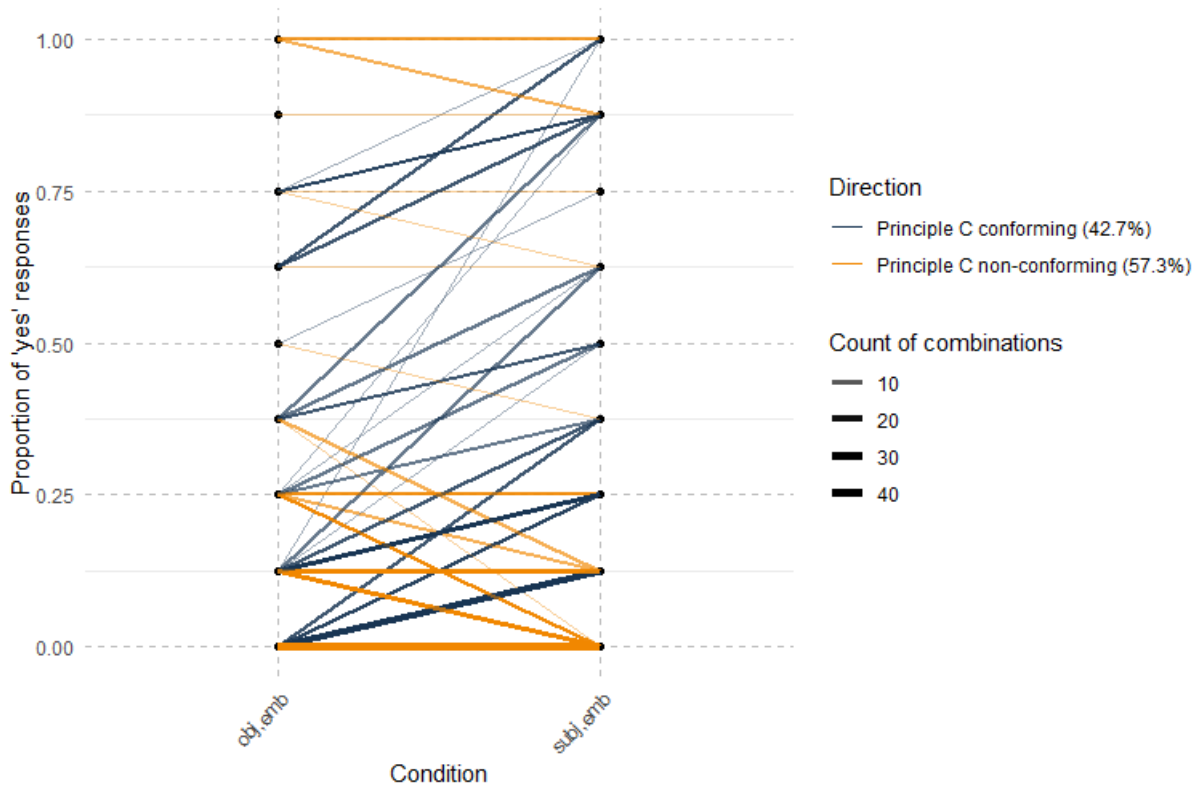


Figure 2.3: Individual participants' ($n = 150$) overall proportions of *yes* responses in experiment 1 by condition. Points on y-axis indicate respective proportion (ranging from 0 to 1 in steps of 0.125 based on eight observations per participant per condition), lines connect individual participants' proportions in the condition *obj,emb* to *subj,emb*. Color indicates whether the direction of the effect conforms to the predictions based on a principle C violation, thickness and opacity indicate how frequent the respective combination of proportions from the two conditions is in the data set.

2.2. Experiments

The data set was thus reduced to the two levels of GRAMFUNC as the conditions *object* and *subject*. A conservative α -level of 0.05 was defined. The model included a random effects structure with varying intercepts and slopes for condition, for both participants and items (Barr et al. 2013). Factors were treatment contrast coded such that the intercept is the estimate of the baseline probability of *yes*-responses in the *subject* condition. The estimated slope indicates the change in log-odds from *subject* to *object* condition.

$$(13) \quad \text{rating} \sim \text{phrase} + \\ (1 + \text{phrase} \mid \text{item}) + (1 + \text{phrase} \mid \text{participant})$$

GLMM	Estimate (SE)
(Intercept)	1.96*** (0.27)
phrase	0.24 (0.16)
AIC	1871.64
BIC	1917.91
Log Likelihood	-927.82
Num. obs.	2400
Num. groups: participant	150
Num. groups: item	32
Var: participant (Intercept)	6.60
Var: participant gramfunc	0.53
Cov: participant (Intercept) gramfunc	-1.88
Var: item (Intercept)	0.55
Var: item gramfunc	0.15
Cov: item (Intercept) gramfunc	-0.29

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 2.2: Coefficients estimated by the maximal model for experiment 1 before log-transformation.

This maximal model revealed a perfect negative correlation of the random intercepts and slopes: the higher the intercept of a participant, the lower their slope. Because the model needs to estimate two values per participant and these values have limited informativity due to the correlation, it is likely overfitting and therefore not providing reliable estimates for the fixed effects. Two more models were fitted. One excluded the correlation between the random effects, the other one reduced them to include estimates for random intercepts only. A Likelihood ratio test using the ‘anova’ function revealed that the more complex model, despite its fitting problems, is superior to the models with a simplified

random effects structure, indicated by a lower AIC and a significant p-value. Still, the estimates suggest that the maximal model is compensating for noise rather than identifying stable and interpretable effects. This is particularly evident from inspecting the log-transformed estimates and confidence intervals, which reveal that the model is estimating proportion of *yes*-responses to decrease from object to subject condition, and it likewise overestimates the intercepts. The log-transformed estimates and confidence intervals of the model including varying intercepts and slopes are reported below using the function *ggemmeans* (Lüdtke 2018), followed by the summary of the estimates from the maximal model in log odds in Table 2.2. Significance testing revealed no significant effect of GRAMFUNC. Estimates for the random effects by participants are extremely high. The following discussion includes log-transformed estimates, see Table 2.3, with the estimate on the log scale as shown in the table given in parentheses. The intercept by participant, that is, the baseline response of individual participants, is estimated to vary by 0.9986 (6.60).

	x	predicted	std.error	conf.low	conf.high	group
1	obj	0.920	0.229	0.880	0.947	1
2	subj	0.852	0.223	0.788	0.899	1

Table 2.3: Summary of log-transformed intercepts and confidence intervals estimated by the maximal statistical model for experiment 1.

Given that we are dealing with proportions, that is, values between 0 and 1, this indicates that participants are estimated to vary along the entire possible range of responses. The corresponding slopes, i.e. the effect of the predictor for individual participants, are estimated to vary by 0.6295 (0.53). On the other hand, the baseline response by item is estimated to vary by 0.6341 (0.55), with slopes estimated to vary by 0.5374 (0.15). The model thus also attributes a fair amount of variability to the items. Notice that the estimates for the random effects, particularly by participant, are much higher than for the fixed effects. To evaluate how dominant the random effects truly are, a fourth generalized linear mixed effects model was fitted, this time completely omitting the fixed effect, estimating only the intercept and the random effects structure:

$$(14) \quad \text{rating} \sim 1 + (1 + \text{phrase} \mid \text{item}) + (1 + \text{phrase} \mid \text{participant})$$

A likelihood ratio test comparing the generalized mixed effects model in (13) to the model in (14) revealed that the latter model, i.e. the one entirely excluding the effect of GRAMFUNC, has the superior fit based on a lower AIC and p-value.⁷ The

⁷To the best of my knowledge, this is a fairly unusual scenario. Even if the predictor is not improving the model's fit, the superiority of the model with random effects only implies that

dominance of random effects by participant found in the model is in line with the data visualization in Figure 2.3.

2.2.2.7 Discussion

The evidence does not suggest that a principle C violation under reconstruction played a role in determining participants' responses in this experiment. Generally, coreference rates are much lower than in the data collected by Salzmann, Wierzba, and Georgi (2023), and in particular very low in the subject condition despite no syntactic violation inhibiting coreference. It may be the case that the presence of a matrix referent has a depressing effect on the coreference rate with the embedded referent. There are two potential reasons: on the one hand, participants may be choosing the embedded referent less frequently because the matrix referent is simply too prominent. It is mentioned first in the sentence, it is the subject, and neither of its positions is c-commanded by the pronoun. On the other hand, it may further be preferred over the embedded referent because it is both structurally and linearly more distant from the pronoun (cf. Adger et al. 2017; Stockwell, Meltzer-Asscher, and Sportiche 2021). The overwhelming prominence of the matrix referent is supported by comments provided by 18 participants in the conditions inquiring about the embedded referent, some of which indicate ambiguity, others specifically reinforcing the interpretation regarding the matrix referent, which was not offered in the respective trials.

There may also be methodological reasons why coreference rates are generally low. The simplified task may be a poor fit to address the research question as it is. Since it is formulated as a forced choice task with the response options *yes* or *no*, participants are essentially asked to evaluate the truth value of only one reading of a potentially ambiguous sentence. This is not supported by the observation that the matrix referent is given a positive response in nearly all observations. However, one could argue that participants establish coreference with the matrix referent much more readily based on the aforementioned prominence factors, and simply do not consider the additional, weak possibility of the reading involving the embedded referent to be problematic for the truth value judgment regarding the matrix referent reading. Again, this is supported by participants' comments explicitly reinforcing the matrix referent reading despite being asked about the embedded referent. Like in the experiments by Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022), participants may have responded based on preferences rather than possibilities.

Closer inspection of the data distribution among participants reveals that the

omitting the predictor leads to a better understanding of the data. In other words, rather than having no effect, the predictor appears to have a negative effect on the model fit.

overall coreference rates are the result of computing the mean over extreme inter-individual variability. This is particularly evident from how different statistical models fit the data. Recall that the maximal model, though indicated to be a better fit than a simplified version by a likelihood ratio test, (i) has severe fitting problems due to a perfect negative correlation for random intercepts and slopes, (ii) overestimates the intercepts in general, and (iii) estimates the predictor to have an effect in the opposite direction than indicated by the overall coreference rates in the observed data. Modeling under omission of the predictor revealed that a model based solely on the random effects structure fit the data best, indicating that the distribution of data is best explained in terms that completely ignore the information of underlying c-command and therefore, a principle C violation under reconstruction. Due to the additional confounds introduced by the design that are discussed above, the conclusion that principle C does not play a role at all may nevertheless be premature. Recall that participants were only asked about one out of two readings per trial, potentially making the *someone else* response dispreferred because the experimental task was poorly chosen altogether considering the fully intended ambiguity of the items.

To summarize, the experiment did not provide data illustrating the previously reported subject-object contrast more straightforwardly, it achieved quite the contrary. Although this may be indebted to the method itself, it is noteworthy that any potential effects traditionally associated with a principle C violation are apparently quite easily overridden, particularly in the subject condition where coreference is not inhibited by any syntactic factors. Under the assumption that there is no principle C reconstruction, a more plausible scenario would involve chance level results. The evidence so far seems to favor a view where relatively low coreference rates are not determined by an underlying principle C violation, but rather by the presence of a competing referent. This issue is explored in the second experiment, which is outlined in the following section.

2.2.3 Experiment 2: omission of the matrix referent

2.2.3.1 Materials

In the second experiment, participants saw an interrogative sentence in each trial. The sentence contained a displaced NP modified by a PP containing the only referent in the clause, followed by a pronoun matching the referent (resembling items by Adger et al. 2017; Stockwell, Meltzer-Asscher, and Sportiche 2021, 2022). The study had a Latin square design.

(15) Object condition

[Welche Geschichte über Hanna] fand sie ____ ärgerlich?
which story-ACC about Hanna-ACC found she-NOM ____ upsetting

‘Which story about Hanna did she find upsetting?’

What is this about?

☐ Hanna found a story upsetting. ☐ Someone else found a story upsetting.

(16) Subject condition

[Welche Geschichte über Hanna] ____ hat sie verärgert?
which story-NOM about Hanna-ACC has she-ACC upset

‘Which story about Hanna has upset her?’

What is this about?

☐ A story upset Hanna. ☐ A story upset someone else.

As in experiment 1, 24 target items exploring a different research question were also presented. 12 unambiguous distractors were included, involving a referent and a morphosyntactically mismatched pronoun, making only the response *someone else* acceptable. Participants were guided through three training trials illustrating the task.

2.2.3.2 Method

Participants were again given a single forced choice task per trial. This time, they were asked what the presented sentence was about and had to decide between two readings. The task thus resembles the task in the experiments by Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022). Crucially, instead of two scales judging the naturalness of the offered readings, it was a simple forced choice task measuring preference between the only referent in the sentence and *someone else*.⁸ Again, metalinguistic terms were avoided by providing full sentences with the intended reading. The order in which the response options were shown was pseudo-randomized across trials. Sentences were not accompanied by individual context. Instead, participants were given the instruction that they should imagine being at a party and picking up snippets of conversations (Stockwell, Meltzer-Asscher, and Sportiche 2021, 2022). Items only appeared in two conditions based on the grammatical function of the displaced wh-phrase,

⁸Forced choice was used as an alternative given concerns by Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022) about participants not using the scales as they were intended.

i.e. *subject* and *object*. Like in experiment 1, both the surface and the base position of the extracted subject c-commands the object pronoun, and, if there is reconstruction, a principle C violation should only occur in the object condition.

2.2.3.3 Participants

A total of 64 participants have been recruited over the platform Prolific, with mean age 32.2, with $n = 39$ identifying as male, $n = 24$ identifying as female, and $n = 1$ identifying as non-binary. Monetary compensation was provided to everyone who successfully completed the experiment. Participants fulfilled the same criteria as in experiment 1 (4.69% from Austria, 92.19% from Germany, 3.12% from Switzerland). It was made sure that none of the participants from experiment 1 took part in experiment 2. Although testing different methods with the same participant pool is beneficial to reduce variability, due to data collection for both experiments taking place within a short time window, the priority was to avoid familiarization or adaptation effects from the first experiment affecting the results of the second experiment. Distractors contained a referent and a pronoun with mismatched ϕ -features. Participants who chose the reading indicating coreference between the two in at least 25% of the respective trials were excluded, leaving 60 participants in the sample.

2.2.3.4 Procedure

The experiment was set up using the platform L-Rex (Starschenko and Wierzba 2024). The instructions were the same as in experiment 1. The sentences were shown one by one, with the question *Worum geht es hier?* 'What is this about?' being shown below them. The answer options were repetitions of the event described in the sentence, as shown in (15) and (16).⁹ The first block included the training items in non-randomized order, showing two cases where the referent in the sentence could corefer with the pronoun and one case where it could not due to a morphosyntactic mismatch. The second block contained target items, pseudodistractors and distractors in pseudo-randomized order, such that two items from the same set of materials were never shown consecutively.

⁹A reviewer asks whether participants were informed about the ambiguity of some sentences, expressing concerns about participants choosing the *someone else* response over the embedded referent due to being certain about its validity. While participants indeed were not explicitly informed about some sentences being ambiguous, they were told that the study was about preferences. Although the reviewer's concerns are valid, we shall see that they do not seem particularly relevant based on the results of the study.

2.2.3.5 Hypotheses

This experiment measures preferences in contrast to the experiment by Salzmann, Wierzba, and Georgi (2023) and experiment 1 reported herein. This has important implications for the expected response rates. If the PP modifier containing the referent reconstructs, coreferent readings should be dispreferred in the object condition, with the embedded referent being chosen well below chance. In the subject condition, chance level performance is expected due to the lack of a syntactic violation. If there is no reconstruction, chance level performance is expected in the object condition, too. Because it is a single forced choice task between two referents, if the proportion to which the embedded referent is chosen increases, the proportion to which the unnamed referent is chosen automatically decreases. Chance level performance indicates no preference for either of the readings. Generally, given the lack of an alternative, higher proportions indicating coreference with the embedded referent are expected. If an increase is observed only in the subject condition, this would indicate an underlying principle C violation. If, however, participants choose the embedded referent above chance across conditions, this would indicate a general bias to resolve pronominal reference with any matching antecedent given in the discourse.

2.2.3.6 Results

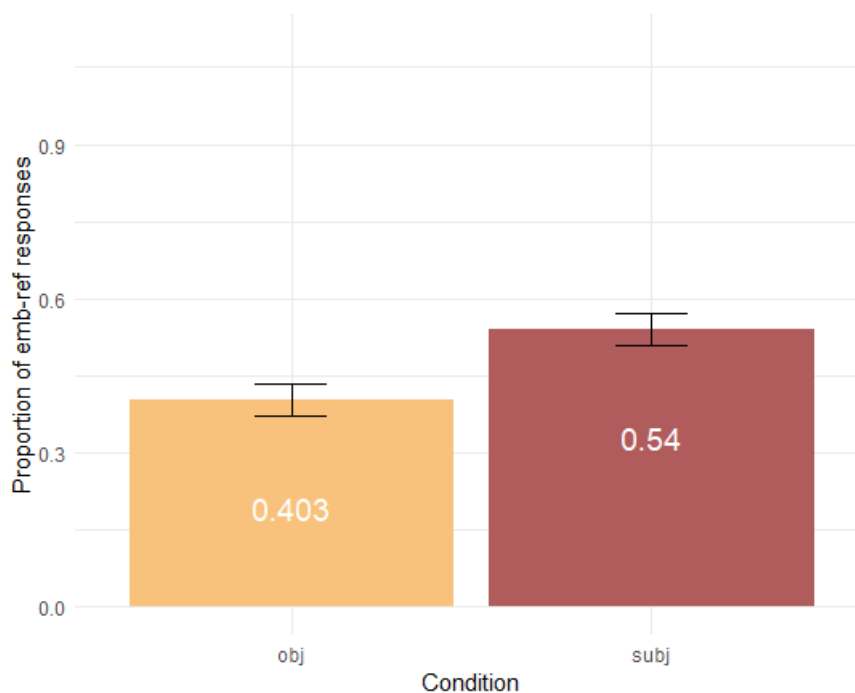


Figure 2.4: Coreference rates across conditions in experiment 2. Error bars indicate standard error.

Proportions of responses indicating coreference with the embedded referent

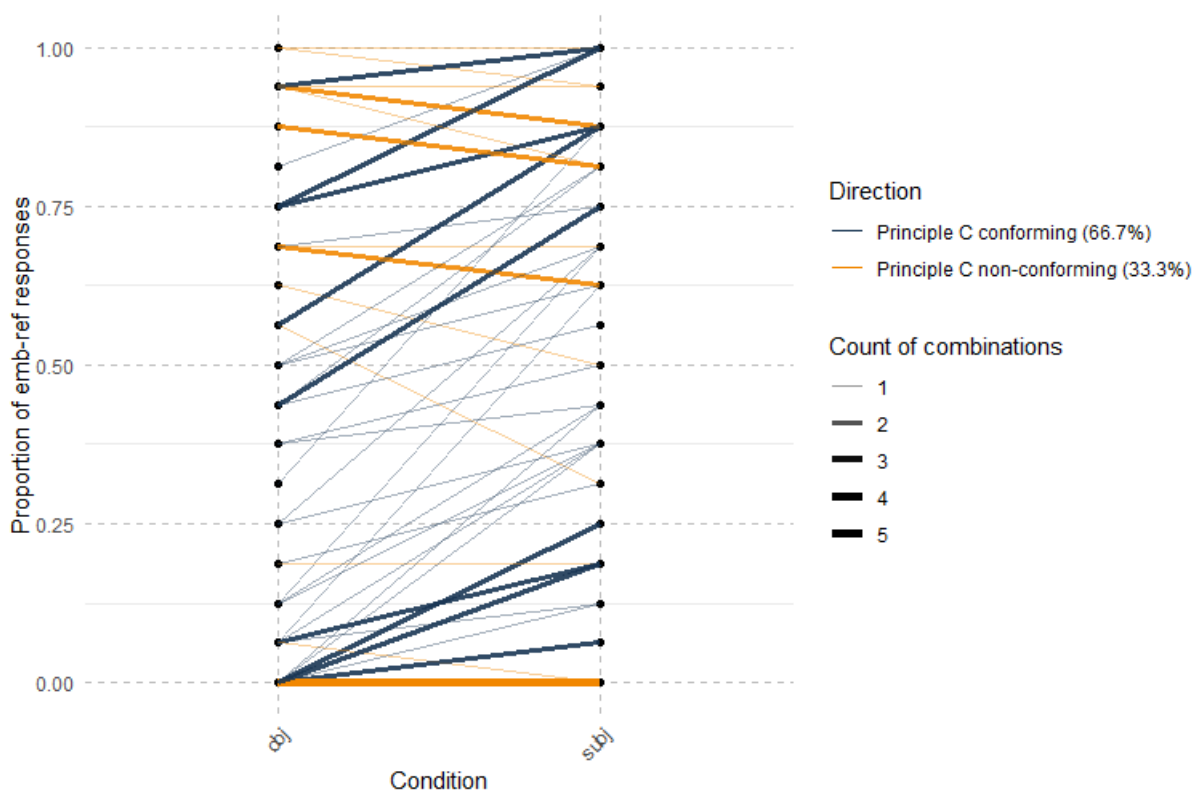


Figure 2.5: Individual participants' ($n = 60$) overall proportions of responses indicating coreference with the embedded referent in experiment 2 by condition. Points on y-axis indicate respective proportion (ranging from 0 to 1 in steps of 0.0625 based on 16 observations per participant per condition), lines connect individual participants' proportions in the condition *obj* to *subj*. Color indicates whether the direction of the effect conforms to the predictions based on a principle C violation, thickness and opacity indicate how frequent the respective combination of proportions from the two conditions is in the data set.

across conditions are illustrated in Figure 2.4. Participants chose the answer indicating coreference with the embedded referent in 40.3% of the observations in the object condition. In the subject condition, the embedded referent was chosen in 54% of the observations, indicating no clear preference. There is no matrix referent in the sentence, meaning coreference with it cannot be assessed as a sanity check. Participants' performance in the subject condition is in line with the lack of a principle C violation. Confronted with the two possible readings, participants show no clear preference for the embedded referent or the unnamed referent (*someone else*). In the object condition, participants show a slight preference for the option *someone else*, though this is not as strong as previously hypothesized.

Figure 2.5 shows the distribution of data among participants. Given the omission of the matrix referent, participants each saw 16 items per condition, meaning that the individual proportion of responses indicating coreference with the embedded referent range from 0 to 1, increasing by 0.0625 at a time.

Based on Figure 2.5, it is clear that participants show a vast degree of variability.

2.2. Experiments

There is no clustering among participants around chance level. Much rather, participants seem drawn toward the extremes. The majority of participants exhibit combinations of response proportions that are in line with the predictions of a principle C violation. However, pronounced slopes indicating a systematic preference for coreference in the subject but not the object condition and thus a syntactic violation, are extremely rare. 33.3% of participants show no effect at all or a reverse effect.

The data were analyzed in R (R Core Team 2024) using a generalized linear mixed effects model *glmer* with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’ (D. Bates et al. 2015) and a conservative α -level of 0.05. The model included a random effects structure with varying intercepts and slopes both per participants and items. The estimates from the maximal model specified in (17) are reported in Table 2.4.

$$(17) \quad \text{rating} \sim \text{phrase} + \\ (1 + \text{phrase} \mid \text{item}) + (1 + \text{phrase} \mid \text{participant})$$

GLMM	Estimate (SE)
(Intercept)	-0.20 (0.33)
gramfunc	1.19*** (0.18)
AIC	1768.23
BIC	1812.71
Log Likelihood	-876.12
Num. obs.	1920
Num. groups: participant	60
Num. groups: item	32
Var: participant (Intercept)	4.86
Var: participant gramfunc	0.72
Cov: participant (Intercept) gramfunc	1.05
Var: item (Intercept)	0.59
Var: item gramfunc	0.02
Cov: item (Intercept) gramfunc	-0.11

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 2.4: Coefficients estimated by the maximal model for experiment 2 before log-transformation.

Significance testing revealed a significant effect of GRAMFUNC. As with the data from experiment 1, this model too had fitting problems, revealing a perfect negative correlation of random slopes and intercepts by item, but not by participant. The

estimates for the random effects are quite large, i.e. the intercept by participant is estimated to vary by 0.9923 (4.86), while the accompanying slopes are estimated to vary by 0.6735 (0.72). By item, the intercept is estimated to vary by 0.6433 (0.59) and the slopes by 0.5049 (0.02). Model comparison with a simplified model excluding random slopes revealed that the more complex model was nevertheless superior based on a lower AIC and a significant p-value. However, this was again accompanied by large confidence intervals, a slope that is estimated to be much larger than the observed data, and a sign reversal. Upon suspicion that the model was relying on the random effects rather than the fixed effect to derive the estimates, another generalized linear mixed effects model was fitted omitting the predictor `GRAMFUNC` altogether (identical to the formula in (14)). However, a likelihood ratio test revealed that the model including the predictor was superior to the random effects only model ($p = 1.316e-08$), indicating that `GRAMFUNC` is informative to the model and improves its fit.

	x	predicted	std.error	conf.low	conf.high	group
1	subj	0.449	0.330	0.299	0.609	1
2	obj	0.727	0.393	0.553	0.852	1

Table 2.5: Summary of log-transformed intercepts and confidence intervals estimated by the maximal statistical model for experiment 2.

The significance test using the maximal model, though possibly inflated, is therefore not entirely misleading. The log-transformed estimates from the maximal model including confidence intervals are summarized in Table 2.5.

2.2.3.7 Discussion

In this experiment, coreference rates increased compared to the previous studies reported in section 2.2.1 and 2.2.2, respectively. This increase is particularly dramatic because the experimental task requires a different interpretation of the results. While in the previous experiment, participants were given a forced choice task between *yes* and *no* concerning one interpretation of a sentence, experiment 2 forced participants to choose between two possible interpretations. In other words, while in prior experiments each referent could score a 100% coreference rate independently (within a single trial in the study by Salzmann, Wierzba, and Georgi 2023 and across trials in experiment 1 reported herein), in experiment 2, a 100% choice rate of one referent automatically leads to a 0% choice rate of the other. Participants are at chance level in the subject condition. There is a statistically significant effect leading to 14% fewer responses indicating coreference with the embedded referent in the object condition compared to the subject condition. This shows that a principle C violation decreases the proportion of responses indicating

2.2. Experiments

coreference under reconstruction. The overall coreference rate of 40.3% also indicates that isolating c-command as an experimental factor is crucial, since a violation does not rule out the reading completely. Table 2.6 summarizes the coreference rates of experiments 1 and 2.

	obj (emb)	subj (emb)
Experiment 1	19.4%	26.4%
Experiment 2	40.3%	54%

Table 2.6: Overall proportion of responses indicating coreference by condition in experiments 1 and 2.

The simplification of the task was an attempt to approximate the setup used by Stockwell, Meltzer-Asscher, and Sportiche (2021, 2022). The alternative referent, which intended to ease resolution of pronominal reference in prior experiments, was omitted due to the hypothesis that its presence was distracting from the embedded referent. This turned out to be correct, participants were much more willing to choose the embedded referent in this experiment, where they were given no salient alternative. This is indicative of the importance of non-syntactic factors, such as a trivial bias to assign reference to a referent already present in the discourse (cf. Gordon and Hendrick 1998). The effect of GRAMFUNC is significant, indicating that participants’ behavior is at least partially informed by the presence or absence of an underlying principle C violation. The model indicating the significance of this effect nevertheless had severe fitting problems, manifesting in overestimation, large random effects, and a sign reversal. Further inspection of variability through visualizing individual participants’ proportions of responses by condition, see Figure 2.5, confirmed that while there is a tendency to respond according to principle C, participants are generally scattered all across the range of possible responses in both conditions. Inspecting individual participants’ data is imperative for understanding the implications of this experiment. It is not the case that participants are generally behaving close to chance level, or that there is a clear clustering. The patterns observed in these data are not merely noisy, they reflect that coreference resolution is guided by a number of factors. A principle C violation under reconstruction, while appearing to be one of them based on experiment 2, appears to play a variable role from participant to participant.

In sum, the experiment yielded a slightly more pronounced contrast between subjects and objects and a significant effect, providing evidence in favor of principle C reconstruction. This is further supported by individual participants’ slopes that show an increase in responses indicating coreference from object to subject condition. Consider that the magnitude of this effect does not compare to previous hypotheses about an underlying principle C violation ruling out coreference

entirely. While the coreferent reading is chosen less given an underlying principle C violation than in the absence of such a violation, we observe a general bias to resolve coreference with whatever referent is given in the sentence. The principle C reconstruction effect does not hold universally across participants either. The subsequent section places these findings in a broader context.

2.3 General discussion

2.3.1 Methodological comparison

We start with a discussion of the impact different methods have on the outcome, which is the key issue I aimed to address with the two novel experiments. In the original study by Salzmänn, Wierzba, and Georgi (2023), the authors used two forced choice tasks per trial assessing the possibility of either intended reading of the sentence. There were two R-expressions matching the pronoun, one of which was expected to be in disjoint reference with the pronoun in the object condition, given a principle C violation under reconstruction. The authors report a significant effect of `GRAMFUNC`, indicating that reconstruction has a negative effect on the coreference possibility between the referent contained in the wh-extracted phrase and the pronoun in the object condition, as predicted by principle C. The authors conclude that principle C is a violable constraint in German, and that there is evidence for reconstruction both for adjunct and argument PPs. However, coreference in conditions without the relevant c-command relation in the underlying structure likewise failed to elicit maximal coreference rates. In experiment 1 reported herein, the experimental task was simplified to include only one possible reading, aiming to minimize participants' cognitive load and to facilitate the consideration of both intended readings. The simplifications were hypothesized to yield a more clear-cut result. Instead of increasing coreference rates, this simplification further depressed them overall. The statistical model, which was revealed to be a poor fit for the data under the inclusion of the predictor `GRAMFUNC`, did not reveal a significant effect. While these findings are clearly in conflict with the reports by Salzmänn, Wierzba, and Georgi (2023), the simplification of the task introduced a further confound due to asking participants to judge a truth value for a single reading of an ambiguous sentence, thus not being a suitable task for the provided item structure. Experiment 2 saw further simplifications by omitting the matrix referent and providing participants with a forced choice task between the reading where the pronoun referred to the embedded referent and the reading where it referred to *someone else*, i.e. a

referent not present in the discourse.¹⁰ Here, chance level performance is likely indicative of the equal availability of both referents due to assessing the two readings in relation to one another.¹¹ We observe near chance performance in this experiment, with a significant effect of GRAMFUNC indicating the involvement of principle C in guiding participants' responses. The drastic increase in responses indicating coreference with the embedded referent, however, also suggests that participants prefer to resolve coreference with any referent given in the discourse, even if it is dispreferred in a setup where a prominent alternative is available.

The overall fluctuation in coreference rates across experiments demonstrates how participants' choices are affected by the method that is chosen. This may help understand why the experimental work on English principle C reconstruction led to conflicting conclusions given the diverging methods and setups that were employed (as discussed in section 2.2.3). The observation that coreference rates can be manipulated via external, non-syntactic factors supports the view that an underlying principle C violation is at best one out of many factors determining coreference in \bar{A} -dependencies. This has far-reaching consequences for the field of theoretical syntax, where principle C reconstruction is widely used for determining c-command relations among constituents as well as testing the properties of movement types, under the assumption that only \bar{A} -movement reconstructs. Considering that such important classifications hinge on coreference judgments often collected from lesser studied languages and using small samples of speakers as well as items, whether the conclusions drawn from such tests tell us anything at all is unlikely in light of the current findings. What is more, as the next section highlights, the conclusions also differ substantially depending on the number and type of speakers one examines.

2.3.2 Inter-speaker variability

This section is devoted to the discussion of variability between participants. The statistical analysis revealed that inter-speaker variability contributes extensively

¹⁰A reviewer objects to the claim that experiment 2 is a simpler than experiment 1, arguing that it does not reduce complexity, but rather relocates it from the item to the responses by including an additional referent there in the form of *someone else*. This is a valid perspective since the task itself is indeed not necessarily simpler. However, navigating the experiment as a whole likely is simpler nonetheless. A named referent in the sentence, especially with highly prominent features such as subjecthood and being mentioned first, is likely a greater distraction from the structure we are actually interested in, i.e. the movement dependency, than an unnamed referent among the response options. A consequence of reducing the item structure is that upon reading the sentence, participants do not necessarily think of an alternative referent by default, making the trials ultimately less demanding.

¹¹Note, however, that it may also be a result of uncontrolled pragmatic influences impacting individual participants' responses and canceling each other out in the overall proportions. Section 2.3.2 discusses this possibility in more detail.

to the understanding of the data composition. Crucially, cases of participants showing the predicted significant increase in coreference rate from object to subject condition are sparse. Many of the participants even show entirely flat slopes, indicating no difference at all between the presence or absence of an underlying principle C violation, while some even show the reverse effect (as shown in Figures 2.3 and 2.5). Experiment 1 revealed that only 42.7% of participants showed an effect, no matter how minor, in the predicted direction, while in experiment 2, 66.7% of participants did.

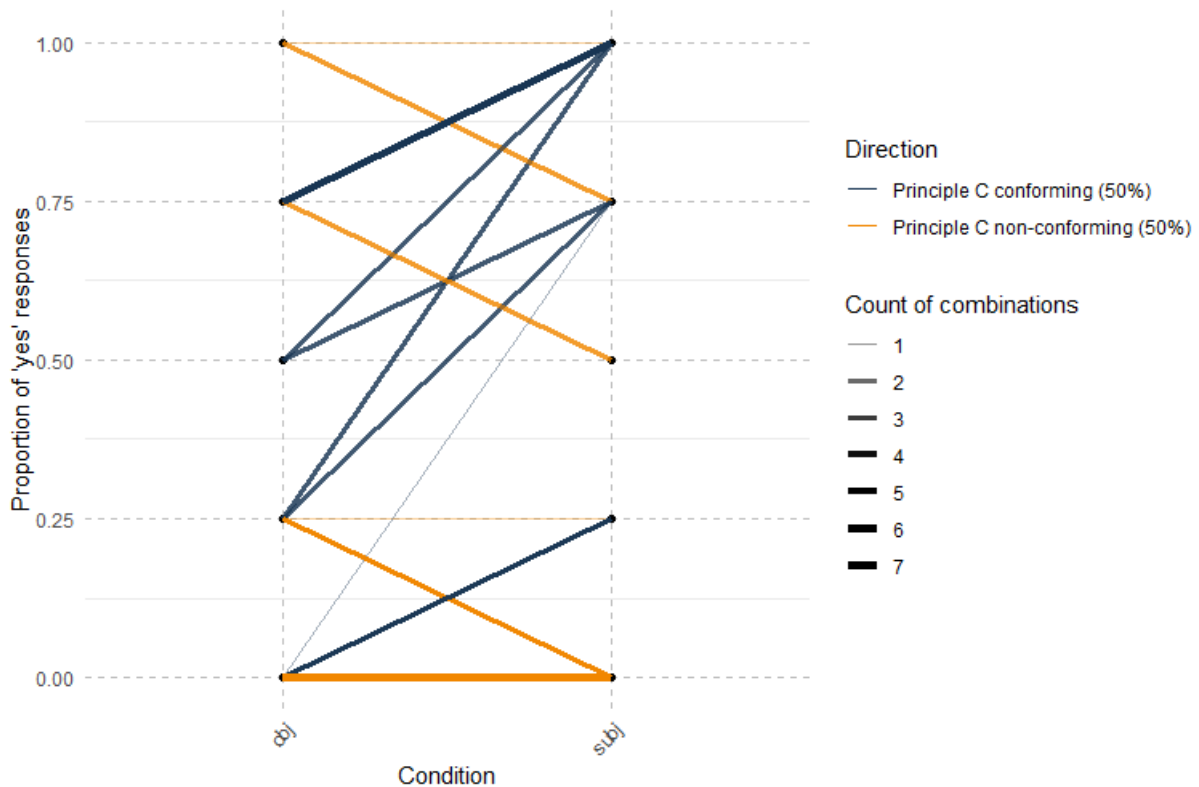


Figure 2.6: Individual participants' ($n = 32$) overall proportions of *someone else* responses reported by Salzmann, Wierzba, and Georgi (2023) by condition (experiment 2 in the original paper, singling out the conditions *object, argument, moved* and *subject, argument, moved* and the task pertaining to the embedded referent). Points on y-axis indicate respective proportion (ranging from 0 to 1 in steps of 0.25 based on 4 observations per participant per condition), lines connect individual participants' proportions from the condition *obj* to *subj*. Color indicates whether the direction of the effect conforms to the predictions based on a principle C violation, thickness and opacity indicate how frequent the respective combination of proportions from the two conditions is in the data set.

The data by Salzmann, Wierzba, and Georgi (2023) is re-examined in Figure 2.6. Using the same visualization and criteria, it shows that despite a decrease in overall coreference rates in the presence of a principle C violation, only 50% of participants show an effect in the predicted direction. This measure of directionality is a rather crude test, and yet large portions of the data fail it.

Despite its coarseness, the fact that this categorization splits participants into two large groups rather than filtering out a few outliers, provides particularly valuable insights to inter-individual variability. This type of information is entirely lost when simply inspecting overall proportions by condition and whether the predictor is significant in the modeling process. The extremely wide distribution of participants as well as the high percentage of them showing an effect in the opposite direction supports the idea that while an underlying principle C violation plays a role, it is not a particularly strong predictor of coreference resolution in *wh*-dependencies, and by far not a universal one. If reconstruction were the driving force behind coreference resolution, participants would exhibit a much more pronounced contrast between object and subject conditions, and should, at the very least, by majority show the effect in the predicted direction. The overall spread of participants furthermore indicates the involvement of non-syntactic, but not necessarily non-linguistic factors. They could be prosodic, pragmatic or even linear—and, importantly, have not been controlled for in this experiment, meaning that participants may have constructed prosodic or certain pragmatic cues freely, contributing to the variability in the data. It could also be the case that participants belong to different speaker profiles depending on how relevant principle C reconstruction is in how they resolve coreference. A future longitudinal study could investigate the existence and nature of such speaker profiles—whether they stay stable across methodological and/or pragmatic manipulations and if not, how they vary depending on the type of non-syntactic influence, allowing us to rank the strength of syntactic and non-syntactic cues on coreference resolution. It would be particularly interesting to explore whether certain participants simply do not ever conform to principle C reconstruction, as suggested by the data presented in the experiments herein, and are therefore only sensitive to pragmatic factors. In the next section, I discuss the idea that principle C reconstruction could be optional.

2.3.3 Optionality of reconstruction

In the following, I touch on what I take the results to indicate, namely that the reconstruction of PP modifiers is optional, including not only inter- but potentially also intra-individual variability based on pragmatic and semantic factors and how these factors are weighted. We saw in section 2.2.1 that coreference is available to R-expressions in addition to binding and is, according to the more widely accepted view, not determined by syntactic binding principles but by discourse. This observation goes back to the distinction between strict and sloppy readings under ellipsis, showing that R-expressions can bind, but they can also merely

refer to the same referent as the pronoun, which does not require any syntactic relation between the two. The standard view is that whenever both operations are available, binding wins over coreference (Grodzinsky and Reinhart 1993). Notice, however, that the prohibitive nature of principle C requires a rather contrary view. It states that R-expressions need to be free, i.e. they need to *avoid* being bound. A syntactic principle C violation would rule out establishing coreference via means of discourse under the minimalist assumption that pragmatics cannot override syntax.¹² The nature of the phenomenon, however, prompts the question whether reconstruction, if it is optional, can simply be avoided in case it would lead to a syntactic violation. This view would not require us to relativize principle C as such. Late Merger (Lebeaux 1991, 2009) and Neglect (Sportiche 2016) traditionally presuppose an argument-adjunct asymmetry, the relevance of which is quite limited with modifiers of nominals (Salzmann, Wierzba, and Georgi 2023 for German, Adger et al. 2017 vs. Stockwell, Meltzer-Asscher, and Sportiche 2022 for English). Principle C and even PPs in general are not the problem, however: there is experimental evidence not discussed here which suggests that principle C reconstruction is more stable with PPs of predicates (Adger et al. 2017, Salzmann, Wierzba, and Georgi 2023 corroborating Heycock 1995).

A radical alternative is to rethink the theory of movement, i.e. by assuming Parallel Merge instead of a change of syntactic positions leaving behind copies or traces (Citko 2005, 2011; Johnson 2018, 2025). Under such a multidominant theory of movement, the phrase occupies its base and final positions simultaneously, and the effects boil down to which position the phrase is interpreted in by the speaker. Reconstruction effects and their optionality essentially come for free, and the approach may have the added benefit of allowing speakers to differ with respect to how much importance they assign to the non-syntactic cues determining which position to interpret the phrase in.¹³ For some, avoiding to refer back to an antecedent with a pronoun that immediately follows it may be a priority (Arnold 2001; Gordon, Grosz, and Gilliom 1993). Others may be more lenient with respect

¹²A reviewer argues that pragmatics can override syntax because even ungrammatical sentences may be assigned an interpretation and get a boost in acceptability. To pinpoint what it means for pragmatics to override syntax, it is crucial to distinguish acceptability from grammaticality. Grammaticality is one of multiple factors influencing the acceptability of a sentence. That is, as the reviewer points out, an ungrammatical sentence can be acceptable because its acceptability is not exclusively determined by its syntactic grammaticality, but also pragmatic and processing related factors. However, these additional factors, while improving the acceptability of the sentence, cannot override its grammaticality status. See Fanselow (2007) for a series of conceptual arguments defending the view that grammaticality is categorical. Under this view, one would need to argue that cases where coreference is acceptable despite an underlying principle C violation are necessarily grammatical illusions. While the reviewer is absolutely right that these interactions exist, I am uncertain if modeling them as acceptable ungrammaticalities is desirable.

¹³The tricky part for multidominance is linearizing the structure correctly, ensuring that the element is pronounced in the higher position (Gračanin-Yüksek 2013).

to this inhibition, assigning more importance to avoiding the underlying principle C violation. Note, however, that it is not only difficult but also unsolicited to deem these ideas anything more than speculative and a potential incentive for future research. We now take a look at the limitations of the experiments reported herein and how future research could proceed to answer the questions that are left open.

2.3.4 Limitations

This section addresses the limitations of the current study. First of all, the number of participants and therefore the number of observations decreased from experiment 1 to experiment 2. This is visible in Tables 2.2 and 2.4. The reason behind this was strictly financial. Notice, however, that despite analyzing the data of 90 participants fewer in experiment 2, there are only around 500 datapoints missing compared to experiment 1. This is indebted to the modifications in experimental design. Recall that experiment 2 omitted the factor `REFERENT`, which in experiment 1 encoded which referent in the sentence the experimental task was about. Experiment 2 hence has a total of 960 observations, while experiment 1 has a total of 1200. In experiment 1, the 32 items were presented in four conditions, meaning each participant saw eight items per condition. In experiment 2, the number of conditions was reduced to two, meaning each participant saw 16 items per condition. This counterbalances the loss of statistical power to some extent. The result of this is a more uniform sample in experiment 2, which is important to note given the relevance of inter-speaker variability. Experiment 2 has more observations *within* subjects, while experiment 1 has more observations *between* subjects. Assuming that participants do not vary randomly across trials but are systematically guided by specific cues left untested here, the data may show a more pronounced effect of reconstruction due to this reduction in inter-individual variability.

Second, the potential confoundedness of the experimental task employed in experiment 1 likely has an impact on the comparison. Although the matrix referent was given a positive response quite consistently across trials and participants, participants may have been unsatisfied agreeing with the interpretation of the sentence involving the embedded referent because it was much less salient than the matrix referent interpretation. The results would likely turn out differently if participants were confronted with a question phrased in the same manner as done by Salzmann, Wierzba, and Georgi (2023): *Can the sentence be understood such that...?*, but only for one of the interpretations per trial.

This brings us to the third major limitation, namely that the possibilities regarding experimental tasks and instructions are virtually endless. Ideally, one

would conduct the same experiment over and over again manipulating even more minimal factors. One obvious option would be to omit the task inquiring about the matrix referent from the study by Salzmänn, Wierzbā, and Georgi (2023) and keep the phrasing exactly the same. Further options include prompting participants to give a truth value judgment such as in experiment 1, but omitting the matrix referent. One could also omit the context introduced in experiment 1. As a reviewer points out, experiment 2 did not have a context (due to only featuring a single referent), nor did the experiment by Salzmänn, Wierzbā, and Georgi (2023). Although the context was introduced with the intention of increasing the prominence of the embedded referent compared to the matrix referent by establishing both in the discourse before reading the sentence, the effect may have been canceled out in that both referents could have received an equal boost. It may have even caused the complete opposite of the intended effect and boosted the prominence of the matrix referent even further without having the desired impact on the embedded referent.

The current investigation does not exhaust the full array of possibilities by any means, but rather aims to provide a starting point, providing evidence that experimentation needs to be carried out carefully and with a special focus on non-syntactic factors. This allows us to use variability in the data as a valuable source of information not only to learn about the phenomena themselves, but to likewise improve the methods we use to study them. As outlined in section 2.3.2, a longitudinal study on principle C reconstruction across different experimental designs and under the control of pragmatic factors might give us a better understanding of what determines inter-individual differences. With this potential for future investigations in mind, the following section concludes.

2.4 Conclusion

The experiments reported herein demonstrate that the principle C reconstruction effect is detectable under a specific set of experimental circumstances and, particularly, in large sample sizes. Crucially, for the first time ever, we have data that eliminate the uncertainty associated with comparisons across variable lexical items and experimental designs. While one of the experiments revealed a significant effect, between 33.3% and 57.3% of participants show no effect at all or an effect in the opposite direction. The study comparison revealed that coreference rates vary systematically based on the experimental task and the presence of an alternative referent. This is confirmed by the reexamination of previously published conflicting experiments which vary immensely with respect to these properties. While the isolation of c-command as an experimental factor

is crucial to ensure the validity of the test, experiment 1 indicates that even then, the effect may fail to show up due to unrelated manipulations. The study allows us to reevaluate the empirical picture with much more nuance: A principle C violation under reconstruction is at most one out of many factors determining coreference resolution in \bar{A} -dependencies. Whether it is a relevant factor at all seems to depend on the individual speaker. It appears that to pragmaticists and psycholinguists, this is no novel empirical contribution. Rather, the study should prompt theoretical syntacticians to critically rethink their current practices in which principle C reconstruction is, based on this study, falsely attributed a high diagnostic value across situations, speakers and typologically unrelated languages.

Chapter 3

Experiments on German multi-gap dependencies

The findings presented in chapter 2 improve our understanding of the empirical conflicts discussed in chapter 1. To reiterate, the experiments do not support a view under which principle C reconstruction universally yields the unacceptability of coreference. Instead, speakers vary not only with respect to the size of the effect, but also regarding its presence. Furthermore, the presence of an alternative referent, context and the complexity of the experimental task significantly decrease the likelihood of detecting an overall significant effect at all. It is quite likely that increasing the complexity of the dependency, i.e. simply by testing principle C reconstruction in multi-gap rather than single gap dependencies, de-stabilizes the effect even further. Speakers may also be biased to find coreference more acceptable when the pronoun precedes one rather than the other gap simply due to the necessarily greater distance between the filler and the pronoun in the non-initial than the initial domain. In this chapter, we take a critical look at multi-gap dependencies in German.

We start with a pre-study on the acceptability of German parasitic gaps, i.e. experiment 3 in section 3.1. As touched on in section 1.3.1.2 of chapter 1, parasitic gaps are a highly variable phenomenon cross-linguistically. Many sources claim them to be rather marginal in German, and crucially, researchers seem to focus on parasitic gaps licensed by scrambling more than less controversial cases involving wh-movement. There exist no reliable data clarifying the status of parasitic gaps in German at all. The aim is to explore whether speakers find basic parasitic gap structures licensed by wh-movement acceptable before considering the application of any further diagnostics to test their underlying syntax. As it turns out, the majority of participants finds parasitic gaps marginal or unacceptable. As such, due to the low probability of getting interpretable results, the study of principle C

reconstruction in German parasitic gaps is set aside.¹

Section 3.2 presents a series of experiments on principle C reconstruction under ATB movement. Experiment 4, is based on the design employed by Salzmann, Wierzba, and Georgi (2023) for simple *wh*-dependencies, while experiment 5 and 6 are modeled after experiments 1 and 2 presented in chapter 2. We shall see that the effect of principle C reconstruction under ATB movement paints a similarly complex picture as under simple *wh*-movement. Whether a principle C reconstruction effect is detectable at all and whether it shows up in both gaps is highly dependent on the experimental task and setup. Furthermore, the experiments demonstrate significant inter-speaker variability with respect to the size and presence of the effect. The principle C reconstruction test in German therefore does not deliver conclusive outcomes regarding the underlying syntax of ATB movement.²

3.1 On the acceptability of German parasitic gaps

3.1.1 Existing reports

The literature on German parasitic gaps is dominated by cases licensed by scrambling. Scrambling itself is a much debated subject—essentially all conceivable types of explanations have been offered as to how scrambled word order is derived, specifically since it exhibits both characteristics of A-movement as well as \bar{A} -movement. Theories have analyzed scrambling either as A-movement (Fanselow 1987) or \bar{A} -movement (Müller 1993); as two separate operations, i.e. A-scrambling and \bar{A} -scrambling depending on the distance at which it applies (Déprez 1989; Mahajan 1990); a third independent type of movement (Webelhuth 1989) or base generation (Fanselow 1993, 2001). More recently, the mixed properties of scrambling have been embraced, with the idea emerging that it is not the landing positions that have conflicting A- and \bar{A} -properties but the features triggering movement (van Urk 2015). Crucially, scrambling seems to license parasitic gaps, constituting a prominent argument in support of an \bar{A} -movement approach:

- (1) a.

dass er Maria	[_{CP} ohne	— _{pg}	anzuschauen]	—	geküsst hat.
that he Maria	without		to-look-at		kissed has
'that he kissed Maria without looking at (her).'					

¹In future research, one could pre-screen relevant participants who accept parasitic gaps and investigate how they respond to principle C reconstruction. In the study on parasitic gaps in English purpose clauses in chapter 4 participants who did not accept the relevant constructions were filtered out, however only after testing.

²All supplementary materials from this thesis, including experimental items, raw data and analysis scripts can be accessed on OSF at <https://osf.io/hvm8u/> (Szarvas 2025b).

- b. *dass er [_{CP} ohne — pg anzuschauen] Maria geküsst hat
 that he without to-look-at Maria kissed has
 Intended: ‘that he kissed Maria without looking at (her).’

(Fanselow 2001, p. 411)

In (1a), the object *Maria* scrambles to a higher position preceding the adjunct clause and supposedly licenses a parasitic gap by doing so. In (1b), the object stays in its base position, and thus no parasitic gap is licensed. However, scrambling also exhibits some benchmarks of A-movement: It feeds Binding Principle A, for example. \bar{A} -movement reconstructs and thus cannot alter binding options, while A-movement does not reconstruct. In (2), *die Gäste* can only bind the reciprocal when interpreted in its surface position, assuming that direct objects are base generated below indirect objects in these examples:³

- (2) ?Peter hat die Gäste [ohne — pg anzuschauen] einander —
 Peter has the guests without to.look.at each.other
 vorgestellt.
 introduced
 ‘Peter introduced the guests to each other without looking at (them).’

(Webelhuth 1992, p. 209)

The literature on parasitic gaps in German is heavily preoccupied with the question whether these gaps licensed by scrambling constitute ‘real’ instances of parasitic gaps or not, and as such, whether these two gaps should be treated on a par. The following sections summarize previous reports about German parasitic gaps, focusing on the perspective taken by the respective author regarding their nature and distribution.

3.1.1.1 Pseudoparasitic vs. parasitic gaps in German

Fanselow (2001) argues that the gap in the adjunct clause is not a parasitic one, since the construction is also acceptable with non-referential DPs and inherent reflexive pronouns (3a), which standardly do not license parasitic gaps (Cinque 1990; Postal 1994). Further, there can be more than one gap in the structure (3b), i.e. both the direct object *das Buch* as well as the indirect object *dem Kind* leave a ‘gap’.

³This argument is based on the data provided by Grewendorf (1988). Note that this is controversial, as it has likewise been argued that the canonical base order is actually the opposite, i.e. IO > DO (Twiner and Lee-Schoenfeld 2019). For the earlier stages of the debate, see Fanselow (2007), Featherston (2007), and Featherston and Sternefeld (2003).

3.1. On the acceptability of German parasitic gaps

- (3) a. dass er sich anstatt (sich) um Maria zu kümmern mit Büchern
that he REFL instead REFL of Maria to care with books
beschäftigte
occupied
'that he occupied himself with books instead of caring for Maria'
b. dass er dem Kind das Buch anstatt zu leihen verkaufte
that he the.DAT child the.ACC book instead to lend sold
'that he sold the book to the child instead of lending it to him/her'
(Fanselow 2001, p. 412)

Instead, the apparent gap licensed by scrambling is not a result of \bar{A} -movement, but rather an instance of forward deletion in a coordinate structure (cf. Wilder 1997). Fanselow (2001) motivates this assumption by pointing out that the properties in conflict with canonical parasitic gaps happen to characterize coordinate structures. In (4a), we see deletion within an inherently reflexive construction, and in (4b), we see deletion of more than one phrase.

- (4) a. dass er sich [[um Maria kümmert] und [mit Büchern beschäftigt]]
that he REFL of Maria cares and with books occupies
'that he cares for Maria and occupies himself with books'
b. dass er dem Kind das Buch [erst lieh] und [dann verkaufte]]
that he Maria knows and Maria loves
'that he knows and loves Maria'
(Fanselow 2001, p. 412)

PF-deletion is enabled if nothing but heads that are coordinate conjunctions intervene between the identical antecedent A and the deletable constituent B. Furthermore, these heads must c-command B but not A. This prerequisite only holds if the preposition introducing the adjunct clause, i.e. *ohne* 'without' (and likewise *anstatt* 'instead', but not *bevor* 'before') behaves like a coordinating conjunction in the syntactic sense. Further evidence for the validity of this assumption comes from the fact that the complementizer *dass* 'that' can combine with the conjunctions, a trait that only applies to coordinating but not subordinating ones (5).

- (5) a. Es regnet ohne dass es schneit.
it rains without that it snows
'It rains without snowing.'
b. Er sagt, dass es regnet und dass es schneit.
he says that it rains and that it snows
'He says that it rains and that it snows.'
c. Es regnet bevor (*dass) es schneit.
it rains before that it snows

Intended: ‘It rains before snowing.’ (Fanselow 2001, p. 413)

It is noteworthy that (5a) and (5b) are impeccable in spite of the Doubly Filled Comp Filter—even dialects that do not tolerate violations thereof allow for these constructions. The forward deletion strategy is not applicable to similar English constructions. The heads *knowing* and *kissed*, which are not conjunctions, intervene between the two instances of Mary.

- (6) a. *because he kissed *Mary* without knowing *Mary*
 b. *He saw *Mary* and kissed *Mary*. (Fanselow 2001, p. 413)

Crucially, Fanselow’s aim is to show that the apparent parasitic gaps licensed by scrambling are fundamentally different from those accompanying \bar{A} -movement. He does so by providing evidence for their parallel behavior to conjunct reduction in coordinate structures. The account does not claim that parasitic gaps in German do not exist per se. Quite the contrary is implied: It is argued that the gap licensed by scrambling is not a parasitic one, but no claims are made about actual parasitic gaps otherwise.

3.1.1.2 Only pseudoparasitic gaps in German

Further illustrating the level of controversy that there is to this subject, we now turn to another approach that claims to show precisely that there is no such thing as a parasitic gap in German despite relying on the same kind of data as Fanselow (2001). Kathol (2001) argues that parasitic gaps in German do not exist. According to him, they are ungrammatical in extraposed clauses as in (7a), in cases where pronouns would become dislocated from postpositional structures as in (7b), as well as in relative clauses (7c), suggesting that the construction must be underlyingly different from its English equivalent because of its restricted distribution.

- (7) a. *Hans hat Maria geküsst [ohne _____{pg} anzuschauen].
 Hans has Maria kissed without to-look-at
 Intended: ‘Hans kissed Maria without looking at (her).’
 b. *weil mich da [ein Bild [_____{pg} von]] [____ vor] gewarnt
 because me there a picture of against warned
 hat
 has
 Intended: ‘because a picture of (it) warned me of (it)’
 c. *Dies ist das Buch welches [jeder [der _____{pg} liest]] ____
 this is the book which everyone who reads
 bewundert
 admires
 Intended: ‘This is the book which everyone who reads admires.’

(Kathol 2001, pp. 320–321)

It is claimed that the constructions like (1a), despite resembling English parasitic gaps on the surface, involve a special non-coordinate case of Left Node Raising (LNR, cf. Postal 1994).⁴ The relevant property is thus that these German gaps, although non-coordinate on the surface, behave like coordinations syntactically. By adopting this hypothesis, a number of otherwise mysterious properties fall into place. One of them is that constructions with multiple verbal heads are only acceptable if the heads are in clause-final position (8c), whereas constructions with verb-second (8a) or verb-initial word order (8b) are unacceptable.

- (8)
- a. *Hans sah Maria [erst lange an] und [dann leidenschaftlich geküsst
Hans looked Maria first long at and then passionately kissed
hat]
has
Intended: ‘Hans first looked at Maria for a long time and then kissed
(her) passionately.’
 - b. *Hans sah Maria [erst lange an] und [küsste dann leidenschaftlich]
Hans looked Maria first long at and kissed then passionately
Intended: ‘Hans looked at Maria first for a long time and then kissed
(her) passionately’
 - c. Hans hat Maria [erst lange angesehen] und [dann leidenschaftlich
Hans has Maria first long looked-at and then passionately
geküsst]
kissed
‘Hans looked at Maria first and then passionately kissed (her).’

(Kathol 2001, p. 325)

That these gaps are degraded when the finite main verb is fronted also readily follows from assuming a coordinate structure. The placement of the verb should be irrelevant because it is not involved in licensing the gap, contrary to fact:

- (9)
- a. #Hans küsste Maria [ohne —_{pg} anzusehen].
Hans kissed Maria without to-look-at
Intended: ‘Hans kissed Maria without looking at (her).’
 - b. *Hans sah Maria [ohne —_{pg} zu küssen] lange an.
Hans looked Maria without to kiss long at
Intended: ‘Hans gazed at Maria without kissing (her).’

⁴Kathol (2001) adopts the term ‘non-coordinate’ from Postal (1994), arguing that despite their appearance, the constructions are to be analyzed like coordinate constructions where the object is shared among the conjuncts (said conjuncts being the matrix and adjunct clause). I assume that he takes them to be quasi-coordinate in the sense that Williams (1987) does, i.e. a coordinate structure with, in this case, both syntactically and semantically asymmetrical conjuncts. Note however that he does not use an abstract coordinator, nor does he claim that the adverbial behaves like one. It is thus not entirely clear how and where the coordinate structure is actually realized.

(Kathol 2001, p. 326)

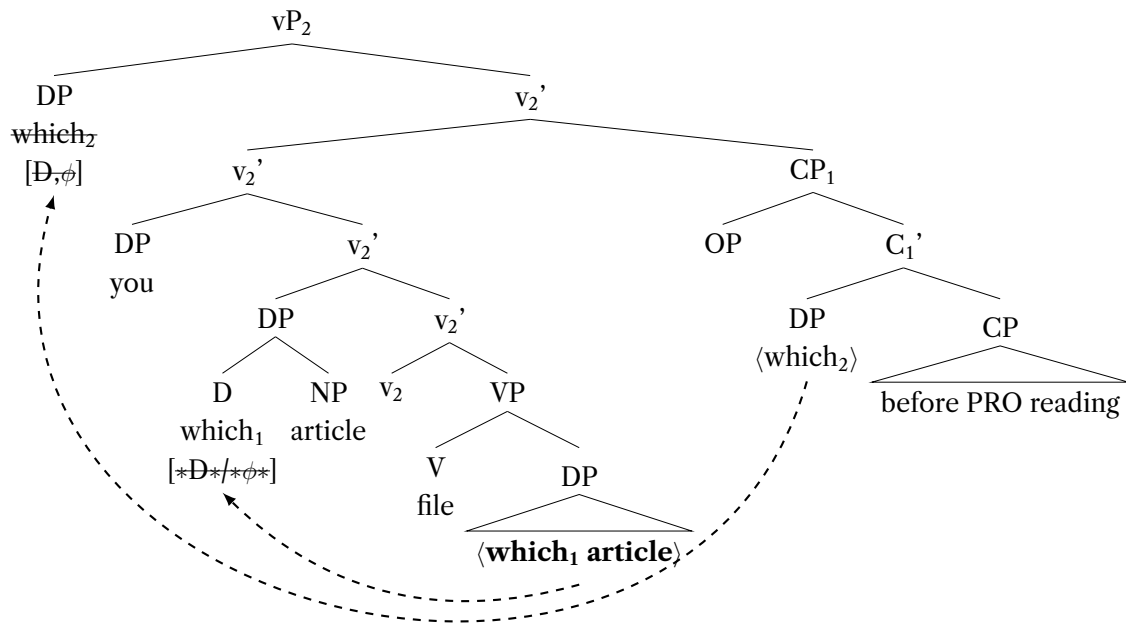
The paper goes on to highlight similarities between coordination and the examined structures regarding the behavior of ditransitives and datives, the licensing of multiple gaps and the fact that the gaps can have controllers beyond NPs. What does not fit the picture is that the choice of preposition seems to affect the acceptability of the clauses. Kathol (2001) offers a solution in the framework of Head-driven Phrase Structure Grammar (HPSG) where the central role of the preposition is built into the analysis. The prepositions fuse their argument structures (or parts thereof) with the V' category they modify, forming a complex predicate. By introducing the requirement that the prepositional head match the valence requirements of the verbal predicates, these differences are supposedly accounted for. Thus, Kathol (2001) assumes that the properties of the gaps licensed by scrambling represent the properties of all parasitic gaps in German. More explicitly, based on his observations, he concludes that parasitic gaps in German do not exist at all, claiming that all instances are merely pseudoparasitic.

3.1.1.3 Both types are parasitic gaps

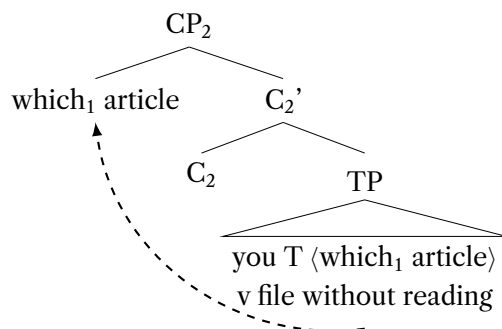
A more general, cross-linguistic approach is pursued by Assmann (2012). She accounts for English and German parasitic gaps and proposes an approach exhibiting both asymmetrical and symmetrical properties. Crucially, she treats cases licensed by scrambling and wh-movement to be uniform and does not postulate either of them to involve a coordinate structure. (10) and (11) illustrate a sample derivation the sentence *Which article did you file before reading?*. Under her analysis, parasitic gaps result from an inherently flawed numeration with fewer items than necessary to successfully check and delete all (probe and edge) features. To fix this, the wh-element fissions into two separate items, one of them bearing only the syntactically required probe and edge features to avoid the derivation from crashing (cf. Agbayani and Ochi 2007, 2022). This reduced item (*which*₂ in the derivation illustrated in (10)) eventually agrees with the wh-item bearing the phonological and semantic features (*which*₁) via c-command and is deleted under recoverability. There is only one wh-phrase fully equipped with all features, with its reduced counterpart, *which*₂, giving rise to the parasitic gap. Based on this analysis, a new definition of a parasitic gap emerges: a parasitic gap is a gap whose antecedent is the deleted part of a formerly fissioned lexical item, rather than an empty operator or the antecedent of the licensing gap that has been extracted from the matrix clause. (10) shows the derivational steps inside the matrix vP and the adjunct clause attached to it, while (11) illustrates that the DP *which*₁ *article* is ultimately extracted from the matrix vP.

3.1. On the acceptability of German parasitic gaps

(10) Fission and Fusion (Assmann 2012, p. 58)



(11) Sketch of matrix clause (Assmann 2012, p. 58)



The account is symmetrical in the sense that there is an instance of the antecedent present at the parasitic gap site as well as the licensing gap site. It also bears asymmetrical properties, since only the item originating in the matrix gap site (i) bears phonological and semantic features and (ii) raises to SpecCP of the matrix clause. The approach is particularly explicit about the technical details of the derivation, successfully accounting for most properties of parasitic gaps, but leaving open the issue why it is the *wh*-element specifically that fissions, and not, say, the NP (which is pied-piped in the final step of the derivation). Nevertheless, this is the most explicit approach from a technical perspective and the only one focusing on actual parasitic gaps in German licensed by (undisputed instances of) \bar{A} -movement. The following section provides a discussion on why the study of parasitic gaps should preferably focus on cases accompanying \bar{A} -movement, and why the conclusions drawn from approaches doing otherwise, yet claiming to account for parasitic gaps overall, are problematic.

3.1.2 Empirical complications

With the exception of Assmann (2012), none of the aforementioned literature on German parasitic gaps features typical gaps due to clear cases of \bar{A} -movement. The discussion of parasitic gaps in German is heavily focused on the parasitic gap-like structures licensed by scrambling despite the lack of clarity about the exact mechanism. This goes back to Felix (1985) who argues that the licensing of parasitic gaps is evidence for scrambling as (\bar{A} -)movement. Felix (1985) treats the two gap types as equal in German, arguing in favor of the empty operator approach (Chomsky 1986). While Felix (1985) primarily examines the gaps arising alongside scrambling Assmann (2012) focuses on parasitic gaps licensed by wh-movement and additionally subsumes the scrambled cases under her proposal. I will attempt to highlight why focusing on scrambling, such as Kathol (2001) does, inhibits the validity of within-language and cross-linguistic comparisons of parasitic gaps.

The main problem is that we are trying to learn about a phenomenon by limiting ourselves to its co-occurrence with yet another phenomenon exhibiting conflicting behavior. That is, scrambling has mixed A- and \bar{A} -properties, but parasitic gaps per definition are licensed exclusively by \bar{A} -movement. It is therefore not surprising that the gaps arising due to scrambling show a similarly peculiar set of characteristics as the operation that gives rise to their occurrence. With the exception of Fanselow (2001) who argues against this specifically, all publications on German parasitic gaps implicitly take scrambling to be \bar{A} -movement. What follows from this is that cross-linguistically, we are comparing apples and oranges: English does not have scrambling and thus no construction equivalent to the apparent gap(s) licensed by it in German, so naturally, differences between the two languages may be predicted on the basis of this alone. Assmann (2012), for example, acknowledges that the properties of parasitic gaps in German and English are not the same, but they do share a significant number of peculiar properties, prompting her to conclude that both should be derived by the same mechanism.

Crucially, German allows for typical parasitic gaps licensed by, for example, wh-movement. It appears that one can only conclude that parasitic gaps in German do not exist if one ignores the cases occurring in the presence of uncontroversial cases of \bar{A} -movement, i.e. the cases that actually match the definition of the phenomenon (cf. Kathol 2001). Examining parasitic gaps that are licensed by cross-linguistically available types of \bar{A} -movement is the only method that enables us to make reliable claims about their distribution. To summarize, the claim that German does not have parasitic gaps is thus based on two implicit, though strongly related assumptions. The first is that scrambling is \bar{A} -movement, having to explain away its A-properties. Second, one needs to assume that the distribution of gaps

licensed by scrambling is entirely equal to the gaps arising alongside genuine \bar{A} -movement such as wh-movement. Kathol's (2001) examination highlights the properties that the scrambled gaps share with coordinate structures. At the same time, his data and account do not provide any evidence for or against the existence, distribution and status of 'regular' parasitic gaps in German. Based on the terminology coined by Postal (1994), Kathol concludes that German only has *pseudoparasitic* gaps. Crucially, this conclusion is reached without taking a single look at gaps that unanimously fit the bill for regular parasitic gaps. As such, it may simply be the case that German has both parasitic and pseudoparasitic gaps, the latter ones licensed by scrambling.

Fanselow (2001) pursues a discussion in this direction. As summarized above, he concludes that no claims about scrambling can be derived from the fact that it seems to license parasitic gaps in German as the respective structure does not seem to correspond to a typical parasitic gap, but is argued to be an instance of forward deletion in a coordinate structure. Although both Kathol and Fanselow point out similarities to coordinate structures, they differ with respect to the conclusions they draw from the data. While Kathol infers that any candidate for a parasitic gap in German shares the properties of the gaps licensed by scrambling, Fanselow implies that the gaps licensed by scrambling are a different construction than parasitic gaps licensed by \bar{A} -movement.⁵ I follow this assessment by Fanselow (2001) and assume that the gaps arising due to scrambling are likely different from canonical parasitic gaps. As for scrambling itself, no strong opinions will be defended here as it goes well beyond the scope of this thesis. As long as the central aim is to study the properties of parasitic gaps, claims about them should be based on the instances appearing when an element is \bar{A} -moved, and not the fringe cases where the involvement of \bar{A} -movement is subject to debate.

The remainder of this section is concerned with an experimental investigation of German parasitic gaps in sentences with wh-movement. As such, the study will establish whether parasitic gap constructions in German are suitable for further investigation of principle C reconstruction in multi-gap dependencies and omit any confounds introduced by scrambling. In multiple sources, it is noted that they are rather marginal cross-linguistically, and German is no exception to this observation (Assmann 2010, 2012; Fanselow 2001; Felix 1985; Himmelreich 2017; Kathol 2001). I report on a rating experiment designed to assess the acceptability of parasitic gaps compared to the minimal pairs with a coreferent pronoun in place of the gap. Furthermore, the study explores whether extraposing the adjunct clause has an effect on the acceptability of the parasitic gap construction. It is reported that gaps licensed by scrambling are only acceptable as long as the adjunct clause

⁵Indeed, Fanselow does not explicitly comment on his views on regular parasitic gaps.

is not extraposed (Kathol 2001), while no such effect is reported for parasitic gaps licensed by *wh*-movement Assmann (2012). If the extraposed parasitic gaps are not outright unacceptable, we will have preliminary empirical support for the view that the gaps licensed by scrambling in German are underlyingly different from the parasitic gaps licensed by \bar{A} -movement (cf. Fanselow 2001).⁶

3.1.3 Experiment 3: German parasitic gaps

3.1.3.1 Design

Since the status of regular German parasitic gaps is somewhat unclear based on the existing literature, an acceptability judgment study is in order. An experiment was designed manipulating the factors *GAP* with levels *pg* and *prn*; and the factor *PLACEMENT* with levels *incorp* and *extra* (determining the placement of the adjunct clause). The factors were fully crossed, yielding a 2×2 Latin Square design and four conditions.

PLACEMENT / GAP	<i>pg</i>	<i>prn</i>
<i>incorp</i>	<i>incorp, pg</i>	<i>incorp, prn</i>
<i>extra</i>	<i>extra, pg</i>	<i>extra, prn</i>

Table 3.1: 2×2 design, Latin Square distribution.

The factor *GAP* tests whether adjunct clauses containing a parasitic gap are perceived as significantly less acceptable than adjunct clauses containing an inanimate third person singular pronoun instead. The factor *PLACEMENT*, on the other hand, addresses whether participants perceive adjunct clauses that are extraposed as less acceptable than incorporated ones. The interaction of the two factors would indicate an impact of the placement of the adjunct clause on the acceptability of parasitic gaps.

3.1.3.2 Materials

All experimental items included a matrix clause and an embedded clause with an extracted *wh*-object. The adjunct clause is attached to the embedded clause and either follows it (*extra* conditions) or follows the subject of the embedded clause (*incorp* conditions). Further, the adjunct clause either contains a parasitic gap (*pg* conditions) or a coreferent pronoun (*prn* condition). The experimental items of this study aimed to get as close as possible to the intended structure of the items employed in the other experiments reported in this thesis given that the aim was

⁶To definitively answer this question, a follow-up investigation of the acceptability of gaps licensed by scrambling is necessary. I leave this open for future research.

3.1. On the acceptability of German parasitic gaps

to assess whether further investigations would be reasonable. A sample item is illustrated in (12):

- (12) a. *Incorp, pg*
Marie hat erzählt, welches Buch Clara ohne zu lesen ____ pg
Marie has told which book Clara without to read
verschenkt hat ____ .
gifted.away has
'Marie told (us) which book Clara gifted (to someone) without reading.'
- b. *Incorp, prn*
Marie hat erzählt, welches Buch Clara ohne es zu lesen
Marie has told which book Clara without it to read
verschenkt hat ____ .
gifted.away has
'Marie told (us) which book Clara gifted (to someone) without reading it.'
- c. *Extra, pg*
Marie hat erzählt, welches Buch Clara verschenkt hat ____ ohne
Marie has told which book Clara gifted.away has without
zu lesen ____ pg .
to read
'Marie told (us) which book Clara gifted (to someone) without reading.'
- d. *Extra, prn*
Marie hat erzählt, welches Buch Clara verschenkt hat ____ ohne
Marie has told which book Clara gifted.away has without
es zu lesen.
it to read
'Marie told (us) which book Clara gifted (to someone) without reading it.'

The grammatical gender of the names used in the items mismatched the grammatical gender of the filler, which was invariably neuter. The reasoning behind this is to avoid potential coreferent readings between the names and the pronoun in the conditions illustrated by (12b) and (12d). The experiment included 36 distractors which were taken from work by Brown et al. (2021) based on Featherston (2009), where they were structured into six levels based on their grammaticality.⁷ Six sentences were chosen from each level to ensure an even

⁷These distractors were designed and extensively tested by Featherston (2009) in an effort to identify anchoring points in acceptability judgment experiments. They cover a range of normed acceptability levels from unacceptable to acceptable, but the sentences within levels do not (intentionally) share any structural properties. As such, participants are exposed to a variety of sentence types and acceptability levels.

proportion of sentences invoking a certain acceptability judgment. This ensures that participants stay alert by being presented with various degrees of acceptability and to keep them from forming strategies. The exceptionally ungrammatical distractors also served as attention checks as participants who consistently gave them high ratings were excluded.

3.1.3.3 Method and procedure

The experiment employed a repeated-measures acceptability judgment task on a 7-point Likert scale ranging from completely unacceptable (1) to completely acceptable (7). The experiment was set up on the platform L-rex (Starschenko and Wierzba 2024). In each trial, the item appeared on the screen alongside the scale and the question *Wie akzeptabel ist dieser Satz?* ('How acceptable is this sentence?'). Participants were instructed to rate the sentences based on their gut feeling, and not to spend a long time dwelling on individual sentences. They were informed that the completion of the study should take them around ten minutes. It was also emphasized that the point is not to test their knowledge of German but to use the information they provide to learn something about language. Participants were asked to provide information about any second languages they spoke proficiently, the state of Germany they spent most of their lifetime living in, and whether they considered themselves a speaker of a certain dialect of German. The latter two variables were assessed so that one could trace back potential inter-speaker variability to specific regions. A short training sequence involved one unacceptable and one acceptable sentence, illustrating how to navigate the scale for the rest of the experiment.

3.1.3.4 Participants

A total of 80 native speakers of German (mean age 35.1, SD = 11.73) were recruited via the online platform *Prolific* to participate in the study. Using the platform's prescreening options, participants had to be born and raised in Germany, raised monolingually to avoid influences of another native language, and must have spent most of their lifetime in Germany. Participants' states of origin were grouped into rough geographical regions, i.e. South, West, North and East, respectively, to increase the number of data points available in each region.⁸ Participants hailed from the following regions: $n = 24$ from the South, $n = 34$ from the West, $n = 12$ from the North, and $n = 10$ from the East. This information was merely intended

⁸The states were grouped as follows: Baden-Württemberg and Bavaria comprised the South; Hesse, Rhineland Palatinate, Saarland and North-Rhine Westphalia comprised the West; Hamburg, Bremen, Lower Saxony and Schleswig-Holstein comprised the North; and Berlin, Brandenburg, Saxony, Saxony-Anhalt and Thuringia comprised the East.

to serve as a post-hoc indicator for any inter-speaker variability that could arise and as such was not used as a predictor in the modeling process. All participants received monetary compensation.

3.1.3.5 Hypotheses

The research question of this preliminary study is how acceptable German parasitic gaps in adjunct clauses are in comparison to adjunct clauses without parasitic gaps. The study further investigates the effect that the placement of the adjunct clause containing the parasitic gap has on the acceptability of the sentence. Based on reports from the literature that parasitic gaps are a rather marginal phenomenon, a main effect of *GAP* is expected, indicating that the presence of a parasitic gap leads to a decrease in acceptability. Anecdotal evidence indicates that speakers vary with respect to the effect of *PLACEMENT*. Some speakers dislike extraposition in general, in which case we expect a main effect of *PLACEMENT*. Others seem to prefer it specifically with parasitic gaps, in which case we would expect an interaction between *GAP* and *PLACEMENT*. Assmann (2010, p. 110) notes that the position of the adjunct clause does not affect the grammaticality of cases *with* parasitic gaps, but if there is no parasitic gap but a pronoun as in the condition *incorp, prn*, incorporation becomes less acceptable. In that case, we would likewise expect an interaction between the two factors. However, given these mixed judgments from individual speakers, it is questionable whether we can expect either of the tendencies to dominate. Rather, we expect high inter-speaker variability and no significant effect of *PLACEMENT*.

3.1.3.6 Data Analysis

Data analysis was carried out in the frequentist framework. All calculations were executed using R (R Core Team 2024) via RStudio (RStudio Team 2021). Visualizations were also generated in RStudio, using the package *ggplot2* (Wickham 2016). The factors *GAP* and *PLACEMENT* were treatment contrast coded with *pg* and *extra* as the base levels, respectively. Statistical modeling was carried out via ordinal regression using a Cumulative Link Mixed Model (Christensen 2019). For significance testing, a conservative alpha level of 0.05 was defined. Prior predictor checks were carried out by comparison of the minimal models including none or each of the predictors separately, using the ‘*anova*’ function (R Core Team 2024). They revealed a significant main effect of both predictors. Since the model with the maximal random effects structure did not turn out as a significantly better fit based on an ‘*anova*’, the final model includes the main effects of *GAP* and *PLACEMENT*, their interaction, and the random slopes and intercepts per item and participant

for each factor but not their interaction (cf. Barr et al. 2013).

$$(13) \quad \text{acceptability} \sim \text{gap} * \text{placement} + \\ (1 + \text{gap} + \text{placement} | \text{participant}) + (1 + \text{gap} + \text{placement} | \text{item})$$

A post-hoc analysis was carried out to further include the factor trial in the analysis, upon suspecting that participants may have become accustomed to the rating task or the complexity of the construction over the course of the experiment. The number of trials was centered and included in the model in addition to the factors named above.

$$(14) \quad \text{acceptability} \sim \text{ctrtrial} * \text{gap} * \text{placement} + \\ (1 + \text{ctrtrial} + \text{gap} + \text{placement} | \text{participant}) + (1 + \text{ctrtrial} + \text{gap} + \text{placement} | \text{item})$$

3.1.3.7 Results

CLMM	Estimate (SE)
gap	3.18*** (0.35)
placement	0.69*** (0.19)
gap:placement	-1.82*** (0.25)
Log Likelihood	-1476.82
AIC	2991.64
BIC	3084.11
Num. obs.	960
Groups (participant)	80
Groups (item)	12
Variance: participant: (Intercept)	5.70
Variance: participant: gap (prn)	6.16
Variance: participant: placement (incorp)	0.62
Variance: item: (Intercept)	0.11
Variance: item: gap (prn)	0.03
Variance: item: placement (incorp)	0.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3.2: Estimates of the cumulative link mixed effects model given in (13).

The estimates of the model in (13) are reported in Table 3.2. The model reveals a significant main effect of GAP and PLACEMENT as well as a significant interaction of the two. Based on the treatment contrast coding (base level of GAP being *pg* and of PLACEMENT *extra*), the presence of a pronoun has a positive effect on the acceptability of the items, i.e. a parasitic gap makes the items significantly less acceptable. Likewise, the incorporation of the adjunct clause has a significant

3.1. On the acceptability of German parasitic gaps

effect on acceptability. The significant interaction, taken together with Figure 3.1, indicates a crossover interaction. That is, incorporation has a positive effect in the presence of a parasitic gap while it has a negative effect in its absence. The impact of variable slopes for GAP by participant dominates among the random effects.



Figure 3.1: Mean acceptability ratings based on levels of GAP and PLACEMENT.

This indicates that participants vary mostly with respect to how the presence vs. absence of a parasitic gap affects their response, whereas they do not vary as much with respect to how the placement of the adjunct clause impacts it. Likewise, we see that there is much less variance with respect to items, i.e. responses vary primarily based on participant and do not fluctuate as much based on which item

is presented. The mean ratings of the items are depicted in Table 3.3. Overall, the type of parasitic gap sentence standardly reported in the English literature where the adjunct clause follows the licensing gap, i.e. *extra*, *pg* is the least acceptable out of the four conditions in German with a mean rating of 3.86.

The condition *extra*, *prn*, where the adjunct clause is extraposed containing a pronoun, scored the highest mean rating of 5.49, illustrating the significant crossover interaction between the factors GAP and PLACEMENT. The conditions where the adjunct clause preceded the licensing gap are a bit closer together regarding mean ratings, with the *incorp*, *prn* condition scoring a bit higher (4.95) than the *incorp*, *pg* condition (4.25). Standard error is the lowest in the condition

Condition	Mean rating (SE)
<i>incorp</i> , <i>pg</i>	4.25 (0.110)
<i>incorp</i> , <i>prn</i>	4.95 (0.098)
<i>extra</i> , <i>pg</i>	3.86 (0.110)
<i>extra</i> , <i>prn</i>	5.49 (0.085)

Table 3.3: Mean ratings by condition.

that scored the overall highest, i.e. *extra*, *prn*; whereas it is equally high in the two conditions featuring parasitic gaps. Sentences with parasitic gaps received a combined mean rating of 4.05, while the sentences without the gap received a combined mean rating of 5.20. As a comparison, filler sentences of the highest of six acceptability categories (level A) received a mean rating of 6.91, whereas distractors from the two mediocre categories (levels C and D) received a mean

rating of 4.03 and 2.97, respectively. The plot in Figure 3.2 illustrates the exact distribution of responses in experimental conditions. The spread of judgments for parasitic gaps is greater than for the sentences without them, suggesting that variability is particularly rampant within the factor level *pg*.

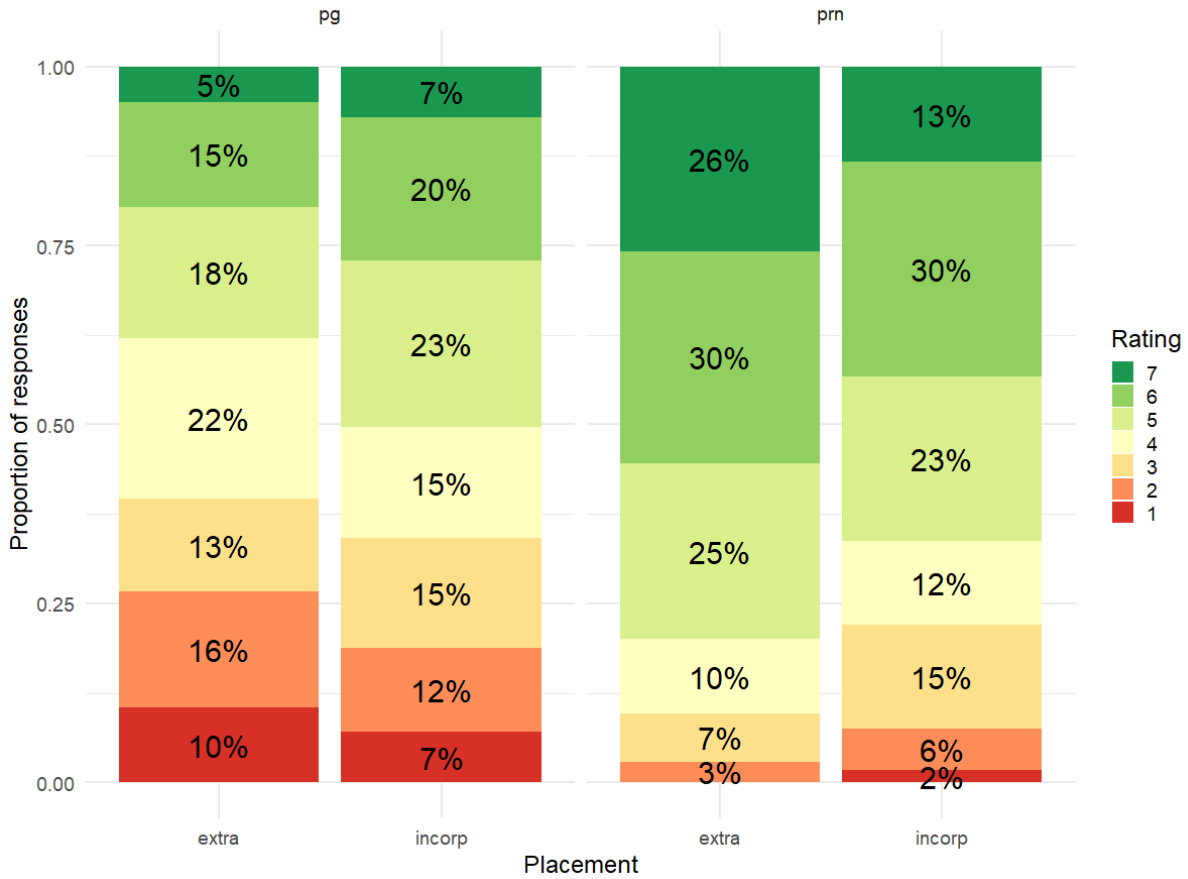


Figure 3.2: Distribution of responses across factor levels of GAP and PLACEMENT. Percentages indicate proportion of responses within each condition.

The conditions *extra, pg* and *incorp, pg* facilitated responses all along the 7-point-scale. In the former condition, only 38% of responses are a 5 or higher, indicating that the majority of trials, i.e. 62%, elicited ratings of 4 or lower. Again, this illustrates that the most canonical cases of parasitic gaps known from English, i.e. parasitic gaps in extraposed adjunct clauses, are rated rather poorly in German. Notice also that 26% of responses to this condition are a 2 or lower, indicating that roughly a quarter of trials elicited responses indicating outright unacceptability. The positive effect of adjunct clause incorporation is visible in the second bar plot from the left, illustrating the proportion of responses to the condition *incorp, pg*. Ratings of 5 or higher make up exactly 50% of responses here, indicating that 50% of responses are a 4 or lower. However, again, 19% of responses indicate unacceptability in the range of 1 and 2. Turning to the control conditions without parasitic gaps, we see that 81% of responses in the condition *extra, prn* are a 5 or

3.1. On the acceptability of German parasitic gaps

higher, whereas for *incorp,prn*, this proportion is at 66%, illustrating the crossover interaction between GAP and PLACEMENT compared to *pg* conditions. Only 3% of responses indicate unacceptability (rating 2 or lower) in the condition *extra,prn*, whereas 8% of responses do in the condition *incorp,prn*.

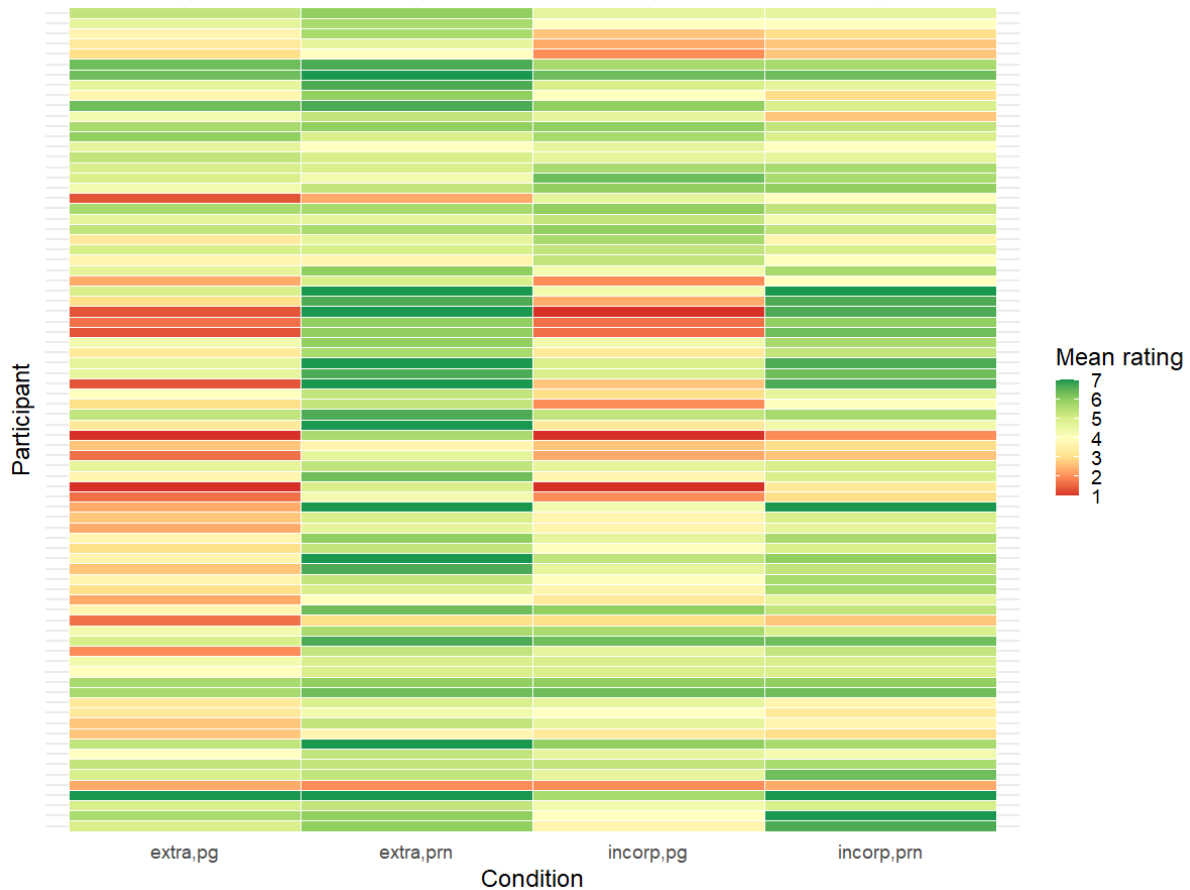


Figure 3.3: Heatmap of participants clustered by their mean ratings per condition. Each cell represents one participant's mean rating in the respective condition, each row represents one participant's ratings across conditions. Red cells indicate low means, green cells indicate high means. Clustering based on raw means.

Another interesting piece of information is how individual participants responded and what patterns their responses exhibit. That is, whether there are participants who categorically rejected or accepted every experimental condition, and more importantly, whether there are certain participant (clusters) who specifically reject parasitic gap structures above all. The heatmap in Figure 3.3 illustrates participants' response patterns based on their mean ratings for each condition. Participants are clustered based on how similar their raw responses are. That is, each participant is assigned a vector consisting of their mean ratings for each condition, and the patterns are determined by measuring the (Euclidean) distance between these vectors. For each participant, the participant with the most similar absolute rating level is determined. Each pair of most similar participants is merged to form a cluster. Then, the closest cluster of participants is determined,

merged again, and so on. The ordering of participants' lines in the heatmap depicts this clustering. Straight away, we see a clear vertical pattern indicating the effect of the experimental manipulation: conditions without parasitic gaps elicited more positive ratings than conditions with them. We further see that there are participants who rated every condition rather favorably indicated by mostly green rows, participants who found every condition mediocre overall indicated by yellow rows. Only one participant gave every condition low ratings overall, indicated by the single orange row in the bottom section of the plot. We likewise see that for some participants, the placement of the adjunct clause was a more relevant factor in determining the response pattern than the presence of a parasitic gap, as indicated by orange cells in *extra* conditions and pale green cells in *incorp* conditions. Overall, most participants exhibiting low ratings in *pg* conditions, qualified by red cells, show a remarkable improvement of their ratings in *prn* conditions. Overall, the plot illustrates a multitude of response patterns with some major trends, but no clear dominant pattern.

Of the participants who rated parasitic gaps poorly with mean ratings between 1 and 2, some provided comments along with their rating. These comments in (15)–(19) reveal that these participants prefer to either insert a pronoun or avoid the use of an adjunct clause entirely by suggesting paraphrases.⁹

- (15) Participant 6, Schleswig-Holstein, no dialect
“ohne ES anzuprobieren” klänge für mich natürlicher
 “‘without trying IT on” would sound more natural to me’
- (16) Participant 12, Hesse, no dialect
‘ohne “es” zu reparieren’
 ‘without repairing “it”’
- (17) Participant 15, Baden-Württemberg, Swabian
“ohne aufzuschieben” ist eingeschoben. Richtig: “...welches Problem Lisa zeitnah gelöst hat.”
 “‘without postponing” is interjected. Correct: “...which Problem Lisa solved timely.”
- (18) Participant 22, Baden-Württemberg, Palatine
‘Referenz “es” zu Dokument fehlt’

⁹It is to be noted that some of the participants did not seem to be entirely naive regarding grammar theory, some of them using terms common among scholars of German studies or teachers of German—*Prädikat* to refer to the verb as well as *Verbendstellung* and *Verbzweitstellung* when referring to its position. These comments only pertained to the distractors, and there is no reason to assume that participants understood which items were relevant to the study based on their potential linguistic knowledge.

‘reference to document “it” is missing’

(19) Participant 72, Baden-Württemberg, no dialect

‘Ohne Anprobe oder ohne es anzuprobieren’

‘without try-on or without trying it on’

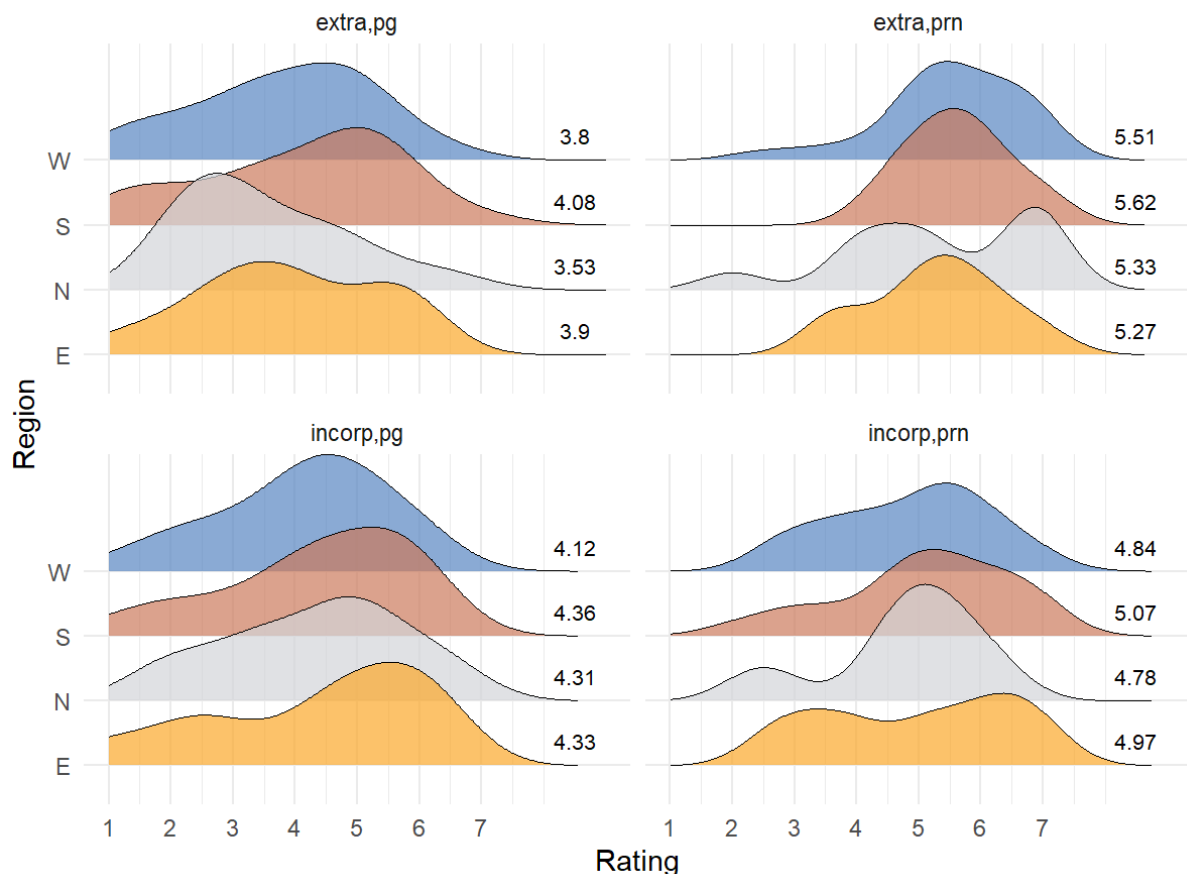


Figure 3.4: Density of mean ratings per participant per condition based on participants' region of origin. Mean ratings per condition per region are given next to the curves on the right.

Figure 3.4 illustrates the distribution of mean ratings based on condition and region of origin, showing the mean rating for the respective condition across all participants of a single region. The clearest and possibly only distinctive tendency that emerges is that in the condition *extra, pg*, participants from the west and south seem to pattern together in being skewed slightly toward higher ratings, while the peak for participants from the north and east is skewed toward lower ratings. All regional mean ratings correspond to the pattern qualifying the significant crossover interaction—conditions are ranked as follows: *extra, pg* is rated the lowest, followed by *incorp, pg*, then *incorp, prn* and then *extra, prn*. Most dialect speakers are from the south (n = 8 participants Swabian or Alemmanic, n = 6 Bavarian or Franconian), followed by the west (n = 4 Palatine, n = 3 Hessian, n = 1

Saarland dialect) and the east ($n = 3$ Saxonian, $n = 1$ Berlin German). The sample of identifiable dialect speakers is thus quite rather small, as $n = 54$ participants stated that they only spoke Standard German. Recall also that the sample is not balanced with respect to participants' geographical region either (regardless of whether they indicated to speak a dialect or not) as there are 24 participants from the south and 34 from the west, but only 12 from the north and 10 from the east as discussed in section 3.1.3.4. On the one hand, this explains the dips in the density curves of northern and eastern participants, and likewise that they show a greater difference in means between conditions than the other two groups. Due to this imbalance, claims about any regional or dialectal influences are unjustified at this point.

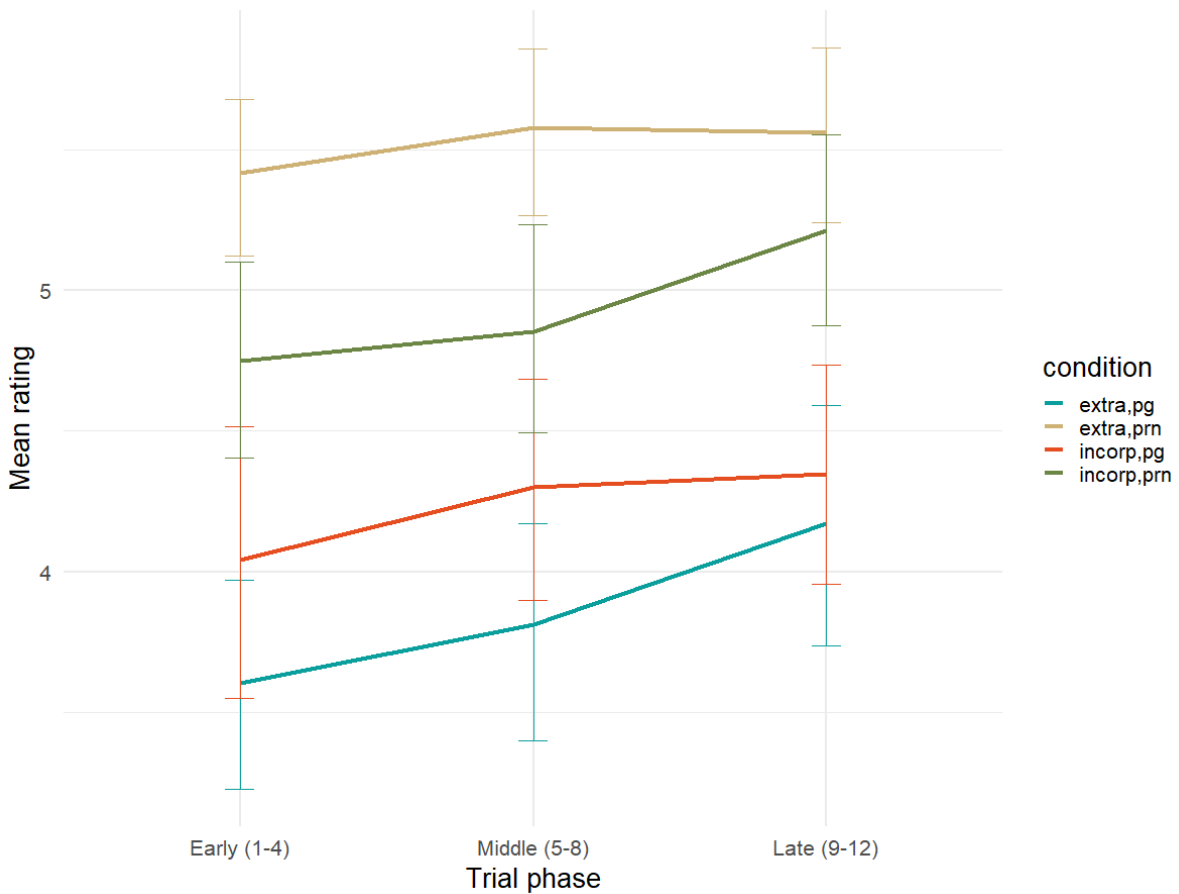


Figure 3.5: Effect of trial number on acceptability of experimental conditions. Points indicate mean ratings across participants and early, middle and late trials in respective condition, error bars indicate standard error.

Finally, we turn to the possibility of a satiation effect. Figure 3.5 shows participants' mean ratings for trials 1 to 4 (early phase), 5 to 8 (middle phase), and 9 to 12 (late phase). Recall that another model was fitted to include information about trial number and its interactions with the factors `GAP` and `PLACEMENT`, see (14). Trial number was included as a numerical variable centered around 0 called `CTRIAL`. The cumulative link mixed model with random effects for slopes and intercepts by participant and item for each factor individually (without interactions) yielded

a significant effect of `CTRIAL` ($\Pr(>|z|) \sim 0.00617$) (along with a significant effect of `GAP` and `PLACEMENT` as well as their interaction) but no significant interaction with any of the other factors. That is, increasing trial number seems to have a positive effect on ratings overall, however, this does not seem to coincide with the presence of a parasitic gap. Thus, parasitic gaps are not subject to a satiation effect in this experiment in the sense that otherwise unacceptable or mediocre syntactic structures became more acceptable. Rather, participants' ratings generally became slightly more positive in later trials, which is indicative of an adaptation effect. That is, rather than becoming accustomed to the construction, participants may have adapted to the experimental task by judging sentence after sentence in a very repetitive pattern (cf. Baayen et al. 2017). However, note that this trend seems to track the significant crossover interaction between `GAP` and `PLACEMENT`: the conditions where `PLACEMENT` has a negative effect overall, i.e. *extra,pg* and *incorp,prn*, receive a particular boost in the final four trials indicated by the upward trend of the graph between the middle and late phase. On the other hand, conditions where `PLACEMENT` has a positive effect overall, i.e. *extra,prn* and *incorp,pg*, show a slight increase in earlier trials. Note, however, that the error bars indicating standard error overlap in multiple conditions, and although the effect is significant, it is rather small.

3.1.4 Discussion

The results of the experiment confirm the hypothesis that sentences with parasitic gaps are less acceptable than sentences without parasitic gaps in German. Furthermore, the hypothesis that inter-speaker variability will overshadow any significant effect of the placement of the adjunct clause does not receive support. Instead, adjunct clause incorporation seems to have a positive effect on the acceptability of parasitic gaps, whereas it has a negative effect on the acceptability of constructions without parasitic gaps (cf. Assmann 2010). These findings are qualified by significant main effects of the factors `GAP` and `PLACEMENT` as well as a significant interaction between the two. The experiment did not reveal any clear influence of participants' region of origin or dialect on the acceptability of parasitic gaps. There might be regional differences with respect to the acceptability of parasitic gaps in extraposed adjunct clauses as indicated by Figure 3.4, but the sample is not balanced enough to allow for conclusions beyond this slight tendency. We further find that increasing trial number has a significantly positive effect on the ratings participants give. However, this is not a satiation effect indicating that participants become 'accustomed' to otherwise unacceptable parasitic gaps under repeated exposure, but rather that they become more lenient with their judgments

over the course of the experiment in general. Having to judge sentence after sentence in a very repetitive pattern for 10 to 20 minutes on average, participants may become bored or inattentive and end up giving more lenient judgments than when they are still ‘fresh’ at the beginning of the study. This is indicated by a significant effect of `CTRIAL` in the absence of an interaction with any of the two factors. The next section explores a potential explanation for the crossover interaction of `GAP` and `PLACEMENT`.

3.1.4.1 Crossover interaction of adjunct clause placement

To explore the possible implications of the crossover interaction between `GAP` and `PLACEMENT`, I invite the reader on a de-tour to the processing of filler-gap dependencies. Recall that adjunct clause extraposition has a negative effect on the acceptability of parasitic gaps, while it has a positive effect on the acceptability of conditions without a parasitic gap, crucially featuring a pronoun referring to the filler instead. Since extraposition only affects the linear position of the adjunct clause, this section focuses on an explanation in terms of linear rather than hierarchical structure.

We can assume that encountering a displaced *wh*-phrase in a clause prompts a search for a gap (based on the first resort model of gap filling and Active Filler Hypothesis by Clifton and Frazier 1989; Frazier 1987). The parser supposedly associates the filler to a gap as soon as possible, meaning that it actively posits gaps in syntactically and lexically licit positions. This hypothesis has received widespread support over the past decades, most notably based on studies showing the ‘filled gap effect’ (see summary by Ness and Meltzer-Asscher 2017). Slower reading times are detected in the critical region of sentences where a potential gap site is unexpectedly filled (20a), suggesting that a gap was posited as the direct object of *bring*, compared to sentences where this was not the case (20b):

- (20) a. My brother wanted to know who Ruth will bring us home to ____ at Christmas.
 b. My brother wanted to know if Ruth will bring us home to Mom at Christmas. (Ness and Meltzer-Asscher 2017, p. 1354)

McElree and Griffith (1998) tested how the resolution of filler-gap dependencies interacts with island constraints, finding that when assigning gaps to fillers, this type of syntactic information is considered 200-400 ms sooner than local lexical constraints. This suggests that the parser considers syntactic islandhood before positing a gap based purely on lexical information. Now consider the structures below in (21), corresponding to simplified versions of the experimental items.

3.1. On the acceptability of German parasitic gaps

- (21) a. Welches Fahrrad hat Lars ohne — pg zu reparieren verkauft
 which bicycle has Lars without to fix sold
 — ?
 ‘Which bicycle did Lars sell without fixing?’
- b. Welches Fahrrad hat Lars ohne es zu reparieren verkauft — ?
 which bicycle has Lars without it to fix sold
 ‘Which bicycle did Lars sell without fixing it?’
- c. Welches Fahrrad hat Lars verkauft — ohne — pg zu
 which bicycle has Lars sold without to
 reparieren?
 fix
 ‘Which bicycle did Lars sell without fixing?’
- d. Welches Fahrrad hat Lars verkauft — ohne es zu reparieren?
 which bicycle has Lars sold without it to fix
 ‘Which bicycle did Lars sell without fixing it?’

The most acceptable condition, i.e. *extra, prn*, is represented by (21d). In these cases, the adjunct clause contains no parasitic gap and follows the matrix clause. From a syntactic perspective, it is unsurprising that this condition received the highest mean rating out of all (5.49)—it is syntactically well-formed and does not contain a parasitic gap. If we assume the Active Filler Hypothesis, potential contrasts between this condition and *incorp, prn* represented by (21b), i.e. where incorporation of the adjunct clause led to a decrease in mean rating (4.95), may also receive an explanation. In (21d), the filler-gap dependency is resolved within the matrix clause when the real gap is encountered. The parser does not have a reason to posit a gap in the adjunct clause: There is no filler in need of being assigned to a gap, and additionally, the adjunct clause is a syntactic island in which gaps should not be posited in the first place (McElree and Griffith 1998). Hence, there is nothing unlikely or inherently problematic about this particular dependency. In the type of items represented by (21b), however, the parser encounters an adjunct clause instead of a continuation of the matrix dependency. Having established that the parser does not posit gap sites in syntactic islands, the reason for the decrease in acceptability may be the introduction of additional material that increases the distance between the filler and the potential gap. What we observe here may thus simply be a proximity effect, whereby sentences that resolve the filler-gap dependency earlier on are preferred over ones where the distance between the two elements is increased (Chaves and Putnam 2020 based on Alexopoulou and Keller 2007; Hofmeister and Norcliffe 2013).

Contrasts between items represented by (21a) and (21c), which scored a mean rating of 4.25 and 3.86, respectively, may also receive an explanation based on the

same hypothesis. In a case like (21a), again, the parser is looking for structures in which to posit a gap to assign to the filler, but instead encounters an adjunct clause. This time, it is not only the increase in distance that may cause difficulty, but also the presence of a gap in a syntactic island. We can hypothesize that another gap in the matrix clause is posited in order to repair the island violation. In other words, the search for a gap to match the filler is likely not terminated by the parasitic gap, and recognizing it as such, the parser will posit another gap in the matrix clause to resolve the dependency. In the parasitic gap condition where the adjunct clause is extraposed shown in (21c), the additional decrease in acceptability may be the result of yet another unexpected property that the adjunct clause exhibits. The filler-gap dependency is resolved in the matrix clause, much like in the case of the highest rated condition represented by (21d). After this, not only does the parser encounter a gap at a point in the sentence where there is no active filler to associate with it; on top of that, the gap is entirely unexpected because it occurs in a syntactic island. In other words, it is only the last chunk of the sentence that reveals the parasitic gap dependency to the parser. While on-line measurements are necessary to test these speculations, the crossover interaction between GAP and PLACEMENT may follow from the way that filler-gap dependencies are processed rather than grammatical constraints.

3.1.4.2 Implications for the study of reconstruction

Although participants rated parasitic gaps to be rather mediocre, sentences without parasitic gaps were also rated to be less acceptable than distractors on the second-highest level of acceptability (5.20 for items in *prn* conditions vs. 5.94 for level B distractors). This could likely reflect the overall complexity of the items and the added effect of the placement (and sheer presence) of the adjunct clause. Although the distractors vary in length, their syntactic complexity is lower, which may have made the target items stand out as more complex and thus less acceptable. One of the aims of this experiment was to identify whether parasitic gaps are acceptable enough in German to proceed with the study of reconstruction patterns. This is to ensure that participants are not distracted by the mediocre acceptability of the construction itself when judging different readings of the sentence, especially because the sentences will become longer and more complex once multiple referents and a potentially ambiguous pronoun are included on top of a parasitic gap construction. These simple cases of parasitic gap constructions lacking any ambiguity received a mean rating of 4.05 overall (4.25 if the adjunct clause was incorporated, 3.86 if not). The extraposed condition is of particular relevance for the study of reconstruction effects to control for interfering effects of linear vs. structural distance and order (when compared to ATB movement).

As such, no large scale experiment will be conducted on reconstruction effects in German parasitic gaps due to their compromised overall acceptability and the risk that their marginal nature is further amplified when combined with the potential ambiguity of reconstruction. It is nevertheless to be acknowledged that a non-negligible proportion of participants gave parasitic gaps a rather favorable rating, and that these speakers could be pre-screened in future studies to ensure the validity of their responses to reconstruction data. Nevertheless, the general tendency is clearly that the acceptability of parasitic gaps is degraded. Future research could aim to narrow down the factors impacting inter-speaker variability not only with respect to parasitic gaps, but also the placement of adjunct clauses in general as there appear to be clearer patterns in this variability than previously expected.

3.1.4.3 Differentiating parasitic and pseudoparasitic gaps

Based on the experiment, there are certain claims from the literature on German parasitic gaps (and what counts as one) that we need to re-evaluate. First, the claim that German parasitic gaps do not exist at all is unsolicited (Kathol 2001). This conclusion was reached based on the study of the gaps arising via scrambling and the observation that they are subject to different licensing conditions than English parasitic gaps. What was not discussed is that they differ from German genuine parasitic gaps just as much (Assmann 2010), and that these genuine parasitic gaps, though eliciting a wide range of acceptability ratings, receive mediocre to high ratings from a non-negligible proportion of participants. As such, German parasitic gaps *do* exist, and as discussed in section 3.1.2, so do gaps that resemble parasitic gaps. The question then is whether these two gap types require distinct syntactic derivations.

The assumption that the gaps arising via scrambling are derived by a different mechanism than typical parasitic gaps seems particularly likely given that the parasitic gaps tested in the experiment are not ruled out if the adjunct clause is extraposed. For the gaps arising via scrambling, on the other hand, extraposing the adjunct clause is unanimously reported to yield an unacceptable outcome. This implies that the derivation of pseudoparasitic gaps must involve operations that are ruled out by extraposition. At the same time, the derivation of parasitic gaps, being a phenomenon of core syntax, is entirely independent of extraposition. The proposal by Fanselow (2001) for the gaps licensed by scrambling straightforwardly targets the restriction on extraposition (22). Since the account treats the phenomenon as PF-deletion, it makes reference to the linear distance between the antecedent and the deleted copy based on the number and nature of intervening heads (recall section 3.1.1.1).

- (22) a. dass sie Hans [ohne Hans zu umarmen] begrüßt hat
 that she Hans without Hans to.hug greeted has
 ‘that she greeted Hans without hugging (him)’
 b. *dass sie Hans begrüßt hat [ohne Hans zu umarmen]
 that she Hans greeted has without Hans to.hug
 Intended: ‘that she greeted Hans without hugging (him)’
 (cf. Fanselow 2001, p. 412)

The non-coordinating head *begrüßt* ‘greeted’ intervenes between the two instances of *Hans* in (22b), thus the occurrence in the adjunct clause cannot be deleted. The reportedly unacceptable construction in (22b) becomes (marginally) acceptable if we swap scrambling for wh-movement, as in (23):

- (23) Wen_i hat sie begrüßt ____i [ohne pg_i zu umarmen]?
 whom has she greeted without to hug
 ‘Whom did she greet without hugging?’ (cf. Kathol 2001, p. 320)

Now, in order for PF-deletion to be applicable, the adjunct clause must be underlyingly coordinated with the matrix clause. The acceptability of a gap appears to be directly tied to the type of adjunct clause hosting it. Though this is widely discussed and acknowledged, there appears to be no proper explanation as to why this is the case. Notice that *um* ‘(in order) to’ does not allow for pseudoparasitic gaps. However, the construction seems to improve if there is a wh-dependency in the sentence:¹⁰

- (24) a. *dass sie Hans [um zu begrüßen] umarmt hat
 that she Hans in.order.to to greet hugged has
 b. ?Wen_i hat sie [um pg_i zu begrüßen] umarmt?
 whom has she in.order.to to greet hugged
 ‘Whom did she hug in order to greet?’
 c. ?Wen_i hat sie umarmt [um pg_i zu begrüßen]?
 whom has she hugged in.order.to to greet
 ‘Whom did she hug in order to greet?’

This follows from the forward deletion account if *um* does not behave like a coordinate head, since this is merely a prerequisite for pseudoparasitic but not proper parasitic gaps. Indeed, we find that unlike *anstatt* and *ohne*, it cannot combine with a complementizer:

- (25) Es regnet ohne/anstatt/*um dass es schneit.
 it rains without/instead that it snows
 ‘It is raining without/instead of snowing.’

¹⁰These judgments are my own and need further verification.

The derivation of pseudoparasitic gaps thus seems to make reference to a certain property of the preposition that the derivation of parasitic gaps does not make reference to. Following the assessment by Fanselow (2001), the relevant property could be whether the preposition can act as a coordinate head. As such, the gaps in scrambled constructions are most likely pseudoparasitic rather than proper parasitic gaps, possibly featuring some type of coordinate structure in their underlying syntax. The evidence in favor of distinguishing pseudoparasitic and parasitic gaps in German requires us to reassess approaches that subsume pseudoparasitic gaps under genuine parasitic gaps (Assmann 2012; Felix 1985; Kathol 2001). I conclude that there is no convincing evidence against the assumption that there exist parasitic gaps in German, but that the empirical data suggests the parallel existence of pseudoparasitic gaps.

3.1.5 Conclusion

The experiment presented herein demonstrates that parasitic gaps in German do exist and that they are marginally acceptable overall, while eliciting a wide range of judgments based on the speaker. While some participants rejected them, the majority found them mediocre, and some even judged them to be acceptable. The experiment further shows that the acceptability of the parasitic gap interacts with the placement of the adjunct clause in a significant way—the German counterparts of the most canonical English examples where the adjunct clause is extraposed are less acceptable if they contain a parasitic gap. On the other hand, if the adjunct clause does not contain a parasitic gap but a pronoun referring to the filler instead, participants prefer extraposition over incorporation. The findings allow us to draw a distinction between parasitic and pseudoparasitic gaps in German, as extraposition is unanimously reported to yield unacceptability for the latter while this is not the case for the former based on the experiment. The crossover interaction between GAP and PLACEMENT raises further interesting questions about the processing of incorporated vs. extraposed adjunct clauses with and without parasitic gaps. For the purposes of this thesis, the study of reconstruction in German parasitic gaps is not pursued any further. The reason is, on the one hand, the overall marginal acceptability, and, more importantly, that this marginal acceptability would be further exaggerated by multiple referents and an ambiguous pronoun. Future research should aim to clarify factors determining this variability among speakers. Thus, the following section of this chapter delves into the study of principle C reconstruction in German ATB movement, setting the study of German parasitic gaps aside.

3.2 Principle C reconstruction in German ATB movement

In chapter 1, we have established the theoretical foundations underlying the current investigation. First, we saw that approaches to multi-gap dependencies differ with respect to how they model the one-to-many relation between the filler and the gap. In particular, they differ in their assumptions about which position(s) the moving filler originates from, and if the relation between the filler and the gaps is the same in parasitic gap constructions as it is in ATB movement. Second, we saw that principle C is reported to reconstruct ‘all the way down’, therefore revealing the base position of an extracted element, crucially distinguishing it from intermediate positions. Third, we saw that despite the clarity of the predictions derived from most approaches and the clarity of the predictions principle C reconstruction makes, the empirical data reported so far are as conflicting as they could be. Chapter 2 set out to study how robust principle C reconstruction itself is in simple *wh*-dependencies given the existence, but strong ambivalence of experimental data on the matter. I concluded that principle C reconstruction yields a significant effect, but that this effect does not lead to the unacceptability of the coreferent reading between the *R*-expression contained in the filler and the pronoun *c*-commanding its base position. Instead, we saw that the presence of the effect is manipulable based on the experimental task and setup, that only some participants show it at all, and that its magnitude likewise varies from speaker to speaker. This puts previous empirical reports about principle C reconstruction in multi-gap dependencies into a new context: It is quite likely that neither of them is technically inaccurate, instead capturing the natural inter-speaker variability of coreference resolution we saw in chapter 2. Principle C reconstruction could be a relevant factor, but likely not a universal factor that retains its significance across speakers and situations, thus undermining the results of the diagnostic. This section sets out to explore how likely this assessment is by conducting multiple experiments with varying setups and factors found to impact coreference in chapter 2.

Having concluded that parasitic gaps in German are too marginal a phenomenon to further complicate via reconstruction, this section deals with principle C reconstruction in German ATB dependencies. The aim of this section is to assess what coreference pattern the data exhibit, and how much we can infer from these data with respect to the pattern of principle C reconstruction and thus the path of movement in ATB dependencies. This question will be assessed in three experiments varying the experimental setup and task slightly, just like the experiments on principle C reconstruction in chapter 2 did. Special attention will

be paid to the response patterns of participants to assess the effect of principle C reconstruction on an individual level.

As we have seen in section 1.3.2.7 of chapter 1, the evidence is rather messy and primarily based on individual judgments, with some reporting asymmetry between the two gaps, claiming that reconstruction only targets the initial one (Citko 2005). Others report that principle C reconstruction is symmetric (Nissenbaum 2000). Experimental evidence for English has challenged the claim that reconstruction for principle C is asymmetric (Bruening and Al Khalaf 2017), while failing to provide interpretable evidence due to an inaccurate experimental design. This empirical conflict highlights the need to test the predictions of existing theories systematically to better understand and evaluate the evidence. This section will show that principle C reconstruction in German ATB movement appears to display a weak asymmetry, arguing that this is not due to the syntactic structure of ATB movement, but the lack of robustness found in principle C reconstruction as established in chapter 2. I will conclude that principle C reconstruction is not a reliable test for the underlying structure of ATB movement in German and that the differences between binding phenomena in ATB movement most likely arise from their own nature rather than a complex property of ATB movement itself (cf. Citko 2005; Bruening and Al Khalaf 2017).

3.2.1 Testing principle C reconstruction in ATB movement experimentally

All experiments are based on the setup created by Salzmann, Wierzba, and Georgi (2023). Experiment 4 applies the original setup of their study to ATB dependencies, while experiments 5 and 6 track the modifications of experiments 1 and 2 reported in chapter 2. Recall that in the original study by Salzmann, Wierzba, and Georgi (2023), each item was accompanied by two forced choice tasks inquiring about the coreference possibility with the R-expression in the displaced NP (called embedded referent in the following) or the R-expression in the matrix clause (called the matrix referent), respectively. Participants were explicitly asked about the possibility of coreference and were instructed to consider each reading carefully. The setup including a sample item and the accompanying tasks is illustrated by (26) and (27).

(26) *Object condition*

Lisa erzählt, welche Geschichte über Hannah sie ____
 Lisa.NOM tells which story.ACC about Hannah.ACC she.NOM
 ärgerlich fand.
 upsetting found

‘Lisa tells (us) which story about Hannah she found upsetting.’

Can the sentence be understood such that...

- a. ...Lisa found a story upsetting? yes/no (matrix referent)
- b. ...Hannah found a story upsetting? yes/no (embedded referent)

(27) *Subject condition*

Lisa erzählt, welche Geschichte über Hannah ____ sie
 Lisa.NOM tells which story.NOM about Hannah.ACC she.ACC
 verärgert hat.
 upset has

‘Lisa tells (us) which story about Hannah upset her.’

Can the sentence be understood such that...

- a. ...a story upset Lisa? yes/no (matrix referent)
- b. ...a story upset Hannah? yes/no (embedded referent)

In (26), the extracted constituent *welche Geschichte über Hannah* ‘which story about Hannah’ is the object of the verb in the embedded clause, while in (27), it is its subject. Under subject extraction, the R-expression c-commands the object pronoun, therefore not giving rise to a principle C reconstruction effect, whereas there should be a reconstruction effect under object extraction due to the subject pronoun c-commanding the R-expression. Studies comparing moved vs. unmoved conditions, i.e. surface vs. underlying principle C violations, predict a null effect under reconstruction, which is difficult to interpret. The confound of linear order, i.e. anaphoric vs. cataphoric reference, makes it unlikely that an underlying principle C violation will elicit the same response as a surface violation, thus risking that the effect the test is intended to measure is overshadowed. Contrasting subject vs. object extraction, however, allows for more straightforward predictions where a difference between the two conditions would indicate successful reconstruction. While the surface order of constituents is the same, the principle C violation only occurs in object conditions and only under reconstruction. Subject extraction should show no such effects and always allow for coreference. The experiment only tested displaced nouns with PP arguments, not adjuncts. All three experiments reported in this section have the same 2×2 Latin Square design varying the factors GRAMFUNC (grammatical function of the extracted phrase, object or subject), and

POSITION (of the conjunct containing the matching pronoun, initial or non-initial).

3.2.2 Experiment 4: Two forced choice tasks

3.2.2.1 Method

The first experiment is the application of Salzmänn, Wierzbä, and Georgi's method to ATB movement without modifications (see also section 2.1.2 of chapter 2). In each trial, participants saw a sentence with three referents and a pronoun. The pronoun matched the features of the matrix and the embedded referent, yielding a principle C violation under reconstruction of the latter. Participants were given two forced choice tasks per item, asking about the possibility of coreference with either of the matching referents. Participants saw each item in only one out of the four conditions. Metalinguistic terms were avoided by repeating the sentence with the intended reading (28). Participants were instructed to read the sentences carefully, but to make decisions based on their first impression.

- (28) Can the sentence be understood such that...
- | | |
|-------------------------|----------------------------|
| Marie overheard a joke? | yes/no (matrix referent) |
| Ute overheard a joke? | yes/no (embedded referent) |

3.2.2.2 Participants

A total of 300 participants with mean age 31.2 (standard deviation of 9.92) were recruited over Prolific. Participants were native speakers of German located in Germany with a monolingual upbringing and no language related disorders. The data from 33 participants were excluded from the analysis based on failed attention checks. All participants who successfully completed the study received monetary compensation, regardless of whether their data were used in the analysis or not.

3.2.2.3 Materials

Sentences involved a referent in the matrix clause (*I asked Helen...*), a displaced wh-phrase with a PP argument containing another referent (*...which joke about Ute...*), followed by the coordinate structure with two conjuncts, one of them containing a pronoun matching both referents (*...she heard...*), and the other a mismatched referent (*...and Mats made up.*). Each sentence contained a context introducing all referents to counteract factors influencing the prominence of either in the target sentence alone. The experiment manipulated two factors with two levels each: GRAMFUNC, denoting the grammatical function of the displaced constituent with levels *object* and *subject*, and POSITION denoting which conjunct the matching pronoun is in with levels *initial* vs. *non-initial*. Given the 2×2 design,

there were four conditions in total. Participants saw 12 target items, each appearing in only one of the four conditions. Additionally, participants were exposed to 48 distractor items. These included sentences testing relative clause attachment height; the effect of (linear) distance on coreference with unambiguous pronouns in ATB constructions; and principle A reconstruction in ATB constructions. An example of a target item is given in (29) and (30).

- (29) Während der Pause haben Helen und Ute sich über die Witze
 during the break have Helen and Ute self about the jokes
 unterhalten, die unter anderem von Mats in der Schule verbreitet
 conversed that among others by Mats in the school spread
 werden.
 are
 ‘During the break, Helen and Ute had a chat about the jokes that are
 disseminated at their school by Mats and others.’

Ich habe Helen gefragt, welchen Witz über
 I have-1SG Helen.ACC ask.PST which-ACC joke.NOM/ACC about
 Ute...
 Ute.ACC

‘I asked Helen which joke about Ute ...

- a. *object, initial*

...sie — mitgehört und Mats — erfunden hat.
 she.NOM — overhear.PST and Mats.NOM — invent.PST has
 ...she overheard and Mats made up.’

- b. *object, non-initial*

Mats — erfunden und sie — mitgehört hat.
 Mats.NOM — invent.PST and she.NOM — overhear.PST has
 ...Mats made up and she overheard.’

- (30) Helen, Ute und Mats haben ihren Kommilitonen in der Pause
 Helen Ute and Mats have their fellow.student in the break
 belauscht und dabei die neuesten Witze gehört, die momentan
 eavesdrop.on and thereby the latest jokes heard that currently
 die Runde machen.
 the round make
 ‘Helen, Ute and Mats have eavesdropped on their fellow student(s) during
 the break and have thereby learned about the latest jokes currently making
 the rounds.’

Ich habe Helen gefragt, welcher Witz über Ute...
 I have-1SG Helen.ACC ask.PST which-NOM joke about Ute.ACC

‘I asked Helen which joke about Ute ...

a. *subject, initial*

___ sie irritiert und ___ Mats amüsiert hat.
she.ACC irritate.PST and Mats.ACC amuse.PST has

...irritated her and amused Mats.’

b. *subject, non-initial*

___ Mats amüsiert und ___ sie irritiert hat.
Mats.ACC amuse.PST and she.ACC irritated has

...amused Mats and irritated her.’

The logic of the design is as follows: Reconstruction of the extracted constituent to the gap c-commanded by the pronoun should rule out coreference between the pronoun and the referent in the reconstructing phrase, i.e. *Ute* in the examples above. The research question at hand is whether this effect occurs in both conjuncts, as predicted by symmetric extraction approaches, or only one of them, as predicted by asymmetric approaches. Each sentence was accompanied by a neutral context introducing the referents to avoid making either of them more prominent than the other. Making the referent in the matrix clause the direct object of *fragen* ‘to ask’ was a conscious choice to avoid prominence effects associated with subjects and topics (Cowles, Walenski, and Kluender 2007; Kaiser 2011), trying to minimize the factors distracting from the embedded referent.

3.2.2.4 Procedure

The experiment was set up using the platform L-Rex (Starschenko and Wierzba 2024). The sentences were displayed simultaneously with the context, the latter in italics. The two questions were shown below the sentence with the answer options *yes*, indicating coreference, and *no*, indicating disjoint reference. The first block included two training items in non-randomized order, showing sentences with two referents matching the pronoun, in one case allowing for coreference with both and in one case allowing for coreference with only one of them. The second block contained target items and distractors in pseudo-randomized order, such that two items from the same set of materials were never shown consecutively.

3.2.2.5 Predictions

In the following, the term *coreference rate* indicates the proportion of *yes* responses to the forced choice task asking about coreference with the embedded referent. High coreference rates in subject conditions are predicted if the impossibility of coreference is determined by asymmetrical c-command between a pronoun and a

matching R-expression. Both the surface position as well as the base position of the extracted subject c-command the pronoun, yielding no principle C violation on the surface nor in the underlying configuration. This should be qualified by significant main effect of `GRAMFUNC`.

There are three distinct reconstruction patterns predicted by the theories of ATB movement with respect to object conditions. Symmetric approaches, where the displaced element is base generated in and extracted from both conjuncts, predict no difference between the gaps. The constituent should reconstruct symmetrically and a violation should occur symmetrically as well. No additional significant effects are expected if this is the case, since the position of the pronoun should be irrelevant. A significant main effect of `GRAMFUNC` alone would indicate that there is a difference between subject and object extraction (effect of principle C reconstruction) but not between gaps (symmetry of effect across gaps).

In asymmetric approaches where the shared element only moves from the initial gap, the predictions are determined by how the non-initial gap is created. If there is a full copy of the filler in the non-initial gap site, a principle C violation is still expected even without reconstruction (Munn 1993, et. seq). That is, such an asymmetrical extraction approach would nevertheless predict a symmetrical pattern for a principle C reconstruction effect. In the forward ellipsis approach (Salzmann 2012), the aforementioned asymmetric reconstruction pattern is predicted due to the possibility of vehicle change in the non-initial conjunct, avoiding a principle C violation (Fiengo and May 1994). In that case, a significant interaction between `GRAMFUNC` and `POSITION` is expected, with higher coreference rates in the condition *object, non-initial* than *object, initial*.

The backward ellipsis approach as well as the sideward movement approach predict that reconstruction should only target the non-initial gap. Although the extracted element moves through the initial gap in sideward movement, based on the observation that principle C is only evaluated in base positions but not intermediate ones, the prediction is asymmetric. Coreference should be unconstrained in the condition *object, initial* but not in *object, non-initial*, yielding the exact opposite of the pattern predicted by forward ellipsis. Again, a significant interaction between `GRAMFUNC` and `POSITION` is expected.

It is further to be noted that non-syntactic effects will likely play a role. We saw in chapter 2 that principle C reconstruction is not a particularly strong nor universal predictor of coreference. There are two conceivable ways in which non-syntactic biases could impact coreference under reconstruction. First, the relative surface position of the referent and the pronoun could influence coreference, regardless of reconstruction. Authors diverge with respect to their hypotheses and findings. While for English, increasing linear and structural distance between the referent

and the pronoun is argued to facilitate coreference equally (Adger et al. 2017; Bruening and Al Khalaf 2019), the opposite has been claimed for German based on recency effects (Salzmann, Wierzba, and Georgi 2023). If it is indeed recency that plays a role in German, coreference should be facilitated in initial conditions. Second, it needs to be considered that in multi-gap dependencies, the robustness of reconstruction could be affected by the distance between the filler and the gap. That is, the variable distance between the filler and each gap may yield stronger reconstruction effects in closer gaps than distant ones. This would predict a lower coreference rate in the condition *object, initial* than *object, non-initial*. If the relative surface position plays a role, there should be a significant main effect of POSITION. The proportion of responses will help determine whether the effect of proximity is positive or negative.

In sum, a significant interaction between GRAMFUNC and POSITION would be indicative of a reconstruction effect targeting the conjuncts asymmetrically. A significant main effect of GRAMFUNC would indicate reconstruction to both gaps. A significant main effect of POSITION would indicate the involvement of linear factors, i.e. that the position of the pronoun determines the coreference pattern regardless of the pronoun c-commanding the base position of the extractee.

3.2.2.6 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using generalized linear mixed effects models for binary data from the package ‘lme4’ (D. Bates et al. 2015) with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’ (D. Bates et al. 2015). Modeling was carried out only for the task about the embedded referent due to the lack of theoretical value in hypothesizing about coreference with the matrix referent. A conservative α -level of 0.05 was defined. The model in (31) includes fixed effects for both factors, GRAMFUNC and POSITION, the interaction of the two, and a random effects structure with varying intercepts and slopes for both participants and items (Barr et al. 2013).

$$(31) \quad \text{rating} \sim \text{gramfunc} * \text{position} + \\ (1 + \text{gramfunc} + \text{position} \mid \text{item}) + (1 + \text{gramfunc} + \text{position} \mid \text{participant})$$

The model in (31) is the final model after simplifications were applied to the maximal random effects structure due to a perfect negative correlation of random effects for GRAMFUNC by item. Participants’ attention and whether they understood the task was assessed through inspecting coreference rates with the matrix referent, i.e. one of the questions in target trials, as a sanity check. Failure to indicate coreference with the matrix referent in 25% of cases or more led to exclusion. This was taken to indicate that participants either responded based on preferences

rather than possibilities, or that they did not complete the task responsibly. After applying this exclusion criterion, data from $n = 277$ participants were analyzed.

3.2.2.7 Results

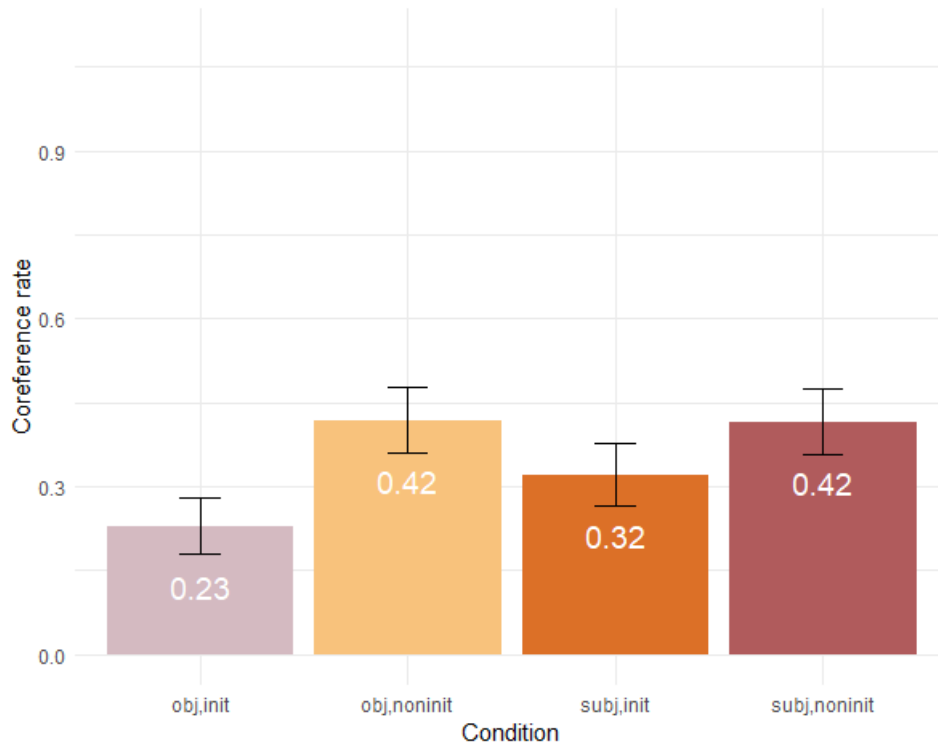


Figure 3.6: Coreference rates for the embedded referent across conditions.

The remaining participants correctly indicated that coreference between the pronoun and the matrix referent was possible in 87% to 89% of cases depending on the condition. This sanity check indicates that participants have understood the task, although responses are not at ceiling. Coreference rates with the embedded referent are shown in Figure 3.6. In particular, the hypothesis that coreference should always be allowed in subject conditions is not borne out, similar to what we observed in simple wh-dependencies. Not only is the proportion below chance level in subject conditions, but crucially, coreference rates differ by 10% based on which conjunct the pronoun is in. Turning to object conditions, responses deviate from all possible predictions. Again, there is a clear difference between the condition testing reconstruction to the initial vs. the non-initial gap, but the contrasts are weaker than expected based on the syntactic predictions. Participants found that coreference was possible more frequently when the pronoun was in the non-initial conjunct than in the initial one, but the proportions do not indicate a clear tendency towards coreference being ruled in or out by a syntactic constraint. The significant interaction between `GRAMFUNC` and `POSITION` reveals that the contrast between subject and object conditions only holds if the pronoun is in the

initial conjunct. In the non-initial conjunct, this contrast vanishes completely.

GLMM	Estimate (SE)
(Intercept)	1.00*** (0.15)
gramfunc	0.73*** (0.19)
position	-0.65*** (0.17)
gramfunc:position	-0.77*** (0.19)
AIC	3157.94
Num. obs.	3048
Num. groups: participant	254
Num. groups: item	12
Var: participant (Intercept)	3.57
Var: participant gramfunc	0.06
Var: participant position	1.05
Var: item (Intercept)	0.00
Var: item gramfunc	0.21
Var: item position	0.06

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3.4: Estimates of the generalized linear mixed effects model given in (31).

The estimates of the model in (31) are reported in Table 3.4. Data analysis revealed a significant main effect of GRAMFUNC and POSITION as well as a significant interaction between the two. That is, while the pronoun *c*-commanding the R-expression's base position is estimated to have an effect on coreference in both conjuncts (main effect of GRAMFUNC), the effect is particularly pronounced in the initial conjunct (interaction). The relevance of distance, whether it is structural or linear, is further shown by the significant main effect of POSITION, meaning that the model estimates that there is generally a decrease of coreference rate in the initial conjunct compared to the non-initial conjunct regardless of the factor GRAMFUNC. Notice that the coreference rates are very low in the subject condition where no syntactic violation is in sight, indicating that factors beyond principle C reconstruction give rise to the pattern. Because the model had problems estimating random slopes and intercepts by item due to insufficient data, two more models were fitted. One of them attempted to address the issue by dropping the correlation between the random intercepts and slopes by item, the other one omitted the random effects structure by item altogether. Nevertheless, a likelihood ratio test revealed that the model reported in (31) has the lowest AIC out of the three and a significant p-value. The reported estimates are on the log-scale and factors are treatment contrast coded (levels *initial* and *subject* treated as the base, coded 0).

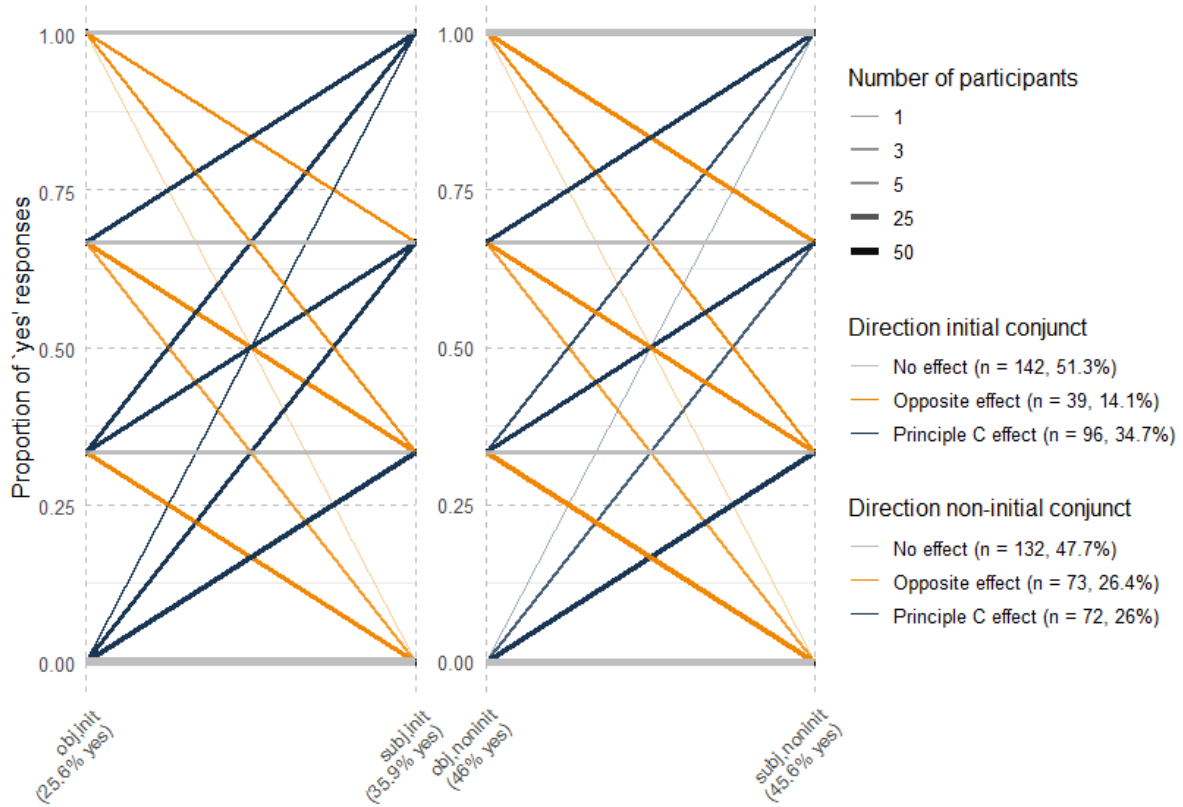


Figure 3.7: Individuals' ($n = 277$) overall proportions of responses indicating coreference with the embedded referent in experiment 4 by condition. Y-axis indicates proportion (0-1 by 0.33 based on 3 observations per participant per condition), lines connect individual participants' proportions in the object and subject conditions of the initial (l) and non-initial conjunct (r), respectively. Color indicates the direction of the effect (blue: subject > object, orange: subject < object, gray: subject = object), thickness and opacity indicate the number of participants. X-axis indicate overall percentage of yes responses in condition.

Figure 3.7 illustrates the various slopes by participant for the effect of GRAMFUNC, i.e. the proportion of responses indicating coreference between the embedded referent and the pronoun in object vs. subject conditions. That is, the steepness of the slopes indicates the magnitude of the principle C reconstruction effect. The slopes for the initial gap are shown on the left, those for the non-initial conditions on the right. In the initial conditions, 51.3% of participants show no effect at all, indicating that coreference was equally possible regardless of the presence of a principle C violation. For the majority of participants this manifested in 0% coreference rate in both conditions. Only 34.7% of participants show the predicted pattern, i.e. higher coreference rates in the subject condition than the object condition, and only very few of them display an absolute pattern, where coreference is not chosen in the object condition at all and chosen (almost) always in the subject condition. Remarkably, 14.1% of participants display the opposite pattern, i.e. coreference is chosen more frequently in the object than the subject

condition.

On the surface, as illustrated by Figure 3.6, it appears that there is no difference between the coreference rates in the non-initial object and subject condition—both are at 42%. However, inspecting individual participants' patterns reveals that this is not caused by a uniform distribution of flat slopes, but rather by rampant inter-individual variability with an equal number of participants showing the effect in the predicted direction and in the opposite direction, as seen in Figure 3.7. In non-initial compared to initial conditions, there is a remarkable increase in participants showing the opposite effect (26.4%), especially to the detriment of the principle C effect group's size (34.7% in initial but only 26% in non-initial condition). Among those participants showing no effect, higher coreference rates are more frequent in the non-initial than the initial conditions, reflecting the significant effect of POSITION manifesting in a positive effect on coreference in *non-initial* conditions. As such, it is clear that although a portion of speakers seem to respond in line with the predictions of a principle C reconstruction effect, the majority of them does not.

3.2.2.8 Discussion

The significant interaction of GRAMFUNC and POSITION taken together with the observed coreference rates indeed suggests an asymmetry between the initial and non-initial gaps of ATB movement. Coreference was chosen more frequently in the non-initial condition, showing at least the tendency reported by Citko (2005) for English. However, once we depart from the overall pattern and inspect individual participants' responses, it becomes clear that the significant interaction does not arise from a uniform, weak tendency across participants, but rather from vast inter-individual variability both with respect to the direction as well as the size of the effect. As such, only 34.7% of speakers exhibit a response pattern that matches the predictions of a principle C reconstruction effect in the initial gap, which is largely at odds with the assumption that principle C should be a universal syntactic constraint across dependency types and complexities.

The main effect of POSITION indicates that the surface configuration plays a role, too, which is in line with the observation that even in subject conditions where no reconstruction takes place, there is a difference in coreference rate between the initial and the non-initial condition. In line with findings by Adger et al. (2017) and Bruening and Al Khalaf (2019), increasing the distance between the referent and the pronoun boosts coreference. The pattern is thus not a recency effect whereby the most recently mentioned referent relative to the pronoun is favored, since this wrongly predicts that coreference rates should increase in initial conditions. Rather, we observe that increasing the distance has a positive effect, potentially

because it allows the referent to ‘decay’ in memory before a pronoun refers back to it. The interference of a mismatched referent in non-initial conditions may also make coreference more felicitous. Thus, the asymmetrical pattern reported for principle C reconstruction in ATB movement is most likely the result of an unstable reconstruction pattern paired with a positive effect of distance, rather than a reliable indicator of the presence or absence of the displaced constituent in the underlying representation.

It remains to be explored whether the presence of the matrix referent and the complexity of the experimental task have a depressing effect on coreference rates, particularly in the subject condition. To address this, the experiment will be repeated with slight changes. In experiment 5, the context, matrix referent and phrasing of the task will remain the same. However, participants will only have to respond to one task in each trial. In target items, participants will always be asked about coreference possibilities with the embedded referent. Distractors will be used for counterbalancing, i.e. the respective trials will only inquire about the possibility of coreference with the matrix subject. In experiment 6, the matrix referent, embedding sentence and context will all be omitted. Participants will be given a general context applicable to all sentences and asked whether the pronoun refers to the embedded referent or ‘someone else’ (cf. Stockwell, Meltzer-Asscher, and Sportiche 2021, 2022). Both of these experiments will include the experimental items tested by Salzmann, Wierzba, and Georgi (2023) as pseudodistractors (i.e. the target items from experiments 1 and 2 in chapter 2), allowing for a comparison of participants’ behavior across dependencies with varying complexity. Lexical variability across conditions will also be eliminated by using psych verbs, contrasting the pattern ‘X delighted Y’ with ‘Y found X delightful’, X being the displaced constituent containing the embedded referent and Y the pronoun, instead of using distinct verbs (Salzmann, Wierzba, and Georgi 2023). The prediction is that removing the matrix referent or merely the question about it from the experimental items should boost coreference at least in subject conditions. If the boost is observed across conditions again, this will further support the view that the pattern is not a result of a c-command-based principle C violation.

In the next section, the second experiment on principle C reconstruction in German ATB movement is presented. The aim of the experiment is to test whether the effect observed in the first experiment could be amplified by simplifying the experimental design. The idea is that participants may have been distracted by the additional question about the matrix referent, which only served as a sanity check to ensure that they understood the task, but seems to have impacted the willingness of participants to consider the embedded R-expression as a referent. This is particularly evident from the below chance coreference rates found in

subject conditions where nothing is predicted to inhibit coreference. Furthermore, the experiment aims to eliminate the confound of using different verbs in object vs. subject conditions to make the items more uniform.

3.2.3 Experiment 5: One forced choice about possibility of embedded referent

3.2.3.1 Method

For this experiment, the complexity of the experimental task was reduced to avoid distracting participants from considering the embedded referent, suspecting that being presented with two questions per task made the matrix referent stand out too much against the embedded referent. As such, participants only saw one instead of two questions per trial, assessing whether the question could be understood such that the pronoun referred to the embedded referent. To ensure that people did not exclusively see trials inquiring about the embedded referent, the experiment adjusted the distractors and pseudodistractors. Distractors featured multiple referents and always inquired about the possibility of coreference with the matrix referent, while for pseudodistractors, the referent in question was encoded as an experimental factor, asking about the embedded referent in half the trials and about the matrix referent in the other half. As such, participants were asked about coreference with the embedded referent in 40 trials (24 targets + 16 pseudodistractors) and about coreference with the matrix referent in 28 trials (the remaining 16 pseudodistractors + 12 distractors). The pseudodistractor trials inquiring about coreference with the matrix referent served as attention checks, with failure to indicate coreference with the matrix referent in more than 20% of cases leading to exclusion. This threshold was also implemented to assess whether participants understood the task and considered coreference possibilities rather than preferences.

3.2.3.2 Participants

175 participants with mean age 32.08 (SD = 10.26) were recruited over Prolific. Participants were native speakers of German located in Germany with a monolingual upbringing and no language related disorders. The data from 25 participants were excluded from the analysis based on failed attention checks. All participants who successfully completed the study received monetary compensation, regardless of whether their data were used in the analysis or not.

3.2.3.3 Materials

The materials from experiment 4 of this chapter were used, with the correction that verbs across respective subject and object conditions now matched. Instead of using regular transitive verbs in the object conditions and psych verbs in the subject conditions, the psych verbs from the subject conditions were rephrased in the object conditions. As such, a subject condition such as *which X irritated Y* was re-modeled to *which X did Y find irritating* for the object condition. This study featured 24 targets, 32 pseudodistractors and 12 distractors.

3.2.3.4 Procedure

The procedure matched the procedure of experiment 4. Given the simplification of the task, participants were shown one question per trial instead of two regarding the interpretation of the sentence. Target items were interspersed with pseudodistractors and distractors in pseudo-randomized order.

3.2.3.5 Predictions

The theoretical predictions with respect to the coreference patterns under reconstruction in ATB movement are the same as in experiment 4, as the manipulated factors and logic of the experiment remain the same. Compared to experiment 4, however, the prediction is that participants should choose the embedded referent more readily in subject conditions when only being confronted with a single question assessing the possibility of coreference with the embedded referent (without additionally asking about the matrix referent, like in experiment 4). The hypothesis is that any contrast between object and subject conditions, i.e. the effect of `GRAMFUNC`, should be strengthened in this experiment: The omission of the matrix referent is expected to boost participants' willingness to accept the embedded referent in the absence of a principle C violation, that is, in subject conditions.

3.2.3.6 Data analysis

Statistical modeling matched the procedure of experiment 4. Data from 150 participants were analyzed. The model in (32) includes fixed effects for both factors, `GRAMFUNC` and `POSITION`, the interaction of the two, and the maximal random effects structure with varying intercepts and slopes for both participants and items for each fixed effect as well as the interaction (Barr et al. 2013).

$$(32) \quad \text{reference} \sim \text{gramfunc} * \text{position} + \\ (1 * \text{gramfunc} * \text{position} \mid \text{item}) + (1 + \text{gramfunc} + \text{position} \mid \text{participant})$$

Participants' attention and whether they understood the task was again checked by inspecting coreference rates with the matrix referent as in experiment 4. After applying this exclusion criterion, the data from $n = 150$ remaining participants were analyzed.

3.2.3.7 Results

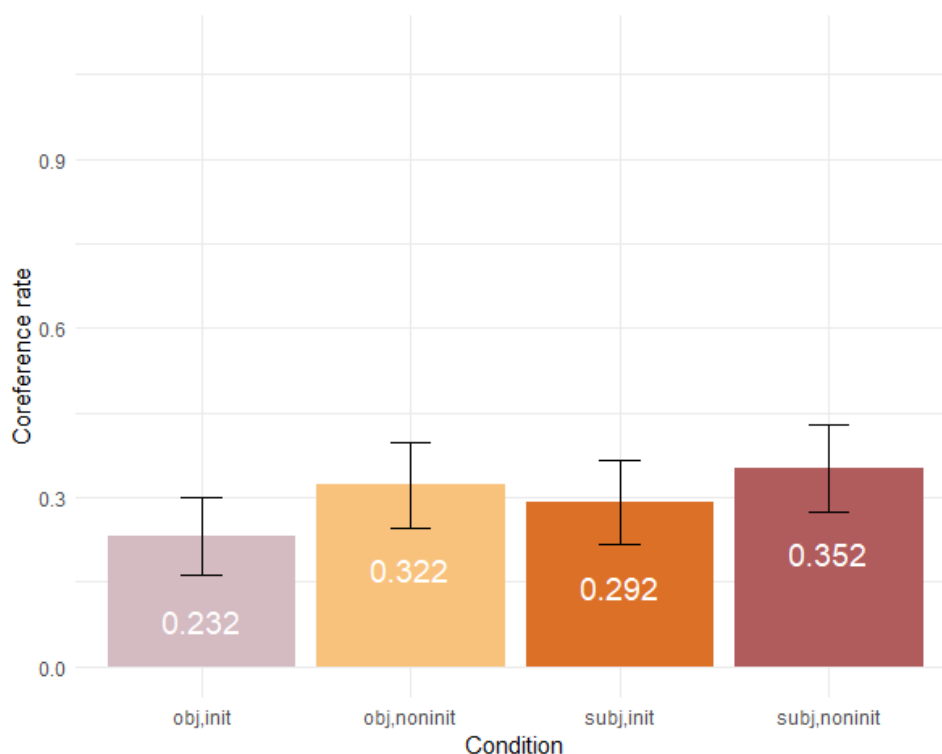


Figure 3.8: Coreference rates for the embedded referent across conditions.

The overall coreference rates with the embedded referent in the relevant experimental conditions are shown in Figure 3.8. Although the pattern remains the same in experiment 5 as it was in experiment 4, coreference rates in this experiment are lower across conditions. That is, the intention of boosting coreference rates by simplifying the task backfired. The modification instead depressed coreference rates even in conditions where coreference should always be possible, thereby reducing the strength of the manipulations rather than increasing it. While overall coreference rates in the first experiment differ by 19% between object and subject conditions testing the initial gap and 10% testing the non-initial gap, this contrast decreased to 9% and 6%, respectively.

The estimates of the model reported in (32) are reported in Table 3.5. Despite the contrast becoming weaker, significance testing revealed a significant main effect of `GRAMFUNC` and `POSITION`. The interaction of the two is not significant. This indicates that both principle C reconstruction as well as the linear distance between the pronoun and the R-expression remain to have an effect on the coreference

GLMM	Estimate (SE)
(Intercept)	1.47*** (0.32)
gramfunc	0.60*** (0.17)
position	-0.52** (0.18)
gramfunc:position	-0.38 (0.23)
AIC	3142.02
BIC	3290.55
Log Likelihood	-1547.01
Num. obs.	3600
Num. groups: participant	150
Num. groups: item	24
Var: participant (Intercept)	4.56
Var: participant gramfunc	0.28
Var: participant position	0.07
Var: participant gramfunc:position	0.34
Var: item (Intercept)	1.41
Var: item gramfunc	0.12
Var: item position	0.33
Var: item gramfunc:position	0.33

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3.5: Estimates of the generalized linear mixed effects model given in (32).

rate. However, this time, there is no evidence for principle C reconstruction to differ between conjuncts. Aside from, again, mismatching the predictions of all approaches due to the relatively low coreference rates in the subject conditions, the coreference pattern and the significance of the effects fit the predictions of symmetrical approaches best, as there is no significant interaction between the effect of principle C reconstruction and the position of the pronoun.

The results of experiment 5 are thus different from the results of experiment 4. While there is a significant interaction in experiment 4, there is none in experiment 5. We saw in experiment 4 that only a subset of participants showed a response pattern that was in line with the predictions of a principle C reconstruction effect at all, and we now turn to the discussion of this inter-individual variability in experiment 5.

Figure 3.9 illustrates inter-speaker variability in the direction and size of the effect of GRAMFUNC. In this experiment, too, we see a lot of variability among participants. Interestingly enough, the proportion of participants showing an effect in line with principle C reconstruction has increased overall, although the size of

3.2. Principle C reconstruction in German ATB movement

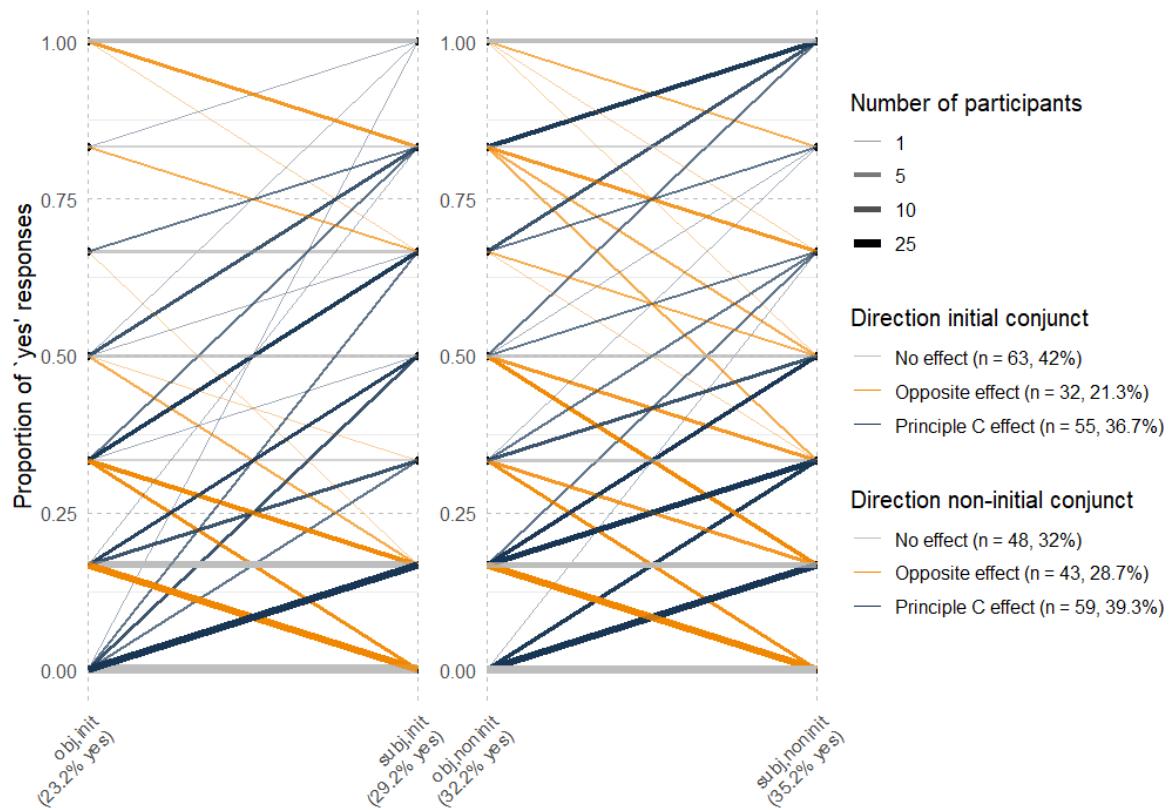


Figure 3.9: Individuals' (n = 150) overall proportions of responses indicating coreference with the embedded referent in experiment 5 by condition. Y-axis indicates proportion (0-1 by 0.167 based on 6 observations per participant per condition), lines connect individual participants' proportions in the object and subject condition of the initial (l) and non-initial conjunct (r), respectively. Color indicates the direction of the effect (blue: subject > object, orange: subject < object, gray: subject = object), thickness and opacity indicate the number of participants. X-axis indicate overall percentage of yes responses in condition.

the effect decreased. That is, there are more participants whose slopes increase in the predicted direction, but the increase is smaller. Note that doubling the number of observations per participant per condition compared to experiment 4 due to having tested 24 instead of 12 items has likely contributed to this.¹¹ In both the initial and non-initial conjunct, group sizes are a bit more balanced than in experiment 4. More participants responded in line with principle C reconstruction in the non-initial than the initial gap, though this effect is not significant, contrary to the contrast between the initial and non-initial gap in experiment 4.

¹¹Experiment 4 used only 12 items because participants should have been tested in two sessions to decrease the complexity of the experiment. However, after the first session, data collection was put on hold upon suspecting that something went wrong with the experimental setup due to the small and inconsistent effect size of GRAMFUNC. Experiments 5 and 6 of this chapter and experiments 1 and 2 of chapter 2 were attempts to identify the setup to detect the 'true' principle C effect, ultimately resulting in a dissertation with a heavy focus on methodology. These investigations revealed that the 'problem' was not fixable—coreference resolution is simply a much more complex topic than Binding Theory alone could ever do it justice.

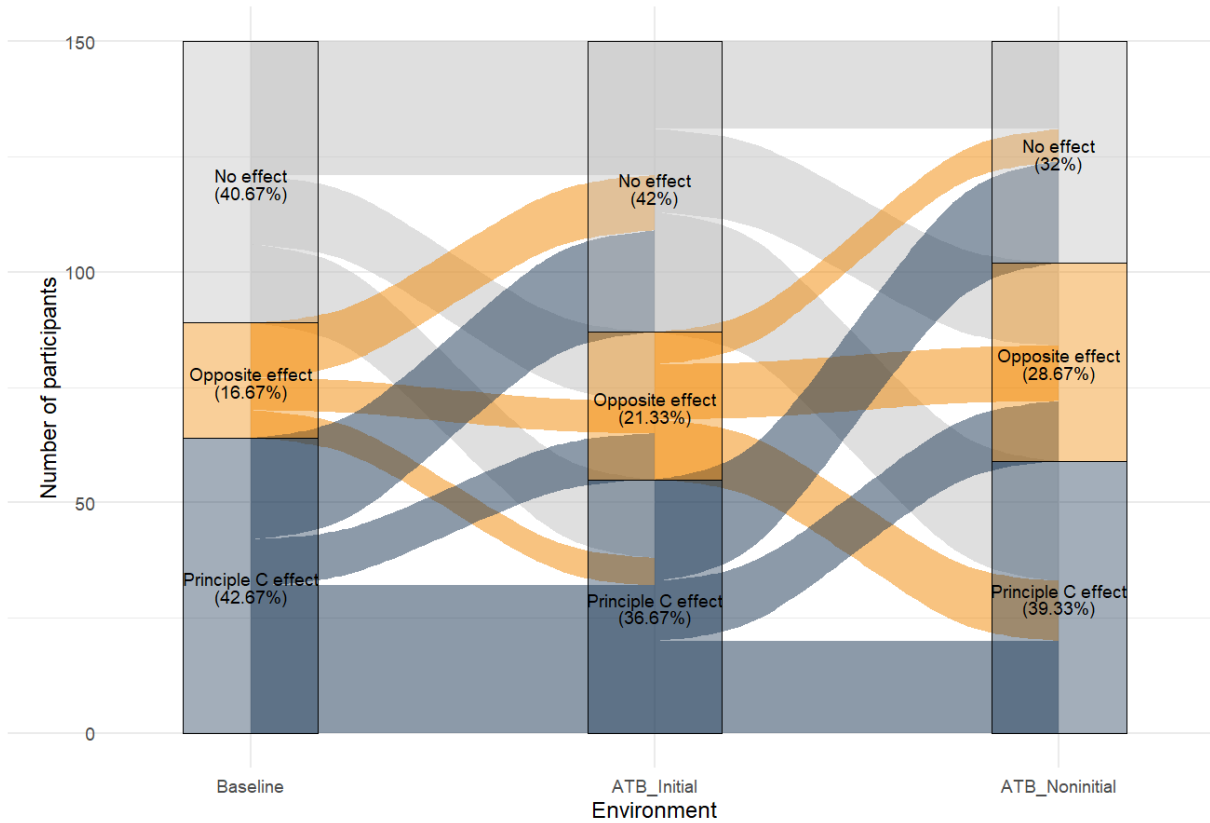


Figure 3.10: Proportion of participants corresponding to effect groups across different test environments (baseline = experiment 1 in chapter 2 on simple wh-dependencies, ATB_Initial = initial conjunct of ATB movement in current experiment, ATB_Noninitial = non-initial conjunct of ATB movement in current experiment. Flow indicates migration of individual participants from one effect group to the other.

Let us turn to the overlap between the participants who did show an effect in line with principle C in the initial vs. non-initial gap, and crucially, whether the same participants showed the effect in the experiment on simple wh-dependencies in experiment 1 of chapter 2. Figure 3.10 shows the migration of participants across effect groups. We see that participant groups are incongruent across environments, i.e. simple wh-dependencies and the different conjuncts of ATB movement. While there are participants who responded in line with a principle C reconstruction effect in simple wh-dependencies and either conjunct of ATB movement ($n = 19$ for the initial and $n = 13$ for the non-initial conjunct), there are few participants who are in line with the effect across all three conditions ($n = 13$). Crucially, there is variability among the remaining 105 participants, some of them not showing an effect in either environment, others showing it in only one of them. We can thus see that the relatively consistent group *sizes* are not indicative of group *identity* in this experiment. That is, principle C reconstruction informs less than 10% of participants' behavior across all syntactic environments that were tested. This

strongly suggests that participants responded based on more general pragmatic factors that may vary at item level. Turning to the modification of the task, it had the unwanted effect of further depressing coreference rates. Recall that the aim was to make the task easier for participants, hoping that any effect that was already present would become enhanced by inquiring about each referent individually rather than both of them in a single trial. However, it is quite likely that presenting an item with two possibly competing referents, but then only inquiring about one of them, especially if it is the less favored referent, put participants in the dissatisfactory position of not being able to voice their judgment about the most prominent reading. Participants' comments support this: Trials inquiring about the embedded referent caused irritation due to the overwhelming prominence of the matrix referent. Participants commented that they preferred the matrix referent over the embedded referent regardless of the experimental condition. See section 2.2.2.7 of chapter 2 for a detailed discussion of participants' comments (recall that the items discussed in experiment 1 of chapter 2 served as pseudodistractors in the current experiment).

3.2.3.8 Discussion

The significant main effects of `GRAMFUNC` and `POSITION` in the absence of a significant interaction suggest symmetry between the initial and non-initial gaps of ATB movement, contrary to the findings of experiment 4. Inspecting the individual patterns exhibited by participants we find that only 36.7% of participants responded in line with the predictions of principle C reconstruction in initial gap conditions, while 39.3% of them did in the non-initial gap conditions. Given that the subject conditions are sensitive to the manipulation of `POSITION` despite the absence of an effect of reconstruction, we can conclude that superficial factors play a role in determining the pattern rather than a fundamental, universal syntactic effect. This is further underpinned by the main effect of `POSITION` as well as the extent of inter-individual variability regarding the direction of the effect of `GRAMFUNC`.

Crucially, omitting one of the questions from each trial did not make the intended task more straightforward. Instead, it led to a mismatch between task and item structure, as there were multiple matching referents in the sentence despite only assessing coreference with one. Especially in the relevant trials inquiring about the embedded referent, i.e. the potential principle C violation, the prominence of the matrix referent and the impossibility to indicate their preference for it led to a decrease in overall coreference rates. The final experiment aims to address this issue by omitting the matrix referent altogether.

3.2.4 Experiment 6: One forced choice between embedded and unnamed referent

3.2.4.1 Method

The third experiment eliminated two crucial components of the item structure: There was no embedding sentence and thus no competing referent. These modifications were made upon suspicion that the matrix referent was too prominent, keeping participants from considering the embedded referent at all. Furthermore, the context sentence was omitted because there were not multiple matching referents in the sentence whose prominence had to be controlled. Participants saw interrogative sentences, i.e. modifications of the materials from experiment 5. The task was also altered so that participants had to respond to the prompt *What is this about?*, offering the coreferent with the embedded referent (*a joke that irritated Ute*) and an unnamed referent (*a joke that irritated someone else*) as response options. While the maximum response proportion was 100% per reading in experiments 4 and 5, in experiment 6, the proportion of responses indicating one reading depended on the proportion of responses indicating the other reading. If there is no preference, proportions of 50% for each reading are expected. Participants' attention was checked through distractors where coreference was ruled out due to gender mismatches, with failure to indicate disjoint reference in more than 3 out of 12 cases leading to exclusion.

3.2.4.2 Participants

64 participants with mean age 32.2 (SD = 9.37) were recruited over Prolific. Participants were native speakers of German located in Germany with a monolingual upbringing and no language related disorders. The data from 4 participants were excluded from the analysis based on failed attention checks. All participants who successfully completed the study received monetary compensation, regardless of whether their data were used in the analysis or not.

3.2.4.3 Materials

The materials from experiment 5 were modified such that the embedding sentence was omitted entirely as well as omitting the context. Participants thus saw interrogative sentences. Like in experiment 5, there were 24 targets, 32 pseudodistractors and 12 distractors.

- (33) Welchen Witz über Ute...
 which-ACC joke.NOM/ACC about Ute.ACC

‘Which joke about Ute ...

a. *object, initial*

hat sie ___ mitgehört und Mats ___ erfunden?
has she.NOM overheard and Mats.NOM invented
...did she overhear and Mats make up?’

b. *object, non-initial*

hat Mats ___ erfunden und sie ___ mitgehört?
has Mats.NOM invented and she.NOM overheard
...did Mats make up and she overhear?’

- (34) Welcher Witz über Ute...
 which-NOM joke about Ute.ACC
 ‘Which joke about Ute ...

a. *subject, initial*

___ hat sie irritiert und ___ Mats amüsiert?
 has she.ACC irritated and Mats.ACC amused
...irritated her and amused Mats?’

b. *subject, non-initial*

___ hat Mats amüsiert und ___ sie irritiert?
 has Mats.ACC amused and she.ACC irritated
...amused Mats and irritated her?’

3.2.4.4 Procedure

The procedure matched the procedure of experiments 4 and 5. Target items were interspersed with pseudodistractors and distractors in pseudo-randomized order.

3.2.4.5 Predictions

Again, the theoretical predictions remain the same as in experiments 4 and 5. Coreference rates in subject conditions are expected to increase compared to the first two experiments due to the omission of the matrix referent.

3.2.4.6 Data analysis

Statistical modeling matched the procedure of experiments 4 and 5. Data from 60 participants were analyzed. The maximal model was fitted initially, then simplified because a likelihood ratio test using the ‘anova’ function in R favored a simpler model. The final model is given in (35).

- (35) $\text{reference} \sim \text{gramfunc} * \text{position} +$
 $(1 + \text{gramfunc} + \text{position} \mid \text{item}) + (1 + \text{gramfunc} + \text{position} \mid \text{participant})$

3.2.4.7 Results

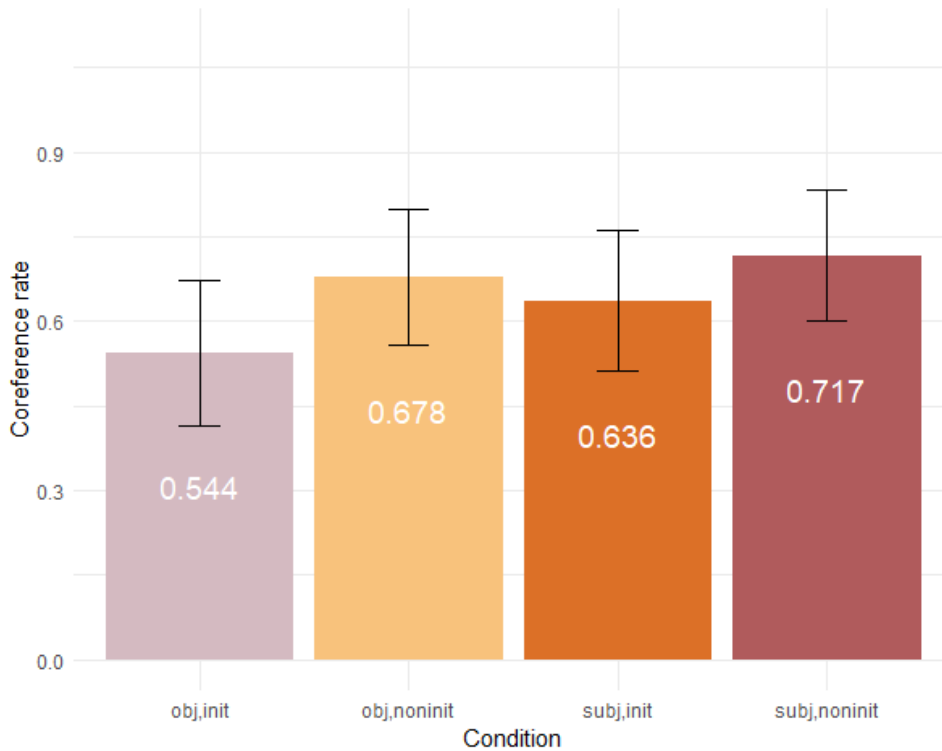


Figure 3.11: Coreference rates for the embedded referent across conditions.

The overall coreference rates scored in the entire experiment are shown in Figure 3.11. Compared to experiments 4 and 5, we see a drastic increase in coreference rates. Across conditions, the coreference rate is consistently above chance level. Again, the relative ranking of conditions is the same as we observed in experiments 4 and 5: Initial conditions elicited lower coreference rates than non-initial conditions, and the contrast between object and subject conditions in non-initial items is smaller than in initial items. However, the manipulations once again failed to increase the effect size, instead merely increasing coreference rates overall.

The estimates of the model defined in (35) are given in Table 3.6. The experiment revealed a significant effect of GRAMFUNC but no significant main effect of POSITION and no significant interaction between the two. That is, the significance of GRAMFUNC indicates that principle C reconstruction has an effect and that this effect is present in both gaps. Since the interaction of the two is not significant, the experiment does not provide evidence for a syntactic difference between the two conjuncts. Despite the overall difference in coreference rates between initial and non-initial conditions across levels of GRAMFUNC, significance

GLMM	Estimate (SE)
(Intercept)	-1.03** (0.39)
gramfunc	0.75*** (0.23)
position	-0.42 (0.26)
gramfunc:position	-0.35 (0.28)
AIC	1458.45
BIC	1542.81
Log Likelihood	-713.22
Num. obs.	1440
Num. groups: participant	60
Num. groups: item	24
Var: participant (Intercept)	6.39
Var: participant gramfunc	0.10
Var: participant position	1.38
Var: item (Intercept)	0.49
Var: item gramfunc	0.19
Var: item position	0.03

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3.6: Estimates of the generalized linear mixed effects model reported in (35).

testing did not reveal a significant effect of POSITION, i.e. the distance between the R-expression and the pronoun. Note, however, that the data of this experiment gave rise to the largest variance estimate for intercepts by participant, indicating high inter-speaker variability.

Once again, we shall take a look at inter-individual performance rather than just the overall ratings. The data reveal that 46.7% of participants' responses are in line with the predictions of a principle C effect in the initial conjunct, while this rate drops to 36.7% in the non-initial conjunct. The size of the participant group showing an effect in the opposite direction increases from 15% in the initial conjunct to 23.3% in the non-initial conjunct, while the size of the group showing no effect in either remains rather stable with 38.3% in the initial conjunct and 40% in the non-initial. Participants' slopes in this experiment are overwhelmingly flat. There are more participants with lower coreference rates overall in the initial conditions than the non-initial conditions, corresponding to the significant effect of POSITION. Likewise, we can inspect whether the participants who did show an effect of principle C reconstruction in simple wh-dependencies in experiment 2 of chapter 2 are the same participants who show it in this experiment under ATB movement. Furthermore, we shall also probe whether the participants who did

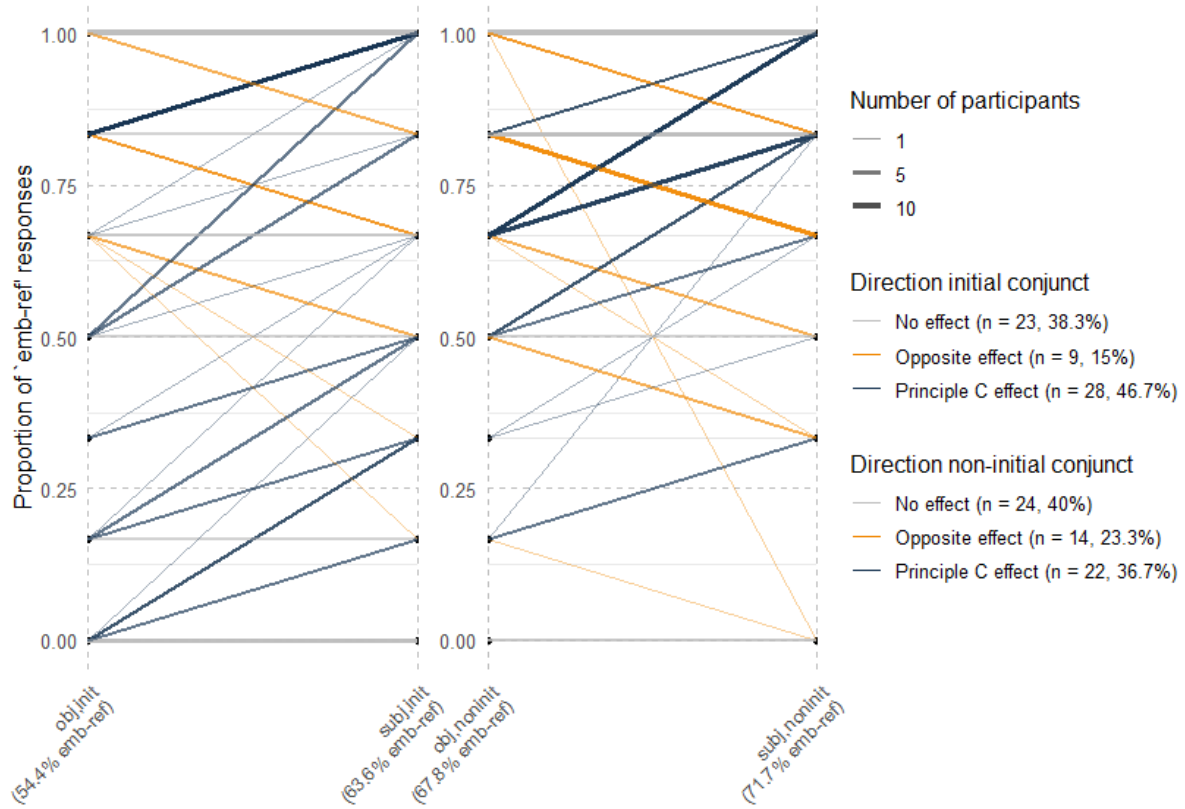


Figure 3.12: Individuals' ($n = 60$) overall proportions of responses indicating coreference with the embedded referent in experiment 6 by condition. Y-axis indicates proportion (0-1 by 0.167 based on 6 observations per participant per condition), lines connect individual participants' proportions in the object and subject conditions of the initial (l) and non-initial conjunct (r), respectively. Color indicates the direction of the effect (blue: subject > object, orange: subject < object, gray: subject = object), thickness and opacity indicate the number of participants. X-axis indicate overall percentage of yes responses in condition.

show an effect in the non-initial conjunct overlap with those who did in the initial conjunct. Figure 3.13 illustrates the migration of participants between effect groups across the different test environments. There is some overlap between participants exhibiting the effect in simple wh-dependencies as well as either the initial ($n = 15$) or non-initial conjunct in ATB ($n = 6$), respectively, and some participants show the effect consistently across all three environments ($n = 10$). There is no uniform pattern: Aside from the participants for whom the effect vanishes in ATB dependencies altogether, some participants responded in line with a principle C effect in simple wh-dependencies and only the initial conjunct of ATB. Others responded in line with principle C in simple wh-dependencies and only in the non-initial conditions of ATB; and yet again, there are certain participants who only responded in line with principle C in the initial conjunct of ATB, but neither in the baseline wh-movement cases nor the non-initial conjunct.¹²

¹²Interestingly, not a single participant shows the effect exclusively in the non-initial conjunct.

3.2. Principle C reconstruction in German ATB movement

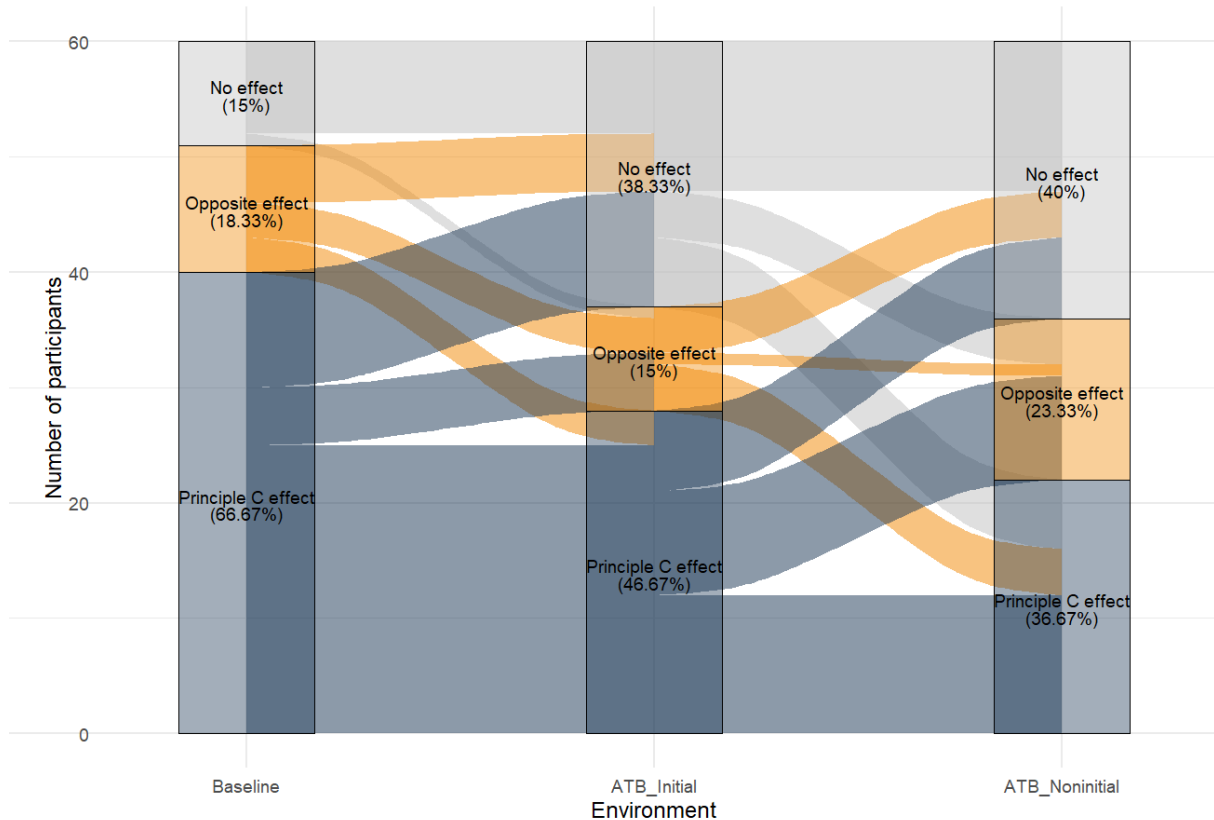


Figure 3.13: Proportion of participants corresponding to effect groups across different test environments (baseline = experiment 2 in chapter 2 on simple wh-dependencies, ATB_Initial = initial conjunct of ATB movement in current experiment, ATB_Noninitial = non-initial conjunct of ATB movement in current experiment. Flow indicates migration of individual participants from one effect group to the other.

3.2.4.8 Discussion

The significant main effect of GRAMFUNC supports the view that the effect of principle C reconstruction does not vary based on the position of the pronoun. Interestingly, not only is there no significant interaction between the two factors, but no significant main effect of POSITION either. As such, the results of experiment 6 are at odds with the findings of experiments 4 and 5, meaning that all three experiments that were conducted on principle C reconstruction in German ATB dependencies supplied distinct results.

Experiment 6 modified the task and item structure in an attempt to eliminate the distraction that the presence of a matrix referent may have introduced. As a result, coreference rates increased drastically overall compared to the two prior experiments, with participants picking the coreferent reading in the forced choice task above chance level. This is surprising considering that the two response options were dependent on one another given the change of task in this experiment

to a single forced choice task between two possible referents. It illustrates that participants are highly biased to resolve pronominal reference with whatever ϕ -matching referent is available. It further provides evidence for the hypothesis that the embedded referent's availability is not ruled out by a strong syntactic constraint, but is rather dispreferred in certain scenarios if a more prominent alternative is available, and preferred, on the other hand, if no specific alternative is available at all.

3.2.5 General discussion

3.2.5.1 Differences across experiments

The overall coreference rates show the same pattern across experiments, see Table 3.7. The conditions can be ranked as follows, in increasing order of coreference chosen in experiments (with a minor deviation of 0.4% between experiment 4): *object, initial* \succ *subject, initial* \succ *object, non-initial* \succ *subject, non-initial*.

	<i>obj, initial</i>	<i>subj, initial</i>	<i>obj, non-initial</i>	<i>subj, non-initial</i>
Experiment 4	25.6%	35.9%	46%	45.6%
Experiment 5	23.2%	29.2%	32.2%	35.2%
Experiment 6	54.4%	63.6%	67.8%	71.7%

Table 3.7: Overall proportions of responses indicating coreference between pronoun and R-expression in PP modifier of wh-phrase in experiments 4-6.

After finding a rather small difference between subject and object extraction in the initial and a no difference at all in non-initial conditions of experiment 4, experiments 5 and 6 set out to study how the manipulations of pragmatic factors might increase the detected effect of the intended principle C manipulation. These manipulations included the presence of context, how much attention the experimental task drew to each available referent as well as the availability of an alternative referent at all. Manipulating these non-syntactic and thus possibly distracting factors was expected to increase the strength of the syntactic manipulation in case pragmatic and syntactic factors interacted, an idea entertained in section 2.3.3 of chapter 2. However, the experiments clearly show that the applied manipulations do not increase the size of a principle C reconstruction effect in the sense that coreference becomes uninhibited in subject conditions and nearly unavailable in object conditions under the 'right' pragmatic conditions. While the overall coreference rates fluctuate drastically, their pattern remains the same, modulo variable results of the significance tests. That is, a pragmatically neutral setting in which we would expect stronger syntactic effect is difficult if not impossible to achieve. If there is a matching, syntactically available referent in the

matrix clause like in experiments 4 and 5, it will obviously outrank the embedded referent in prominence. If there is no referent in the matrix clause, however, like in experiment 6, the embedded referent will be the most obvious choice as it is easier to pick the given referent rather than construing an alternative out of nothing.¹³ Recall also that the size of the participant pool varies a lot across experiments, and therefore the number of observations between experiments varies both with respect to each condition as well as with respect to each participant and item. Table 3.8 summarizes them.

	Participants	Items	Obs./cond.	Obs./part./cond.
Experiment 4	254	12	762	3
Experiment 5	150	24	900	6
Experiment 6	60	24	360	6

Table 3.8: Comparison of number of observations across experiments per condition and per participant per condition.

Experiment 4 had half as many items as experiments 5 and 6 but over a 100 more participants than experiment 5. That is, experiment 4 has only half as many observations per participant per condition as experiment 5, although the two are much closer together in terms of the number of observations per condition. This imbalance may have had an impact on the findings, assuming that there may be effects associated to trial number (cf. experiment 3 in section 3.1 of this chapter). A post-hoc analysis of the effect of trial number revealed that it did not have a significant effect in experiments 4 and 6, while it did in experiment 5 ($\text{Pr}(> |z|) = 1.25\text{e-}06$). Additionally, the interaction between trial number and POSITION is marginally significant in experiment 5 ($\text{Pr}(> |z|) = 0.090022$). Overall, this indicates that there was a familiarization effect in experiment 5, i.e. that participants were more inclined to pick the coreferent reading during later trials than earlier ones. Overall, while the pragmatic manipulations altered the outcomes of the significance tests, they did not result in an increase of effect size revealing the one ‘true’ pattern of principle C reconstruction in ATB movement.

3.2.5.2 Implications for theories of ATB movement

What do the experiments reveal about the different theoretical proposals to ATB movement? The findings of experiment 4 are consistent with theories predicting asymmetrical reconstruction to the initial gap, while those of experiments 5 and 6 are consistent with a symmetrical reconstruction pattern. In experiment 6, even the effect of distance between the R-expression and the pronoun vanishes,

¹³A third possibility would be to have an additional matching referent in the context but not in the target sentence. I leave this issue open for future research.

qualified by a significant effect of POSITION in other experiments. However, all experiments have in common that the size of the principle C effect is smaller than previously hypothesized in the literature, in line with the findings from chapter 2, and that a large proportion of participants' responses are not in line with the predictions of principle C reconstruction at all. Furthermore, coreference was chosen more frequently in experiment 6 of this chapter compared to experiment 2 in chapter 2. This may indicate an increased bias to resolve pronominal reference with increasing dependency complexity (cf. Gordon and Hendrick 1998).

In sum, the data from the three experiments suggest the following: (i) The principle C reconstruction diagnostic using PP modifiers, as employed in the previous literature, does not support a specific syntactic derivation of ATB movement. The fact that the interaction between GRAMFUNC and POSITION is significant only in experiment 4 speaks against an underlying syntactic asymmetry between the conjuncts, and likewise, the fact that experiments 5 and 6 revealed no such interaction does not supply clear evidence in favor of symmetrical reconstruction. (ii) Depending on the task and the experimental condition, the size of the participant group showing no effect at all or an effect in the opposite direction, i.e. an increase of coreference under object extraction compared to subject extraction, varies between 53.3% and 74%, indicating that the diagnostic does not reliably measure principle C reconstruction at all. (iii) The differences across experiments illustrate how sensitive coreference resolution is to the presence of context, a competing referent, and the phrasing of the experimental task. The findings are consistent with previous studies suggesting that the reconstruction of PP modifiers is compromised, although there is no unanimous explanation as to how this variability should be modeled. A comprehensive overview of this problem is given in section 2.1 chapter 2. The findings are consistent with the claim by Bruening and Al Khalaf (2017) that the asymmetrical principle C reconstruction pattern in ATB is not necessarily indicative of an asymmetrical syntactic derivation. Rather, it appears that the outcome of the diagnostic is orthogonal to questions about the underlying syntax of multi-gap dependencies.

The variable pattern of principle C reconstruction with PP modifiers reported previously in the literature therefore seems to be a superficial phenomenon. The increased distance between the R-expression and the pronoun in non-initial conditions is a more likely cause for potential asymmetries than syntactic structure. The evidence from this series of experiments is not conclusive regarding theories of ATB movement, it merely highlights that principle C reconstruction, at least of PP modifiers, is a poor choice to study underlying syntactic relations. The main takeaway from the experiments is the importance of including baselines to check

whether the test is valid. Previous reports in the literature only compared structures corresponding to the conditions *object*, *initial* and *object, non-initial*. Here, subject and object extraction were contrasted, manipulating whether pronoun c-commanded the base position of the R-expression, while keeping linear order constant. Crucially, this manipulation gave rise to a significant effect of GRAMFUNC across all experiments, indicating that there is a residual effect of reconstruction informing at least some participants' responses. Given that in two out of three experiments POSITION also had a significant effect, the next section focuses on possible ways to determine the exact nature of this effect of distance.

3.2.5.3 Effect of distance

Based on chapter 2, we know that the principle C reconstruction effect is detectable in simple wh-dependencies, though not particularly robust. The principle C reconstruction data previously claimed to reflect the underlying syntax of ATB movement proved to be inconclusive upon experimental scrutiny, at least in German. While the overall coreference rates revealed an asymmetry between the two conjuncts, this asymmetry also surfaces in the absence of a configuration where the pronoun c-commands the R-expression, i.e. in *subject* conditions where coreference should be unrestricted. Therefore, rather than a syntactic difference between the material occupying each conjunct, the experiments demonstrate that the robustness of principle C reconstruction is compromised under increased dependency complexity. Crucially, coreference rates increase across levels of GRAMFUNC in non-initial conditions, confirming an effect of distance (Adger et al. 2017; Salzmann, Wierzba, and Georgi 2023; Stockwell, Meltzer-Asscher, and Sportiche 2021). Whether the effect in multi-gap dependencies arises due to linear or structural distance cannot be answered based on the ATB data—the linear order of the conjuncts always corresponds to their hierarchical order. In principle, the same paradigm could be adapted to parasitic gaps to tease apart the difference. While the adjunct clause is more deeply embedded, it can be linearized either as incorporated or extraposed in German, allowing for the free manipulation of the surface order as shown in (36).

- (36) a. Welchen Witz über Ute hat Mats ____
 which joke-NOM/ACC about Ute-ACC has Mats-NOM
 erfunden ohne ihr _____{pg} zu erzählen?
 made.up without she.DAT to tell
 ‘Which joke about Ute did Mats make up without telling her?’
- b. Welchen Witz über Ute hat Mats, ohne
 which joke-NOM/ACC about Ute-ACC has Mats-NOM without

ihr — pg zu erzählen, — erfunden?
 she.DAT to tell.INF make.up
 ‘Which joke about Ute did Mats, without telling her, make up?’

Although promising, some confounds make this test difficult to carry out experimentally. First of all, we have already seen in section 3.1 of this chapter that parasitic gaps are a highly variable and rather marginal phenomenon in German, such that even simple versions of the constructions are judged to be rather awkward by most participants, some rejecting them altogether. Second, because the relevant adjunct clauses in German do not introduce new subjects, instead of the subject-object contrast, one would have to switch to indirect vs. direct objects. This gives rise to new problems, such as ambiguity about the direct object of the verb in the adjunct clause: Some ditransitive verbs can be plausibly interpreted as missing the direct object gap, thus allowing for a parse where there is no parasitic gap to begin with (*lesen* ‘to read’ for example, can plausibly be interpreted as intransitive, meaning ‘doing some reading’ rather than ‘reading something’). On the other hand, ditransitive reportative verbs such as *erzählen* ‘to recount/tell’ can also be interpreted such that they take the entire proposition of the matrix clause as their direct object (i.e. the actual parasitic gap reading, ‘Mats did not tell Ute about a joke’ vs. ‘Mats did not tell Ute about making up a joke’). Further, the c-command relation between indirect and direct objects in German is debated (Featherston and Sternefeld 2003; Grewendorf 1988; Twiner and Lee-Schoenfeld 2019). An investigation along these lines in English manipulating the placement of purpose clauses is pursued in chapter 4.

Finally, given the current findings on ATB as well as prior ones on simple wh-movement, there is a very simple preliminary explanation for why reconstruction patterns in ATB movement seem to differ across phenomena. As noted in section 1.3.2 of chapter 1, most types of binding show a symmetric reconstruction pattern, with principle C being one of the noteworthy exceptions seemingly only targeting the initial gap (Citko 2005; Salzmann 2012). The present data, in line with Bruening and Al Khalaf (2017), are taken to indicate that the allegedly asymmetric principle C effect in ATB movement is not an effect of reconstruction. I hypothesize that this is due to the instability of principle C reconstruction in and of itself. On the one hand, it has long been noted that coreference or the lack thereof does not require c-command whereas binding does (Reinhart 1983a,b). On the other hand, even if we follow the idea that binding principles are responsible for regulating coreference as well (Bruening 2021; Chomsky 1981; Heim 2007), the data entail that PP arguments do not reconstruct alongside the head noun, as claimed by many (Bianchi 1995; Henderson 2007; Kuno 2004; Lasnik 1998; Safir 1999), rendering the principle C reconstruction test in its standard form irrelevant. The innovation

of the current experiment on ATB, following work by Salzmann, Wierzba, and Georgi (2023) on simple wh-dependencies, is that the subject conditions serve as a sanity check, showing that distance plays a crucial role in determining coreference patterns in ATB dependencies.

3.2.6 Conclusion

The experimental findings indicate that the principle C reconstruction test in ATB movement is inconclusive. The reason for this, on the one hand, is the weakness of the contrast between responses in the conditions *object, initial* and *subject, initial*, and on the other hand, the presence of a contrast between the conditions *subject, initial* and *subject, non-initial*. One of the experiments provided evidence for a syntactic asymmetry between the gaps, while the other two experiments found no significant interaction between the manipulated factors, suggesting both gaps to be equal. This inconsistency suggests that coreference rates are not primarily determined by principle C reconstruction. The distance between the pronoun and the embedded referent plays a significant role based on the presence of the effect in subject conditions where the offending c-command configuration never arises. The view that the main driving force behind the pattern is not a syntactic violation is further solidified by the inspection of individual participants' slopes, which indicate that the majority of them did not respond in line with the predictions of a principle C reconstruction effect.

The current findings are in line with at least some claims by Bruening and Al Khalaf (2017) about English. The experiment compared surface violations of principle C in RNR to underlying ones under reconstruction in ATB, assuming that RNR does not 'bleed' principle C (Levine 1985). Coreference rates in the RNR conditions were found to be around 5% and thus 20-25% lower than in ATB conditions. The authors conclude that the reported asymmetry is not real. For English, the contrast between the initial and non-initial gap is reported to be lower than found for German in the current experiment, suggesting that there may be cross-linguistic variability. Note, however, that comparing surface violations of principle C to reconstruction effects makes interpreting the results rather difficult due to the confound of linear order and the overall compromised strength of the reconstruction effect. In the three experiments reported in this chapter, we saw how much the overall coreference rates can fluctuate depending on the experimental task and item structure. It is therefore very easy to manipulate the outcome based on the presence and prominence of alternative referents and the phrasing of the task, showing that a solid experimental design including control conditions (where the absence of any effect is expected) is crucial to interpret the

results correctly. While this was not the case in the experiments presented by Bruening and Al Khalaf (2017), the experiments reported herein give us more clarity about the factors influencing principle C reconstruction in multi-gap dependencies.

We can conclude that the principle C reconstruction test in German is insufficient to make any claims about the underlying structure of ATB movement. The experiments show that coreference possibilities are affected by the position of the pronoun even in the absence of a principle C violation. Furthermore, the outcomes of the three experiments do not converge: Based on significance testing, experiment 4 indicates an asymmetry between the gaps with respect to principle C reconstruction and an additional effect of linear distance; experiment 5 indicates symmetry with respect to principle C reconstruction and an effect of distance; and experiment 6 only indicates symmetry between the gaps for reconstruction but no effect of distance. The manipulation of underlying c-command via subject vs. object extraction is entirely novel for ATB movement and provides evidence against the view that the employed test reliably measures underlying syntactic relations. This is further supported by the rampant inter-speaker variability. Especially in experiments 5 and 6 for which data from the same participants' in simple wh-dependencies is available, we see that although group sizes with respect to the effects are similar, the identity of participants within each group varies across conditions. That is, although a similar number of participants may show the effect in each conjunct of ATB, it is not the same participants who show the effect across conjuncts.

3.3 Interim summary

The experiments on multi-gap dependencies in German illustrate the importance of controlled and thorough data collection before making far reaching theoretical claims. First, I set out to examine parasitic gaps, finding that the literature disagrees about what is to be considered a parasitic gap in German and whether the language even has parasitic gaps in the first place. The acceptability judgment experiment reported in section 3.1.3 made it clear that parasitic gaps in adjunct clauses do exist in the language, but do not enjoy widespread acceptability among speakers. In extraposed adjunct clauses, parasitic gaps were judged to be acceptable in only 38% of trials (eliciting a rating of 5 out of 7 or higher). The variant where the adjunct clause was incorporated performed better, reaching a judgment of 5 or higher in 50% of trials. Significance testing revealed a significant effect of GAP, indicating that gapless minimal pairs of parasitic gap constructions were judged to be significantly more acceptable; a significant effect of PLACEMENT, indicating that incorporation of the adjunct clause has

a significant effect on the acceptability; and crucially, a significant crossover interaction between the two factors. This crossover interaction indicates that while extraposition has a negative effect in parasitic gap constructions, it has a positive effect otherwise and vice versa, incorporation has a positive effect on parasitic gaps, while it has a negative effect otherwise. I hypothesized that this may follow from the way filler-gap dependencies and islands are processed, though online experimentation is necessary to verify these speculations. We saw a high degree of inter-speaker variability, with certain participants rejecting parasitic gaps altogether and the vast majority judging them to be quite mediocre. This variability could not be tied to participants' geographical region of origin. We also saw that the experiment was subject to a familiarization effect, with participants' judgments increasing across conditions as the experiment progressed. In sum, given that the tested structures were among the simplest cases of parasitic gaps, I concluded that further experimentation with the added complications of principle C reconstruction—multiple competing referents and the ambiguity of pronoun resolution—would likely further decrease the overall acceptability of parasitic gap constructions even among those participants who accepted them in the first place. This would in turn compromise the interpretability of responses as it is questionable how revealing judgments about the meaning of a structure are if it is unacceptable to begin with.

Having set parasitic gaps aside in, I then investigated what coreference pattern principle C reconstruction in German ATB movement gives rise to. As it turns out, the conflicting patterns reported in the literature for English, i.e. asymmetrical reconstruction to the initial gap only vs. symmetrical reconstruction to both gaps, were reproduced by the three experiments. In one of them, there was a significant interaction between *GRAMFUNC* and *POSITION*, which is expected if reconstruction is asymmetric (to the initial gap, as response proportions revealed). However, the two follow-up experiments found no significant interaction between the factors *GRAMFUNC* and *POSITION*, indicating that the effect of principle C reconstruction was equal no matter which gap the pronoun c-commanded. One of these two even failed to find a main effect of *POSITION*, indicating that the distance between the referent and the pronoun did not have a significant effect on participants' responses either. As such, the results of the three experiments do not converge, leading us to question what this implementation of the principle C reconstruction test actually measures.

We have seen in chapter 2 that principle C reconstruction is by far not as robust of a phenomenon as its ubiquity in the formal syntactic literature makes us believe. The data presented in section 3.2 illustrate how the validity of the test is further compromised with increasing dependency length or even complexity,

regardless of the experimental design. In spite of a consistently significant effect of *GRAMFUNC*, we see that less than 50% of participants' responses are in line with the predictions of principle C reconstruction across experiments, the rest either showing no effect or an effect in the opposite direction. The fact that significance testing does not paint a uniform picture may stem from a difference in statistical power, as we are dealing with a distinct number of observations in each experiment and likewise a different composition of item and participant numbers. Recall Table 3.8 summarizing these imbalances.

Nonetheless, variable statistical power cannot be the sole source of the diverging results. Participants' behavior across conditions based on the expected effect of *GRAMFUNC* is not robust either, indicating that there is no evidence for a division of speakers into a group showing a principle C reconstruction effect vs. a group being led by surface order only. Crucially, as in the experiments reported in chapter 2, the effect of *GRAMFUNC* is fairly small, and the manipulations targeting this issue only led to an overall increase or decrease of coreference rates across all conditions rather than a strengthening of the effect. The patterns in experiments 4 and 5 are possibly best explained by a general disliking of using a pronoun to refer to a referent that was just named (Arnold 2001), indicated by a significant effect of *POSITION* which seems to improve coreference when the pronoun is in the non-initial conjunct. At the same time, the bias to resolve pronominal reference with whatever referent is available in the discourse seems to inform the results of experiment 6, where the effect of *POSITION* is not significant. Recall that participants chose the coreferent option in the subject conditions below chance, especially in experiment 4 and 5 measuring possibility of coreference rather than preference like experiment 6. This is particularly puzzling from a purely syntactic standpoint since there is no syntactic condition inhibiting coreference between the pronoun and the referent in subject conditions—the reconstructing constituent containing the R-expression is the subject of the clause, and therefore it will never end up in the c-command domain of the object pronoun, regardless of reconstruction. The consistency of subject conditions not scoring maximal coreference rates supports the idea that the competition between pragmatic factors on coreference resolution—the presence of an alternative referent and context, the bias to resolve pronominal reference, the disliking of referring to an R-expression with a pronoun immediately following it—can easily overshadow the outcome of the principle C reconstruction test.

It is quite clear that variability in the experiments, although rampant, is not random by any means. Across all three of them, the pattern arising among conditions is robust. However, the findings illustrate that there are many independent factors at play here which are typically not accounted for under

the application of the principle C reconstruction diagnostic, indicating the need for more controlled means of data collection. I conclude that the principle C reconstruction test in German ATB movement does not reliably measure underlying syntactic relations, therefore proving to be an ill suited diagnostic to evaluate the accuracy of the many existing syntactic derivations in complex syntactic dependencies. In the next chapter, the findings from chapter 2 and 3 inform the investigation of principle C reconstruction in English multi-gap dependencies, or rather, how much of the coreference pattern in multi-gap dependencies can be attributed to principle C reconstruction.

Chapter 4

Experiments on English multi-gap dependencies

Having provided evidence that the principle C reconstruction test does not reliably map to underlying syntactic relations in German ATB movement, this chapter addresses the same research question for English multi-gap dependencies. The chapter discusses five experiments. Section 4.1 presents two foundational studies: Experiment 7 deals with principle C reconstruction in simple wh-dependencies in section 4.1.1, establishing the counterpart to the experiments reported in chapter 2 on German; and experiment 8 investigates the acceptability of English parasitic gaps in finite and infinitival adjunct clauses in section 4.1.2. The latter experiment serves as a sanity check and is based on existing reports about clause finiteness compromising the acceptability of parasitic gaps, since the design of the principle C reconstruction test employed here requires the use of finite adjunct clauses. Section 4.1 thus establishes the necessary foundation for further investigations of principle C reconstruction in multi-gap dependencies of English.

Section 4.2 then moves on to coreference in English multi-gap dependencies. The investigation is split into two parts since testing the complete paradigm that crosses all factors is ruled out due to the unavailability of subject parasitic gaps (Davis 2025; Engdahl 1983). Experiment 9 in section 4.2.1 compares coreference patterns in parasitic gaps to those in ATB movement. Moving on to the investigation principle C reconstruction proper, experiment 10 in section 4.2.2 investigates how much of the coreference pattern observed in 4.2.1 can be attributed to a principle C reconstruction effect, providing data for ATB movement and inferring what a likely pattern in parasitic gap constructions might look like. Two follow-up investigations are then presented in section 4.3. Experiment 11 is a final explorative experiment presented in section 4.3.1, aiming to clarify which type of distance constitutes the effect of POSITION across all of the experiments reported herein. Section 4.3.2 discusses a post-hoc statistical analysis of how the experimental method has

influenced the outcomes of the experiments in this chapter. We shall see that parasitic gaps and ATB movement give rise to the same overall coreference pattern, but that the involvement of principle C reconstruction only receives limited support in the initial gap site of ATB movement. Furthermore, post-hoc analyses reveal crucial problems associated with the experimental task(s), imposing limitations on how robust the conclusions are that we can draw from them.¹

4.1 Foundations

4.1.1 Experiment 7: Principle C reconstruction in English

We begin our investigation on English with an experiment on principle C reconstruction in simple wh-dependencies. Given that we learned from German that (i) the effect of principle C reconstruction is detectable but rather small, and (ii) it varies based on the experimental design, it is crucial to collect data on participants' responses to the diagnostic itself. The design of the experiment presented in this section most closely resembles experiment 2 section 2.2.3 of chapter 2. As such, we expect coreference rates to be around chance level in any case, but crucially also that the effect will be rather easily detectable based on the simplicity of the forced choice task and the presence of only one referent.

4.1.1.1 Design

The experiment manipulates a single experimental factor, GRAMFUNC, with the two levels *object* and *subject*. As in previous experiments presented throughout chapters 2 and 3, this indicates the grammatical function of the extracted element, i.e. whether it is the object or the subject of the clause. The idea is that under object extraction, reconstruction of the constituent to its base position should yield a principle C violation given the presence of a subject pronoun c-commanding said base position. Under subject extraction, the constituent's base position c-commands the object pronoun, therefore not giving rise to a principle C reconstruction effect. The manipulation of this factor makes for clearly interpretable results as it targets the very core of principle C reconstruction: The c-command relation between a pronoun and an R-expression's base position without confounding it with the linear order of the two.

¹All supplementary materials from this thesis, including experimental items, raw data and analysis scripts can be accessed on OSF at <https://osf.io/hvm8u/> (Szarvas 2025b).

no language related disorders. After applying the exclusion criterion, data from $n = 6$ participants were excluded. All participants received monetary compensation, regardless of whether their data were used in the analysis.

4.1.1.5 Procedure

The experiment was set up using the platform L-Rex (Starschenko and Wierzba 2024). The sentences were displayed simultaneously with the two tasks, one of them testing coreference with the embedded referent and the other measuring the acceptability of the item. The first block included three training items, one where the pronoun could refer to the only matching referent in the sentence, one where the pronoun could not refer to any of the referents in the sentence due to a gender mismatch, and one where there was no pronoun and the only available reading of the sentence was assessed. The second block contained target items and distractors in pseudo-randomized order, such that two items from the same set of materials were never shown consecutively.

4.1.1.6 Predictions

Principle C reconstruction predicts a significant effect of `GRAMFUNC` such that the reading where the pronoun refers to the R-expression in the sentence should be chosen more often in the *subject* condition than the *object* condition. Moreover, given what we know from previous experiments, coreference is not expected to be ruled out entirely in the condition *object* either. Given that the task measures preferences, the coreferent reading should be chosen around chance level in the condition *subject* to indicate that the referent is available, with a higher proportion of responses indicating a bias to resolve pronominal reference with whatever matching referent is given in the sentence. The presence of a principle C reconstruction effect is expected to manifest in below chance coreference rates in the *object* condition, again, factoring in the possibility that participants will be biased to choose the coreferent reading for lack of a salient alternative.

4.1.1.7 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using a generalized linear mixed effects model for binary data from the package ‘lme4’ (D. Bates et al. 2015) with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’. Modeling was carried out only for the forced choice task. A conservative α -level of 0.05 was defined. The model in (2) includes the fixed effect of `GRAMFUNC` as well as a random effects structure with varying intercepts and slopes for both participants and items (Barr et al. 2013). Data from $n = 54$ participants were analyzed.

(2) $\text{reference} \sim \text{gramfunc} + (1 + \text{gramfunc} \mid \text{item}) + (1 + \text{gramfunc} \mid \text{participant})$

4.1.1.8 Results

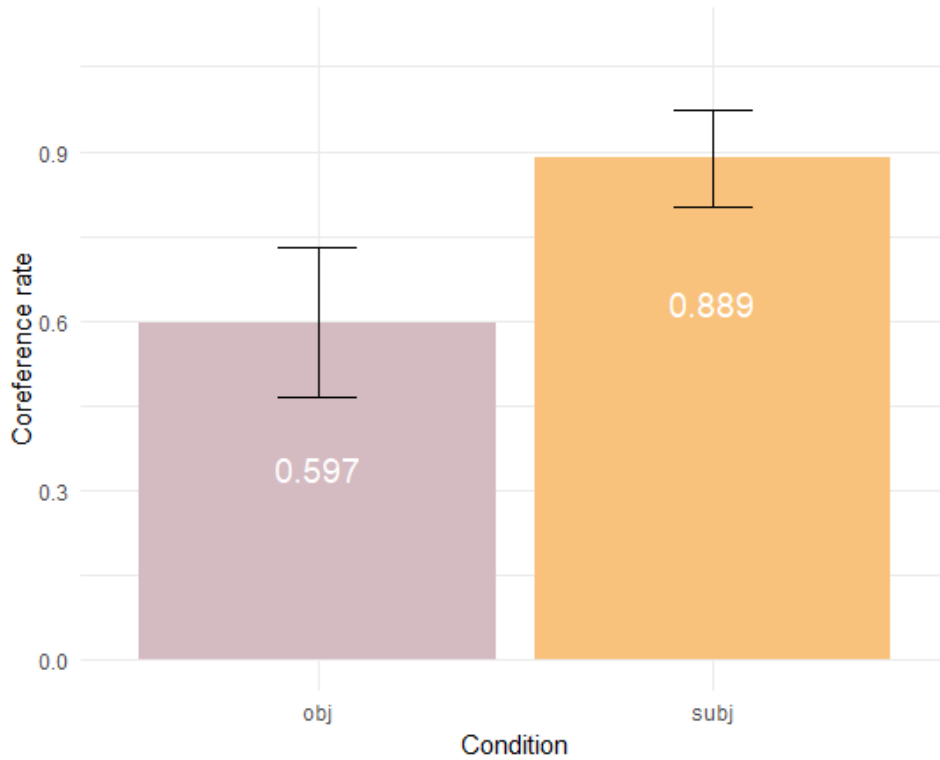


Figure 4.1: Coreference rates across conditions in experiment 7. Error bars indicate standard error.

The overall coreference rates are summarized in Figure 4.1. Coreference rates are above chance level in both conditions and notably almost at ceiling in the *subject* condition. Crucially, coreference rates in the two conditions differ by 29.2%, which is far beyond the 13.7% found in the German counterpart of the experiment reported in chapter 2, section 2.2.3. Therefore, although the effect does not rule out coreference in the *object* condition, the absence of a principle C reconstruction effect led participants to choose the coreferent option in 88.9% of trials, indicating a strong preference to assign the pronoun to the presented R-expression rather than an arbitrary unnamed referent.

The estimates of the model defined in (2) are given in Table 4.1. Significance testing revealed a significant effect of GRAMFUNC, rendering the effect of principle C reconstruction significant. Participants' responses are in line with the prediction that the coreferent reading should be chosen less frequently in the object than subject condition. There is a considerable amount of variance between participants' intercepts as well as the effect of GRAMFUNC.

Figure 4.2 illustrates participants' slopes to explore this variability. In line with

GLMM	Estimate (SE)
(Intercept)	-3.05*** (0.38)
gramfunc	2.39*** (0.41)
AIC	402.30
BIC	434.85
Log Likelihood	-193.15
Num. obs.	432
Num. groups: participant	54
Num. groups: item	8
Var: participant (Intercept)	2.97
Var: participant gramfunc	0.90
Var: item (Intercept)	0.06
Var: item gramfunc	0.31

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.1: Estimates of the generalized linear mixed effects model for experiment 7, including fixed effect for GRAMFUNC and random intercepts and slopes for items and participants.

the rather large effect of GRAMFUNC compared to previous experiments, 63% of participants responded in line with the predictions of principle C reconstruction, i.e. indicating coreference with the embedded referent more often in the subject than in the object condition. Only 5.6% of participants responded in the opposite direction, and 31.5% of participants show no effect at all, indicating coreference equally as often in the subject as in the object condition. We thus find that although the overall proportion of responses varies quite a bit, the majority of participants respond in line with the predictions of the manipulations, with roughly a third of participants not showing any effect. These participants, all of them choosing coreference in 100% of trials with target items, seem to be exclusively led by the items' surface structure.

4.1.1.9 Discussion

The first study in the series of experiments in English revealed a more robust principle C reconstruction effect than the German counterpart of the experiment presented in section 2.2.3 in chapter 2 did. The overall coreference rates between conditions differ by 29.2% compared to the mere 13.7% found in German. This is accompanied by a significant effect of GRAMFUNC. While this may in part be indebted to cross-linguistic differences between the languages, it could also be a result of using canonical transitive verbs taking an animate agent and patient,

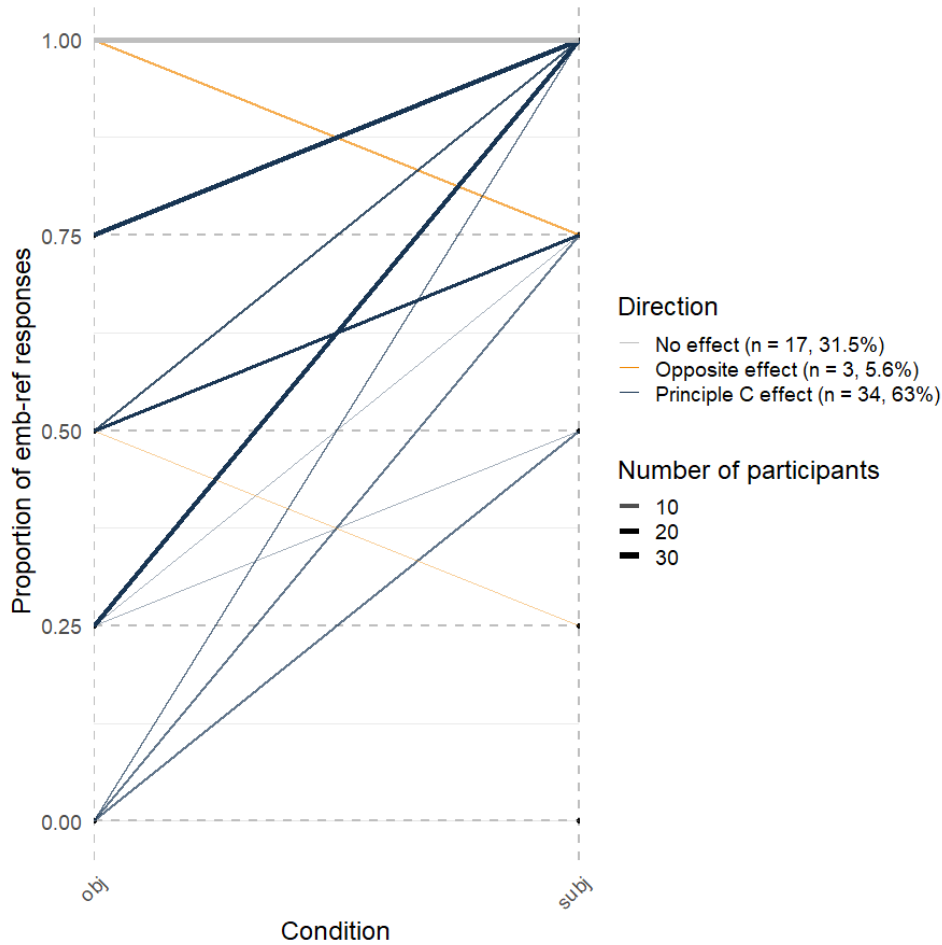


Figure 4.2: Individual participants' ($n = 54$) overall proportions of responses indicating coreference with the embedded referent in experiment 7 by condition. Points on y-axis indicate respective proportion (ranging from 0 to 1 in steps of 0.25 based on 4 observations per participant per condition), lines connect individual participants' proportions in the condition *object* to *subject*. Color indicates whether the direction of the effect conforms to the predictions based on a principle C violation, thickness and opacity indicate how frequent the respective combination of proportions from the two conditions is in the data set.

thus yielding a particularly even item structure across conditions due to allowing for simply flipping the arguments rather than rephrasing the verbs. The use of animate extractees furthermore made for items that exhibit a more argument-like relationship between the NP and its modifier, such as kinship terms for example, potentially strengthening the syntactic requirement for reconstruction and thus the robustness of the effect.

Of the 54 participants whose data remained in the sample for analysis, 34 responded in line with the predictions of principle C reconstruction, while 17 chose the coreferent reading equally often across conditions. The remaining 3 participants chose the coreferent reading in one more trial in the object than the subject condition, yielding an effect in the opposite direction. However, it seems

clear that the majority of participants is led by the underlying structure and thus a principle C reconstruction effect, even if it only yields a degradedness of the respective reading and not the impossibility of coreference. As Figure 4.2 shows, only few participants indicated the coreference was entirely impossible in the object condition. This further supports the conclusions reached in chapter 2 that principle C reconstruction is only one out of many contributing factors influencing coreference resolution.

The next section turns to the status of parasitic gaps in finite vs. infinitival adjunct clauses given the use of finite adjunct clauses in experiment 9.

4.1.2 Experiment 8: The effect of finiteness on the acceptability of parasitic gaps in English

Experiment 8 serves to establish another one of the basic pillars necessary to continue the investigation of multi-gap dependencies in English—the acceptability of parasitic gaps. In order to test principle C reconstruction to a parasitic gap, the clause hosting it has to have a pronominal constituent in a position c-commanding the parasitic gap. The easiest way to allow for this without complicating the syntactic structure any further through the use of causative or ECM verbs, is to use finite clauses. Previous research by Phillips (2006) has established that parasitic gaps in English are judged to be more acceptable in an experimental setting than reports from the literature lead us to suspect, scoring a mean rating around 3.5 out of 5.0 in infinitival complement clauses (around the same rating as the control condition scored with a single gap in the main clause and a finite relative clause with no parasitic gap). Parasitic gaps in finite relative clauses, however, scored a mean rating of 2.6 out of 5.0. The study found a significant main effect of whether there was a parasitic gap in the sentence and whether the clause was finite, as well as a significant interaction between the two factors. These data crucially show that the acceptability of parasitic gaps is significantly compromised in finite clauses compared to infinitival clauses.

We have seen in section 3.1 of chapter 3 that the acceptability of parasitic gaps in German is rather marginal, only scoring a mean rating of 3.86 in extraposed adjunct clauses and 4.25 in incorporated adjunct clauses (on a 7-point Likert scale). Based on these findings, the study of principle C reconstruction in German parasitic gaps was set aside considering that further complicating the structure through a setup testing reconstruction would compromise the acceptability even further. The same question arises now for English, i.e. whether parasitic gaps in finite clauses in general are too marginal to probe for principle C reconstruction in their underlying syntax. Consider that the previous experiment by Phillips (2006)

tested parasitic gaps in infinitival complement clauses comparing them to finite relative clauses, i.e. cases where the parasitic gap precedes the licensing gap, and rather complex sentences, see the discussion in section 4.1.2.7. The experiments herein, however, deal with parasitic gaps in adjunct clauses, i.e. cases where the parasitic gap follows the licensing gap, and furthermore, sentences that are not quite as complex as the ones tested by Phillips (2006). Since the constructions we are dealing with here may elicit a different acceptability rating, the following experiment explored this possibility.

4.1.2.1 Design

Upon suspicion that parasitic gaps in finite vs. infinitival adjunct clauses may diverge less with respect to their acceptability than the ones tested by Phillips (2006), an acceptability judgment experiment with a 2×2 Latin Square design was conducted. The two tested factors were GAP, consisting of the two levels *prn* and *pg*, and the factor FINITENESS, consisting of the levels *inf* and *fin*.

4.1.2.2 Materials

Participants saw 8 experimental items in 4 conditions, seeing each item in only one of the conditions. Like in the other experiments reported on in this chapter, target items were interrogative sentences presented without context. Sentences contained an extracted filler, an inanimate D-linked wh-phrase, an R-expression as the subject and regular transitive verbs. Adjunct clauses were temporal adjunct clauses introduced by *before*, see (3). In *finite* conditions, the subject of the adjunct clause was a pronoun most saliently interpreted to corefer with the subject of the matrix clause.

- | | | | |
|-----|----|--|-----------------|
| (3) | a. | Which movie did Sarah watch before she recommended? | <i>fin, pg</i> |
| | b. | Which movie did Sarah watch before she recommended it? | <i>fin, prn</i> |
| | c. | Which movie did Sarah watch before recommending? | <i>inf, pg</i> |
| | d. | Which movie did Sarah watch before recommending it? | <i>inf, prn</i> |

Again, participants saw 12 distractors, and a total of 20 targets from experiments 7, 9 and 10 of this chapter.

4.1.2.3 Method, participants and procedure

The experiment used the same method as the other experiments discussed in this chapter, see section 4.1.1.3 for details. The same participants described in section 4.1.1.4 participated in this experiment, too, and the procedure matched the description in section 4.1.1.5.

4.1.2.4 Predictions

Based on the study by Phillips (2006), we expect a significant main effect of `GAP` and a significant interaction between `GAP` and `FINITENESS`. That is, while parasitic gaps should decrease the acceptability of adjunct clauses in general, the effect should be particularly pronounced in finite adjunct clauses. Furthermore, parasitic gaps in English are expected to be more acceptable than parasitic gaps in German, i.e. mean ratings should be higher than those in experiment 3 of chapter 3.

4.1.2.5 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using a cumulative link mixed effects model with symmetric thresholds. The model was fitted including fixed effects for `GAP` and `FINITENESS` as well as the maximal random effects structure, including random intercepts and slopes by participant and item for both fixed effects and their interaction.

$$(4) \quad \text{acceptability} \sim \text{gap} * \text{finiteness} + \\ (1 + \text{gap} * \text{finiteness} \mid \text{item}) + (1 + \text{gap} * \text{finiteness} \mid \text{participant})$$

4.1.2.6 Results

Condition	Mean rating (SE)
<i>fin, pg</i>	5.76 (0.15)
<i>fin, prn</i>	5.94 (0.12)
<i>inf, pg</i>	6.18 (0.11)
<i>inf, prn</i>	6.11 (0.11)

Table 4.2: Mean ratings by condition.

The experiment revealed that parasitic gaps in English are more acceptable than parasitic gaps in German. The mean ratings per condition are summarized in Table 4.2 and visualized in Figure 4.3. Parasitic gaps in finite adjunct clauses scored the lowest mean rating of 5.76, followed by the finite adjunct clauses with a pronoun scoring a mean rating of 5.94. Infinitival clauses with and without parasitic gaps scored near identical mean ratings of 6.18 and 6.11, respectively. The highest rated condition in this experiment is *inf, pg*, indicating that parasitic gaps were judged to be more acceptable than the control condition including a third person singular pronoun. This is a strong departure from the data we saw in chapter 3, section 3.1 for German parasitic gaps. Statistical modeling revealed no significant effect of `GAP` and a marginally significant effect of `FINITENESS`. The interaction is not significant, see Table 4.3 reporting the model estimates. Variance components reveal particularly high variability with respect to participants' as well as items' slopes for the interaction.

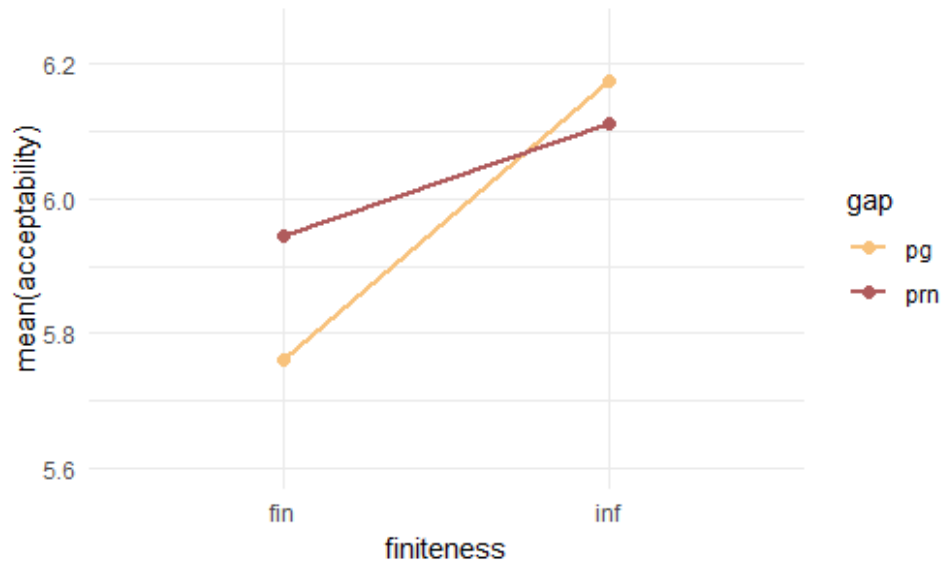


Figure 4.3: Mean acceptability ratings based on levels of GAP and FINITENESS.

CLMM	Estimate (SE)
finiteness	0.41 (0.24)
gap	0.04 (0.29)
finiteness:gap	0.05 (0.74)
Log Likelihood	-533.67
AIC	1121.34
BIC	1231.19
Num. obs.	432
Groups (participant)	54
Groups (item)	8
Variance: participant: (Intercept)	2.64
Variance: participant: finiteness	0.13
Variance: participant: gap	0.70
Variance: participant: finiteness:gap	2.27
Variance: item: (Intercept)	0.64
Variance: item: finiteness	0.02
Variance: item: gap	0.16
Variance: item: finiteness:gap	2.41

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.3: Estimates of the cumulative link mixed effects model with symmetrical thresholds including fixed effects for FINITENESS, GAP, and the interaction of the two, as well as random intercepts and slopes for items and participants.

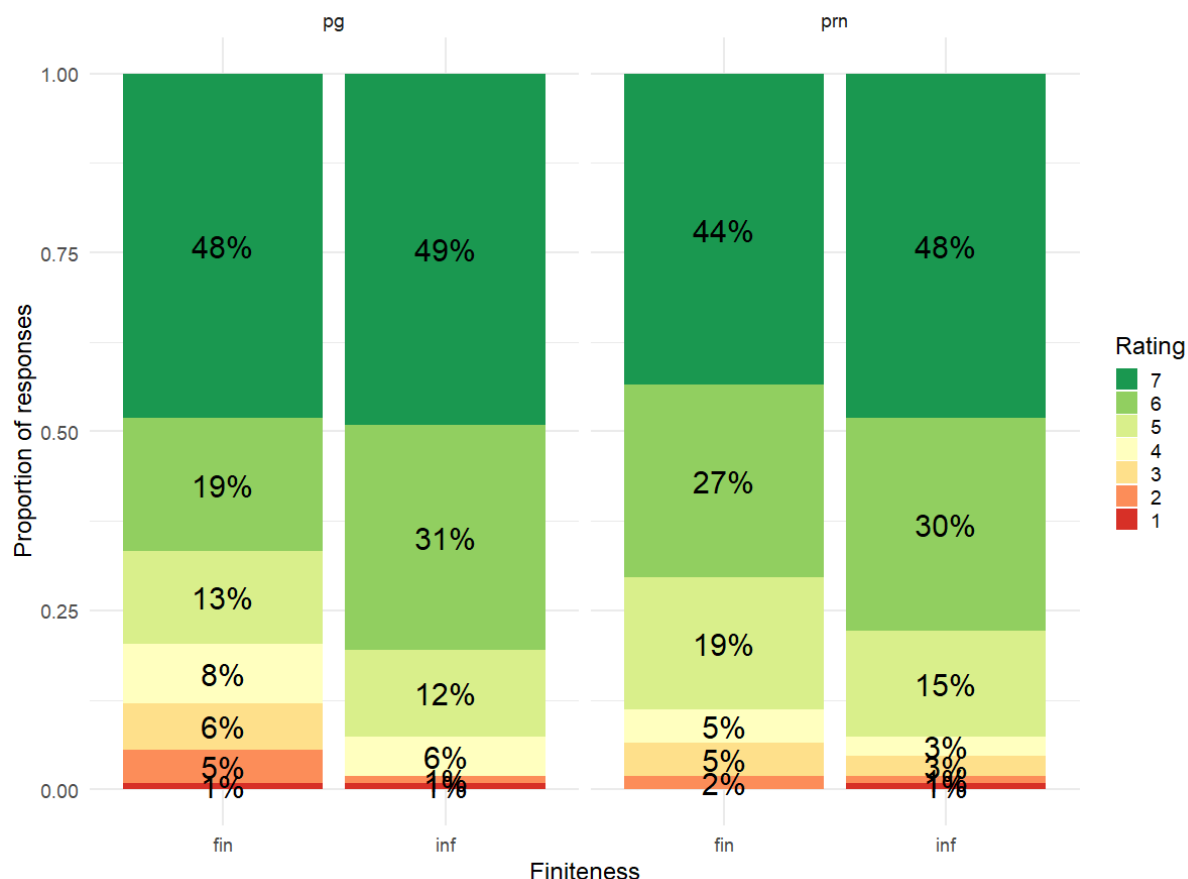


Figure 4.4: Distribution of responses across factor levels of GAP and FINITENESS. Percentages indicate proportion of responses within each condition.

The prediction that parasitic gaps are less acceptable in finite adjunct clauses than infinitival adjunct clauses is not borne out, as illustrated by the distribution of ratings across trials shown in Figure 4.4 and supported by the estimates of the statistical model. Parasitic gaps in finite and infinitival adjunct clauses elicited a rating of 5 or higher in 80% and 92% of trials, respectively (compared to 38% in the condition *extra, pg* and 50% in the condition *incorp, pg* of the German experiment). In the control condition with pronouns in place of parasitic gaps, 90% of trials in the finite and 93% of trials in the infinitival condition were given a rating of 5 or higher. That is, despite the lack of a significant effect, parasitic gaps in finite adjunct clauses are nevertheless judged to be worse based on the raw data, while those in infinitival adjunct clauses are judged more similarly to constructions without parasitic gaps. The heatmap in Figure 4.5 further illustrates that there is a lot less variability among participants in English than in German, supporting the conclusion that parasitic gaps in English are a widely accepted phenomenon. We see that the majority of participants gave high ratings above 5.0 across conditions, with exceptions found mostly in the condition *fin, pg*, as previously hypothesized.

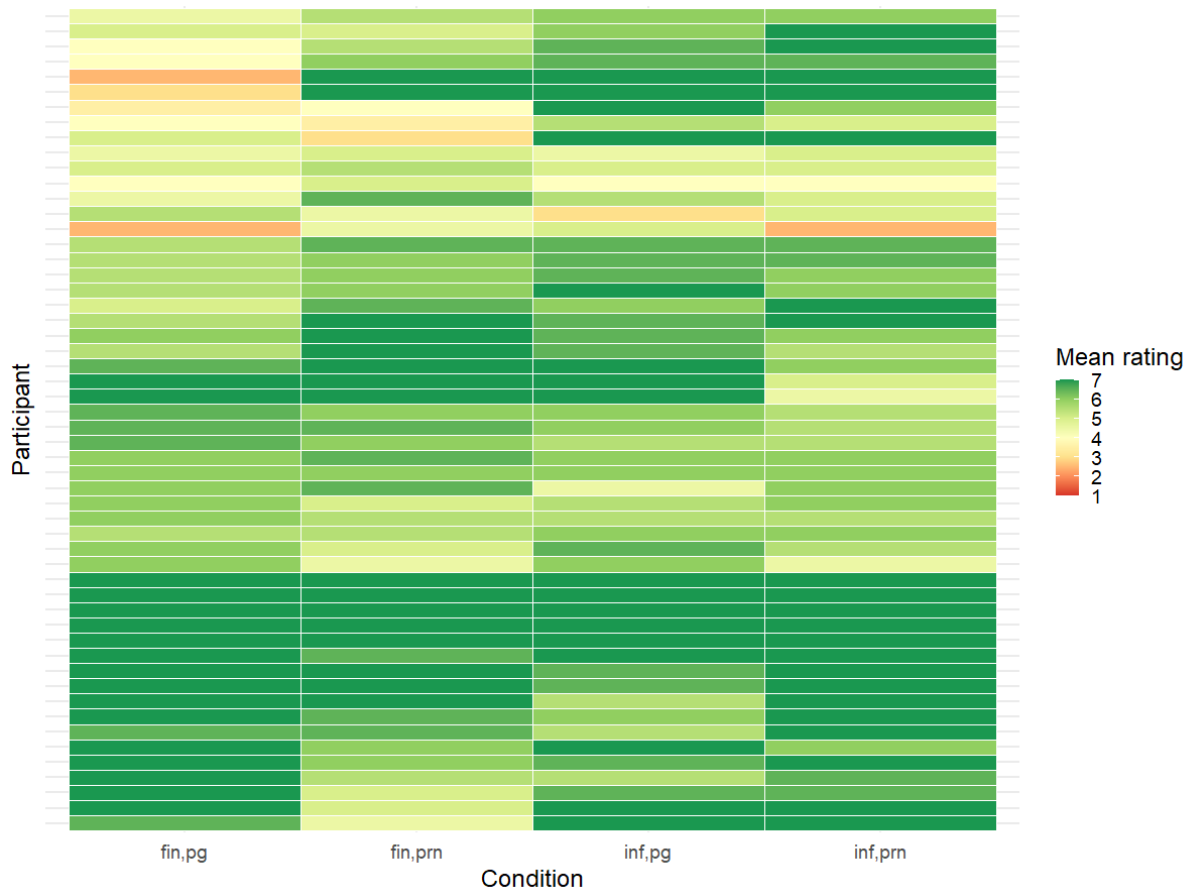


Figure 4.5: Heatmap of participants clustered by their mean ratings per condition. Each cell represents one participant’s mean rating in the respective condition, each row represents one participant’s ratings across conditions. Red cells indicate low means, green cells indicate high means. Clustering based on raw means.

4.1.2.7 Discussion

Parasitic gaps in English finite and infinitival adjunct clauses do not differ significantly in their acceptability. Contrary to prior hypotheses, the experiment did not find a significant effect of *GAP*, nor *FINITENESS* or their interaction. As such, it appears that the presence of a parasitic gap did not impair the acceptability of the items per se. While inspecting the raw data revealed fewer ratings above 5.0 for the condition *fin, pg* compared to the other three conditions, the interaction between the two factors is not significant. The acceptability of parasitic gaps does not seem to be inhibited by the finiteness of the adjunct clause, allowing us to proceed with the investigation of coreference judgments in English parasitic gap constructions.

These findings do not align with previous conclusions about parasitic gaps based on an acceptability judgment experiment reported by Phillips (2006). However, note that there are several differences between the two experiments that may explain the diverging results. The study by Phillips (2006) featured parasitic

gap constructions where the parasitic gap precedes the licensing gap, and items were quite complex compared to the those in the experiment reported herein as illustrated by an example of infinitival and finite parasitic gap items in (5).

- (5) a. (*inf, both*) The outspoken environmentalist worked to investigate what the local campaign to preserve _____{pg} had harmed ____ .
b. (*fin, both*) The outspoken environmentalist worked to investigate what the local campaign that preserved _____{pg} had harmed ____ .

The fact that the infinitival condition featured complement clauses and the finite condition finite relative clauses is probably negligible. On the one hand, the items only differed in a single word, otherwise being completely identical, and on the other hand, control conditions featuring a single gap either in the matrix clause (regular \bar{A} -movement) or the complement/relative clause (island violation) revealed no significant difference between the two clause types. That being said, it could be the case that the complexity of the items and the linear precedence of the parasitic gap enhanced the effect in Phillips' study, while the simplicity and canonicity of the items tested in the experiment herein depressed it. In the next section, we turn to an interim discussion before continuing the investigation of principle C reconstruction in English multi-gap dependencies.

4.1.3 Interim discussion

The current section reviewed the two fundamental requirements to proceed with the study of principle C reconstruction across English multi-gap dependencies. On the one hand, the validity and strength of principle C reconstruction in simple wh-dependencies was assessed. Employing the design used in experiment 2 in section 2.2.3 in chapter 2 on the phenomenon in German, the results of experiment 7 of this chapter are promising in that they illustrate a higher sensitivity of English participants to principle C reconstruction than of German participants. Future testing should confirm whether the difference between the languages is a result of cross-linguistic variability or the result of the improvements to the item design implemented in the English experiment, such as the use of transitive verbs taking animate agents and patients rather than psych verbs with an inanimate argument. Considering that the experiments in the next section feature the same changes, the comparison between principle C reconstruction in simple wh-dependencies and multi-gap dependencies within English is unaffected by these changes.

On the other hand, after finding the acceptability of parasitic gaps in German to be rather compromised, this section set out to establish the acceptability of parasitic gaps in English. The focus was placed on parasitic gaps in finite and

infinitival adjunct clauses, testing how finiteness interacted with the presence of a gap. Surprisingly, the experiment found no significant effect of GAP and only a marginally significant effect of FINITENESS. Crucially, the interaction between the two factors is not significant either. These results are at odds with previous reports about other cases of finiteness alternations in parasitic gap constructions, where such an interaction was indeed significant (Phillips 2006). Overall, the experiment revealed a much higher acceptability of the items in English compared to German. This prompted the conclusion that parasitic gaps in English, including the ones in finite adjunct clauses, are suitable for the study of principle C reconstruction in multi-gap dependencies. Moving forward, we can thus pursue the full comparison of principle C reconstruction into each gap in ATB as well as parasitic gap constructions. The next section turns to this matter, starting with a comparison of the reconstruction of extracted objects in ATB and parasitic gaps, then looking into the effect of principle C reconstruction in ATB movement, i.e. comparing object vs. subject extraction.

4.2 Principle C reconstruction in English multi-gap dependencies

We now turn to the investigation of principle C reconstruction in multi-gap dependencies. In chapter 2, the importance of isolating c-command as an experimental factor in the study of reconstruction was established. By doing so, relying on threshold based reasoning can be avoided, instead drawing conclusions from the presence of an contrast between conditions. The point of isolating c-command is that we can directly compare items that eliminate the confound of linear order. That is, the R-expression linearly precedes the pronoun across conditions, the only difference being whether the pronoun c-commands the base position of the R-expression or not. This can be achieved by comparing subject vs. object extraction—if the object of a monotransitive clause is questioned and the subject is a pronoun, we should observe a principle C effect if there is reconstruction; if the subject of the clause is questioned and the pronoun is the object, there should be no effect, regardless of reconstruction.

This logic allowed us to quantify the principle C reconstruction effect in German in chapter 2 more precisely than previous experiments, paving the way for turning to more complex cases in chapter 3 as well as English in the current chapter. While no comparison between ATB and parasitic gaps was drawn in German due to the extremely variable and rather marginal status of the latter, experiment 8 on English parasitic gaps revealed that such a comparison should yield interpretable results in

ATB dependencies	PG dependencies
ATB, init, obj	PG, init, obj
ATB, noninit, obj	PG, noninit, obj
ATB, init, subj	*PG, init, subj
ATB, noninit, subj	*PG, noninit, subj

Table 4.4: Fully crossed (optimal) design including manipulations of the factors CONSTRUCTION, POSITION, and GRAMFUNC.

the language. That is, the experiment(s) carried out on ATB in German in chapter 3 should be applicable not only to ATB but also parasitic gaps in English. There is one pitfall, however: The fully crossed design shown in Table 4.4 comparing the construction type (ATB or parasitic gap), the position of the pronoun as well as the grammatical function of the extractee, i.e. the effect of principle C reconstruction, cannot be pursued due to the anti c-command condition on parasitic gaps. That is, in parasitic gap dependencies, the main gap may not c-command the parasitic gap. In the object extraction conditions, this is not a problem since the object gap in the main clause does not c-command the object gap in the adjunct clause. However, in the subject extraction conditions, the subject gap in the main clause does c-command the subject gap in the adjunct clause.³ Thus, the configuration that would be necessary for the clear interpretability of the principle C reconstruction effect in parasitic gap constructions is ruled out on independent grounds.

The examples in (6) illustrate the attempt to form a parasitic gap construction

³Generally, the problem associated with parasitic gaps licensed by subject extraction is that the subject licensing gap c-commands the subject parasitic gap. This has been argued to be about anti c-command since subjects can license parasitic gaps per se, as shown by Engdahl (1983, pp. 21–22) based on the following examples. The extracted element serves as the subject of the main clause and the object of the adjunct clause, but the licensing subject gap does not c-command the parasitic object gap due to the attachment site of the adjunct clause being at a higher VP:

- (i) a. Which caesar did Brutus imply ___ was no good while ostensibly praising ____{pg} ?
b. Who did you say John's criticism of ____{pg} would make us think ___ was stupid?

Crucially, for the paradigm in Table 4.4 to work, we need cases where the parasitic gap itself takes on the function of the subject, regardless of the grammatical function of the licensing gap. These cases should likewise be ruled in by the anti c-command condition as regardless of the attachment site of the adjunct clause, the object gap cannot c-command out of its VP. Engdahl's example above indeed seems possible even if we flip the grammatical functions of the gaps:

- (ii) Which caesar did Brutus ostensibly praise ___ while implying ____{pg} was no good?

However, constructing examples akin to the conditions tested in the experiments herein yields an unacceptable outcome:

- (iii) a. *Which cousin of Max's did he greet ___ before ____{pg} hugged Mary?
b. *Which cousin of Max's did Mary greet ___ before ____{pg} hugged him?

What is at stake here is a violation of anti-locality, see Davis (2025). Another level of embedding would make this configuration testable, albeit rather complex to judge for naive participants.

via subject extraction. In (6a) and (6b), we see the attempt with a finite adjunct clause. Since *which cousin of Max's* is the subject of both the main and the adjunct clause, the constructions are ruled out. It is important to note that the counterparts with infinitival adjunct clauses are acceptable for independent reasons, namely that infinitival adjunct clauses in English do not introduce new subjects. That is, while the reading we get corresponds to the intended reading in (6a) and (6b), the reason that the sentences are acceptable is that their underlying structure is different. Because no new subject can be introduced, the subject of the adjunct clause is PRO, controlled by the subject of the main clause. Thus, the underlying structure does not contain a parasitic gap in the first place.

- (6) a. *Which cousin of Max's ____ greeted him before _____{pg} hugged Mary?
 b. *Which cousin of Max's ____ greeted Mary before _____{pg} hugged him?
 c. Which cousin of Max's ____ greeted him before PRO hugging Mary?
 d. Which cousin of Max's ____ greeted Mary before PRO hugging him?

As such, while the use of finite adjunct clauses allows us to test how participants respond to cases where we predict reconstruction to have an effect, i.e. the object extraction cases with a pronominal subject in the adjunct clause, there is no way to get around the restrictions imposed on subject parasitic gaps without substantially changing the structure of the items across conditions. The current section is thus divided into two parts. In the first part, we look at the effects of CONSTRUCTION and POSITION while keeping the level of GRAMFUNC constant at *object*. That is, experiment 9 in section 4.2.1 investigates whether there is a difference between the two gaps in multi-gap dependencies when a pronoun c-commands either gap site, and moreover, whether the relationship between the two gaps in ATB constructions differs from that of the two gaps in parasitic gaps dependencies. Experiment 10 in section 4.2.2 investigates whether the pattern can be traced back to a principle C reconstruction effect by looking at the effects of GRAMFUNC and POSITION, keeping the level of CONSTRUCTION constant at *ATB*.

Ultimately, the experiments will not give us total certainty about the effect of principle C reconstruction in parasitic gap dependencies due to the impossibility to test the conditions involving subject extraction. However, the evidence from the two following experiments allows us to realistically evaluate the likelihood of reconstruction to the parasitic gap site. Recall that the parasitic gap site is most often argued not to show reconstruction effects with the exception of SCO. Indeed, in prior empirical reports and theoretical proposals, there is more robust evidence for a syntactic link between the filler and the non-initial gap in ATB dependencies than between the filler and a parasitic gap. That is, while reports and theories vary with respect to whether this link is stronger in ATB or equal in both constructions,

as far as I am aware, no data illustrate the reverse—no phenomenon suggests that the link between the filler and the parasitic gap is stronger than between the filler and the non-initial gap in ATB. As such, finding a robust principle C reconstruction effect in the non-initial gap of ATB dependencies would be less informative than finding no effect of reconstruction. If there is no convincing evidence for principle C reconstruction in the non-initial gap of ATB dependencies, it is highly unlikely that this evidence could be found in the parasitic gap site given an adequate configuration to do so. Although drawing inferences is not as valuable as collecting data first hand, moving forward with the described logic is the next best thing to do given the aforementioned theory-internal limitations.

4.2.1 Experiment 9: Difference between ATB and parasitic gap constructions

Because of the limitations on parasitic gap items where the base position of the filler c-commands the pronoun, experiment 9 looks at a subset of 4 out of the 6 conditions shown to be possible in Table 4.4. The current experiment aims to answer the question whether ATB and parasitic gap constructions differ with respect to the coreference pattern they elicit based on whether the pronoun precedes the initial or the non-initial gap.

4.2.1.1 Design

The experiment manipulates two factors with two levels each: CONSTRUCTION, i.e. whether the presented item involves an ATB or a parasitic gap construction, and POSITION, i.e. whether the pronoun whose referent is assessed precedes the initial or non-initial gap. Given the anti c-command condition on parasitic gaps, the factor GRAMFUNC is held constant at *object*. Therefore, the experiment features 4 conditions in a 2×2 Latin Square design with each participant seeing each item in only one of the conditions.

4.2.1.2 Materials

Participants saw 12 experimental items in 6 conditions given that the experiment shares the materials of the experiment 10 in section 4.2.2, and that we are looking at the subset of 4 conditions relevant for the current experiment, turning to another subset of 4 conditions in the next experiment. Target items were interrogative sentences presented without context, see (7). Sentences contained an extracted filler with a PP that contains an R-expression modifying it and either an ATB or parasitic gap dependency. Both verbs in the respective dependency are transitive,

taking an animate agent and patient argument. Parasitic gap dependencies involved finite temporal adjunct clauses introduced by *before*, just like the items testing the acceptability of parasitic gaps in experiment 8 in section 4.1.2.

- (7) a. Which cousin of Max’s did he greet and Mary hug? *ATB,init,obj*
b. Which cousin of Max’s did Mary greet and he hug? *ATB,noninit,obj*
c. Which cousin of Max’s did he greet before Mary hugged? *PG,init,obj*
d. Which cousin of Max’s did Mary greet before he hugged? *PG,noninit,obj*

4.2.1.3 Method, participants and procedure

The experiment used the same method as the other experiments discussed in this chapter, see section 4.1.1.3 for details. Participant details are given in section 4.1.1.4, the procedure details in section 4.1.1.5.

4.2.1.4 Predictions

A significant interaction of CONSTRUCTION and POSITION is expected if there is a difference in the relationship of the two gaps in an ATB dependency compared to the two gaps in a parasitic gap dependency. If this is the case, the overall response pattern will most likely indicate a symmetrical coreference pattern in ATB and an asymmetric pattern in parasitic gaps, manifesting in a higher coreference rate in the condition *PG,noninit,obj* compared to all other conditions. On the other hand, if the pattern is asymmetric in both constructions, this should manifest in a significant main effect of POSITION—the construction type would not matter in this case, only the position of the pronoun. Coreference rates should be similarly low in the conditions *ATB,init,obj* and *PG,init,obj*, while they should be similarly high in the conditions *ATB,noninit,obj* and *PG,noninit,obj*. If the pattern is symmetric in both constructions, none of the effects should be significant and all conditions should elicit roughly equal coreference rates. A significant main effect of CONSTRUCTION alone is not predicted by either theory. It would indicate a symmetric pattern within constructions, but crucially such that there is reconstruction to both gaps in one construction type and no reconstruction to either gap in the other. Given the observations in prior experiments using this design, we expect participants to choose the coreferent reading, i.e. the reading ruled out by principle C, around chance level due to the lack of an alternative referent.

4.2.1.5 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using a generalized linear mixed effects model for binary data from the package ‘lme4’ (D. Bates et al.

2015) with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’. Modeling was carried out only for the forced choice task. A conservative α -level of 0.05 was defined. The model in (8) includes the fixed effects of CONSTRUCTION and POSITION as well as their interaction, and a random effects structure with varying intercepts and slopes for the fixed effects and their interaction for participants and items (Barr et al. 2013).

$$(8) \quad \text{reference} \sim \text{construction} * \text{position} + (1 + \text{construction} * \text{position} \mid \text{item}) \\ + (1 + \text{construction} * \text{position} \mid \text{participant})$$

Participants’ attention was assessed as described in section 4.1.1.7.

4.2.1.6 Results

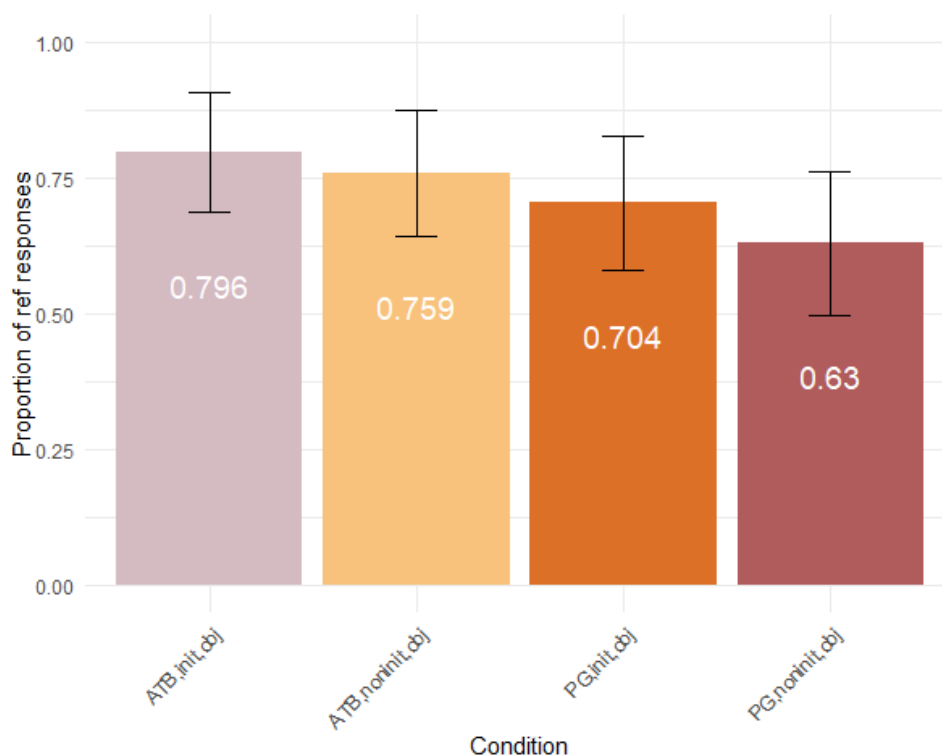


Figure 4.6: Overall coreference rates across conditions. Error bars indicate standard error.

Figure 4.6 summarizes the overall coreference rates found across conditions. We see that coreference rates are lower in non-initial conditions than initial conditions, which is indicative of some asymmetry between the gaps in both constructions, but crucially in the opposite direction than in the experiments on German in chapter 3 and previously reports from the literature for English. While coreference should be more widely available in *non-initial* condition based on introspective reports, the experiment found the reverse, i.e. participants choosing coreference less frequently in the *non-initial* condition. Moreover,

GLMM	Estimate (SE)
(Intercept)	-2.45*** (0.50)
position	1.06* (0.46)
construction	0.93 (0.51)
position:construction	-0.22 (0.60)
AIC	508.78
BIC	606.42
Log Likelihood	-230.39
Num. obs.	432
Num. groups: participant	54
Num. groups: item	12
Var: participant (Intercept)	4.34
Var: participant position	2.15
Var: participant construction	0.22
Var: participant position:construction	0.38
Var: item (Intercept)	0.77
Var: item position	0.18
Var: item construction	1.20
Var: item position:construction	1.08

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.5: Estimates of the generalized linear mixed effects model for experiment 9, including fixed effect for POSITION, CONSTRUCTION, and the interaction of the two, as well as random intercepts and slopes for items and participants.

coreference rates are generally lower in the parasitic gap conditions than the ATB conditions, which, again, is entirely unexpected. The overall proportions of coreference rates found in this experiment are couched between the rates found in the two conditions of the baseline experiment on simple wh-dependencies in section 4.1.1.

Significance testing revealed a significant main effect of POSITION and a marginally significant main effect of CONSTRUCTION. The interaction of the two effects is not significant. The model estimates are reported in Table 4.5. The results are in line with the view that reconstruction to both constructions is asymmetric, indicated by the significant effect of POSITION. However, taken together with the observed coreference rates, the direction of the effect goes against the asymmetrical data reported in the literature. That is, if the results are to be interpreted as a genuine principle C reconstruction effect, the lower coreference rate in the non-initial conditions would indicate a stronger reconstruction effect to the non-initial gap than the initial gap, which would translate to a stronger

4.2. Principle C reconstruction in English multi-gap dependencies

syntactic link between the non-initial gap and the filler than the initial gap and the filler. This pattern would go against all existing reports.

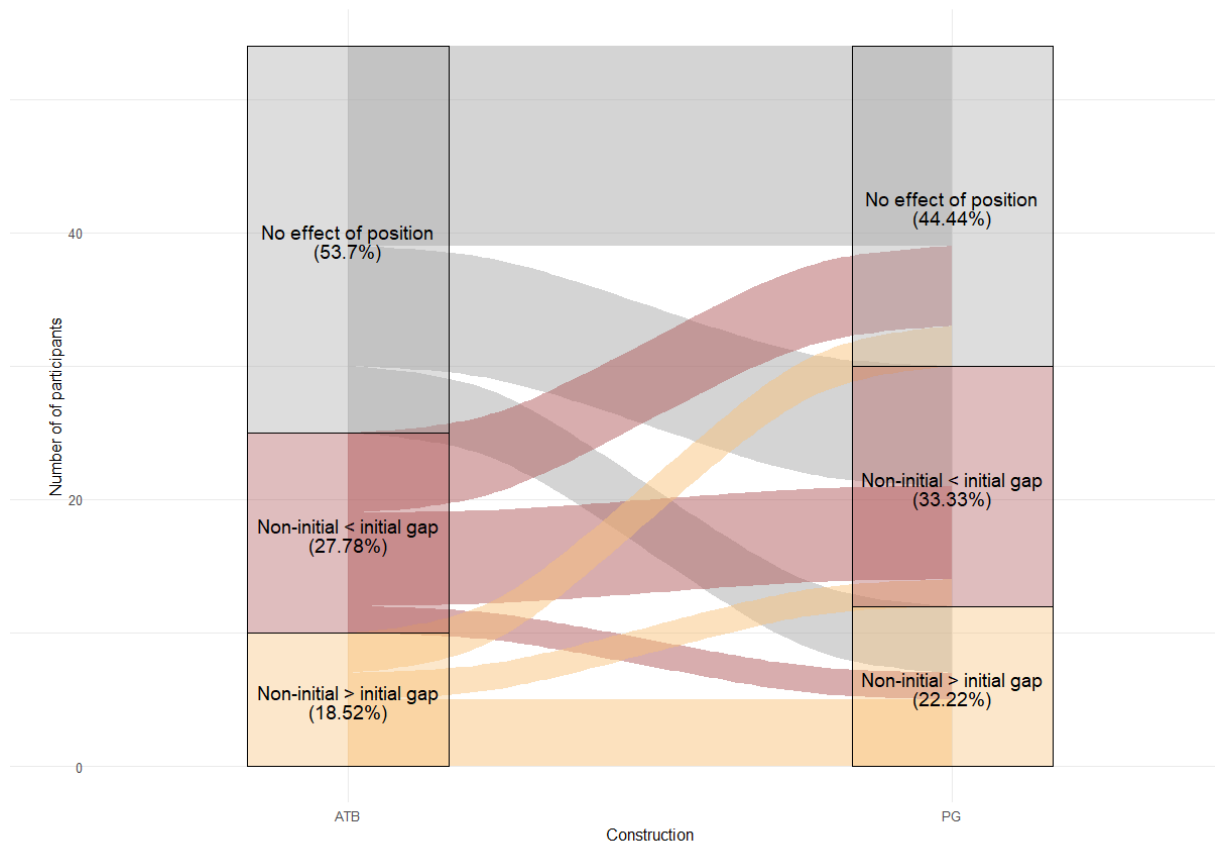


Figure 4.7: Participant migration between effect groups of POSITION in ATB and PG dependencies. Flow indicates change in individual participants' response pattern from ATB conditions to PG conditions.

Given this hitherto unattested pattern, it is once again important to clarify how much participants differ with respect to the effect. Figure 4.7 illustrates participant groups based on the effect of POSITION. The proportion of participants showing no effect is greater in the ATB conditions than the parasitic gap conditions. In turn, the groups showing an effect at all in either direction is greater in the parasitic gap conditions; however, the relative size of each group is roughly equal in both construction types. The figure illustrates that the overlap between the groups across constructions is limited. That is, the effect of POSITION is the same in both construction types for only half of participants, with the other half changing effect groups between construction types. The effect is most likely a mixture of multiple factors, resulting in a non-uniform pattern both across and within participants.

4.2.1.7 Discussion

The experiment found evidence for a difference between the initial and non-initial gap across dependency types, which manifests in higher coreference rates in initial conditions. The evidence for a difference between ATB and parasitic gap dependencies is limited. Although coreference rates are lower in parasitic gap dependencies based on the raw proportions, the effect of CONSTRUCTION is only marginally significant. Furthermore, there seems to be an effect of construction complexity on coreference rates when comparing the results of this experiment to the findings in the baseline study in simple wh-dependencies, i.e. experiment 7 in section 4.1.1. While the coreference rate in the object extraction condition of experiment 7 was found to be at 59.7%, it increased in all conditions in this experiment (with an increase of 3.3-19.9% depending on condition). This could be interpreted as a stronger bias to resolve pronominal reference with any referent available the more complex the dependency is. The increased coreference rate in initial compared to non-initial conditions may follow from a superficial effect of distance between the R-expression and the pronoun, though contrary to previous experiments, increased distance seems to negatively affect the coreference rate.

Crucially, the patterns found in this experiment are quite different from previous reports on principle C reconstruction in multi-gap dependencies. While significance testing revealed a pattern supporting approaches treating the two construction types on a par, the direction of the effect of POSITION is at odds with previously reported data, whether the source is introspection, the experiment conducted by Bruening and Al Khalaf (2017) or the experiments on German ATB movement reported in chapter 3. This begs the question whether something else may have influenced participants in the rating they gave. It is plausible that requiring a response to two independent tasks within a single trial was the problem. More precisely, it is quite likely that participants did not rate the acceptability of the sentence independently of the coreference judgment, instead choosing to rate the acceptability of the respective reading they chose, or judging the reading of the sentence based on how acceptable the sentence was. In whichever way, the two tasks may have accidentally bled into one another.

The ranking of conditions based on coreference rates indeed overlaps with the ranking of conditions based on mean acceptability ratings. ATB conditions were judged to be more acceptable than parasitic gap conditions, and initial conditions within construction types were rated higher than non-initial conditions. The parasitic gap conditions may be disfavored for syntactic reasons, considering the raw mean ratings in the experiment reported in section 4.1.2 on the acceptability of parasitic gaps in finite vs. infinitival clauses.

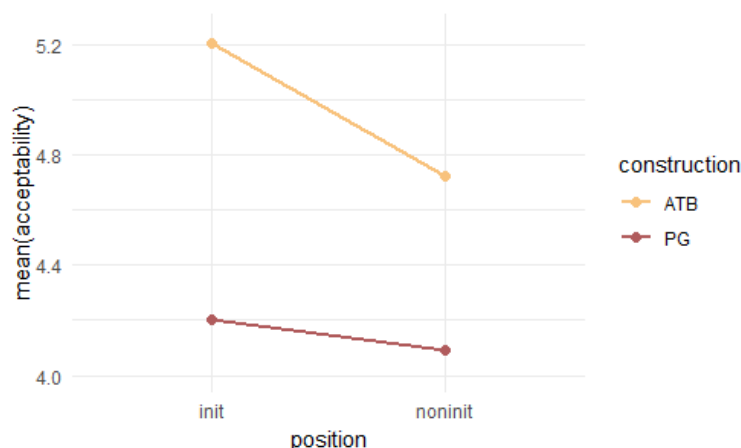


Figure 4.8: Mean acceptability ratings based on levels of CONSTRUCTION and POSITION.

However, there is no syntactic reason for the non-initial conditions to be judged worse than the initial ones, and even considering pragmatics, it is difficult to find a cause for this pattern. Figure 4.8 illustrates the mean acceptability rating for each condition. Compare this to Figure 4.6, showing a decrease in coreference rates in the order from conditions

rated the most to least acceptable. Potential problems surrounding the methodology are explored in detail in section 4.3. First, we turn to investigating the involvement of principle C reconstruction by looking at the effect of GRAMFUNC and its interaction with POSITION within ATB items in the next section.

4.2.2 Experiment 10: Principle C reconstruction in ATB dependencies

The following experiment sets out to explore the effect of principle C reconstruction under ATB movement in English. The aim of this experiment is to assess how much of the difference between the two gaps observed in experiment 9 in section 4.2.1 can be attributed to an underlying principle C violation. Therefore, we are again focusing on a subset of 4 out of the 6 conditions shown to be possible in Table 4.4 in section 4.2.

4.2.2.1 Design

The experiment has a 2×2 Latin Square design which corresponds exactly to the design of the experiments on German ATB movement reported in chapter 3. The factor GRAMFUNC manipulates whether the filler is the object or the subject of the clause, modulating the presence of a principle C reconstruction effect. The factor POSITION manipulates the position of the pronoun, testing the effect in the initial or the non-initial conjunct, respectively.

4.2.2.2 Materials

Participants saw 12 experimental items in 6 conditions given that the experiment shares the materials of experiment 9 in section 4.2.1. This time, we are looking at a different subset of conditions. Target items were interrogative sentences presented without context, see (9). The filler was either the subject or object of the sentence, with a PP that contained an R-expression modifying it, and all sentences contained an ATB dependency. Both verbs are transitive, taking an animate agent and patient argument.

- (9) a. Which cousin of Max's did he greet and Mary hug? *ATB,init,obj*
 b. Which cousin of Max's did Mary greet and he hug? *ATB,noninit,obj*
 c. Which cousin of Max's greeted him and hugged Mary? *ATB,init,subj*
 d. Which cousin of Max's greeted Mary and hugged him? *ATB,noninit,subj*

4.2.2.3 Method, participants, and procedure

The experiment used the same method as the other experiments discussed in this chapter, see 4.1.1.3 for details. Participant details are given in section 4.1.1.4, the procedure details in section 4.1.1.5.

4.2.2.4 Predictions

A significant interaction of `GRAMFUNC` and `POSITION` is expected if there is evidence for principle C reconstruction into one gap, but not the other. If this is the case, the pattern is most likely expected to favor coreference in non-initial gap conditions, indicating a reconstruction effect to the initial gap only. If we find a significant main effect of `GRAMFUNC` alone, this would indicate symmetrical reconstruction to both gaps. A significant main effect of `POSITION` alone would indicate an effect of distance between the pronoun and the R-expression regardless of reconstruction. A significant main effect of `POSITION` in addition to `GRAMFUNC` would indicate that the distance between the pronoun and the R-expression has an effect on coreference in addition to the effect of principle C reconstruction in both conjuncts. In the presence of an effect of `GRAMFUNC`, whether it is a main effect or part of the interaction, we expect lower coreference rates in subject than object conditions. Based on previous experiments in German, lower coreference rates are expected in initial than in non-initial conditions if there is a significant effect of `POSITION`, indicating a positive effect of distance on coreference between the pronoun and the R-expression. However, given that we observed the opposite pattern in experiment 9 in section 4.2.1 for the object conditions, the pattern should surface in the subject conditions as well. This would further support the

assumption that the task had an unwanted effect on the outcome.

4.2.2.5 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using a generalized linear mixed effects model for binary data from the package ‘lme4’ (D. Bates et al. 2015) with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’. Modeling was carried out only for the forced choice task. A conservative α -level of 0.05 was defined. The model in (10) includes the fixed effects of GRAMFUNC and POSITION as well as their interaction, and a random effects structure with varying intercepts and slopes for the fixed effects and their interaction for participants and items (Barr et al. 2013).

$$(10) \quad \text{reference} \sim \text{gramfunc} * \text{position} + \\ (1 + \text{gramfunc} * \text{position} \mid \text{item}) + (1 + \text{gramfunc} * \text{position} \mid \text{participant})$$

Participants’ attention was assessed as described in section 4.1.1.7.

4.2.2.6 Results

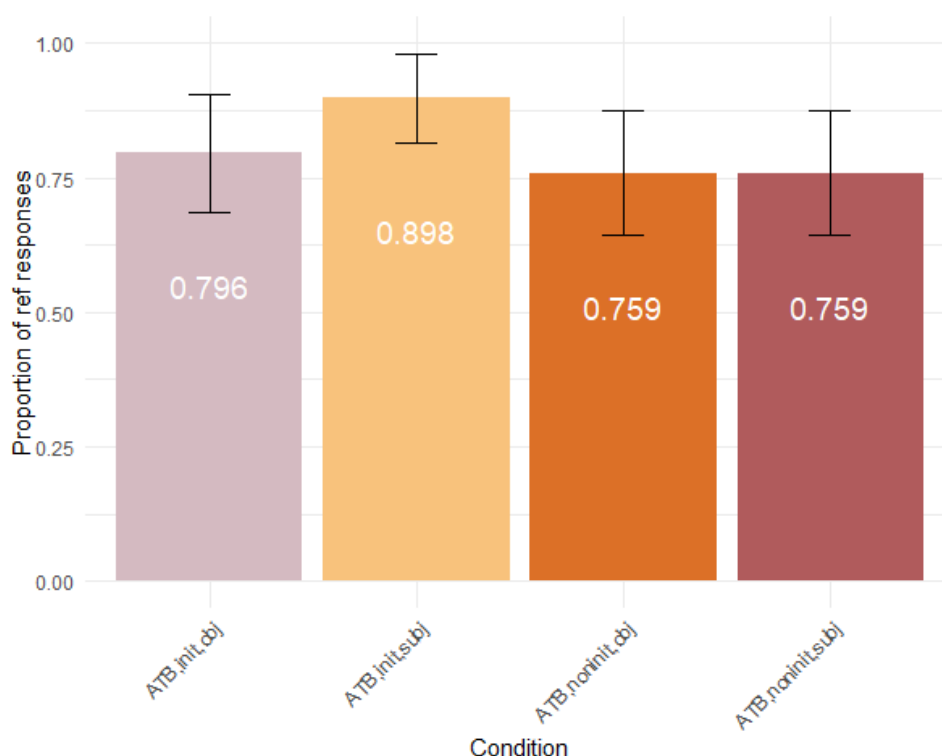


Figure 4.9: Overall coreference rates across conditions.

The overall coreference rates are shown in Figure 4.9. Coreference rates across conditions are once again fairly high above chance, between 75.9% and 89.8%. The initial conditions show a pronounced difference according to the

GLMM	Estimate (SE)
(Intercept)	-5.26*** (0.98)
position	3.42** (1.13)
gramfunc	2.90** (1.04)
position:gramfunc	-2.52* (1.13)
AIC	418.09
BIC	515.73
Log Likelihood	-185.05
Num. obs.	432
Num. groups: participant	54
Num. groups: item	12
Var: participant (Intercept)	14.67
Var: participant position	9.02
Var: participant gramfunc	13.81
Var: participant position:gramfunc	10.52
Var: item (Intercept)	0.08
Var: item position	4.16
Var: item gramfunc	0.60
Var: item position:gramfunc	2.05

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.6: Estimates of the generalized linear mixed effects model for experiment 10, including fixed effects for GRAMFUNC, POSITION, and the interaction of the two, as well as random intercepts and slopes for items and participants.

levels of GRAMFUNC, eliciting a higher coreference rate in the subject than the object condition as previously hypothesized to occur given the effect of principle C reconstruction. Notably, this contrast vanishes in the non-initial conditions. Like in experiment 9 reported in section 4.2.1, here, too, the overall coreference rate is higher in the initial conditions than non-initial conditions, which goes against the results of the experiments on German and every previous data report in the literature.

Significance testing revealed significant main effects of GRAMFUNC and POSITION as well as the interaction of the two. The estimates for the random effects display an enormous degree of variance for virtually all components by participant, and a notably high degree of variance for the effect of POSITION and the interaction of the fixed effects by item. A closer look is taken at variability between participants in the following, starting with the effect of GRAMFUNC in each conjunct depicted in Figure 4.10. The vast majority of participants, i.e. 64.8% and 55.6% do not show any effect of GRAMFUNC in the initial and non-initial conjunct,

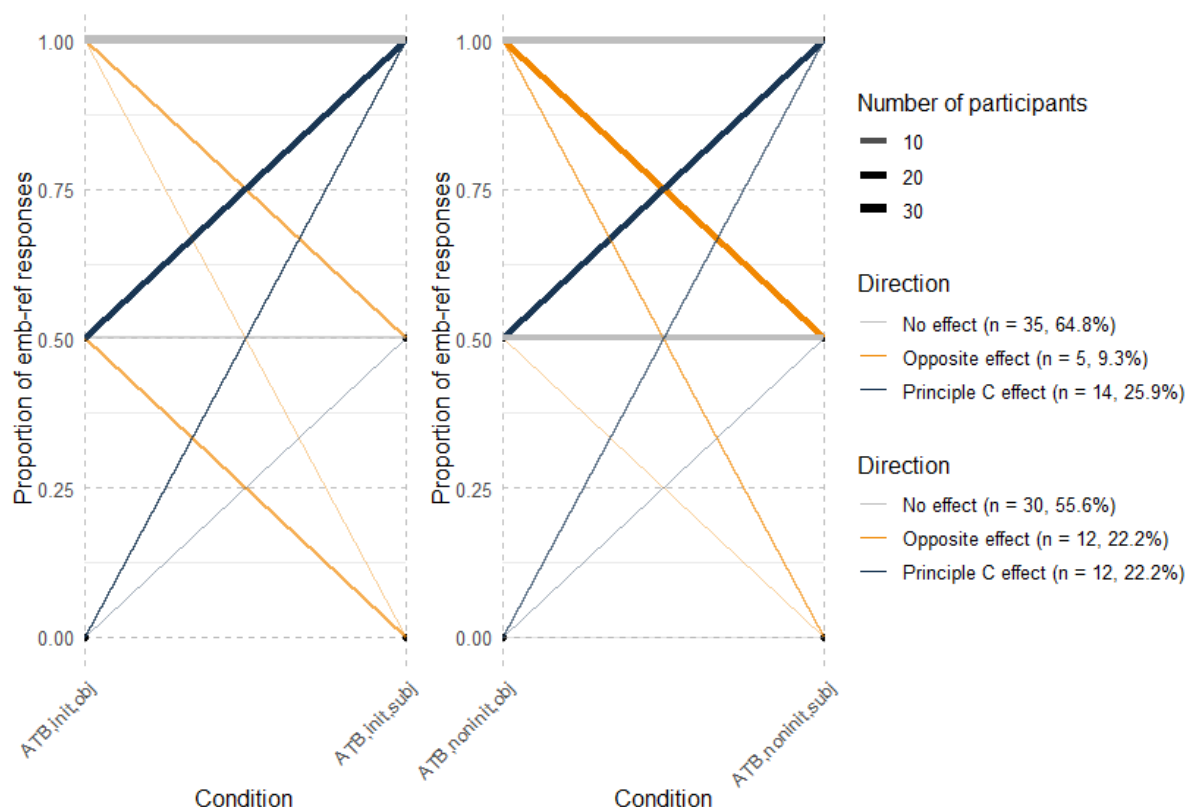


Figure 4.10: Individuals' ($n = 54$) overall proportions of responses indicating coreference with the embedded referent by condition. Y-axis indicates proportion (0-1 by 0.5 based on 2 observations per participant per condition), lines connect individual participants' proportions in the object and subject conditions of the initial (l) and non-initial conjunct (r), respectively. Color indicates the direction of the effect (blue: subject > object, orange: subject < object, gray: subject = object), thickness and opacity indicate the number of participants. X-axis indicate overall percentage of yes responses in condition.

respectively. The proportion of participants showing the opposite of a principle C effect, i.e. a higher coreference rate in the respective object than subject condition, is fairly low at 9.3% in the initial conjunct, increasing to 22.2% in the non-initial conjunct. The proportion of participants responding in line with the predictions of a principle C effect is low but fairly stable across conjuncts, at 25.9% in the initial and 22.2% in the non-initial conjunct. That is, responses are more variable in the non-initial than the initial conjunct. The proportion of participants responding in line with principle C reconstruction is considerably lower in this experiment than the experiment on simple wh-dependencies in English, i.e. experiment 7 in section 4.1.1. These proportions are also lower than in the comparable German experiment reported in 3.2.4.

Turning to the investigation of speaker profiles, Figure 4.11 shows the migration of participants between the effect groups (no effect of GRAMFUNC; opposite effect, i.e. more coreference under object than subject extraction; and predicted

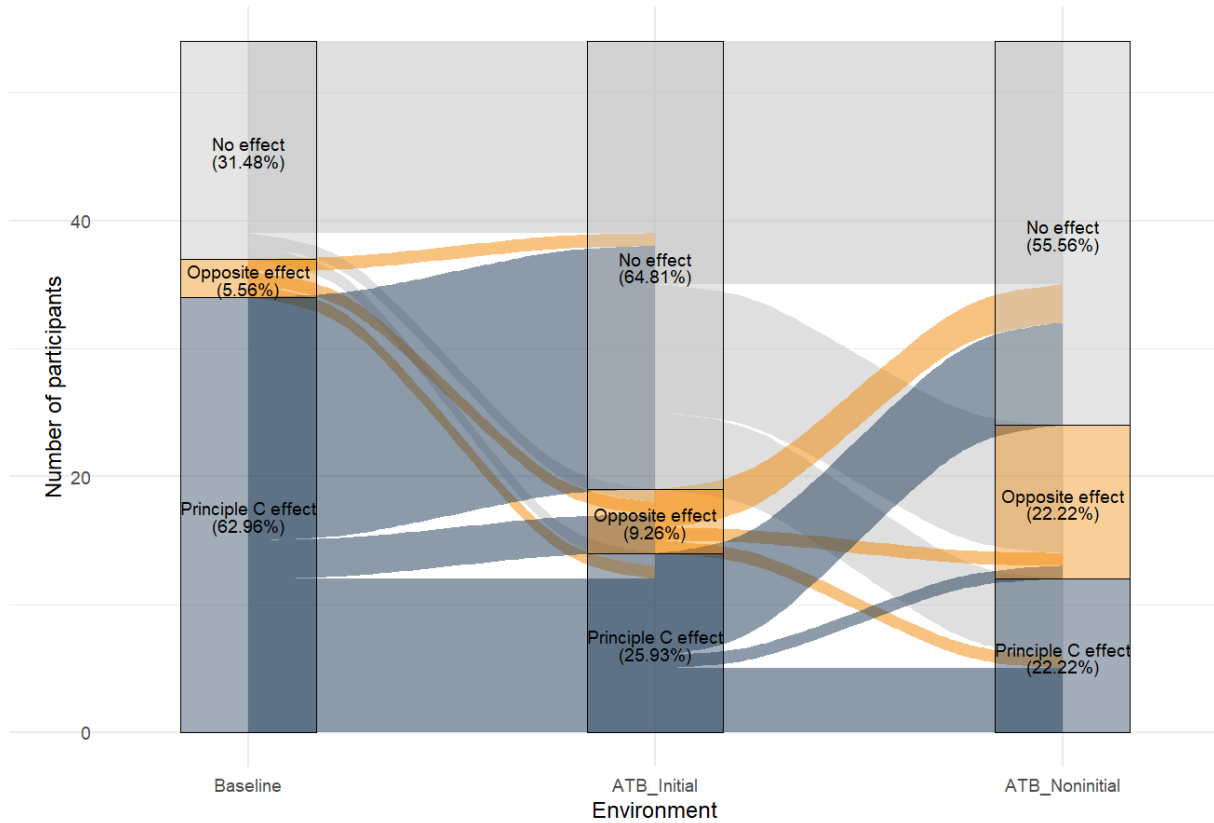


Figure 4.11: Proportion of participants corresponding to effect groups across different test environments (baseline = experiment in section 4.1.1 on simple wh-dependencies, ATB_Initial = initial conjunct of ATB movement in current experiment, ATB_Noninitial = non-initial conjunct of ATB movement in current experiment). Flow indicates migration of individual participants from one effect group to the other.

effect, i.e. more coreference under subject than object extraction). Participants' behavior is compared across the different syntactic environments tested in this series of experiments, with the simple wh-dependencies tested in experiment 7 of section 4.1.1 serving as the baseline compared to the conjuncts in ATB. As noted, Figure 4.10 revealed that the proportion of participants showing an effect in the predicted direction does not differ drastically among conjuncts of ATB. The crucial question concerns the overlap between the groups showing the effect in either conjunct compared to simple wh-dependencies. We see that 62.96% of participants responded in line with the predictions of a principle C reconstruction effect in simple wh-dependencies, and that the majority of participants showing the effect in either conjunct belongs to this group. Of the 14 participants showing the effect in the initial conjunct, 12 do so in the baseline as well; while of the 12 participants showing the effect in the non-initial conjunct, 9 do so in the baseline as well. Importantly, however, only 5 participants show the effect consistently in all three syntactic environments. Not a single participant exhibits the effect in both

conjuncts of ATB without also showing it in the baseline. Based on Figure 4.11, most participants who did show the effect in simple wh-dependencies migrate to the group showing no effect under ATB movement.

4.2.2.7 Discussion

The experiment revealed remarkably high coreference rates across conditions and particularly in the condition *ATB, initial, subject*. Although we know that the employed setup used in the experiment has a particularly dramatic effect on coreference rates, increasing them consistently compared to other setups used in chapters 2 and 3, it is nevertheless notable that the coreference rate is above 70% across all conditions. This supports the view that any effect of principle C reconstruction gets even smaller with increased dependency complexity.

The factor *GRAMFUNC* has a significant effect on the coreference rates, and so does its interaction with *POSITION*. As such, the observations are most in line with a theory of ATB movement where there is no reconstruction to the non-initial gap. However, one has to consider the vast inter-speaker variability present in all of the conducted experiments on multi-gap dependencies so far. This experiment marks the lowest proportions of participants responding in line with the predictions of principle C recorded across the experiments reported herein, i.e. 25.93% and 22.22% for the initial and non-initial conjunct, respectively. That is, while significance testing did reveal a significant effect, it is noteworthy that this effect plays a very non-uniform role across participants.

There are several aspects of this series of experiments that deserve further clarification. The next section presents two follow-up investigations. The first one aims to dig deeper and clarify what type of distance exactly gives rise to the effect of *POSITION* across experiments, while the second addresses concerns about the method employed in the experiments of this chapter.

4.3 Follow-up investigations

With the exception of experiment 5, the main effect of *POSITION* is significant across the experiments on multi-gap dependencies presented in this thesis, both in English and in German. That is, whether the pronoun precedes the initial or non-initial gap has a significant effect on the observed coreference rates, regardless of the effect of *GRAMFUNC* and thus principle C reconstruction. While it is clear that the distance between the R-expression and the pronoun has an effect on its own, it is rather unclear what exactly this effect is based on—it could either be the linear or the structural distance between the R-expression and the pronoun.

Experiment 11 addresses this question.

Furthermore, in experiments 9 and 10 in section 4.2, the coreference rate is lower in the non-initial than the initial conditions. This is at odds with the findings of the experiments on German in chapter 3 and differs from previous data reports from the literature as well. The unexpected pattern solidifies the suspicion that the method may have led to unwanted effects on the coreference rates. Section 4.3.2 addresses this issue, taking into account experiments 9, 10 and 11 of this chapter.

4.3.1 Experiment 11: The effect of distance on coreference

In this section, we turn to how the placement of the adjunct clause in parasitic gap constructions influences the coreference pattern, aiming to clarify what the effect of POSITION that is observed across experiments is rooted in. So far, the working hypothesis has been that the effect of POSITION can be traced back to the distance between the R-expression and the pronoun. However, little has been said about the exact type of distance, whether it is structural or purely linear. In chapter 2, it was discussed that previous authors studying principle C reconstruction in simple wh-dependencies found that both types of distance facilitated coreference, i.e. both the increase of linear distance as well as the increase of level of embedding, such as under long \bar{A} -movement. Crucially, however, structural distance in these cases is by default confounded with linear distance: while linear distance does not equal structural distance (11a), structural distance does equal linear distance in the case of long \bar{A} -movement (11b).

- (11) a. [Which statue of Barack_i in Michelle's study] does he_i dislike ____?
b. [Which statue of Barack_i] does Michelle think [he_i dislikes ____]?

(Adger et al. 2017, p. 6)

That is, while there is evidence showing that participants' response patterns are identical under both of these conditions, the only thing that we can conclude from this observation is that linear distance has an effect. It is impossible to tell whether structural distance alone, without the manipulation of linear distance, likewise does. The following section aims to investigate this issue by teasing apart the two types of distance, using a coreference judgment task in parasitic gap dependencies while manipulating the placement of the adjunct clause.

4.3.1.1 Design

The experiment has a 2×2 Latin Square design, manipulating the factors PLACEMENT, i.e. whether the adjunct clause is incorporated or extraposed, and

4.3. Follow-up investigations

POSITION, i.e. whether the pronoun is in the main or the adjunct clause, yielding the four conditions illustrated by (12) in the following subsection. The study uses purpose clauses introduced by *in order to* (cf. Pesetsky 1982). Since little is known about the acceptability of purpose clauses and in particular, how it interacts with their placement, a set of control items was designed. These control items were likewise distributed over a 2×2 Latin Square, manipulating the factors PLACEMENT (incorporated or extraposed) and GAP (whether there is a parasitic gap or an inanimate third person singular pronoun in place of it). Examples are given in (13) in the following subsection. The aim of these controls was to assess how participants responded to adjunct clause placement in general and whether it interacted with the presence of a parasitic gap.

4.3.1.2 Materials

Following concerns about participants disliking the non-initial conditions of the experiment reported in sections 4.2.1 and 4.2.2 for their interrogative prosody, the experiment featured embedded parasitic gaps dependencies.⁴ All experimental items started with *Everyone knows...* followed by the filler containing the PP modifier with the R-expression, followed by the matrix or adjunct clause depending on the condition (12).

- (12) Everyone knows which brother of Peter's...
- a. ...Julia called ____ in order for him to talk to ____ pg . *adj,extra*
 - b. ...he called ____ in order for Julia to talk to ____ pg . *matrix,extra*
 - c. ..., in order for him to talk to ____ pg , Julia called ____ . *adj,incorp*
 - d. ..., in order for Julia to talk to ____ pg , he called ____ . *matrix,incorp*

The 8 control items were likewise embedded dependencies, however with an R-expression as their subject rather than a quantifier. The purpose clause was either extraposed or incorporated, and either featured a parasitic gap or an inanimate third person singular pronoun.

- (13) Sally knows which bike...
- a. ...Andrew fixed ____ in order to sell ____ pg . *pg,extra*
 - b. ...Andrew fixed ____ in order to sell it. *prn,extra*
 - c. ..., in order to sell ____ pg , Andrew fixed ____ . *pg,incorp*
 - d. ..., in order to sell it, Andrew fixed ____ pg . *prn,incorp*

⁴This concern was raised by Brian Dillon (p.c.).

4.3.1.3 Method, participants, and procedure

The method corresponded to the method employed in previous experiments of this section, see 4.1.1.3. The same participants who participated in the prior experiments were contacted again via Prolific to participate in this study. Of the original $n = 54$ participants, $n = 46$ returned (mean age 45.42, $SD = 14.68$). The same participants were tested to determine complete speaker profiles across all syntactic environments: Simple wh-dependencies, multi-gap dependencies, and the manipulation of linear vs. structural order pursued in this experiment.

The eligibility of participants' data was determined based on several criteria. First, participants who responded inaccurately to distractors with gender mismatches in at least 2 out of the 6 relevant trials were excluded. Second, participants who only ever indicated disjoint reference were likewise excluded. Third, participants' with a mean acceptability rating below 4.25 for incorporated conditions, i.e. across respective target and control items, were excluded. This threshold was determined based on the decision to terminate the investigation of German parasitic gaps due to a mean acceptability rating of parasitic gap conditions of only 4.05. After excluding one participant based on the first criterion and 30 based on the last, only 15 participants remained in the sample. The procedure matched the procedure of previous experiments on English, see section 4.1.1.5 for a description.

4.3.1.4 Predictions

In target items, the focus is on coreference ratings given that we are interested in the effect of linear vs. structural distance between the referent and the pronoun on coreference between them. The predictions rely on the assumption that linear distance facilitates coreference (as found in the experiments in chapter 3 and in prior investigations by Adger et al. 2017 and Stockwell, Meltzer-Asscher, and Sportiche 2021). If the effect of distance is structural, it should hold regardless of the placement of the adjunct clause because the matrix clause is structurally closer to the filler than the adjunct clause. Distance having a positive effect, we thus expect increased coreference ratings in adjunct conditions manifesting in a significant main effect of POSITION. If the effect of distance is linear, however, the placement of the adjunct clause should interact with the pronoun's position. The ranking of coreference rates across conditions would then be as follows: *adj*, *incorp*; *matrix*, *extra*; *adj*, *extra*; *matrix*, *incorp*. That is, we expect a significant crossover interaction between POSITION and PLACEMENT, i.e. a positive effect of extraposition in adjunct conditions and a negative effect in matrix conditions. A significant main effect of PLACEMENT is not predicted.

4.3.1.5 Data analysis

Statistical modeling was carried out in R (R Core Team 2024) using a generalized linear mixed effects model for binary data from the package ‘lme4’ (D. Bates et al. 2015) with the family ‘binomial’ (logit link) and the optimizer ‘bobyqa’. Modeling was carried out only for target items and only for the forced choice task. A conservative α -level of 0.05 was defined. The model in (14) includes the fixed effects of `PLACEMENT` and `POSITION` as well as their interaction, and a random effects structure with varying intercepts and slopes for the fixed effects and their interaction for participants and items (Barr et al. 2013).

$$(14) \quad \text{reference} \sim \text{placement} * \text{position} + \\ (1 + \text{placement} * \text{position} \mid \text{item}) + (1 + \text{placement} * \text{position} \mid \text{participant})$$

4.3.1.6 Results

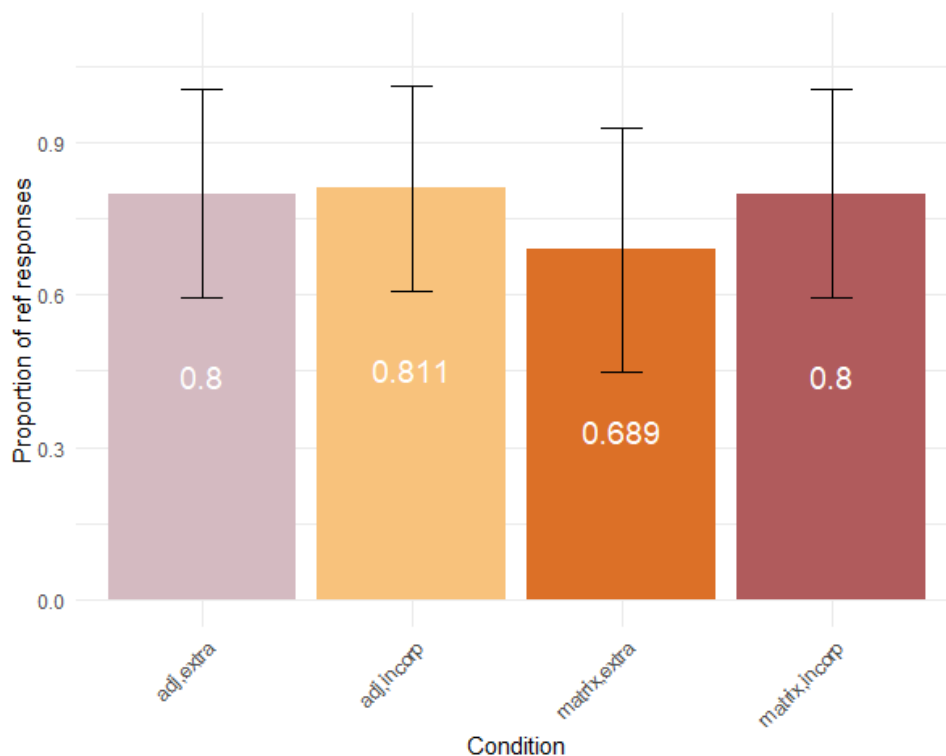


Figure 4.12: Overall coreference rate across conditions.

While the small number of participants is compensated to some degree by the relatively high number of items, the results are nevertheless to be taken with a grain of salt given that we have observed a high degree of inter-speaker variability in all experiments reported in this thesis. Figure 4.12 illustrates the overall coreference rates in the experiment. While the conditions *adjunct, extra*; *adjunct, incorp*; and *matrix, incorp* scored coreferent responses in around 80% of the respective trials, the condition *matrix, extra* did so in 68.9% of them. The fact that *matrix, extra*

stands out among the conditions is not particularly informative, as both the linear and the structural distance are the smallest in this condition. It thus merely supports the tendency observed in previous experiments that coreference is less felicitous the smaller the distance between the R-expression and the pronoun. Embedding the interrogative sentences appears to have corrected the reverse effect of distance observed in experiments 9 and 10 of this chapter (a suspicion presupposing that the acceptability judgments impacted the coreference judgments, assuming that non-initial conditions were less acceptable than initial conditions for prosodic reasons). Note that significance testing did not reveal either of the effects or their interaction to be significant, indicating that the tendency observed in the raw data is not robust.

GLMM	Estimate (SE)
(Intercept)	-2.40** (0.76)
position	1.30 (0.73)
placement	-0.96 (0.80)
position:placement	-0.18 (0.95)
AIC	340.01
BIC	433.28
Log Likelihood	-146.01
Num. obs.	360
Num. groups: item	24
Num. groups: participant	15
Var: item (Intercept)	2.05
Var: item position	3.51
Var: item placement	5.28
Var: item position:placement	9.13
Var: participant (Intercept)	5.05
Var: participant position	2.29
Var: participant placement	1.50
Var: participant position:placement	0.41

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.7: Estimates of the generalized linear mixed effects model for experiment 11, including fixed effects of POSITION, PLACEMENT, and the interaction of the two, as well as random intercepts and slopes for items and participants.

Estimates are reported in Table 4.7. Given that this experiment likewise combined a coreference judgment task with an acceptability judgment task within trials, the risk that the two tasks influenced each other is imminent. This issue is addressed in section 4.3.2.

4.3.1.7 Discussion

The results of the follow-up experiment are unfortunately orthogonal to the question about the type of distance that affects coreference between an R-expression and a pronoun. First of all, 30 out of 46 participants had to be excluded due to scoring low on the acceptability rating task for control and target items involving incorporated purpose clauses. Given the written presentation and that such items require a specific intonation, their compromised acceptability was expected. However, the exclusion of two thirds of participants due to mean ratings below 4.25 (a threshold defined based on the results of the German experiment in section 3.1) was not anticipated, leaving us with a rather small sample size. It goes without saying that the interpretability of the data is thus compromised due to the study likely being underpowered. For what it is worth, there appears to be a mild penalty on coreference in the condition where the pronoun immediately follows the R-expression, i.e. the condition where linear and structural distance coincide in being the smallest (cf. Arnold 2001). The other three conditions do not differ in overall coreference rates, and neither of the effects were found to be statistically significant. There exists the proposal that coreference between the R-expression and the pronoun is more felicitous when the pronoun is in the adjunct clause due to its backgroundedness, compared to when both are in the matrix clause (Harris and E. A. Bates 2002). However, this should lead to a decrease in coreference in both matrix conditions which is not the case. Given that no significant effect could be detected and that two thirds of participants had to be excluded from the analysis, the aim of constructing full speaker profiles across the different syntactic environments tested in this chapter could not be pursued.

This experiment likewise combined an acceptability judgment task with a coreference task within trials. Since there is no difference between the overall coreference rates unless the R-expression and the pronoun immediately follow one another, any existing effect may have further been depressed by problems associated with the method. The following section addresses the issue of how the tasks may have bled into one another, including data from experiments 9, 10 and 11 of this chapter.

4.3.2 Correlation of tasks

Experiments 9 and 10 of this chapter investigated principle C reconstruction in English multi-gap dependencies and yielded rather unexpected results. It is reasonable to assume that these odd patterns resulted from non-linguistic factors,

and this section explores this idea.⁵ In both experiments, coreference rates in non-initial conditions were found to be lower than in initial conditions, which, if principle C were at play here, would indicate a slight asymmetry with respect to reconstruction favoring the non-initial gap. Such a reconstruction pattern would suggest that the syntactic link of the filler is stronger to the non-initial gap than the initial gap. Despite the previously observed variability and weakening of the principle C effect in multi-gap dependencies, this finding is especially peculiar as it goes against all prior data reports from the literature. It likewise marks a sharp departure from the tendency observed in the German experiments of chapter 3.⁶

In addition to this, experiment 9 in section 4.2.1 found that coreference rates were generally lower in parasitic gap conditions than ATB conditions, regardless of the position of the pronoun. There is no straightforward syntactic explanation as to why this should be the case. The only viable explanation at this point is that coreference rates may have been impacted by external factors. In experiment 8 testing the fundamental acceptability of parasitic gaps in English in section 4.1.2, parasitic gaps in finite adjunct clauses turned out less acceptable compared to other conditions, despite significance testing not revealing a significant interaction between FINITENESS and GAP. Inspecting the acceptability ratings of the items in the experiment comparing ATB to parasitic gaps reveals that here, too, parasitic gap items were judged to be less acceptable than ATB items as shown in Figure 4.8. Given that the ranking of conditions by mean acceptability reflects their ranking based on overall coreference rate, an obvious hypothesis is that the scalar acceptability judgment task and the forced choice coreference judgment task have somehow bled into each other. Post-hoc modeling of the relationship between acceptability ratings and coreference judgments in both directions—i.e. whether the acceptability rating predicts the response to the coreference judgment task and vice versa—revealed a significant effect in both cases.⁷ The models and their output are shown in (15) and in Tables 4.9 and 4.8.

- (15) a. $\text{acceptability} \sim \text{reference} +$
 $(1 + \text{reference} \mid \text{participant}) + (1 + \text{reference} \mid \text{item})$
 b. $\text{reference} \sim \text{acceptability} +$
 $(1 + \text{acceptability} \mid \text{participant}) + (1 + \text{reference} \mid \text{item})$

⁵This section has immensely benefited from discussions at the Syntax Workshop at UMass Amherst and the Linguistics Colloquium at Bielefeld University.

⁶Prior piloting of an experiment on principle C reconstruction under ATB in English using the setup reported in section 3.2.2 of chapter 3 provided data in line with the German pattern, further supporting the view that the odd pattern is caused by external factors. The data and code for the pilot can be found under the following link: https://osf.io/r79tq/?view_only=93f39856964b4cf4a31444b67d45f9b5.

⁷The reported models included target trials from both experiments on English multi-gap dependencies discussed in this section, i.e. 12 items in 6 conditions.

4.3. Follow-up investigations

CLMM	Estimate (SE)
reference	0.99*** (0.22)
Log Likelihood	-1094.53
AIC	2211.06
BIC	2260.27
Num. obs.	648
Groups (participant)	54
Groups (item)	12
Variance: participant: (Intercept)	1.56
Variance: participant: reference	0.50
Variance: item: (Intercept)	0.14
Variance: item: reference	0.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.8: Estimates of the generalized linear mixed effects model reported in (15a), i.e. the predictability of acceptability ratings based on coreference ratings.

GLMM	Estimate (SE)
(Intercept)	-0.24 (0.35)
acceptability	0.38*** (0.08)
AIC	667.59
BIC	703.38
Log Likelihood	-325.79
Num. obs.	648
Num. groups: participant	54
Num. groups: item	12
Var: participant (Intercept)	0.68
Var: participant acceptability	0.02
Var: item (Intercept)	0.06
Var: item acceptability	0.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.9: Estimates of the generalized linear mixed effects model reported in (15b), i.e. the predictability of coreference ratings based on acceptability ratings.

Computing Spearman's correlations between the two measures for individual participants does not reveal a homogeneous pattern, although most participants show a moderate positive correlation, see the histogram in Figure 4.13. Thus, the unusual pattern found in the two experiments on English multi-gap dependencies is likely caused by the setup, including the possibility that participants may have

misunderstood the acceptability rating task as inquiring about the acceptability of the interpretation of the sentence that they indicated with their coreference rating, rather than the target sentence itself. Such effects could easily be avoided in the future by reserving different blocks of the experiment for each measurement.

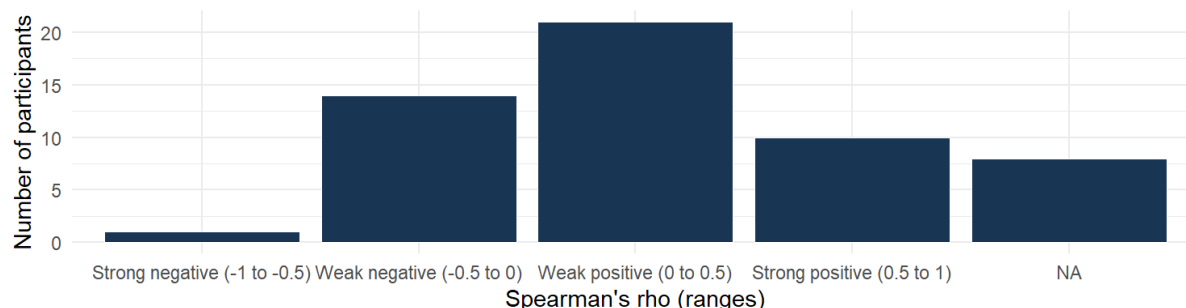


Figure 4.13: Distribution of Spearman's correlations between coreference and acceptability. Participants in the *NA* group did not vary in their responses for at least one of the measurements.

Ultimately, while parasitic gaps in finite adjunct clauses are slightly less acceptable than ATB movement, the question remains why participants rated non-initial conditions more poorly than initial conditions across the two construction types since there is no syntactic or pragmatic reason for them to do so. One tentative hypothesis is that non-initial conditions scored low acceptability ratings due to the 'hanging' nature of the non-initial domain in the interrogative sentences of the respective conditions (Brian Dillon, p.c.). That is, participants may have disliked non-initial conjuncts and adjunct clauses with pronouns for prosodic reasons (*Which cousin of Max's did Mary greet and he hug/before he hugged?*), causing the low acceptability to bleed into the coreference judgment.

Given that experiment 11 investigating the placement of the adjunct clause in parasitic gap dependencies used the same setup, the experiment could have suffered a similar fate to experiments 9 and 10 in terms of the correlation of tasks. Statistical modeling of reference based on acceptability and vice versa was carried out for target and control items jointly. Estimates are reported in Tables 4.10 and 4.11; the model definitions match those given in (15). Participants' responses to the coreference task are significant predictors of their responses to the acceptability judgment task and vice versa. In addition to the decimated number of participants, this may explain why none of the fixed effects was significant.

Finally, although the task-related confounds make it a bit more difficult to draw syntactic conclusions from the experiments, we can still summarize the findings as follows: First, the coreference pattern exhibited in initial and non-initial conditions of experiment 9 is largely identical across construction types, i.e. non-initial gaps in object conditions differ from initial gaps in object conditions in the same manner across multi-gap dependencies. Second, we found evidence

4.3. Follow-up investigations

CLMM	Estimates (SE)
reference	-0.74* (0.30)
Log Likelihood	-717.04
AIC	1456.08
BIC	1501.99
Num. obs.	480
Groups (item)	24
Groups (participant)	15
Variance: item: (Intercept)	0.25
Variance: item: reference	0.15
Variance: participant: (Intercept)	1.56
Variance: participant: reference	0.45

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.10: Estimates of the cumulative link mixed effects model fitted on target and control items of the experiment in section 4.3.1.

GLMM	Estimates (SE)
(Intercept)	0.05 (0.91)
acceptability	0.26* (0.13)
AIC	474.29
BIC	507.68
Log Likelihood	-229.14
Num. obs.	480
Num. groups: item	24
Num. groups: participant	15
Var: item (Intercept)	0.67
Var: item acceptability	0.02
Cov: item (Intercept) acceptability	-0.11
Var: participant (Intercept)	6.79
Var: participant acceptability	0.07
Cov: participant (Intercept) acceptability	-0.70

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4.11: Estimates of the generalized linear mixed effects model fitted on target and control items of the experiment in section 4.3.1.

for principle C reconstruction to the initial gap of ATB dependencies based on a significant interaction of GRAMFUNC and POSITION in experiment 10. However, this is not the result of participants not showing the contrast in the non-initial

gap—the proportions of people exhibiting the effect in each conjunct differ by only 3.7%. Rather, the effect in the non-initial gap is overshadowed by increased inter-speaker variability. Crucially, out of 54 participants tested in these experiments, 34 responded in line with a principle C reconstruction effect in simple *wh*-dependencies. Of these 34, 12 show the effect in the initial conjunct of ATB and 9 in the non-initial conjunct. There are 5 participants showing the effect across all three syntactic environments. The fact that there is a significant interaction of the factors suggesting reconstruction only to the initial conjunct is caused by an increase of the proportion of participants showing the opposite effect in the non-initial conjunct, and not because there are drastically fewer participants who do respond in line with principle C reconstruction. The experiments further demonstrate the unreliability of the principle C reconstruction test, yielding no clear outcome supporting a specific syntactic derivation. Taking into account that there are only two observations per participant for each condition, it is also quite likely that the patterns would become more stable under an increased number of repeated trials.

4.4 General discussion

This section takes stock of the findings presented in the current chapter on principle C reconstruction in English multi-gap dependencies. The first pair of experiments dealt with the foundations, principle C reconstruction itself and the acceptability of parasitic gaps in adjunct clauses based on their finiteness. Experiment 7 in section 4.1.1 confirmed that principle C reconstruction has a significant effect on the coreference rate in English simple *wh*-dependencies, revealing a significant effect of *GRAMFUNC* and a difference in overall coreference rates in the *object* vs. *subject* condition of 29.2%. Furthermore, 63% of participants responded in line with the predictions of a principle C effect, while 31.5% of participants showed no effect. Only 5.6% of participants responded in the opposite direction, indicating coreference more frequently in the *subject* than the *object* condition.

The results of experiment 7 indicate that principle C reconstruction has a stronger effect in English than in German, since in the latter, the contrast between the conditions was merely 13.7%. However, it is to be kept in mind, on the one hand, that experiment 2 on German in section 2.2.3 of chapter 2 had 16 observations per condition compared to the mere 4 observations in the matching English experiment. On the other hand, the experiment in English corrected potential confounds within items based on the use of psych verbs and inanimate extractees, using transitive verbs with two animate arguments whose theta roles could easily be flipped. Nevertheless, given that the experiments used the very

same design and both revealed a significant effect of `GRAMFUNC`, we concluded that principle C informed participants' behavior in both languages, simultaneously revealing a bias to resolve pronominal reference with whatever referent is available, as indicated especially by the above chance performance across conditions in English.

In section 4.1.2, experiment 8 found that contrary to previous reports, the finiteness of the adjunct clause alone did not have a significant effect on the acceptability of parasitic gaps. This was attributed to the simplicity of the experimental items compared to previous investigations and the order of the licensing and parasitic gaps (cf. Phillips 2006). The experiment found that parasitic gaps in adjunct clauses of English are more acceptable than their German counterparts discussed in section 3.1, with English parasitic gap conditions scoring a mean rating of 5.96 and the German ones a mean rating of 4.05. As such, inter-speaker variability was likewise much lower among English than German participants. The successful sanity check constituted by these two experiments ensured that the investigation of principle C reconstruction in English multi-gap dependencies was a fruitful endeavor.

Experiments 9 and 10 reported in sections 4.2.1 and 4.2.2 went on to extend the investigation to ATB and parasitic gaps. Recall that the investigation was split into two sub-experiments due to the impossibility of constructing parasitic gap items under subject extraction. Experiment 9 in section 4.2.1 compared object extraction in ATB and parasitic gap dependencies (`CONSTRUCTION`), varying the position of the pronoun (`POSITION`). The experiment found a significant effect of `POSITION` and a slight decrease of coreference rates in the non-initial conditions, but crucially no significant effect of `CONSTRUCTION` or the interaction of the two factors. As such, the experiment does not provide evidence for a difference in coreference patterns between ATB and parasitic gaps, merely showing a difference between the two gaps across constructions. We also saw that the effect of `POSITION` was not uniform within, let alone across participants. Although 27.78-33.33% of participants favored coreference in non-initial over initial conditions, 18.52-22.22% of participants showed the opposite preference. Furthermore, 44.4-53.7% of participants showed no effect of `POSITION` at all, and again, only around half of the participants exhibited the respective effect consistently across construction types.

Whether the difference between the two gaps within constructions could be attributed to principle C reconstruction or if it was merely a surface phenomenon was explored in experiment 10 in section 4.2.2. The effect of `GRAMFUNC` was tested by varying subject and object extraction as well as the effect of `POSITION` by varying which conjunct of ATB the pronoun was in. The experiment revealed

significant main effects of both factors as well as a significant interaction. The overall coreference rate was at 89.8% in the condition *ATB, initial, subject*, while it was 75.9-79.6% in the remaining conditions. This suggests that the principle C reconstruction effect was detectable to some degree in the initial conjunct but not in the non-initial conjunct, crucially revealing an asymmetry manifesting in lower coreference rates in both non-initial conditions. Again, the inspection of inter-individual variability is where the robustness of the pattern (or rather, the lack thereof), became evident. Only 25.93% of participants responded in line with a principle C effect in the initial conjunct and 22.22% in the non-initial conjunct. The proportion of participants responding in the opposite direction is higher in the non-initial conjunct than the initial one, manifesting in the significant interaction of $\text{GRAMFUNC} \times \text{POSITION}$. The low proportion of participants being led by principle C reconstruction under ATB movement confirmed suspicions about the limited effect of principle C reconstruction in more complex dependencies, just like we observed in German across experiments in chapter 3. However, this experiment marks a departure from the studies on German in the sense that while a very similar proportion of participants respond in line with principle C reconstruction in the baseline (i.e. simple wh-dependencies, 62.96% in English vs. 66.67% in the equivalent experiment on German), the proportion of participants doing so in ATB drops from 36.67-46.67% in German to 22.22-25.93 in English. Recall also that determining the effect groups is not based on effect *size* but merely its *direction*—i.e. any participant responding with coreference more frequently in the respective *subject* than the *object* condition is counted as showing an effect in line with principle C, even if this means doing so in only one more trial. As such, the syntactic conclusions that we can draw are quite limited.

The chapter then went on to pursue a follow-up investigation concerning the effect of POSITION found across experiments in this thesis. Experiment 11 in section 4.3.1 tested the effects of adjunct clause PLACEMENT , i.e. whether it was incorporated or extraposed, and pronoun POSITION , i.e. whether the pronoun preceded the matrix or the adjunct clause. It aimed to clarify whether the previously observed effect on coreference rates in multi-gap dependencies was caused by linear or structural distance between the R-expression and the pronoun. However, given that the majority of participants indicated incorporated adjunct clauses to be marginal to unacceptable, the data of only 15 remaining participants were analyzed. The experiment revealed no significant effect of either effect or their interaction, showing a dip in overall coreference rates in the *matrix, extra* condition where the pronoun immediately followed the R-expression both linearly and structurally.

Finally, the reason why coreference rates were lower in non-initial than initial conditions of experiments 9 and 10 was studied through post-hoc analyses in

section 4.3.2. Not only did this effect contradict previous findings from German, but also prior reports from investigations on the effect of distance in simple wh-dependencies. Notably, coreference rates were generally lower in both parasitic gap conditions than ATB conditions, giving rise to the suspicion that the reversal of the effect had to do with the experimental setup itself, i.e. with the combination of an acceptability judgment and coreference judgment task within trials. Statistical modeling revealed that in the aforementioned experiments, participants' responses to either task had a significant effect on their responses to the other. Upon suspicion that the uninterpretable results of the follow-up study on parasitic gaps suffered the same fate, such post-hoc analyses were carried out with those data, too, revealing a very similar picture. This scenario illustrates that inquiring about two independent measures within a single trial can very easily bias participants, and it unfortunately means that the data are rather uninformative.

As such, the study of principle C reconstruction in English multi-gap dependencies undertaken in this chapter proved to be an endeavor riddled with confounds, yielding no conclusive results. On the one hand, coreference rates in simple wh-dependencies suggest that principle C reconstruction has an undeniable effect, but the contrast is far less pronounced in ATB. There is a significant effect in the initial gap of ATB, with the condition *ATB, initial, object* reaching an overall coreference rate of 79.6% and the respective subject condition 89.8%. That is, although the effect is significant, it is accompanied by a strong bias to resolve coreference with any referent that is available. Such an effect was not at all detectable in the non-initial gap based on coreference rates and significance testing. Inter-individual variability is once again rampant, with 25.93% of participants responding in line with principle C reconstruction in the initial gap of ATB and 22.22% doing so in the non-initial gap (and the equal proportion of participants showing the opposite in the non-initial gap). Furthermore, the coreference pattern in parasitic gap constructions between initial and non-initial gaps is the same as in ATB, indicating no difference between constructions. Interestingly enough, POSITION also did not have a consistent effect across and not even within participants when inspecting both construction types. Together with the results of the investigation on German in chapter 3, the data strongly suggest that the principle C reconstruction diagnostic, as employed here, does not produce data reliably mapping to the underlying syntactic structure of complex movement dependencies, and as such does not yield valid results to draw conclusions about whether ATB and parasitic gaps are derived via the same mechanism, nor to conclude whether the two gaps within either construction are equally complex syntactically. The next section wraps up this chapter by exploring alternative means to approach principle C reconstruction in the future.

4.5 Beyond PP modifiers

This chapter, together with chapter 3 on German ATB movement, aimed to investigate principle C reconstruction in multi-gap dependencies to clarify whether coreference patterns overlapped within and across languages and construction types. The attempt of this thesis was to keep empirical considerations as consistent as possible to ensure that any effect of principle C reconstruction found in simple *wh*-dependencies would map onto the more complex data from multi-gap dependencies. The experiments demonstrated that the effect of principle C reconstruction with PP modifiers, which undeniably exists based on the findings of experiments 1 and 2 in chapter 2 as well as experiment 7 of this chapter, is substantially weakened in multi-gap dependencies, making the results inconclusive with respect to the underlying syntactic derivation. As such, we failed to draw conclusions about the syntax of multi-gap dependencies from the data. However, we were able to conclude that the principle C reconstruction test in its most frequent implementation does not constitute reliable evidence for far reaching theoretical claims, and as such, syntactic approaches relying on principle C reconstruction being either symmetric or asymmetric lose at least one crucial piece of evidence in their favor.

Based on the discussion in section 2.3.3 in chapter 2, it is reasonable to assume that it is the relationship between the noun phrase and its PP modifier that makes for an unstable effect, and not the relevance of principle C itself. The categories of adjuncts and arguments and their reconstruction have been subject to controversy for decades, and experiments do not deliver the clear outcomes we would expect under a purely syntactic explanation. The need for full reconstruction could simply further decrease with increased dependency complexity, neutralizing underlying principle C violations, instead allowing for the most economical interpretation of the sentence in which the pronoun refers to the (most salient) matching R-expression. The hypothesis then is that modifiers bearing a stronger semantic relationship to the noun phrase should not be able to ‘escape’ reconstruction and deliver more reliable effects in multi-gap dependencies as well. As such, future research could still exploit principle C reconstruction as a tool to learn about complex dependencies: For example, predicates are reported to yield stronger principle C reconstruction effects (Adger et al. 2017; Heycock 1995), though they do not allow for the manipulation of *GRAMFUNC* as an experimental factor. Verbal gerunds with small clause complements could be a suitable alternative (Moulton 2009, p. 51).⁸ The examples in (16) illustrate a possible paradigm.

⁸I am grateful to Kyle Johnson for pointing out these data to me.

- (16) a. Whose painting John_i nude did he_i resent ____ and Mary appreciate ____?
ATB, initial, object
- b. Whose painting John_i nude did Mary appreciate ____ and he_i resent ____?
ATB, non-initial, object
- c. Whose painting John_i nude ____ shocked him_i and ____ delighted Mary?
ATB, initial, subject
- d. Whose painting John_i nude ____ delighted Mary and ____ shocked him_i?
ATB, non-initial, subject
- e. Whose painting John_i nude did he_i resent before Mary appreciated _____{pg}?
PG, initial, object
- f. Whose painting John_i nude did Mary appreciate before he resented _____{pg}?
PG, non-initial, object

Here, again, the comparison is imperfect due to the use of distinct verbs because of animacy restrictions. However, the idea of this particular implementation of the test is that the accusative argument of the gerund (*John*) cannot escape reconstruction via Late Merger (Lebeaux 1991) or Neglect (Sportiche 2016), forcing it to be interpreted in the c-command domain of the pronoun. Coreference should be ruled out in (16a) and (16e) and possible everywhere else if principle C reconstruction is asymmetric to the initial gap in both ATB and parasitic gaps. If it is asymmetric to the non-initial gap in both constructions, which is the scenario that has not been reported before, coreference should be ruled out in (16b) and (16f). Coreference should be ruled out in all cases except (16c) and (16d) if reconstruction is symmetric across construction types. If there is a difference between the two construction types, this would most likely manifest in a symmetrical pattern under ATB, ruling out coreference in (16a) and (16b), and an asymmetrical pattern in parasitic gaps, ruling out coreference only in (16e).

Citko (2005) hypothesizes that the variability of reconstruction phenomena in ATB movement is not caused by the inherent properties of ATB movement, but by the variable robustness of reconstruction phenomena themselves. This chapter as well as chapter 3 provided support for the assumption that the variable patterns for principle C reconstruction under ATB movement and in parasitic gaps result from the compromised reconstruction status of nominal PPs modifiers. Items like the ones in (16) are promising to yield more reliable experimental results than the items with PP modifiers tested in this thesis. Aside from targeting the reliability of principle C reconstruction itself, future research on the underlying syntax of multi-gap dependencies could aim to test less controversial reconstruction phenomena, such as variable binding or crossover effects. The next step could therefore also be to test the introspective judgments from the literature

experimentally for other types of binding as well. I leave both of these issues open for future research, discussing them in the outlook provided in chapter 5.

To summarize, the principle C reconstruction test, which has been reported to give rise to controversial reconstruction patterns in ATB and parasitic gap constructions, could not be verified to map to underlying syntactic relations reliably. The claim that reconstruction in ATB and parasitic gaps only appears to be asymmetric for reasons of linear distance has been made before by Bruening and Al Khalaf (2017). Given the data reported herein, the competition of superficial effects is a likely explanation for the conflicting reports from the literature. Having identified that principle C reconstruction with PP modifiers produces data that is orthogonal to the syntactic properties of the gaps, alternative ways to address our research questions need to be sought after in the future. The final chapter of this thesis summarizes its contribution to the field and provides an outlook for possible subsequent research on the syntax of multi-gap dependencies.

Chapter 5

Synthesis and outlook

I opened up the investigation into the underlying syntax of multi-gap dependencies by presenting a number of reports from the literature disagreeing about the reconstruction patterns found in ATB and parasitic gaps in chapter 1. The chapters that followed investigated which of the patterns reported for principle C reconstruction, if any, could be reproduced in a controlled experimental setting. By gaining quantifiable information about the test's robustness, the aim was to clarify how relevant the outcome of the principle C reconstruction diagnostic is for the syntactic modeling of multi-gap dependencies. In other words, the experiments assessed to what extent the observed coreference rates can be traced back to a principle C reconstruction effect, suspecting that the variable reports from the literature are the result of confounds. The experiments considered multiple possible non-syntactic influences: We dove into chapter 2 with a study on simple wh-dependencies showing that the presence of context and alternative referents, but crucially, also the type of experimental task have a significant impact on the outcome of the diagnostic. I concluded that principle C reconstruction has a detectable effect in the absence of context and alternative referents and when using a simple experimental task associated with a large effect size, such as a single forced choice between two possible referents.

In chapter 3, experiment 3 found that parasitic gaps are not acceptable enough in German to allow for the application of reconstruction diagnostics. I then went on to apply the different experimental designs tested throughout chapter 2 only to ATB movement in German to establish how detectable the principle C reconstruction effect was in complex dependencies. The results demonstrated that the robustness and size of the effect further decreases under ATB movement, with experiment 4 finding significant main effects of `GRAMFUNC` and `POSITION` as well as a significant interaction; experiment 5 finding a significant main effect of each but not a significant interaction, and experiment 6 revealing only a significant main effect of `GRAMFUNC` but nothing else. That is, the significance of the fixed

effects was determined by the non-syntactic manipulations. At the same time, the different conditions followed the same ranking across experiments, indicating that syntactic components as well as (for all we know, linear) distance do lend some stability to the pattern, but crucially not enough to make it universal among or even within speakers across conditions.

Chapter 4 applied the insights from German to English. First, I presented an assessment of principle C reconstruction in simple *wh*-dependencies in experiment 7. The experiment had a positive outcome, finding that principle C reconstruction in simple *wh*-dependencies is fairly robust considering our understanding of it based on chapters 2 and 3. This investigation was followed by experiment 8 finding that parasitic gaps in English, even in finite adjunct clauses, are far more acceptable than their German counterparts. I then conducted a study of coreference patterns in ATB movement and parasitic gaps manipulating the position of the pronoun in experiment 9, followed by an investigation of principle C reconstruction under ATB movement with an adequate control condition in experiment 10. Despite the robustness of principle C reconstruction in simple *wh*-dependencies, the experiments revealed an even weaker effect of principle C reconstruction than reported for German ATB movement in chapter 3. As a follow-up, experiment 11 intended to clarify whether structural or linear distance accounted for the difference between the two gaps, which manifested in a significant effect of *POSITION* in all experiments but one. The investigation did not yield interpretable results due to the limited acceptability of incorporated adjunct clauses, leading to the exclusion of the majority of participants. The use of two tasks per trial across experiments in chapter 4, i.e. one acceptability judgment and one coreference judgment task, seemingly confused participants, with each task impacting the outcome of the other. Despite the difficulties associated with this series of experiments, they demonstrate that ATB movement and parasitic gaps give rise to the same coreference pattern; that weak evidence for a principle C reconstruction effect was found in the initial gap of ATB movement; and that any effect of reconstruction to the non-initial gap was most likely overshadowed by the consistently significant main effect of *POSITION*, as indicated by the high overall coreference rates regardless of experimental condition, and an increase of inter-speaker variability particularly in non-initial gap conditions. This led me to conclude that principle C reconstruction does not yield reliable evidence for any claims about the syntax of multi-gap dependencies: The evidence is orthogonal to the research question and as such, claims based on these data lose their empirical foundation. I discuss the implications of the evidence's neutrality in the next section.

5.1 Empirical neutrality

Based on the evidence presented in chapters 3 and 4, I have argued against the robustness of principle C reconstruction with PP modifiers in multi-gap dependencies. This perspective is not new—way before this thesis was even conceptualized, Bruening and Al Khalaf (2017) have already claimed that the asymmetrical reconstruction pattern is only apparent. What is new, however, is the thorough experimental investigation provided in this thesis demonstrating the level of nuance there is to the data underpinning this claim. Despite supporting the view that asymmetries in coreference do not necessarily arise from an asymmetrical syntax, it is crucial to keep in mind that the authors' proposal that follows, i.e. the unificational (symmetric) treatment of multi-gap dependencies, remains open to debate.¹ Concluding that principle C reconstruction is orthogonal to the syntax of multi-gap dependencies does not allow us to rule out asymmetric approaches, and neither does it allow us to argue in favor of symmetric approaches. It simply means that the data are uninformative. What we can note, however, is that claims about the underlying syntax of multi-gap dependencies that rely primarily on principle C reconstruction data lose their grip in light of these findings.

For example, Nissenbaum (2000) presents evidence against a unificational analysis for ATB movement and parasitic gaps, among them principle C reconstruction, reporting it to target both gaps in ATB but only the main gap in parasitic gap constructions. These findings support the null operator analysis of parasitic gap dependencies while ruling out a unificational analysis of any kind. Having seen that both constructions give rise to the same coreference pattern, displaying a slight asymmetry due to the effect of POSITION, the argument against unification based on principle C reconstruction does not hold.

The forward ellipsis approach put forth by Salzmann (2012) is strongly supported by Citko's report, citing asymmetrical reconstruction to the initial gap in ATB. Under the forward ellipsis approach, the filler is extracted from the initial gap and elided in the non-initial gap. Because elided R-expressions and anaphors can 'vehicle change' to pronouns, ellipsis offers the perfect explanation for the obviation of a principle C reconstruction effect in the non-initial gap, meanwhile sustaining the symmetry of other reconstruction diagnostics.² Given the experimental findings presented herein, the explanation via vehicle change becomes superfluous at least

¹The authors report on further experiments on allegedly asymmetrical reconstruction effects, i.e. principle A reconstruction and WCO. See section 1.3.2 of chapter 1 for a discussion of the problems associated with them.

²Anaphors can likewise vehicle change to pronouns, accounting for the asymmetry of principle A reconstruction, too. Salzmann (2012) provides no discussion of WCO, which does not seem to pose a problem for the ellipsis approach due to the wh-phrase in the non-initial conjunct moving within vP and thus not crossing over the subject.

for principle C, and the forward ellipsis approach loses one of its main selling points.

On the flipside, accounts struggling to depict the alleged asymmetry among the gaps within or across constructions do not face this challenge any longer. Such are symmetric accounts like parallel extraction and multidominance, both in their unificational and their non-unificational implementation. Again, this does not mean that we can eliminate the empty operator and the forward ellipsis accounts by Nissenbaum (2000) and Salzmann (2012) and declare either parallel extraction or multidominance the winner. We have not yet gathered enough information to declutter the toolkit, but we took a first step towards it by quantifying the relevance of a major piece of evidence. Imagine that the empty operator and forward ellipsis approaches were moved to the ‘maybe’ pile while everything else remains the same. On the one hand, we have yet to see robust evidence against asymmetrical principle C reconstruction, which could be supplied by testing gerunds with small clause complements as described in section 4.5 of chapter 4. On the other hand, we have yet to see robust evidence verifying the symmetry of the phenomena that are claimed to behave symmetrically. Moreover, it remains to be clarified which of the remaining reconstruction diagnostics could provide reliable evidence at all, as certain tests seem to be just as compromised as the principle C reconstruction test. While a single thesis does not suffice to settle the issue of multi-gap dependencies once and for all, the next section outlines alternative directions we could take with respect to reconstruction.

5.2 Alternative diagnostics

As the current series of experiments does not allow us to draw conclusions about the underlying syntax of multi-gap dependencies, I end this thesis with a discussion of alternative diagnostics, particularly in the family of reconstruction, discussing which ones may yield decisive evidence allowing us to filter out implausible theoretical approaches. In section 1.3.2 of chapter 1, we have identified that most of the reconstruction diagnostics are controversial, some of the available data points being confounded.

Variable binding data is provided by Nissenbaum (2000) as well as Citko (2005), for both dependencies by the former and only ATB by the latter author. The test assesses whether a variable in the modifier of the filler can be bound by a QP in either domain, i.e. either conjunct of ATB and either clause in a parasitic gap construction. We will focus on the ATB data by Citko (2005) here as the data for both constructions provided by Nissenbaum (2000, p. 34) is confounded due to the need of Quantifier Raising to derive the intended reading, rendering the

argument moot. The data in (1) supposedly show that the variable must have a suitable binder in each conjunct of an ATB dependency, implying that the variable reconstructs to both gaps.

- (1) a. Which picture of his mother did every Italian like ____ and every Frenchman dislike ____?
 b. #Which picture of his mother did every Italian like ____ and Mary dislike ____?
 c. #Which picture of his mother did Mary dislike ____ and every Italian like ____? (Citko 2005, p. 492)

Citko (2005) does not elaborate on the meaning of these data. It is unclear what the exact interpretation of (1a) is, why and how it is different from (1b) and (1c), and why exactly the data provide evidence for a symmetric approach. In Szarvas et al. (under revision), we explore the array of possible readings in ATB constructions, including possible interpretations of (1b) and (1c). One could conceive of an interpretation in which the variable is bound by the QP in the initial or the non-initial conjunct, receiving a free reading in the other conjunct (referring to some discourse salient entity). Essentially, such a reading seems to be indeed ruled out, but not because of a requirement imposed by the syntactic derivation of ATB movement, but due to a general condition on pronoun binding:

- (2) No LF representation (for a sentence of multisentential text) must contain both bound occurrences and free occurrences of the same index.

(Heim and Kratzer 1998, p. 254)

That is, (1) most likely does not demonstrate anything about the underlying syntax of ATB movement—again, the impossibility of a mixed reading, i.e. one where the variable is interpreted as bound in one and as free in the other conjunct, arises from a general LF constraint. There is, however, a different configuration in which there is a suitable binder in only one of the conjuncts, yet the variable appears to receive the interpretation akin to an E-type pronoun in the other.³

- (3) Q: Which of his misdeeds [does no bad student feel guilty about ____] but [do you still hope the university will take seriously ____]?
 A: His AI usage.

While *his* is bound in the first conjunct, i.e. co-varying with *no bad student*, in the second conjunct, the question could refer to the misdeeds of the entire

³The example in (3) intentionally deviates from the original examples in (1): Using a negative quantifier rules out telescoping (Evans 1980).

plurality of bad students. In other words, the non-initial conjunct requires the pronoun to be plural. This mismatch in number between the representations of the pronoun required in each conjunct poses a severe challenge to the multidominant and sideward movement approach. Under multidominance, the two gaps of an ATB dependency are underlyingly represented as a single node with multiple dominators. Under sideward movement, a single instance of the filler is copied from one domain to the other, and eventually, the matrix clause. Given that there is only a single representation of the element present in a multidominant structure, and the fact that sideward movement derives the existence of both gaps via copying, the availability of the intended reading involving a number mismatch in (3) could provide plain evidence against these proposals.

Data from idiom and scope reconstruction in ATB movement is cited by Citko (2005). Focusing on scope, she reports that *how many books* necessarily has to take wide scope in the non-initial conjunct, interpreting the evidence to support symmetry between the gaps. No such data is available for parasitic gaps, and no comparison is drawn between the initial and non-initial gap of ATB. The interaction between the quantifier (*almost*) *every* and *how many* questions provides a rather easy testing ground for both gaps of ATB as well as parasitic gap constructions as shown in (4).

- (4)
- a. How many books did every student like ___ and almost every professor dislike ___?
 - b. How many books did almost every student like ___ and every professor dislike ___?
 - c. How many books did every student like ___ despite almost every professor disliking ___ pg?
 - d. How many books did almost every student like ___ despite every professor disliking ___?

Crucially, what is of interest here is whether the *how many books* can have wide scope in one domain of *every* but narrow scope in the domain of *almost every*. Citko (2005) cites (4a), claiming that a reading where each student liked a distinct number of books and almost every professor disliked the same number of books is unavailable. Though she does not comment on this matter any further, I assume that she takes the data to indicate that the syntactic representations of *how many* necessarily need to be the same in each gap, though it is unclear whether this truly motivates the presence of a single multi-dominated element or merely that the elements need to take the same scope, as in some type of parallelism requirement (possibly similar to (2)).

Strong Crossover was first cited by Williams (1987) and then reiterated by Munn

(1993) to motivate the possibility of reconstruction to the parasitic gap site (5a), just like it is supposedly found in the non-initial gap of ATB structures (5b).

- (5) a. *Whose_i mother did we talk to ____ after he_i saw _____{pg} ?
 b. *Whose_i mother did we talk to ____ and he_i never visit ____ ?

(Munn 1993, p. 53)

The observation that coreference between *he* and *whose* is ruled out in (5) is taken to indicate that the filler crosses over the pronoun, i.e. that a movement step indeed has to be launched in the non-initial gap of ATB and the parasitic gap site itself. Recall that in section 1.4.1 of chapter 1, I briefly touched on the fact that Munn (1994) equates principle C to SCO, claiming that it indicates symmetric reconstruction in both multi-gap dependencies. This observation culminates in an extensive discussion about principles A and C being evaluated at distinct stages of the derivation, explaining why the former produces an asymmetrical and the latter a symmetrical pattern. However, authors treating principle C reconstruction and SCO as independent types of evidence report an asymmetrical pattern for the former and a symmetrical pattern for the latter (Citko 2005 and Salzmann for ATB, Nissenbaum 2000 for parasitic gaps). Throughout this thesis, I have demonstrated that principle C reconstruction, as standardly employed with PP modifiers, is unreliable and therefore does not map to the underlying syntax of multi-gap dependencies. In section 2.3.3 of chapter 2, I discussed the apparent optionality of principle C reconstruction hypothesizing that the PP modifiers containing the offending R-expression do not reconstruct reliably alongside the head noun rather than claiming that principle C is a violable constraint (cf. Salzmann, Wierzba, and Georgi 2023). If SCO is indeed symmetric in both construction types, this would lend further support to the conclusion that it is not principle C itself that is at stake, but the reconstruction of the head noun's dependents. Aside from supplying the straightforward evidence for the syntax of multi-gap dependencies that the experiments herein failed to provide, work on SCO as well as alternative configurations of principle C reconstruction outlined in section 4.5 of chapter 4 could deepen our understanding of the interactions between the size of the reconstructing constituent and the strength of principle C.

Turning to weak crossover, it has been claimed to affect only the initial gap in both ATB and parasitic gap dependencies by Munn (2001). Because the pronoun can corefer with *man* in examples (6a) and (6c), i.e. when it occupies the non-initial conjunct, this is interpreted to indicate that there is no movement targeting the non-initial conjunct.

- (6) a. Which man_i did you hire ____ and his_i boss fire ____ ?
 b. *Which man_i did his_i boss fire ____ and you hire ____ ?
 c. Which man_i did you visit ____ just before his_i boss fired ____ pg ?
 d. *Which man_i did his_i boss fire ____ just after you visited ____ pg ?

(Munn 2001, p. 5)

Bruening and Al Khalaf (2017) challenge this claim and report on an experiment aiming to test the involvement of linear factors. A discussion of this experiment's shortcomings was provided in section 1.3.2.5 of chapter 1, and the aforementioned issues are extremely easy to fix: First, the validity of the linear account to WCO can be verified by comparing adjunct parasitic gaps, i.e. (7a) and (7b) to subject parasitic gaps, i.e. (7c) and (7d). In the former, the parasitic gap must follow the licensing gap, while in the latter, the parasitic gap must precede it. Crossing the factors POSITION (of the pronoun, *initial* or *non-initial*), and GAP (preceded by the pronoun), *parasitic* or *main*, a significant main effect of POSITION would favor the purely linear account to WCO, while a significant main effect of GAP would provide evidence for an asymmetric account of parasitic gaps (where the parasitic gap qualitatively differs from the main gap). An interaction of POSITION and GAP would indicate that both factors matter, and would most likely manifest in a particular boost of coreference when linear distance between the pronoun and the R-expression is increased (*non-initial*) in addition to the pronoun preceding the parasitic gap rather than the main gap.

- (7) a. *Initial, main*
 Sally is wondering which girl her teacher shouted at ____ before the mother took out of school ____ pg .
 b. *Non-initial, parasitic*
 Sally is wondering which girl the teacher shouted at ____ before her mother took out of school ____ pg .
 c. *Initial, parasitic*
 Sally is wondering which girl her teacher's shouting at ____ pg made the mother take out of school ____ .
 d. *Non-initial, main*
 Sally is wondering which girl the teacher's shouting at ____ pg made her mother take out of school ____ .

Second, claims about parasitic gaps *and* ATB movement have to be assessed, naturally, with an experimental design including both dependencies. The following items CROSS CONSTRUCTION (*ATB* or *PG*) and POSITION (of the pronoun, *initial* or *non-initial*), pretty much in the same fashion as experiment 9 did in chapter 4. A

significant interaction between the factors CONSTRUCTION and POSITION would indicate that WCO produces distinct coreference patterns among the gaps between constructions, for example, a symmetric effect in ATB and an asymmetric effect in parasitic gap dependencies to the exclusion of the parasitic gap site. A significant main effect of CONSTRUCTION in the absence of an interaction would indicate that no WCO effect was found in either gap of one construction but in both gaps of the other, while a significant main effect of POSITION would indicate no difference between the construction types, but a difference between the initial and non-initial gap.

- (8)
- a. *ATB, initial*
Sally is wondering which girl her teacher shouted at ____ and the mother took out of school ____ .
 - b. *ATB, non-initial*
Sally is wondering which girl the teacher shouted at ____ and her mother took out of school ____ .
 - c. *PG, initial*
Sally is wondering which girl her teacher shouted at ____ before the mother took out of school ____ pg .
 - d. *PG, non-initial*
Sally is wondering which girl the teacher shouted at ____ before her mother took out of school ____ pg .

Principle A reconstruction is likely subject to a similar fate like principle C reconstruction. Experimental evidence from German suggests that wh-movement leads to at least a partial loss of original binding options and creates new ones, contrary to previous assumptions about principle A being evaluated in base positions like principle C. The crucial difference is that the reflexive may be bound in intermediate or final positions, but the principle must not be violated in any of the viable positions (Sportiche 2017). Since principle A reconstruction may likely be just as compromised as principle C reconstruction, the experimental scrutiny of these patterns could lead to the neutralization of the existing evidence, but probably not to any firm conclusions about the underlying syntax of multi-gap dependencies. Bruening and Al Khalaf (2017) aimed to tackle this problem in their experiment by topicalizing the reflexive itself. Unfortunately, the data are largely uninterpretable due to all experimental conditions but one scoring a mean acceptability rating lower than the ungrammatical distractors. The reason for this could be the use of rather complicated experimental items. Future work should aim to balance the syntactic requirements imposed by our research questions with participants' processing capacity, allowing for the collection of interpretable data

with reflexives.

5.3 Final conclusion

Neither of the existing ‘tools’ could be definitively eliminated from the ‘toolkit’. However, given the data from the series of experiments presented throughout this thesis, we have gained knowledge of the weight we can assign to evidence from principle C reconstruction, and we are in a confident position to question the weight assigned to other reconstruction diagnostics more generally. We have gained an understanding of the importance of designing experiments with care and attention to detail, that the choice of collecting free or forced choice data, providing a context, and an alternative referent can determine not only the size but even the presence of a reconstruction effect in multi-gap dependencies. The variability of the judgments cited by individual researchers that motivated the research conducted in this project—supposedly assessing the outcome of one and the same empirical test—is easily explained once we consider the level of variability exhibited even by individual participants exposed to items in a highly controlled experimental setting. Coreference is simply too complex of a phenomenon and principle C reconstruction with PP modifiers simply too small a part of it to yield meaningful information about the underlying syntax of multi-gap dependencies. The evidence presented in this thesis thus neutralizes the conflicting data reports.

Ultimately, the lack of clarity in our empirical foundations is intertwined with a rather shallow formulation of the hypotheses that the distinct approaches generate. The predictions of existing theories of multi-gap dependencies are certainly much more intricate than predicting symmetry or asymmetry between the gaps. This seems especially true for theories defining syntactic movement in entirely new terms, take the case of evidence from variable binding briefly discussed in this chapter. These intricacies reveal themselves rather quickly in the process of designing experiments—they force the researcher to sit down, think about the predictions they want to test, and also consider what type of experimental design is necessary to gain clearly interpretable data. Crucially, however, they force us to think about which types of data can best discriminate between existing proposals, and facilitate a deeper engagement with what these proposals exactly imply for the entirety of syntactic theory. For this reason alone, we should strive to refine and diversify our methods of data collection. Ideally, this study should mark the beginning of a more thorough investigation of the rest of the previously cited data—focusing on the predictions that already exist, and obtaining reliable data to evaluate them, ultimately contributing to a coherent theory of mental grammar.

Bibliography

- Abramovitz, Rafael (2021). “Topics in the Grammar of Koryak”. PhD thesis. Massachusetts Institute of Technology, 2021.
- Adger, David, Alex Drummond, David Hall, and Coppe van Urk (2017). “Is there Condition C reconstruction?” In: *Proceedings of the 47th Annual Meeting of the North East Linguistics Society (NELS 47)*. Ed. by Andrew Lamont and Katerina Tetzloff. Amherst, MA: GLSA, 2017, pp. 21–31.
- Agbayani, Brian and Masao Ochi (2007). *Split lexical insertion in parasitic gap constructions*. Ms. University of California, Fresno and University of Osaka. 2007.
- (2022). “Split lexical insertion in parasitic gap constructions”. In: *Studia Linguistica*, pp. 1–39.
- Alexopoulou, Theodora and Frank Keller (2007). “Locality, cyclicity, and resumption: at the interface between grammar and the human sentence processor”. In: *Language* 83.1.
- Altshuler, Daniel and Robert Truswell (2022). *Coordination and the syntax-discourse interface*. Oxford: Oxford University Press, 2022.
- An, Duk-Ho (2007). “Asymmetric T-to-C movement in ATB constructions: A PF deletion analysis”. In: *Proceedings of the 14th Conference of the Student Organization of Linguistics in Europe (ConSOLE XIV)*. Ed. by Sylvia Blaho, Luis Vicente, and Erik Schoorlemmer. Leiden, The Netherlands: Leiden University Centre for Linguistics, 2007, pp. 1–19.
- Arnold, Jennifer E. (2001). “The effect of thematic roles on pronoun use and frequency of reference continuation”. In: *Discourse processes* 31.2, pp. 137–162.
- Assmann, Anke (2010). “Parasitic Gaps in Derivational Grammar”. MA thesis. Leipzig: Leipzig University, 2010.
- (2012). “Deriving parasitic gaps by fission and fusion”. In: *Proceedings of the 19th Conference of the Student Organization of Linguistics in Europe (ConSOLE XIX)*. Ed. by Enrico Boone, Kathrin Linke, and Maartje Schulpen. 2012, pp. 49–75.

- Baayen, Harald, Shravan Vasishth, Reinhold Kliegl, and Douglas Bates (2017). “The cave of shadows: Addressing the human factor with generalized additive mixed models”. In: *Journal of Memory and Language* 94, pp. 206–234.
- Bachrach, Asaf and Roni Katzir (2009). “Right node raising and delayed spell-out”. In: *Inter-Phases: Phase-theoretic investigations of linguistic interfaces*. Ed. by Kleanthes Grohmann. Oxford, England: Oxford University Press, 2009, pp. 283–316.
- Bárány, András (2015). “Differential object marking in Hungarian and the morphosyntax of case and agreement”. PhD thesis. Cambridge University, 2015.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal”. In: *Journal of Memory of Language* 68, pp. 255–278.
- Barros, Mathew and Luis Vicente (2011). “Right node raising requires both multidominance and ellipsis”. In: *University of Pennsylvania Working Papers in Linguistics* 17.1, pp. 1–9.
- Barss, Andrew (1988). “Paths, connectivity, and featureless empty categories”. In: *Annali di Ca’Foscari. Rivista della Facoltà di Lingue e Letterature straniere dell’Università di Venezia* 27.4, pp. 247–279.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.
- Belk, Zoë, Ad Neeleman, and Joy Philip (2024). “What Divides, and What Unites, Right-Node Raising”. In: *Linguistic Inquiry* 54.4, pp. 685–728.
- Bianchi, Valentina (1995). “Consequences of antisymmetry for the syntax of headed relative clauses”. PhD thesis. Scuola Normale Superiore, Pisa, 1995. URL: https://www.researchgate.net/publication/245772483_Consequences_of_Antisymmetry_for_the_Syntax_of_headed_relative_Clauses.
- Biskup, Petr (2018). “Case syncretism in Russian, Polish and Czech ATB constructions”. In: *Proceedings of the 25th Meeting of Formal Approaches to Slavic Linguistics (FASL 25)*. Ed. by Wayles Browne, Miloje Despić, Naomi Enzinna, Simone Harmath-de Lemos, Robin Karlin, and Draga Zec. 2018, pp. 36–56.
- Blümel, Andreas (2014). “On forked chains in ATB-movement. Defending and newly implementing a traditional notion”. In: *Proceedings of the 22nd Conference of the Student Organization of Linguistics in Europe (ConSOLE XXII)*. Ed. by Eleanor Dutton Martin Kohlberger Kate Bellamy. 2014, pp. 19–38.

- (2017). *Symmetry, shared labels and movement in syntax*. Berlin, Germany: De Gruyter, 2017.
- Bondaruk, Anna (2003). “Parasitic gaps and ATB in Polish”. In: *Journal of Slavic Linguistics* 11.2, pp. 221–249.
- Bošković, Željko and Steven Franks (2000). “Across-the-Board Movement and LF”. In: *Lingua* 3, pp. 107–129.
- Brown, Jessica, Gisbert Fanselow, Rebecca Hall, and Reinhold Kliegl (2021). “Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm”. In: *PLoS ONE* 16.5.
- Browning, Marguerite (1987). “Null operator constructions”. PhD thesis. Cambridge, MA: MIT, 1987.
- Bruening, Benjamin (2014). “Precede-and-command revisited”. In: *Language* 90, pp. 342–388.
- (2021). “Generalizing the presuppositional approach to the Binding Conditions”. In: *Syntax* 24, pp. 417–461.
- Bruening, Benjamin and Eman Al Khalaf (2017). *Reconstruction and linear order in ATB movement and parasitic gap constructions*. Unpublished manuscript. 2017. URL: <https://udel.edu/~bruening/Downloads/ATBMvmtLinear4.4.pdf>.
- (2019). “No argument-adjunct asymmetry in reconstruction for Binding Principle C”. In: *Journal of Linguistics* 55, pp. 247–276.
- Büring, Daniel (2005). *Binding Theory*. Cambridge: Cambridge University Press, 2005.
- Chaves, Rui P. and Michael T. Putnam (2020). *Unbounded Dependency Constructions*. Oxford: Oxford University Press, 2020.
- Chomsky, Noam (1976). “Conditions on rules of grammar”. In: *Linguistic Analysis* 2, pp. 303–351.
- (1981). *Lectures on Government and Binding*. Dordrecht: Foris, 1981.
- (1986). *Barriers*. Cambridge, MA: MIT Press, 1986.
- (1995). *The minimalist program*. Cambridge, MA: MIT Press, 1995.
- (2001). “Beyond explanatory adequacy”. In: *MIT Occasional Papers in Linguistics* 20, pp. 1–25.

- Christensen, Rune Haubo B. (2019). *ordinal—Regression Models for Ordinal Data*. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>. 2019.
- Cinque, Guglielmo (1990). *Types of A'-dependencies*. Cambridge, MA: MIT Press, 1990.
- Citko, Barbara (2005). “On the Nature of Merge: External Merge, Internal and Parallel Merge”. In: *Linguistic Inquiry* 36, pp. 475–497.
- (2011). *Symmetry in Syntax: Merge, move, and labels*. Cambridge Studies in Linguistics 129. Cambridge, England: Cambridge University Press, 2011.
- (2013). *Size matters: Multidominance and DP structure in Polish*. Talk given at the 44th Poznań Linguistic Meeting, Poznań, Aug 6, 2013. 2013.
- Clifton Jr, Charles and Lyn Frazier (1989). “Successive cyclicity in the grammar and the parser”. In: *Language and cognitive processes* 4.2, pp. 93–126.
- Collins, Chris (2005). “A smuggling approach to the passive in English”. In: *Syntax* 8.2, pp. 81–120.
- Cowles, H. Wind, Matthew Walenski, and Robert Kluender (2007). “Linguistic and cognitive prominence in anaphor resolution: topic contrastive focus and pronouns”. In: *Topoi* 26, pp. 3–18.
- Culicover, Peter W. (2001). “Parasitic gaps: A history”. In: *Parasitic Gaps*. Ed. by Peter W. Culicover and Paul Postal. Cambridge, MA: MIT Press, 2001, pp. 3–68.
- Cummings, Ian, Clare Patterson, and Claudia Felser (2014). “Variable binding and coreference in sentence comprehension: Evidence from eye movements”. In: *Journal of Memory and Language* 71, pp. 39–56.
- Davis, Colin P. B. (2025). “Anti-locality explains the restricted interaction of subjects and parasitic gaps”. In: *Proceedings of the 41st West Coast Conference on Formal Linguistics*. Ed. by Nicolas Webster, Yağmur Kiper, Richard Wang, and Sichen Larry Lyu. 2025, pp. 113–120.
- de Vries, Mark (2017). “Across-the-Board Phenomena”. In: *The Wiley Blackwell Companion to Syntax*. Ed. by Martin Everaert and Henk C. van Riemsdijk. 2nd ed. Hoboken, NJ: John Wiley and Sons Inc., 2017, pp. 1–31.
- den Dikken, Marcel (2009). “On the nature and distribution of successive cyclicity: Adjunction, resumption, and scope marking as the roads to success in long-distance relationship building”. CUNY Graduate Center Manuscript. 2009.

- Déprez, Viviane M. (1989). "On the typology of syntactic positions and the nature of chains: Move α to the specified of functional projections". PhD thesis. Cambridge, MA: MIT, 1989.
- Dyła, Stefan (1984). "Across-the-board dependencies and case in Polish". In: *Linguistic Inquiry* 15.4, pp. 701–705.
- É. Kiss, Katalin (1985). "Parasitic Chains". In: *The Linguistic Review* 5, pp. 41–74.
- (2008). "The structure of the Hungarian VP revisited". In: *Approaches to Hungarian* 10, pp. 31–58.
- Engdahl, Elisabet (1983). "Parasitic gaps". In: *Linguistics and Philosophy* 6, pp. 5–34.
- Evans, Gareth (1980). "Pronouns". In: *Linguistic Inquiry* 11.2, pp. 337–362.
- Fanselow, Gisbert (1987). *Konfigurationsalität*. Tübingen: Narr, 1987.
- Fanselow, Gisbert (1993). "Die Rückkehr der Basisgenerierer". In: *Groninger Arbeiten zur germanistischen Linguistik* 36, pp. 1–74.
- (2001). "Features, θ -roles and free constituent order". In: *Linguistic Inquiry* 32.3, pp. 405–437.
- (2007). "Carrots - perfect as vegetables, but please not as a main dish". In: *Theoretical Linguistics* 33.3, pp. 353–367.
- Featherston, Sam (2007). "Data in generative grammar: The stick and the carrot". In: *Theoretical Linguistics* 33.3, pp. 269–318.
- (2009). "A scale for measuring well-formedness: Why syntax needs boiling and freezing points". In: *The Fruits of Empirical Linguistics. Volume 1: Process*. Ed. by Sam Featherston and Susanne Winkler. Berlin: Mouton de Gruyter, 2009, pp. 47–74.
- Featherston, Sam and Wolfgang Sternefeld (2003). "The interaction of factors in judgments of reflexive structures: Data from object coreference in German". In: *Arbeiten zur Reflexivierung*. Ed. by Lutz Gunkel, Gereon Müller, and Gisela Zifonun. Tübingen: Niemeyer, 2003, pp. 25–50.
- Felix, Sascha W. (1985). "Parasitic Gaps in German". In: *Erklärende Syntax des Deutschen*. Ed. by Werner Abraham. Tübingen: Narr, 1985, pp. 173–201.
- Fiengo, Robert and Robert May (1994). *Indices and identity*. Cambridge, MA: MIT Press, 1994.

- Fox, Danny (1999). “Reconstruction, Binding Theory, and the Interpretation of Chains”. In: *Linguistic Inquiry* 2.30, pp. 157–196.
- (2000). *Economy and scope interaction*. Cambridge, MA: MIT Press, 2000.
- Franks, Steven (1993). “On parallelism in across-the-board dependencies”. In: *Linguistic Inquiry* 24.3, pp. 509–529.
- (1995). *Parameters of Slavic morphosyntax*. New York: Oxford University Press, 1995.
- Frazier, Lyn (1987). “Syntactic processing: evidence from Dutch”. In: *Natural Language and Linguistic Theory* 5, pp. 519–559.
- Freidin, Robert (1986). “Fundamental Issues in the Theory of Binding”. In: *Studies in the Aquisition of Anaphora*. Ed. by Barbara Lust. Reidel: Dordrecht, 1986, pp. 151–188.
- Gazdar, Gerald (1981). “Unbounded Dependencies and Coordinate Structure”. In: *Linguistic Inquiry* 12, pp. 155–184.
- Georgi, Doreen, Andrew Murphy, Johannes Rothert, and Timea Szarvas (forthcoming). “Diagnosing syntactic structure in ATB and RNR constructions: A reply to Larson (2013, 2014)”. In: *Proceedings of the 60th Annual Meeting of the Chicago Linguistic Society*. Chicago, IL: The Chicago Linguistic Society, 2025. URL: <https://lingbuzz.net/lingbuzz/008514>.
- Georgi, Doreen, Lefteris Paparounas, and Martin Salzmann (2025). *Two types of non-structural case: Evidence from ATB movement in Modern Greek*. Paper presented at GLOW 47. 2025.
- Gervain, Judit (2003). “Syntactic microvariation and methodology. Problems and perspectives”. In: *Acta Linguistica Hungarica* 50.3-4, pp. 405–434.
- Gibson, Ted and Evelina Fedorenko (2013). “The need for quantitative methods in syntax and semantics research”. In: *Language and cognitive processes* 28.1/2, pp. 88–124.
- Goodall, Grant (1987). *Parallel Structures in Syntax*. Cambridge: Cambridge University Press, 1987.
- Gor, Vera (2020). “Experimental investigations of Principle C at the syntax-pragmatics interface”. PhD thesis. Rutgers University, 2020.
- Gordon, Peter C., Barbara J. Grosz, and Laura A. Gilliom (1993). “Pronouns, names and the centering of attention in discourse”. In: *Cognitive Science* 17, pp. 311–347.

- Gordon, Peter C. and Randall Hendrick (1998). "The representation and processing of coreference in discourse". In: *Cognitive Science* 22, pp. 389–424.
- Gračanin-Yüksek, Martina (2013). "Linearizing Multidominance Structures". In: *Challenges to Linearization*. Ed. by Theresa Biberauer and Ian Roberts. Berlin: Mouton de Gruyter, 2013, pp. 269–294.
- Grewendorf, Günther (1988). *Aspekte der deutschen Syntax. Eine Rektions-Bindungs-Analyse*. Tübingen: Narr, 1988.
- Grodzinsky, Yosef and Tanya Reinhart (1993). "The innateness of binding and coreference". In: *Linguistic Inquiry* 1.24, pp. 69–101.
- Grosu, Alexander (1973). "On the nonunitary nature of the Coordinate Structure Constraint". In: *Linguistic Inquiry* 4.1, pp. 88–92.
- Ha, Seungwan (2008). "Ellipsis, right node raising, and across-the-board constructions". PhD thesis. Boston, MA: Boston University, 2008.
- Haegeman, Liliane (1994). *Introduction to Government and Binding Theory*. Oxford/Malden: Blackwell, 1994.
- Haïk, Isabelle (1985). "The syntax of operators". PhD thesis. Cambridge, MA: MIT, 1985.
- Halle, Morris and Alec Marantz (1993). "Distributed morphology and the pieces of inflection". In: *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. Ed. by Kenneth Hale and Samuel J. Keyser. Cambridge, MA: MIT Press, 1993, pp. 111–176.
- Harris, Catherine L. and Elizabeth A. Bates (2002). "Clausal backgroundedness and pronominal reference: a functionalist approach to c-command". In: *Language and Cognitive Processes* 17.3, pp. 237–269.
- Hartmann, Jutta, Andreas Konietzko, and Martin Salzmann (2016). "On the limits of non-parallelism in ATB movement: Experimental evidence for strict syntactic identity". In: *Quantitative approaches to grammar and grammatical change*. Ed. by Sam Featherston and Yannick Versley. Berlin, Germany: De Gruyter Mouton, 2016, pp. 51–84.
- Heim, Irene (1998). "Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach". In: *MIT Working Papers in Linguistics* 25. Ed. by Uli Sauerland and Orin Percus. 1998, pp. 205–246.
- (2007). "Forks in the road to rule I". In: *Proceedings of the Thirty-Eighth Annual meeting of the North East Linguistic Society (NELS 38)*. Ed. by Muhammad Abdurrahman, Anisa Schardl, and Martin Walkow. 2007, pp. 339–358.

- Heim, Irene and Angelika Kratzer (1998). *Semantics in Generative Grammar*. Wiley-Blackwell, 1998.
- Hein, Johannes (2025). “Exceptional inflection in German right node raising constructions”. In: *Proceedings of the 55th annual meeting of the North East Linguistic Society*. 2025.
- Hein, Johannes and Andrew Murphy (2020). “Case matching and syncretism in ATB-dependencies”. In: *Studia Linguistica* 74.2, pp. 254–302.
- Henderson, Brent (2007). “Matching and raising unified”. In: *Lingua* 117, pp. 202–220.
- Heycock, Caroline (1995). “Asymmetries in reconstruction”. In: *Linguistic Inquiry*, pp. 547–570.
- Hicks, Glyn (2009). “*Tough*-constructions and their derivation”. In: *Linguistic Inquiry* 40.4, pp. 535–566.
- Higginbotham, James (1980). “Pronouns and bound variables”. In: *Linguistic Inquiry* 11, pp. 679–708.
- Himmelreich, Anke (2017). “Case Matching Effects in Free Relatives and Parasitic Gaps: A Study on the Properties of Agree”. PhD thesis. Leipzig: Leipzig University, 2017.
- Hofmeister, Philip and Elisabeth Norcliffe (2013). “Does resumption facilitate sentence comprehension?” In: *The core and the periphery: data-driven perspectives on syntax inspired by Ivan A. Sag*. Ed. by Philip Hofmeister and Elisabeth Norcliffe. Stanford, CA: CSLI Publications, 2013, pp. 225–246.
- Hornstein, Norbert and Jairo Nunes (2002). “On Asymmetries between Parasitic Gap and Across-the-Board Constructions”. In: *Syntax* 5, pp. 26–54.
- Horvath, Julia (1992). “Anti-C-Command and Case Compatibility in the Licensing of Parasitic Chains”. In: *The Linguistic Review* 9.2, pp. 183–218.
- Huybregts, Riny and Henk C. van Riemsdijk (1985). “Parasitic Gaps and ATB”. In: *Proceedings of the 15th Annual Meeting of the North Eastern Linguistic Society (NELS 15)*. Ed. by Stephen Berman, Jae-Woong Choe, and Joyce McDonough. University of Massachusetts Amherst, MA: GLSA, 1985, pp. 168–187.
- Iatridou, Sabine (1995). “Clitics and island effects”. In: *University of Pennsylvania Working Papers in Linguistics* 2.1, pp. 11–30.
- Järvikivi, Juhani, Roger P.G. van Gompel, Jukka Hyönä, and Raymond Bertram (2005). “Ambiguous Pronoun Resolution: Contrasting the First-Mention and Subject-Preference Accounts”. In: *Psychological Science* 16.4, pp. 260–264.

- Johnson, Kyle (2018). *A multidominant theory of movement*. Lectures presented at CreteLing, July 17-27. 2018.
- (2025). *Opacity and multidominance*. Lectures presented at the University of Massachusetts, Amherst. 2025.
- Kaiser, Elsie (2011). “Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus”. In: *Language and Cognitive Processes* 26.10, pp. 1625–1666.
- Kasai, Hironobu (2010). “Parasitic gaps under multiple dominance”. In: *English Linguistics* 27.2, pp. 235–269.
- Kathol, Andreas (2001). “Nonexistence of Parasitic Gaps in German”. In: *Parasitic Gaps*. Ed. by Peter W. Culicover and Paul Postal. Cambridge, MA: MIT Press, 2001, pp. 315–338.
- Kayne, Richard S. (1983). “Connectedness”. In: *Linguistic Inquiry* 14.2, pp. 223–250.
- (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press, 1994.
- Kearney, Kevin (1983). “Governing categories”. Unpublished manuscript. University of Connecticut, 1983.
- Kuno, Susumu (2004). “Empathy and direct discourse perspectives”. In: *The Handbook of Pragmatics*. Ed. by Laurence R. Horn and Gregory Ward. Blackwell publishing, 2004, pp. 315–343.
- Lakoff, George (1986). “Frame semantic control of the coordinate structure constraint”. In: *Chicago Linguistic Society* 22.2, pp. 152–167.
- Lasnik, Howard (1998). “Some Reconstruction Riddles”. In: *Proceedings of the 22nd Annual Penn Linguistics Colloquium (PLC 22)*. Ed. by Alexis Dimitriadis, Hikyoung Lee, Christine Moisse, and Alexander Williams. 1998, pp. 83–98.
- Lebeaux, David (1988). “Language Acquisition and the Form of Grammar”. PhD thesis. Amherst: University of Massachusetts, 1988.
- (1991). “Relative clauses, licensing, and the nature of the derivation”. In: *Perspectives on phrase structure: Heads and licensing*. Ed. by Susan Rothstein and Margaret Speas. San Diego, CA: Academic Press, 1991, pp. 209–239.
- (2009). *Where does binding theory apply?* The MIT Press, 2009.
- Levine, Robert D. (1985). “Right Node (Non-)Raising”. In: *Linguistic Inquiry* 16.3, pp. 492–497.

- Levine, Robert D. (2001). “The Extraction Riddle: Just What are We Missing?” In: *Journal of Linguistics* 37, pp. 145–174.
- Levine, Robert D., Thomas E. Hukari, and Michael Calcagno (2001). “Parasitic gaps in English: Some overlooked cases and their theoretical implications”. In: *Parasitic Gaps*. Ed. by Peter W. Culicover and Paul M. Postal. Current Studies in Linguistics. Cambridge, MA: MIT Press, 2001, pp. 181–222.
- Lohninger, Magdalena (2025). “Patterns in Chaos: Composite A/A Probes”. PhD thesis. University of Vienna, 2025.
- Lüdecke, Daniel (2018). “ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.” In: *Journal of Open Source Software* 3.26, p. 772. DOI: 10.21105/joss.00772.
- Mahajan, Anoop Kumar (1990). “The A/A’ distinction and movement theory”. PhD thesis. Cambridge, MA: MIT, 1990.
- McElree, Brian and Theresa Griffith (1998). “Structural and lexical constraints on filling gaps during sentence comprehension: a time-course analysis”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.2, pp. 432–460.
- Merchant, Jason (2001). *The syntax of silence*. Oxford, England: Oxford University Press, 2001.
- (2013). “Voice and Ellipsis”. In: *Linguistic Inquiry* 44.1, pp. 77–108.
- Moltmann, Friederike (1992). “Coordination and comparatives”. PhD thesis. Cambridge, MA: MIT, 1992.
- Moulton, Keir (2009). “Natural selection and the syntax of clausal complementation”. PhD thesis. University of Massachusetts, Amherst, 2009.
- Müller, Gereon (1993). “On deriving movement type asymmetries”. PhD thesis. Tübingen: Universität Tübingen, 1993.
- Munn, Alan (1992). “A null operator analysis of ATB gaps”. In: *The Linguistic Review* 9.1, pp. 1–26.
- (1993). “Topics in the syntax and semantics of coordinate structures”. PhD thesis. College Park, MD: University of Maryland, 1993. URL: https://www.researchgate.net/publication/35300961_Topics_in_the_syntax_and_semantics_of_coordinate_structures.
- (1994). “A Minimalist account of reconstruction asymmetries”. In: *Proceedings of NELS 24*. Ed. by Merce Gonzalez. 1994, pp. 397–410.

- (1999). “On the identity requirement of ATB-extraction”. In: *Natural Language Semantics* 7.4, pp. 421–425.
- (2001). “Explaining parasitic gap restrictions”. In: *Parasitic Gaps*. Ed. by Peter W. Culicover and Paul M. Postal. Current Studies in Linguistics 35. Cambridge, MA: MIT Press, 2001.
- Neeleman, Ad and Misako Tanaka (2024). “Extraction asymmetries show that type A coordination is adjunction”. In: *Language* 100.1, pp. 1–39.
- Ness, Tal and Aya Meltzer-Asscher (2017). “Working memory in the processing of long-distance dependencies: interference and filler maintenance”. In: *Journal of Psycholinguistic Research* 46, pp. 1353–1365.
- Nissenbaum, Jonathan (2000). “Investigations of covert phrase movement”. PhD thesis. Cambridge, MA: MIT, 2000. URL: <https://dspace.mit.edu/handle/1721.1/8842>.
- Nunes, Jairo (1995). “The copy theory of movement and the linearization of chains in the Minimalist Program”. PhD thesis. University of Maryland, 1995.
- (2001). “Sideward movement”. In: *Linguistic Inquiry* 32.2, pp. 303–344.
- (2004). *Linearization of chains and sideward movement*. Linguistic Inquiry Monographs 43. Cambridge, Mass: MIT Press, 2004.
- Nunes, Jairo and Juan Uriagereka (2000). “Cyclicity and extraction domains”. In: *Syntax* 3.1, pp. 20–43.
- Pesetsky, David (1982). “Paths and Categories”. PhD thesis. MIT, 1982.
- Phillips, Colin (2006). “The real-time status of island phenomena”. In: *Language* 82.4, pp. 795–823.
- (2009). “Should we impeach armchair linguists?” In: *Japanese/Korean Linguistics* 17. Ed. by S. Iwasaki, H. Hoji, P. M. Clancy, and Sung-Ock Sohn. CSLI Publications, 2009, pp. 49–64.
- Pollard, Carl and Ivan A. Sag (1992). “Anaphors in English and the scope of Binding Theory”. In: *Linguistic Inquiry* 23.2, pp. 261–303.
- (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press, 1994.
- Postal, Paul M. (1971). *Cross-Over Phenomena*. New York: Holt, Rinehard and Winston, 1971.

- Postal, Paul M. (1993). “Parasitic gaps and the across-the-board phenomenon”. In: *Linguistic Inquiry* 24.4, pp. 735–754. URL: <https://dspace.mit.edu/handle/1721.1/15166>.
- (1994). “Parasitic and pseudoparasitic gaps”. In: *Linguistic Inquiry* 25.1, pp. 63–117.
- R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: <https://www.R-project.org/>.
- Reinhart, Tanya (1983a). *Anaphora and semantic interpretation*. Chicago: University of Chicago Press, 1983.
- (1983b). “Coreference and bound anaphora”. In: *Linguistics and Philosophy* 6, pp. 47–88.
- Reinhart, Tanya and Eric Reuland (1993). “Reflexivity”. In: *Linguistic Inquiry* 24, pp. 657–672.
- Ross, John R. (1967). “Constraints on variables in syntax”. PhD thesis. Cambridge, MA: MIT, 1967.
- Rothert, Johannes (2022). *An experimental investigation of syncretism and proximity effects in Polish ATB topicalization and RNR. Evidence for ellipsis-based approaches to sharing constructions*. University of Potsdam master thesis. 2022.
- RStudio Team (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA, 2021. URL: <http://www.rstudio.com/>.
- Ruys, Eddy (1992). “The scope of indefinites”. PhD thesis. Utrecht, The Netherlands: Utrecht University, 1992.
- Safir, Ken (1999). “Vehicle change and reconstruction in \bar{A} -chains”. In: *Linguistic Inquiry* 30.4, pp. 587–620.
- Sag, Ivan A. (1976). “Deletion and Logical Form”. PhD thesis. MIT, 1976. URL: <https://dspace.mit.edu/handle/1721.1/16401>.
- Salzmann, Martin (2012). “A derivational ellipsis approach to ATB-movement”. In: *The Linguistic Review* 29.3, pp. 397–438.
- Salzmann, Martin, Marta Wierzba, and Doreen Georgi (2023). “Condition C in German A'-movement: Tackling challenges in experimental research on reconstruction”. In: *Journal of Linguistics* 59.3, pp. 577–622.

- Sauerland, Uli (1998). “The meaning of chains”. PhD thesis. MIT, 1998. URL: <https://www.leibniz-zas.de/en/research/publications/details/publications/3384-the-meaning-of-chains>.
- Schütze, Carson T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press, 1996.
- Seguin, Luisa and Gary Thoms (2025). “Reconciling ATB and parasitic gaps. A smuggling analysis of P-mismatches in Italian”. In: *Syntax in uncharted territories. Essays in honor of Maria Polinsky*. Ed. by Lauren Clemens, Vera Griбанова, and Gregory Scontras. Irvine: University of California, 2025, pp. 505–523.
- Sportiche, Dominique (2016). *Neglect*. Unpublished UCLA manuscript. 2016.
- (2017). “Reconstruction, Binding, Scope”. In: *The Wiley Blackwell Companion to Syntax, Second Edition*. Ed. by Martin Everaert and Henk C. van Riemsdijk. Hoboken, NJ: John Wiley & Sons, 2017.
- (2019). “Somber Prospects for Late Merger”. In: *Linguistic Inquiry* 50.2, pp. 416–424.
- Sprouse, Jon, Carson T. Schütze, and Diogo Almeida (2013). “A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001-2010”. In: *Lingua* 134, pp. 219–248.
- Starschenko, Alexej and Marta Wierzba (2024). *L-Rex Linguistic rating experiments [software], version 1.0.3*. GNU General Public License v3.0. 2024. URL: <https://github.com/2e2a/l-rex/>.
- Stockwell, Richard, Aya Meltzer-Asscher, and Dominique Sportiche (2021). “There is reconstruction for Condition C in English questions”. In: *Proceedings of the 51st Annual Meeting of the North Eastern Linguistic Society (NELS 51)*. Ed. by Alessa Farinella and Angelica Hill. Amherst, MA: GLSA, 2021, pp. 205–214.
- (2022). “Experimental evidence for the Condition C argument-adjunct asymmetry in English questions”. In: *Proceedings of the 52nd Annual Meeting of the North Eastern Linguistic Society (NELS 52)*. Ed. by Özge Bakay, Breanna Pratley, Eva Neu, and Peyton Deal. Amherst, MA: GLSA, 2022, pp. 145–158.
- Szarvas, Timea (2024a). “On reconstruction in German ATB-movement and the optimization of experimental designs”. In: *Proceedings of the 32nd Conference of the Student Organization of Linguistics in Europe*. Ed. by Federica Longo, Leonardo Russo Cardona, Tommaso Sgrizzi, Atefeh Shahbazi, and Hadis Tamleh. Leiden: Leiden University Centre for Linguistics, 2024, pp. 195–212. URL: <https://ling.auf.net/lingbuzz/008687>.

- Szarvas, Timea (2024b). “Unscrambling German parasitic gaps”. In: *Gisbert Fanselow’s Contributions to Syntactic Theory*. Ed. by Artemis Alexiadou, Doreen Georgi, Fabian Heck, Gereon Müller, and Florian Schäfer. Universität Leipzig: Institut für Linguistik, Leipzig, 2024, pp. 133–144. URL: <https://ling.auf.net/lingbuzz/008679>.
- (2025a). “An asymmetry without theoretical consequences: Revisiting principle C reconstruction in ATB movement experimentally”. In: *Proceedings of the 55th Annual Meeting of the North East Linguistic Society*. Ed. by Duygu Demiray, Roger Cheng-yen Liu, and Nir Segal. Vol. 2. University of Massachusetts, Amherst: GLSA, 2025, pp. 239–248. URL: <https://lingbuzz.net/lingbuzz/008992>.
- (2025b). *Supplementary materials for “Rethinking principle C reconstruction in ATB movement and parasitic gaps”*. Open Science Framework. 2025. URL: <https://doi.org/10.17605/OSF.IO/HVM8U>.
- (2025c). “Why we’re still debating principle C reconstruction. A comparative experimental study”. In: *Glossa: A journal of general linguistics* 10.1, pp. 1–42. URL: <https://doi.org/10.16995/glossa.18707>.
- Szarvas, Timea, Aron Finholt, Doreen Georgi, Andreas Pregla, Mariia Privizentseva, Martin Salzmann, and Stanislao Zompì (under revision). “Reconstruction for variable binding in ATB movement: Interpretive possibilities and their theoretical implications”. In: *Syntactic Theory and Research* (2025).
- Takahashi, Shoichi and Sarah Hulsey (2009). “Wholesale Late Merger: beyond the A/ \bar{A} -distinction”. In: *Linguistic Inquiry* 40.3, pp. 387–426.
- Taraldsen, Knut T. (1981). “The theoretical implications of a class of marked extractions”. In: *Theory of markedness in generative grammar*. Ed. by Adriana Belletti, Luciana Brandi, and Luigi Rizzi. Pisa, Italy: Scuola Normale Superiore, 1981, pp. 475–516.
- Temme, Anne and Elisabeth Verhoeven (2017). “Backward binding as a psych effect: A binding illusion?” In: *Zeitschrift für Sprachwissenschaft* 36.2, pp. 279–308.
- Twiner, Nicholas and Vera Lee-Schoenfeld (2019). “Binding German (in)direct objects: Spell-out strategies for disambiguation”. In: *Proceedings of the Linguistic Society of America* 4. Ed. by Patrick Farrell. New York, 2019, pp. 1–15.
- van Riemsdijk, Henk C. and Edwin Williams (1981). “NP-structure”. In: *The Linguistic Review* 1, pp. 171–217.
- van Urk, Coppe (2015). “A uniform syntax for phrasal movement. A case study of Dinka Bor”. PhD thesis. Cambridge, MA: MIT, 2015.

- Varaschin, Giuseppe, Peter W. Culicover, and Susanne Winkler (2023). “In pursuit of Condition C: (non-)coreference in grammar, discourse and processing”. In: *Information Structure and Discourse in Generative Grammar: Mechanisms and Processes*. Ed. by Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, 2023.
- Wasow, Thomas (1972). “Anaphoric relations in English”. PhD thesis. Massachusetts Institute of Technology, 1972.
- Webelhuth, Gert (1989). “Syntactic saturation phenomena and the modern Germanic languages”. PhD thesis. Cambridge, MA: MIT, 1989.
- (1992). *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press, 1992.
- Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2016. ISBN: 978-3-319-24277-4. URL: <https://ggplot2.tidyverse.org>.
- Wilder, Chris (1994). “Coordination, ATB and ellipsis”. In: *Groninger Arbeiten zur germanistischen Linguistik* 37, pp. 291–329.
- (1997). “Some properties of ellipsis in coordination”. In: *Studies on universal grammar and typological variation*. Ed. by Artemis Alexiadou and Tracy A. Hall. Linguistik Aktuell/Linguistics Today 13. Amsterdam: John Benjamins, 1997, pp. 59–107.
- (1999). “Right node raising and the LCA”. In: *Proceedings of the 18th West Coast Conference on Formal Linguistics (WCCFL 18)*. Ed. by Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest. 1999.
- Williams, Edwin (1978). “Across-the-Board Rule Application”. In: *Linguistic Inquiry* 9, pp. 31–43.
- (1987). “The ATB theory of parasitic gaps”. In: *The Linguistic Review* 6.3, pp. 265–279.
- Zhang, Niina Ning (2010). *Coordination in Syntax*. Cambridge: Cambridge University Press, 2010.