NORTHWESTERN UNIVERSITY

Influences on Parsing Ambiguity

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Lauren Marie Ackerman

EVANSTON, ILLINOIS

December 2015

ABSTRACT

Influences on Parsing Ambiguity

Lauren Marie Ackerman

The primary goal of this dissertation is to characterize the relative strength of two of the influences on the parser's behavior during ambiguity resolution: coreference dependency formation and verb frame preference. I find that coreference dependency formation exerts a stronger influence on the parser than does verb frame preference, even when verb frame preference is maximized in transitively biased frames.

Previous studies have shown local attachment bias initially directs the parser to an embedded object analysis in sentences like (1), in which the DP *Annie's melody* is locally ambiguous between the embedded object (EO)/matrix subject (MS) analyses (Ferreira and Henderson, 1990).

(1) Whenever she was trying to casually hum Annie's melody was beautiful.

Additionally, (1) contains a cataphoric pronoun *she* which triggers an active search for an antecedent, whereby the parser seeks the antecedent only in grammatically sanctioned positions, such as where the antecedent is not c-commanded by the pronoun (Kazanina

et al., 2007; van Gompel and Liversedge, 2003). In (1), the closest potential antecedent is *Annie*. However, it can be the antecedent only if the DP that contains it is analyzed as the MS, thus outside the whenever-clause and not c-commanded by *she*. A bias toward an early cataphoric dependency formation could lead the parser to analyze the ambiguous DP as the MS. In (1), there is a bias toward a MS analysis from the antecedent search in addition to a bias toward the local attachment EO analysis.

I find that, regardless of the transitivity bias of the verb in the position of *hums*, the parser forms a dependency between the pronoun *she* and *Annie*. This indicates that dependency formation can supersede verb frame preferences and any default preference the parser may have toward local attachment (Phillips and Gibson, 1997). Moreover, I also observe effects attributable to both the MS and EO parses. This suggests that the parser builds both alternatives and maintains them in parallel. From this, I conclude that the parser prioritizes information from an ongoing dependency search over lexical properties during ambiguity resolution.

Acknowledgements

The profound gratitude I feel toward everyone who has helped me along the way, in any capacity, cannot be fully expressed in writing. For all of the power of language which we study as linguists, words simply cannot do my emotion justice. It's a pretty amazing feeling to be at this point where writing an acknowledgements section is appropriate, and not pre-emptive. It has been a very bumpy road for me, and I owe my (scholarly) survival to countless people. I will do my best to name those who played a significant role in my graduate career, but this list is by no means exclusive.

First and foremost, none of this would have been possible without the support of my advisor, Masaya Yoshida. His patience with my confusion and his critical comments about my writing and thinking have shaped me as a scholar. He has taught me more about the scientific method and clear, explicit writing than I thought was possible for me to learn. My gratitude goes to Nina Kazanina, whose own dissertation inspired mine, and who has been invaluable in planning and interpreting the experiments herein. Many thanks to Matt Goldrick for his careful, insightful feedback and discussion of my analyses and statistical methods.

Beyond my dissertation committee, many others contributed to my progress and the quality of the final product presented here. Brian Dillon suggested the method for checking the correlation between the two critical regions, which lead me to the surprising conclusion that, in fact, the parser represents both alternatives in real time. Megan Risdal talked me through some particularly sticky statistical analysis and terminology, and offered to read some of the longest and driest sections for clarity. Laura Bailey was an excellent accountability buddy, in which she would tweet at me to start working in the morning, and I would tweet to go to bed at night. Maryam Bakht and Barbara Clark were wonderful professional role models, who both talked me down when I needed nonjudgmental adult supervision, and offered to proofread early drafts. Shiloh Drake has been an amazing friend and supportive writing buddy, who has always had time for me and who introduced me to the fantastic group who I now refer to as GALizona. GALizona (Shiloh Drake, Mary-Caitlyn Valentinsson, Elise Bell, and Megan Figueroa) as a group and as individuals have been wonderful, motivational grad school compatriots, despite the hundreds of miles between us. Ann Burchfield has been a dear friend and colleague with whom I've explored cities and trails, and shared many deep conversations about linguistics and life. Mike Frazier has been a fantastic collaborator and friend. We have battled code and equipment and manuscripts side by side, and it has been an honor to do so. Karen Chu has been the most generous and supportive friend throughout my time in grad school. It was written in the stars that we would cross paths and help each other achieve our goals. Even though we've gone different directions in our careers, it is Karen who I think best understands the scholarly path I've taken (not least because of her skillful editing!). In contrast, my sister in spirit Sonia Saraiya and I have taken very different career paths, yet we find ourselves in similar places. After having such different interests for over a decade, we end up both spending our time writing and finding ways to structure our time so that we don't get sucked into Netflix (albeit for different reasons!).

I would be remiss if I did not acknowledge the incredible community of linguists on Twitter (my so-called twitterlings). This community of sharing knowledge, ideas, insights, institutionally-restricted access to peer-reviewed articles, and mundane happenings has kept me grounded and motivated. Notably, Gretchen McCulloch has, in my mind, served as my foil, inspiration, and a window into the world of popular science communication. She helped me to see how wonderful life can be outside academia and showed me that I can continue my work as an advocate and communicator of linguistics from within academia. (And I can leave the time-consuming projects to her, since she's getting paid to do it!)

My family has been a tremendous source of strength and motivation for me. My grandmother, Sydel Ackerman, asks me to explain what I'm working on, listens with patience and wonder, and with her keen insight, asks me pointed questions that help me find ways to talk about my work without resorting to jargon. Her support, especially within the last (tumultuous) year, has been invaluable. My parents have given me unconditional love and support my whole life, the value of which has become particularly clear throughout grad school. My mother, Adrienne Tymiak, always has time to talk me through dealing with impostor syndrome, based on her own experiences and wisdom. She showed me the value in forging a unique path through life and career, no matter how difficult. Her strength and tenacity taught me how to fight for my dreams. My father, David Ackerman, instilled in me the value of honesty and integrity from a young age, and has kept me curious about the world in all of the infinite ways one can study it, scientifically and artistically. He helped me hone a process for thinking logically (which I have always struggled with) by use of flowcharts, and pushed me to think about problems from new perspectives. My brother, Paul Ackerman, who is also working to finish his dissertation at this moment,

has probably been the most influential teacher in my family. He is so incredibly smart and quick to pick up new concepts, it has been an honor to discuss linguistics and natural language processing with him. In turn, he has taught me about chiral liquid crystals. Our discussions have taught me to think more computationally, and have strengthened my work ethic and tenacity. He is one of the most important role models in my life.

In turning my attention to my most immediate family, I will address Leeroy Jenkins and Loki with a quote from Ross (1967), "Since in my view, cats are as necessary as air or water, I thank our cats … for deigning to stay with us and seasoning our existence." But to return to a more sombre tone, it is Thomas Bergman to whom I am most deeply and sincerely grateful. Without Tom's continuous, loving presence in my life for the last ten years, none of this would have been possible. He has made me laugh when I felt hopeless, fed me when I was overwhelmed, and made sure I got out and interacted with friends occasionally. It is for these reasons, among many, many others, that I dedicate this dissertation to Tom, the most wonderful, loving, open, and honest spouse a person could dream of.

Table of Contents

ABSTRACT 3

Acknowledgements 5

List of Tables 15

List of Figures 16

Chapter 1. Introduction 18

Chapter 2. Survey of Literature 25

2.1. The parser 25

2.1.1. Serial parsers 27

2.1.1.1. Kimball (1973) 27

2.1.1.2. Frazier and Fodor (1978) 27

2.1.1.3. Crain and Steedman (1985) 31

2.1.2. Parallel parsers 33

2.1.2.1. Boland and Cutler (1996) 35

2.1.2.2. van Gompel et al. (2000) 36

2.1.2.3. McRae and Matsuki (2013) 38

2.1.2.4. Konieczny (2005) 39

2.1.2.5. Lewis (2000) 40

2.1.2.6. Gibson and Pearlmutter (2000) 45

2.2. The influence of lexical frequency 49

2.2.1. Tanenhaus et al. (1989) 50

2.2.2. MacDonald et al. (1994) 52

2.2.3. Trueswell and Kim (1998) 58

2.2.4. van Gompel et al. (2012) 59

2.3. Local attachment 61

2.3.1. Minimal Attachment 62

2.3.2. Phillips (1995) 64

2.3.3. Phillips and Gibson (1997) 65

2.3.4. Gibson (1998); Gibson and Pearlmutter (2000); Gibson et al. (1996) 69

2.4. Ambiguity resolution 71

2.4.1. Frazier and Rayner (1982) 72

2.4.2. Ferreira and Henderson (1990) 76

2.4.3. Staub (2007) 77

2.4.4. Ferreira and Henderson (1991) 79

2.4.5. Tabor and Hutchins (2004) 82

2.4.6. Christianson et al. (2001) 84

2.4.7. Sturt (2007) 89

2.4.8. Slattery et al. (2013) 92

2.5. Long-distance dependency formation 96

2.5.1. Wh-dependency formation 96

2.5.1.1. Stowe (1986) 97

2.5.1.2. Traxler and Pickering (1996) 99

2.5.1.3. Frazier et al. (2015) 100

2.5.2. Coreference dependencies 106

2.5.2.1. Cowart and Cairns (1987) 107

2.5.2.2. van Gompel and Liversedge (2003) 108

2.5.2.3. Kazanina et al. (2007) 110

2.5.2.4. Kwon and Sturt (2014) 111

Chapter 3. Influence of equibiased verbs on cataphoric dependency formation 118

3.1. Introduction 118

3.2. Background 126

3.2.1. Garden paths 127

3.2.2. Argument structure ambiguity 130

3.2.3. Models of the Parser 132

3.2.3.1. Serial Models 134

3.2.3.2. Parallel Models 134

3.2.4. Conflicting Evidence for the Influence of Dependency Formation 136

3.2.5. Cataphoric coreference dependencies 138

3.2.6. Interim summary 142

3.3. Local Attachment and Cataphoric Dependency Formation: Predictions 142

3.3.1. Embedded Object/Matrix Subject Ambiguity with No Cataphor 145

3.3.2. Embedded Object/Matrix Subject Ambiguity with Cataphor 153

3.3.2.1. Before reaching the ambiguous region 153

3.3.2.2. Matrix subject parse 162

3.3.2.3. Embedded object parse 164
3.3.3. Parallel parsing, or Multiple Outputs Model 167
3.4. Pilot 1: Offline Coreference Acceptability 177
3.5. Pilot 2: Offline Forced-Choice 182
3.6. Pilot 3: Offline Acceptability with Line-Breaks 188
3.7. Eye tracking task 193
3.7.1. Eye-tracking methodology 193
3.7.2. Materials 194
3.7.3. Predictions 199
3.8. Results 199
3.8.1. Ambiguous NP critical region (R-expression) 201
3.8.2. Matrix verb critical region 205
3.8.3. Matrix verb spillover region 209
3.8.4. Discussion 213
3.9. General Discussion 218
3.10. Conclusion 224

Chapter 4. Influence of transitive-biased verbs on cataphoric dependency formation 227
4.1. Introduction 227
4.2. Background 229
4.2.1. Transitive verbs 229
4.2.2. Intransitive verbs 234
4.2.3. Interim Summary 241
4.3. Pilot 4: Offline Forced Choice Task 242

4.3.1. Materials and Procedure 242

4.3.2. Predictions 245

4.3.3. Analysis 246

4.3.4. Results 250

4.3.5. Discussion 251

4.4. Eye-tracking task 251

4.4.1. Materials and Procedure 252

4.4.2. Predictions 253

4.4.2.1. Hypothesis 1: Verb transitivity is strong 254

4.4.2.2. Hypothesis 2: Verb transitivity is not strong 260

4.4.2.3. Effects at the Matrix Verb 263

4.4.3. Analysis and Results 266

4.4.3.1. Ambiguous NP region 267

4.4.3.2. Matrix Verb region 271

4.4.3.3. Matrix Verb Spillover region 271

4.4.3.4. Second Antecedent region 273

4.4.4. Discussion 276

4.5. General Discussion 277

4.6. Conclusions 279

Chapter 5. General Conclusions 280

5.1. Discussion of results 280

5.1.1. Equibiased verbs 280

5.1.2. Transitive verbs 281

5.2. The parser's behavior 282

5.3. Conclusions 296

5.4. Future directions 298

5.5. Summary 300

References 301

Appendix 312

Equibiased verb stimuli 312

Transitive biased verb stimuli 332

List of Tables

3.1 Means (and Standard Errors) for Eye-tracking Experiment 202

3.2 Combined ANOVA and LME results for Ambiguous NP 203

3.3 Combined ANOVA and LME results for Matrix Verb 211

4.1 GLMER results for Pilot Experiment: FS Forced-Choice 246

4.2 Within-Condition Analysis for Pilot Experiment: FS Forced-Choice 247

4.3 GLMER results for Pilot Experiment: SC Forced-Choice 247

4.4 Within-Condition Analysis for Pilot Experiment: SC Forced-Choice 247

4.5 Means (and Standard Errors) for Eye-tracking Experiment 267

4.6 Combined ANOVA and LME results for Ambiguous DP 268

4.7 Combined ANOVA and LME results for Matrix Verb 272

4.8 Combined ANOVA and LME results for Second Antecedent 275

List of Figures

3.1	Scaled Likert Ratings for Pilot 1: Forced Coreference Acceptability	180
3.2	Proportion of Responses in Pilot 2: Matrix Subj. vs. Embedded Obj.	186
3.3	Likert Ratings for Pilot 3: Forced Coreference Acceptability	192
3.4	Sample presentation screen for eye-tracking task.	198
3.5	First Pass RT (Ambiguous NP Critical region)	204
3.6	Re-read Duration (Ambiguous NP Critical region)	206
3.7	Total Time (Ambiguous NP Critical region)	207
3.8	First Fixation/First Pass Reading Times (Matrix Verb Critical region)	208
3.9	Total Time (Matrix Verb Critical region)	210
3.10	First Pass RT (Matrix Verb Spillover region)	212
3.11	Regression Path RT (Matrix Verb Spillover region)	213
3.12	Reread Duration RT (Matrix Verb Spillover region)	214
3.13	Total RT (Matrix Verb Spillover region)	215
3.14	Correlation between RT at the ambiguous NP and matrix verb	217
4.1	Diagram adapted from van Gompel et al. (2012)	240
4.2	Proportion of FS Responses in Pilot 4: MS vs EO	248

4.3	Proportion of SC Responses in Pilot 4: MS vs EO	249
4.4	First pass reading time at the ambiguous NP *Annie's*	270
4.5	RP durations at the matrix verb spillover region	274
4.6	FP and RP durations at the second antecedent region	276

CHAPTER 1

Introduction

The human linguistic processing mechanism is able to navigate ambiguity with astonishing ease. While there are certainly cases where ambiguity or its associated processing cost overwhelms a listener or reader and makes a grammatical sentence impossible to interpret, these cases are few and uncommon in daily life. Ambiguity, and specifically temporary structural ambiguity, is not uncommon, and often ambiguities pass by unnoticed but interpreted.

The purpose of this dissertation is to investigate how the parser balances different types of linguistic input while it is building the initial structural representation of the incoming sentence. The parser, which is tasked with structure building, has been shown to be sensitive to information beside syntactic category, such as lexical frequency and lexical frame frequency (Frazier and Rayner, 1987; Lapata et al., 2001; MacDonald et al., 1994), semantic attributes like thematic role and plausibility (Traxler and Pickering, 1996), and grammatical constraints (Kazanina et al., 2007; Stowe, 1986). In addition, there seems to be some principle of the parser that prefers to attach incoming material locally, i.e., to the most recently built structure, even across languages (Futrell et al., 2015; Grillo, 2012). This, however, is not a strictly ranked hierarchy of influence, and the combination of multiple sources of information may pull the parser in a different direction from the otherwise “default” structure (Phillips and Gibson, 1997). Based on this, I choose two sources of information that may bias the parser and put an alternate structure in competition with the most recently built structure. Specifically the sources of information used for this purpose are verb frame frequency, which has a body of literature that is inconclusive regarding its strength of influence on the parser, and coreference dependency formation, which is a grammatically non-obligatory but eager process.

Verb frame frequency refers to how often a verb occurs in a particular configuration of arguments. For this study, I will be examining two frames: one with a direct object, and one with no following arguments. These correspond with monotransitive and intransitive argument structures, and I will use the terms for the frames and argument structures interchangeably. The crucial property of these two frames is that a verb may appear before a noun phrase, and if this verb is able to occur in either of the frames, it is ambiguous whether the following noun phrase is the direct object of that verb, or associated with some other structure. In this study, the noun phrase following such an ambiguous verb could be either a direct object of the verb, which is in an embedded clause, as underlined in (1-a), or it could be the subject of the main clause that follows the embedded clause (1-b).

(1) a. Whenever John leaves the house it is dark.
    b. Whenever John leaves the house is dark.

This type of object/subject ambiguity has been investigated in other contexts (Ferreira and Henderson, 1991), and revealed to be sensitive to verb frame frequency (Ferreira and Henderson, 1990; Frazier and Rayner, 1987). It also creates a construction that allows testing of the parser's behavior when receiving another influential source of information: the search for a tail of a dependency.

If the sentences in (1) are slightly modified so that a pronoun is located in the embedded subject position, this pronoun will trigger the search for its antecedent (2), which will look rightward through the parse (van Gompel and Liversedge, 2003).

(2) a. Whenever he leaves John's house it is dark.

b. Whenever he leaves John's house is dark.

Furthermore, this antecedent search is sensitive to and respectful of grammatical constraints, such as Binding Constraints (Chomsky, 1981; Kazanina et al., 2007). That is, the parser will not attempt to form a dependency between locations where such a dependency would violate a Binding Constraint, such as Binding Condition C.

Binding Condition C (BCC): An R-expression (e.g., a lexical noun phrase) must be free, i.e. not c-commanded by a coreferring pronoun.

Thus, in (2-a), a dependency between *he* and *John* is illicit and the parser should not attempt to form a dependency between these two locations, since *he* c-commands *John* and an R-expression like *John* cannot be c-commanded by a pronoun with which it corefers (BCC, Chomsky, 1981). However, the initial string of words in (2-a) is the same as that in (2-b). If the parser will attempt to form a dependency between *he* and *John*, by the time that the parser reaches *John*, the ambiguity has not been resolved. Since *leaves* is equally likely to be monotransitive and intransitive (i.e., it is equibiased), it is not clear whether or not such a dependency would violate Binding Condition C (2-a) or be completely licit (2-b). In this manner, the parser must balance the information available from the lexical properties (e.g., verb frame frequency) with the information about where dependencies can and cannot be formed. If the parser builds a dependency between *he* and *John* before the disambiguating material is reached, it runs the risk of having violated a grammatical constraint. On the other hand, if the parser does not build this dependency, it is not behaving in the eager manner it has been described as having (Kazanina et al., 2007; Omaki et al., 2015; Stowe, 1986).

This is the crux of the issue which this dissertation addresses. The parser may face choices that could result in major violations of grammatical principles, but otherwise follow the behavior trends that have been observed. How, then, will the parser react to being put in such a situation?

If the parser prioritizes information from the verb frame frequency, then when the embedded verb is equibiased, the influence of the dependency search is competing only against the parser's preference to attach incoming material locally. Then, if preference for local attachment is stronger than an active search for the tail of a coreference dependency, the parser will not form a dependency between *he* and *John* because it will place *John* in the embedded object position. Alternatively, if the dependency search is stronger than the preference for local attachment, the parser will form the dependency between *he* and *John* in order to complete the dependency as soon as possible. In fact, I find evidence that the parser builds both alternatives. That is, evidence is found at the name *John* that is indicative of dependency formation only possible in a structure like (2-b). But also, evidence is found at the matrix verb that the parser was not expected to find the matrix verb, which is indicative of a structure like (2-a) where the parser would instead be expecting to find a matrix subject following *John's house*. Furthermore, these two observations were not found to co-vary, which suggests they are present simultaneously. If this is the case, it is very strong evidence that the parser can build and maintain multiple structures in parallel (Boland and Cutler, 1996; Gibson and Pearlmutter, 2000; MacDonald et al., 1994; Pearlmutter and Mendelsohn, 1998, 9; Trueswell and Kim, 1998; van Gompel et al., 2000; Cf. Lewis, 2000).

To further characterize these competing influences, sentences like (3) were adapted from sentences like (2) in order to increase the strength of the influence from the verb frame frequency. In (3), the verb *visits* is highly biased toward a transitive frame. Dependency formation was shown to be stronger than whatever influence is exerted by an equibiased verb and the local attachment preference. So in order to properly characterize the relative strength of the influence from dependency search, the bias introduced by the verb was increased.

(3) a. Whenever he visits John's house it is dark.

b. Whenever he visits John's house is dark.

In fact, the parser still forms a dependency between *he* and *John*, despite the stronger bias from the verb frame frequency information. From this, it is clear that the bias from coreference dependency formation strongly influences the parser, much more so than the verb frame frequency does. One way this could occur is if coreference dependency is a relatively stronger influence than verb frame frequency, and even when the likely verb frame and the local attachment preference collude to pull the parser toward an embedded object representation, the draw from the dependency search is still strong enough to "want" *John* to be within the matrix subject position and serve as the first candidate antecedent.

Furthermore, it is observed that the reading times at the matrix verb in sentences like (3-b) is are fastest when the gender between the pronoun corresponding to *he* and the candidate antecedent corresponding to *John* are incongruent. One reason this may come about is that the parser is so strongly guided by the demands of dependency formation that it predicts the internal structure of the main verb, for instance, building the structure necessary for the verb to take a complement noun phrase before any bottom-up information about the matrix clause has been encountered. This might happen in order to posit the next possible location for a candidate antecedent. This would only take place when the first candidate antecedent was insufficient, i.e., when the gender between the pronoun and name mismatched. Then, when the parser reached the position of the matrix verb in the condition in which the gender of the pronoun mismatched the proper name, the matrix verb was read faster because its internal structure had already been predicted. This would be consistent with studies on other types of dependency formation, in which the parser was shown to predict gaps in advance of the verb (Omaki et al., 2015; Traxler and Pickering, 1996).

These results characterize the way the parser makes use of available information and the way the parser prioritizes sources of information when there is a conflict between them. For instance, the parser is most strongly influenced by dependency formation, so much so that the drive to complete a coreference dependency is attended to even when the verb frame information would strongly suggest such a dependency could violate a grammatical constraint. However, this influence from dependency formation does not force the parser into building just one representation that would satisfy the dependency early. Rather, the parser apparently builds and maintains multiple representations in parallel. These observations depict a parser that is sensitive to grammatical constraints and does not risk violating them in order to satisfy the strong demands of dependency formation. Instead, it follows through with the needs of an active dependency search in parallel with the bias presented from the probabilistic lexical information.

The implications of a parallel parser are particularly relevant for the interpretation of studies on the so-called digging-in effects observed after temporary ambiguities (Christianson et al., 2001; Slattery et al., 2013; Tabor and Hutchins, 2004). These studies observe that readers maintain the parse associated with an initial misinterpretation, even after the reader has completed reading the sentence, past the point of disambiguation. Typically, the discussion of these results assume a serial model of the parser, which has led to placing the responsibility for the lingering misinterpretation on the semantic processing mechanism or modeling the parser as fallible. However, with a parallel model of the parser, this effect can be attributed to the parser building both the globally grammatical parse and the initial misinterpretation, but also maintaining the incorrect parse in memory after disambiguation. In this way, the parser can build and interpret both the initial misrepresentation and globally grammatical representation, without sacrificing the rich structure necessary for the application of grammatical constraints.

CHAPTER 2

Survey of Literature

2.1. The parser

When discussing the way in which humans process language in real time, the mechanism that is responsible for constructing a syntactic representation of an incoming sentence is called the parser. The parser is responsible for building the structure of sentences based on (and constrained by) grammatical principles. These properties of the parser range from strict constraints like Binding Principles (Chomsky, 1981) and selectional restrictions, to broader principles that are hypothesized to be the source of general trends observed in English, such as a bias toward local attachment (Phillips, 1995; Phillips and Gibson, 1997) and the influence of lexical frequency on ambiguity resolution (MacDonald et al., 1994, and references therein). By investigating how these properties interact during real-time structure building, and specifically ambiguity resolution, I will characterize the influence that coreference dependency formation exerts on the parser's behavior relative to the influence exerted by the bias toward building locally attached structures.

Psycholinguistic models of the parser were initially proposed to be relatively naïve, operating autonomously of other hypothetical modules in the language processing mechanism (i.e., without making use of any information in the input beyond broad syntactic category). While it was clear that these strongest versions of the autonomous parser were not sufficient to describe the parser's behavior when it encountered certain salient properties not encoded in broad syntactic category information (i.e., lexical frequency or plausibility), classic models of the parser were able to account for other phenomena observed in reading studies. One of the major classic models of the parser was put forth by Frazier and Fodor (1978), which they affectionately dubbed the Sausage Machine (in reference to the behavior of the first stage of their model). This approach followed from Kimball (1973), which described a model that enumerated seven principles by which the parser should operate. This model had shortcomings whereby it struggled to account for constructions in which the parser seemingly chose to construct a more costly attachment (according to the model), so Frazier and Fodor designed their model to explain the parser's behavior.

This chapter will summarize and discuss what types of information influence the parser's behavior, especially with regard to the processing and resolution of syntactic ambiguity. In particular, I will discuss the differences between serial models of the parser, in which only one representation is built and maintained at a time, thus requiring reanalysis to retrace and rebuild syntactic representations, and parallel models of the parser, in which multiple syntactic representations may be present simultaneously, thereby introducing a different type of reanalysis procedure. Serial parsers have historically been at the center of a very fruitful and popular discussion on the incrementality and construction of syntactic representations. However, modern discussions of the evidence necessary to distinguish between a serial or parallel parser have struggled to find a point at which these classes of models necessarily make different predictions. The following section focuses on the literature that outlines the behavior of the parser, and sets up the following discussion of the ways in which the parser behaves upon encountering structural ambiguities.

2.1.1. Serial parsers

2.1.1.1. Kimball (1973). Kimball's two-stage parser, for instance, was mostly insensitive to the lexical content of the input in the initial chunking stage, with a limited look-ahead system that varied by language in how far ahead it could look before constructing a representation. Incoming lexical items were inserted into a tree based on syntactic category, with the breadth of look-ahead determining how much contextual influence the parser could be aware of. While this type of contextual awareness is similar in function to the interactive and probabilistic serial parsers discussed in recent literature, it differs in that it does not explicitly have access to frequency statistics of the current lexical item in input, but rather uses the upcoming input to determine how to treat ambiguous words. This requires attributing an arbitrary look-ahead distance to the parser, rather than describing the apparent contextual awareness of the parser in terms derivable from other supported properties. In contrast, as will be discussed below, probabilistic parsers make use of the frequency statistics encoded in the lexicon during construction of structurally ambiguous input. Once the parser chunked the input into processing units, each unit would then be shunted to a stage where semantic interpretation and transformational operations could take place. In coordination with look-ahead, this was designed to explain how the parser could fall victim to a garden path while still seeming to be aware of other contextual information.

2.1.1.2. Frazier and Fodor (1978). Frazier and Fodor's model took advantage of the limitations of having a two-stage mostly autonomous approach, and described the first stage as having a naïve chunking method in which the size of each chunk is determined by approximate length (potentially up to six words in English, but this was suggested to vary across languages). By limiting chunks to this certain size, the first stage of Frazier and Fodor's model (the metonymous "Sausage Machine") could account for the parser's difficulty with garden path sentences. It was in the second stage of their parser that the internal structure of the chunks was evaluated. This mechanism was sensitive to semantic and contextual information, but could only build structure for whatever material was in the chunk that the first stage had most recently delivered to it. Thus, if a constituent is much longer than a chunk, it is more likely to be misanalyzed as two constituents and attached to the rest of the parse tree in a globally ungrammatical way. For example in (1), at school is the place at which the boy dropped the model airplane, where as in (1-b), at school is where the boy had been making the model airplane. According to the limitations on the size of a chunk produced by the first stage of the parser, the embedded noun phrase in (1-b) *the delicate model airplane he had so carefully been making at school* would necessarily be divided across chunks, and therefore more likely to be misinterpreted as having the same structure as (1-a).

(1) a. I saw a boy who dropped the delicate model airplane he had so carefully been making at school.

    b. I saw a boy who dropped the delicate model airplane he had so carefully been making at school into a puddle cry.

The garden path phenomenon observed when reading this type of sentence played a pivotal role in the development of theories of the parser, including the Sausage Machine. Bever (1970) discusses how these sentences are grammatical but somehow are altogether unacceptable and almost impossible to parse correctly, even with prior knowledge about the sentence's structure. This type of structure, which is temporarily ambiguous and ultimately resolved in an unexpected way, makes for a rich testing ground for the parser's behavior. In order to model how the human parser seems to get "led down the garden path," many early models of the parser made use of a nave structure building mechanism that could be fooled into building the wrong syntactic structure, and then a more pragmatically or semantically aware mechanism that would evaluate and assign an interpretation to the structure. To this end, Frazier and Fodor's "Sausage Machine" model of the parser tries to streamline the two stages by having a completely nave first stage that just chunks the input into lengths (sausage links) and the second stage puts them together into a structural representation (Frazier and Fodor, 1978).

In Frazier and Fodor (1978) a principle called Minimal Attachment is proposed in which an incoming word attaches at a location in the tree that would require a minimal amount of new structure to be built. This was designed to explain the observed preference for (2) over (3).

(2) I bought the book for Susan. (≈ The book was bought to give to Susan by me.)

(3) I bought the book for Susan. (≈ The book that was for Susan was bought by me.)

Here, the PP *for Susan* is attached to the VP rather than at the most recent possible place as a modifier of *book*. This motivates their claim that each new word is not attached into the most recently built structure, but to a point that requires no additional structure to be built (assuming that the verb *bought* already has an attachment site available for the argument *for Susan*, but the NP *the book* does not). This falls out from their other stipulations about the parser. Given this, a sentence like (4-a) might be chunked as follows in (4-b):

(4) a. Whenever Mia hums Annie's melody gets stuck in her head.

b. Whenever Mia hums — Annie's melody — gets stuck in her head

The second stage of the parser has a few options for how to proceed. Regardless of whether it puts *Annie's* in the first or second chunk, the second stage in which an interpretation of the sentence is calculated will have to figure out the structure of the first chunk. The first chunk contains the word *hums*, which is ambiguous between having a monotransitive frame and an intransitive frame and both frames are equally frequent. The "equibiased" nature of this verb introduces an ambiguity in which there potentially is an object position following *hums*, in which case *Annie's melody* would occupy that object position. However, this object position is not obligatory, the parser recognizes that verbs can have objects but do not always have to, and moreover, another position for a noun phrase may have already been anticipated. Since *Whenever* indicates the beginning of an adverbial clause (which modifies a main clause), the subject position of the main clause may already be made available, thus attaching *Annie's melody* as the obligatory subject of the main clause would require less additional structure to be built. So, in following the predictions of Minimal Attachment, a parser would preferentially attach *Annie's melody* as the subject of the main clause.

(5) a. Whenever Mia hums Annie's melody it sounds gorgeous.

b. Whenever Mia hums — Annie's melody — it sounds gorgeous

In the case of (5), the parser would behave in the same way as in (4), since the words and thus broad syntactic categories are identical in the first chunk (or two). When the parser reaches it, the subject of the main clause, it may boggle, since it has already attached *Annie's melody* to this location, based on Minimal Attachment. At this point, the parser may have to reanalyze the input and representations it had built, or it may simply give up due to the high processing cost of reanalysis (especially in more complicated sentences that would require more working memory). Crucially for this model, the parser is not sensitive to probabilistic or frequency information, so regardless of the equibiased nature of *hums*, the parser will preferentially attach *Annie's melody* to the subject position in the main clause.

2.1.1.3. Crain and Steedman (1985). This type of two-stage autonomous parser seems to account for many different scenarios, but it still is insufficient to describe the speedy way in which readers are able to cope with a wide variety of lexical items that are not all equally likely to pattern with the most common constructions associated with their broad syntactic category. In contrast with the autonomous parsers, there are interactive parsers which are able to identify and react to information beside broad syntactic category (Crain and Steedman, 1985). The class of interactive parsers includes parser that do not necessarily take into account all possible sources of information at every incoming word, but might have access to only a couple of the most useful or most influential information sources during initial structure building. One possible source that has been discussed as being particularly relevant to reading studies is the frequency with which verbs occur in various lexical frames. For instance, if an interactive parser encounters the string in (6), it will have access to the fact that *hums* can have (at least) two frames, one intransitive and one monotransitive.

(6) Whenever Mia hums Annie's melody ...

a. it sounds gorgeous. [Embedded Object parse]

b. gets stuck in her head. [Matrix Subject parse]

The parser would also have access to the fact that guess in (7) is more likely to be intransitive than transitive.

(7) Whenever Mia guessed Annie's melody ...

a. it turned out to be the wrong one. [Embedded Object parse]

b. popped into her head. [Matrix Subject parse]

Upon encountering (7), the parser will be able to make the most probable choice, which means it will be conserving resources by choosing the parse that leads to (7-b). This option is more likely to succeed and therefore is likely to not required the additional cost of reanalysis, based on the higher probability that guessed occurs in a monotransitive frame. In this model of the parser, contextual information is able to steer the parser away from potentially costly mistakes. This is another way that the classic literature motivates the appearance of garden path effects: if the globally grammatical representation is an infrequent one (based on lexical probability), then since the parser apparently chooses the globally ungrammatical, but more frequent representation at first, it must have some access to information besides syntactic category.

Still, these models overlook other potential properties that the parser may have. The parsers discussed by Crain and Steedman (1985) and by Frazier and Fodor (1978) are assumed to be serial. This means that only one potential configuration of the input is considered at a time. Some of the motivations of assuming this type of parser include overly strong claims about the alternatives. The comprehension mechanism must have limited memory capacity, as it is a human capability, so there cannot be an infinite number of potential parses generated and stored in parallel. Since unlimited or unrestricted descriptions of parallel parsers have no way to prevent a parallel parser from generating every possible parse at once, parallel models of the parser have often taken a backseat to serial models. Once a model of the parser (serial or otherwise) has a mechanism for accessing probability information during initial structure building, the breadth of representations generated in parallel can be constrained or limited based on statistical principles, in a similar manner to serial parsers. This makes parallel models of the parser viable alternatives to serial models.

2.1.2. Parallel parsers

Discussion of the properties of serial parsers also bolsters aspects of parallel parsers. For instance, serial and parallel parsers receive input in the same way and must both deal with incremental interpretation. Neither serial nor parallel models of the parser, as classes, are privy to the whole incoming sentence at the start, and neither has the built-in capability of predicting the content of input (i.e., predicting the lexical content of the head of a noun phrase like "the big, red..."). Even in parallel models of the parser, input is processed and incorporated in a comparable way to serial parsers, with the potential for multiple paths being pursued simultaneously.

This type of similarity between serial and parallel parsers is part of the motivation to describe the class of parallel parsers called Ranked Parallel Parsers (Gibson, 1991). In this model of the parser, multiple representations are generated, but only the top-ranked parse is ultimately selected for interpretation. Gibson suggests that the choice of the parse which is ultimately ranked highest and selected for interpretation could be based on which parse requires the least memory cost, thus lightening the parser's processing load. This is not the only way such a ranked parallel model could choose which parse to interpret (e.g., Gorrell, 1987; Kurtzman, 1985; Pritchett, 1992), but to explore the other ways in which such a parser could operate would take the discussion too far afield.

For all the ways in which parallel models may be similar to serial models, there are still issues with how a parallel parser would generate the same patterns of reading time slowdowns as a serial parser, if the correct parse is being generated in parallel models and not in serial models. For instance, if multiple representations of a sentence are being generated, does holding this increased volume of material in memory slow the parser down? Or, alternatively, what mechanism causes the reading time slowdown associated with ambiguity resolution, without also creating a slowdown during the ambiguous region itself (Gibson and Pearlmutter, 2000)? Boland and Cutler (1996) compare the various types of parallel models in contemporary literature and discuss what properties are most plausible for a parallel parser (or "multiple output parser"). They were able to swiftly discard the purely autonomous, deterministic models of the parser, as Crain and Steedman

(1985) did for serial parsers, but find other aspects somewhat more problematic to sort through.

2.1.2.1. Boland and Cutler (1996). Boland and Cutler (1996) explore a lexicalist stance on sentence processing, which was fairly popular in the relevant literature. They examine how lexical retrieval might be a mechanism that could be drawn upon (or directly used) for the identification and retrieval of syntactic structures. This contrasts with the modular theories of the general language processing mechanism that treat syntactic operations as independent from lexical operations, et cetera. This modularity is a key component of the syntax-first style of the serial model in (Frazier and Fodor, 1978). On the other hand, the interactive models allow some communication between modules of the language processing mechanism, meaning that the parser has access to lexical frequency information. This is crucial to many of the more interactive parsing models, and important to distinguish untenable parallel models from more realistic ones.

There must be a balance between the parser's efforts to build a representation and check sources of information available in the input (such as lexical statistics, semantic plausibility, etc.). In other words, it might be possible for the parser to avoid making mistakes if it is able to access and make use of every possible source of information in the input, but it has by nature a limited amount of resources in terms of time and storage, so there must be a way it chooses or prioritizes which sources of information to attend to and which sources to ignore (or deemphasize). For instance, if frequency information is highly prioritized, then one should see that the individual (context independent) frequency information for each word could account for the parser's mistakes and reanalyses. On the other hand, if the parser prioritizes thematic role plausibility (as an example), then one should see infrequent but more plausible constructions be built while the parser avoids being fooled into a garden path analysis by a more frequent frame that would create an implausible thematic role configuration (Crain and Steedman, 1985). There are other sources of information that could influence the ways in which the parser builds structures and optimizes its behavior, but for the purposes of this dissertation, frequency of argument structure is of primary concern.

Boland and Cutler (1996) also discuss the ways in which parallel parsers may be efficient and effective, as competing models to the serial parsers. An unlimited parallel parser is untenable due to the natural limitations of the human mind. So, as mentioned above, Gibson (1991) proposes a ranked parallel parser. The ranking of the parses is another operation that would necessarily draw from the comprehension mechanism's limited resource pool, so the cost associated with reanalysis in serial parsers may be comparable to the cost associated with (equivalent) reranking of parses when one turns out to be unviable and ungrammatical.

2.1.2.2. van Gompel et al. (2000). van Gompel et al. (2000) propose another type of multiple output model that accounts for the observed reading time data better than previous models, specifically of the two-stage variety and constraint-based (lexicalist) varieties. Since two stage models are typically of the syntax-first type, this means that the second stage is the point at which mistakes are gauged and reworked. In such a model, the context or information from the beginning of a sentence shouldn't influence the ease of processing later in the sentence. Yet, it seems that some contexts do indeed make sentences easier to parser (correctly) than other contexts.

In addition, constraint based models do not fit the processing data well either, according to the authors. They claim that constraint based multiple output models should slow down at the region of ambiguity because of the higher processing cost of holding multiple representations in memory at once. However, it is consistently observed that regions of ambiguity (before the disambiguating region) are read equally fast, if not faster, than unambiguous regions. Thus, van Gompel and colleagues claim that constraint based models do not account for the data, either.

They counter these models with one of their own: the unrestricted race model. In this model, the parser experiences difficulty based on how many revisions it must make, and different potential representations of ambiguous regions are adopted proportionally to their likelihood. Plus, in accord with the results of this study, global ambiguity is easier (faster) to process than local ambiguity, so competition cannot be what controls the observed slowdowns associated with processing cost. That is, in a globally ambiguous sentence, constraint-based models might predict that an ambiguous sentence could never be read faster than an unambiguous sentence, since the additional uncertainty would increase the processing cost. However, I am uncertain that the authors are representing the competing models of the parser accurately. It is not explicitly the case that multiple output models must necessarily slow down in ambiguous regions. A reasonable alternative is that the different representations make use of their shared structure. Rather than fully specifying and anticipating the potential ways in which the structure could resolve, it is possible that the parsers under-specify the syntactic structure, at least temporarily. This could account for the faster reading times observed in some ambiguous regions — fewer details represented would take up less space in memory, thus have a lower processing cost.

2.1.2.3. McRae and Matsuki (2013). McRae and Matsuki (2013) consider how a constraint-based parser might work when it does not slow down in regions of ambiguity. In this paper, they discuss what a constraint-based parser would require to function and account for known phenomena. The constraints in a constraint-based parser are akin to the types of information that could influence any interactive parser in that they will push the parser away from or toward building particular structures in particular circumstances. Some of these constraints might be lexical frequency, context (i.e., world knowledge about the entities and actions in the sentence), plausibility of thematic roles, and any number of other things. Anything that could affect the parser could be considered a constraint. Moreover, in the models described here, the information is available very early on, if not immediately. That is, independent of how long it takes the parser to compute how the information source could inform its structure building behavior, the parser will use this information as soon as it becomes available.

This description is vague and raises some issues with how to empirically distinguish it from other models of the parser. For instance, if something like gender agreement takes some time to compute but is available as soon as it's computed, that will not empirically differ from a situation in which gender information is not computed immediately, but is used in some later stage. Since the parser has presumably moved on and not waited for the gender information and since speed (or efficiency) is one of its priorities, then it remains to be explicated how one could differentiate between slow computation of an information source and a later, second stage of processing in which gender information is used, regardless of when this information becomes available. Furthermore, constraint based models are extremely flexible. This might be considered useful because the degree of influence of different constraints can be tweaked and then tested. However, this method of model development is not motivated by theoretical principles. In other words, even if a model were to mimic the human data perfectly, we would not know whether this was due to over-fitting the data, i.e., unnecessarily complex model components, or because we accurately modeled the comprehension mechanism. Moreover, there is no good way to motivate the reasons for each component unless there is a theoretical framework underpinning the constraints. While constraint-based models have had success in modeling empirical data, they fall short in terms of theoretical motivation.

2.1.2.4. Konieczny (2005). With such variety of models of the parser, finding ways in which to distinguish them and test their predictions against human performance is vital for improving understanding of the human parser. One context in which these various models of the parser can be tested is in the manifestation of phenomenon known as "local coherence". Local coherence effects are when the parser has traversed the sentence, presumably building a coherent structural representation. After some time, the parser reaches a string that (in isolation) would form a coherent clause or sentence on its own, and disregarding previously built structure, it then treats the locally coherent string as if it were independent. When the parser builds a representation of the locally coherent string that conflicts with information from earlier in the sentence, this provides evidence that the parser is fallible in its storage of the representation it maintains. In other words, there must be some property of a model of the parser that allows for this phenomenon in order to be accurate.

The consistent presence of local coherence effects in readers, regardless of reading span, is observed by Konieczny (2005). In order for a model of the parser to account for this, it would have to balance any predictive or top-down bias that could affect the parser's interpretation of the incoming words (Tabor et al., 2004; Tabor and Hutchins, 2004) with the bottom-up information such as lexical frequency and lexical frame probability. Konieczny concludes, based on this, that the process cannot be depth-first, since it is consistently missing out on the depth of the representation, independent of reading span. That is, a lack of detail in the representation may lead to the observed local coherence effects, and it's not isolated to effects observed for low-span readers. In addition, he concludes that serial parsers would require a depth-first style of processing, thus parallel parsers would be a better way to account for his results.

2.1.2.5. Lewis (2000). Thus far, I have described some of the processes underlying the development of models of the parser, so I will turn to their evaluation and comparison. The pair of papers by Lewis (2000) and Gibson and Pearlmutter (2000) summarize the situation succinctly, and offer a somewhat open-ended conclusion. To begin, Lewis discusses the pitfalls of testing different types of models of the parser in order to distinguish them. Rather that looking at specific instantiations of the parser, he examines broader classes, in an attempt to narrow the field of possible parsers quickly. He acknowledges that this, however, leaves open the possibility that particular instantiations may have tweaks that are compatible with the human data he uses to evaluate the models.

He introduces the kinds of models in four categories: simple parallel, ranked parallel, serial reanalysis, and probabilistic serial. He steps through the history and problems of each, which I will summarize here.

Lewis discards simple parallel models quickly because they do not provide a mechanism for limiting the number of alternatives calculated. Thus, they would necessarily overload the processing mechanism quickly, if not immediately, upon encountering the first potential ambiguity. This, however, is a very strong take on what a parallel parser would deal with. For instance, the conscious difficulty experienced during a garden path sentence illustrates that not all alternatives are equally available. Indeed, not all alternatives may even be computed, given the example sentence in Bever (1970) that can notoriously overwhelm even experienced linguists.

(8) The horse raced past the barn fell.

a. Garden path: The horse raced past the barn (and maybe the barn fell, if pressed)

b. Grammatical: The horse which was raced past the barn fell.

It is hardly a strong argument against parallel parsers to discount the simplest, strongest version of the simple parallel model. Thus, he next addresses the ranked parallel models, which are a more sophisticated version in which some alternative parses are ranked above others and are more easily accessible or take less processing power to compute. This would account for why mild (or subconscious) garden path sentences can be recovered from without the parser becoming overwhelmed, as it does in (8). For instance, the ambiguity in (9) could constitute a mild garden path, in which the parser prioritizes the interpretation (9-a) over (9-b), but is quickly corrected upon reaching the disambiguating region.

(9) Have the children put on their hats...

a. ...yet? ($\approx$ Is it the case the children have put on their hats yet?)

b. ...tomorrow. (≈ Tomorrow, make the children put on their hats.)

Lewis, however, expresses his doubt in the ranked parallel models. He uses the examples given in Fodor and Inoue (1994) to demonstrate how, according to a ranked parallel model, the ease of processing a sentence shouldn't hinge on the disambiguating material if all parses are already constructed and ranked, which is a contradiction of the observed human data.

(10) Have the troops march to the barracks yet?

(11) Have the troops march to the barracks tomorrow.

(12) Have the troops march to the barracks, will you?

In these examples, the first string of each sentence is identical. The main difference between (10) and (11)/(12) is that (10) is a question, whereas (11)/(12) are imperatives. Thus, in a ranked parallel model, a parser that has difficulty with (11) should have equivalent difficulty with (12), because the structures up until the end of *barracks* are identical. However, this does not seem to be the case. Readers end up having more trouble (i.e., exhibiting more slowdown) in sentences like (11) than in sentences like (12). Thus, by Lewis's reasoning, the ranking of parallel alternatives cannot be set before the disambiguating material is reached. This serves as his way to discount ranked parallel models as viable models of the parser.[1]

[1]Gibson and Pearlmutter (2000) disagree with this evaluation, suggesting that the ranking and re-ranking of parses may also vary in difficulty by the canonicity or clarity of the disambiguating material. For instance, (11) could be a similar structure to (10), but with mismatched tense. For instance, Have the troops march to the barracks today could either be a question or a command. In contrast, the tag question will you? is an unambiguous indication that the previous string was a request or command (an

Within the category of serial parsers, Lewis first addresses a category deemed serial reanalysis. This type of parser calculates the most preferred structure without variation (i.e., in a deterministic manner), and must backtrack or otherwise reanalyze the input when it realizes that it has made a mistake. This type of parser is following the tradition of the two-stage, syntax-first parsers in that it is deterministic. It can be more sophisticated than the original instantiations, in that it can be interactive and thus aware of other sources of information that could bias the parser toward some structures over others (such as information from lexical frequency, thematic roles, etc.).

Lewis (2000) specifically points out problems with deterministic serial parsing algorithms by calling upon work by Pearlmutter and Mendelsohn (1998, 9). In this self-paced reading study, the preferred sentential complement structure is also globally grammatical, and the plausibility of the dispreferred object-extracted relative clause (ORC) parse was manipulated.

(13) The report that the dictator [described]/[bombed (the country)] seemed to be false.

a. ORC1, Plausible: The report that the dictator described seemed to be false.
   RC: "the dictator described the report"

b. ORC2, Implausible: The report that the dictator bombed seemed to be false.
   RC: "the dictator bombed the report"

c. SC1: The report that the dictator described the country seemed to be false.

d. SC2: The report that the dictator bombed the country seemed to be false.

---

imperative), and not the question in (10). This means that (12) might be more easily disambiguated than (11), and should therefore be easier to retrieve, especially if it is a lower ranked alternative.

It was found that the implausible object-extracted relative clause parse (13-b) must have been calculated despite its low probability. Even though the sentential complement parses (13-c) and (13-d) are equivalently plausible to each other, they found that the sentences like (13-d) with implausible locally grammatical relative clause analyses showed increased processing difficulty at the "implausible" verb. This is strong evidence against a deterministic parser because such a model would only calculate the preferred, plausible reading for (13-d), thus never parse sentences like (13) with the dispreferred ORC analysis. Thus, serial reanalysis models of the parser are insufficient.

Finally, Lewis describes the probabilistic serial parser. This parser is similar to the serial reanalysis parser in that it is sensitive to various sources of information outside of broad syntactic category. Its main difference is that rather than always calculating the most likely parse, it computes parses proportionally to their likelihood. So, for instance, a verb that is used in a monotransitive frame 70% of the time will be computed as having a transitive frame 70% of the time. The remaining 30% will be computed as the other 30% of observed/used frames, at the equivalent proportion to the likelihood of each. This could account for how dispreferred parses are occasionally calculated and therefore how effects attributable to the dispreferred parses show up in reading times. Lewis seems partial to this category of parser, and does not provide any evidence that could falsify or conflict with this model's predictions. Moreover, given enough sensitivity to different sources of information, it may be possible for this type of model to correctly predict any resolution on occasion.

In order to test these different models, and specifically the probabilistic serial parser, there are a few things to be done. Since classes of models are too broad and variable to test, evaluation must be done on individual instantiations. This makes it slow work to narrow down what kinds of parsers are feasible. Once the models for testing are chosen, it is useful to focus on their predictions for slowdowns in reading time at the relevant region. Many parallel models would predict a reading time slowdown at the time of disambiguation due to the pruning, discarding, or otherwise dropping of parses that are no longer consistent with the input. Some serial models would predict a speedup at this time of disambiguation since the number of possible resolutions decrease, thus reducing the amount of information necessary to compute a way forward. That is, less ambiguity means the choice is easier to make. However, some race-based parallel models may also predict a speedup in reading time, as the processing cost of maintaining multiple alternatives decreases as parses "drop out of the race". Thus, any studies that specifically address the differences between serial and parallel models must explicitly state what these models predict in order for the results to be informative about the type of models that are feasible (Lewis, 2000).

2.1.2.6. Gibson and Pearlmutter (2000). In response to Lewis (2000), Gibson and Pearlmutter (2000) describe in which ways they disagree with his position. They first criticize Lewis's position on how alternative structures might vary based on the disambiguating information (as illustrated in (10), (11), and (12)). In addition, and more importantly for the current discussion of parsers, Gibson and Pearlmutter touch on the claim that parallel models predict a reading time slowdown during the ambiguous region. If a parallel model is resource-based, Lewis claims being required to maintain many alternative structures at once will cause a slowdown. Even in this case, the parser may only retain the few most likely parses, reducing the storage costs and preventing a measureable slowdown. Moreover, the parser may be able to share structure or representation between the alternatives, further reducing the chance of creating a measurable slowdown due to a high cost of storing multiple parses. Additionally, it may be the process whereby the parser identifies and discards representations that have been resolved to be incompatible with the input that creates a cost-driven slowdown. In this way, the reasons for which Lewis (2000) dismisses parallel models may be unsound.

Rather than discard parallel models as a class, Gibson and Pearlmutter suggest that parallel parsers are still viable given the distinguishing feature being the ability to hold multiple representations in parallel, rather than the constant maintenance of multiple alternatives, akin to Boland and Cutler (1996). In this way, the parallel parser is not obliged to compute representations for every possible continuation at every point, or even the most likely representations. Instead, the parser may have some criteria for when multiple alternatives are considered simultaneously. Given this, they suggest another way of falsifying models of the parser en masse is to divide the categorization into syntax-first parsers compared with parsers that allow multiple sources of input (referred to as interactive parsers, as in Crain and Steedman (1985)). Within each of these two categories, one can make a further division: probabilistic parsers and deterministic parsers.

Gibson and Pearlmutter (2000) agree with Lewis (2000) that syntax-first deterministic (serial) parsers are easily falsified based on the variation in reading times observed for minimal-pair sentences that differ in properties like lexical frequency or thematic role frames (Pearlmutter and Mendelsohn, 1998, 9). Probabilistic syntax-first parsers are not as easy to discard. These parsers are similar to the probabilistic serial parsers Lewis considers, and it is difficult to devise ways in which to falsify them. For instance, one way they consider is looking for a bimodal distribution in reading times at the disambiguation point. In order for such a result to be interpretable, though, they claim a stupendously large amount of data would be needed to distinguish two modes clearly. Moreover, the presence of a bimodal distribution is evidence that for this particular context, the parser acted in a serial manner, and this is not necessarily generalizable. Furthermore, the absence of a bimodal distribution is a null result, and therefore it would be unsound to base any claims on it.

Gibson and Pearlmutter also spell out some of the ways in which Lewis's take on falsifying parallel models of the parser may be limited to only certain instantiations. For instance, it is not necessarily the case that parallel models will predict a slowdown at the disambiguating region, as mentioned above. If the alternative parses share storage in working memory for where their structures are the same, it might be possible to store many more alternatives simultaneously without reaching capacity and showing a noticeable slowdown. It's also possible that the resources uses to build and store the alternative parses may be part of some other mechanism that doesn't tap into the same resources as the top-ranked (or winning) parse, although the details of this possibility are not spelled out. It is also possible that the competition between alternative parses occurs only in the interpretation of the representations (the semantic information), while the structures can coexist without competing with each other. These three examples are just some of the ways in which a parallel parser may not predict such a slowdown.

To back up these claims, Gibson and Pearlmutter reference discuss a study carried out by (Pearlmutter and Mendelsohn, 1998, 9). In this manuscript, they describe a set of experiments in which a dispreferred structure (14-b) is apparently calculated even though it is never adopted. Because such a structure could only influence the reading times if it were computed in parallel with the adopted (more likely, more preferred) structure, this is potentially evidence for a parallel parser. If this is the case, then it seems that serial models as a class might be at a loss to accommodate this data. For a serial parser to predict their results, it would somehow have to compute the dispreferred structure, discard it, and never make use of it. In other words, it would be acting like a parallel parser.

(14) The claim that the cop ignored/shot the informant might have affected the jury.

a. The cop ignored/shot the informant.

b. The cop ignored/#shot the claim.

Thus, Gibson and Pearlmutter (2000) conclude that the only useful way to investigate whether the human sentence processor uses a serial or parallel parser is to individually list the ways one could test each instantiation of the models of the parser. There may always be exceptions (especially in underspecified models), so this method for finding out whether the human parser is serial or parallel is not a fruitful line of research.

From these comparisons of models, it is clear that serial and parallel models may predict much of the same reading time patterns, depending on the detailed properties assumed of parser. Therefore, it will be important to keep in mind how these classes of parser may differ at different points in the structure-building process. But it is also difficult to definitively distinguish serial parsers from parallel parsers because of how much flexibility each class can make use of. While it is important to specifically identify and define what kinds of parser are assumed in order to make explicit predictions, results are not necessarily informative about how the parser is actually functioning in "the real world". This does not mean the results are uninterpretable, only that very few types of results will speak directly to whether the parser is serial or parallel.

2.2. The influence of lexical frequency

In the empirical chapters of this dissertation, I will be testing how the parser deals with ambiguity, so while it is important to explicitly define what kind of parser I will be assuming, it is also important to describe what kinds of ambiguity the parser will encounter. To do so, I will describe what types of studies have been conducted in the past that could inform my predictions and experimental design.

The key type of ambiguity investigated herein is introduced when a verb has the ability to be either intransitive (with no internal arguments) or monotransitive (with a direct object). This type of ambiguity between verb frames allows for a noun phrase following such a verb to temporarily be ambiguous between being in the direct object position of the verb or the subject of an upcoming verb. For example, the sentence in ((15), restated from (6)) is ambiguous until the word melody, at which point it could continue in two distinct ways (illustrated in (15-a) and (15-b)).

(15) Whenever Mia hums Annie's melody...

a. it sounds gorgeous.

b. gets stuck in her head.

Henceforth, this type of ambiguity will be referred to as a Matrix Subject (MS) / Embedded Object (EO) ambiguity. Because the verb is ambiguous between these two frames, it is prudent to discuss how the parser reacts to such ambiguities, both in relation to the models of the parser described in Section 2.1 and independently as empirical data.

One way that the parser's behavior around ambiguous verbs has been discussed throughout the literature is in terms of the verb's "transitivity preference".

2.2.1. Tanenhaus et al. (1989)

Tanenhaus et al. (1989), a classic paper that examines how transitivity preferences manifest, examines after which of two types of verbs are gaps posited. Fodor (1978) suggests that gaps would be posited after verbs which are likely to have transitive argument structures, so Tanenhaus et al. (1989) compares two verb classes: transitive preference verbs and intransitive preference verbs. In transitive preference verbs, an argument is likely to follow the verb, based on corpus frequencies, where as an intransitive preference verb is one in which it is likely that no argument follows (Connine et al., 1984). In conjunction with these two types of verbs, the sentences used in this study had a gap associated with a wh-filler in one of two locations: after the verb in an embedded clause (Early gap condition), or after a preposition following that verb (Late gap condition). An additional plausibility manipulation was included, in which the wh-filler was either plausibly (e.g., which witness) or implausibly (e.g., which church) the direct object of the embedded verb (e.g., asked).

(16) Example stimuli to illustrate the manipulations testing transitivity preference

a. The district attorney found out which witness/church the reporter asked __ about the meeting. (Transitive preference; Early gap)

b. The district attorney found out which witness/church the reporter asked anxiously about __. (Transitive preference; Late gap)

c. The sheriff wasn't sure which horse/rock the cowboy raced down the hill. (Intransitive preference; Early gap)

d. The sheriff wasn't sure which horse/rock the cowboy raced desperately past __. (Intransitive preference; Late gap)

A dependency must be formed between the wh-phrase (e.g., which church) and the gap, although the location of the gap is unknown at the point where the parser encounters the matrix verb. They observe a slowdown in reading time after the embedded verb in transitive preference sentences in which the wh-phrase makes an implausible direct object (as compared to the plausible condition), but not in intransitive preference sentences. The parser must thus have formed a dependency between the wh-phrase and the direct object position of the embedded verb before the plausibility information of the filler-gap pair was evaluated. That this effect was not observed in the intransitive preference conditions indicates that the parser is sufficiently aware of the verb frame frequencies to not posit a gap where it is syntactically unlikely to occur. This indicates that the parser has the ability to make use of syntactic subcategorization information, such as verb frame frequency, but lacks the ability to make use of plausibility information during the initial structure building process.

2.2.2. MacDonald et al. (1994)

If the parser has an awareness of verb frame frequency during its initial structure building, and it can act on that information but cannot act on plausibility information, how is such information stored and retrieved? MacDonald et al. (1994) propose that the mechanism for disambiguating lexical ambiguity and syntactic ambiguity is based on the same underlying architecture. This leads to a formulation in which verb frame information is stored in the lexicon and retrieved as a lexically specific property. They propose that syntactic frames stored in the lexicon are activated based on the lexical recognition process, so as a word is (gradually) identified (e.g., Neighborhood Activation Model: Luce and Pisoni, 1998), the argument structures or lexical frames associated with it are also activated. Furthermore, when more than one frame could be associated with a word (i.e., when the lexical item is ambiguous between two or more entries in the lexicon), those frames are activated, reinforced, and inhibited through the same mechanism as the lexical entries. The frequency with which a frame occurs with a particular form (e.g., orthographic representation), with a lexical (syntactic) category, or in a particular context (e.g., pragmatic or discourse context) could influence how activated a particular argument structure is to begin with.

For example, if it is the case that the sentence processing mechanism has encountered and recognized the word the, then we might assume that the lexical frame of the as started to activate nouns (and do a lesser extent, adjectives) in the lexicon, especially ones that are of high frequency and high frequency given the general context and environment in which the sentence is being processed. That is, it might be more likely for the word construction (a likely noun) to follow the than constructing (ambiguous between noun, verb, and modifier), leading to the phrase in (17-a) to be activated and adopted more slowly than (17-b).

(17) a. The constructing of a sentence can be difficult.

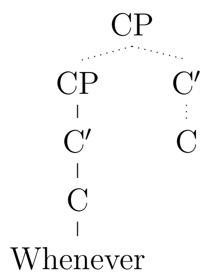
b. The construction of a sentence can be difficult.

This is a lexicalist, constraint-based approach to processing the syntactic structure of sentences, which differs in some crucial ways from the traditional hypotheses for explaining garden path phenomena. MacDonald et al. (1994) suggest that such an activation-based model of sentence processing may do away with the parser all together, although the difficulties of forming unbounded dependencies based on the known properties of the lexicon are difficult to accommodate. As this applies to the coreference dependency in this dissertation, the sentence processing mechanism must identify the pronoun as a trigger for such a dependency, then not only search input for candidate antecedents, but (as will be discussed in Section 2.5) also constrain that search to grammatically licit locations. Since it is not clear how this could happen in the formulation in MacDonald et al. (1994), I will instead examine how their account might handle a simple MS / EO ambiguity.

(18) Whenever John leaves the house ... [restated from (1), above]

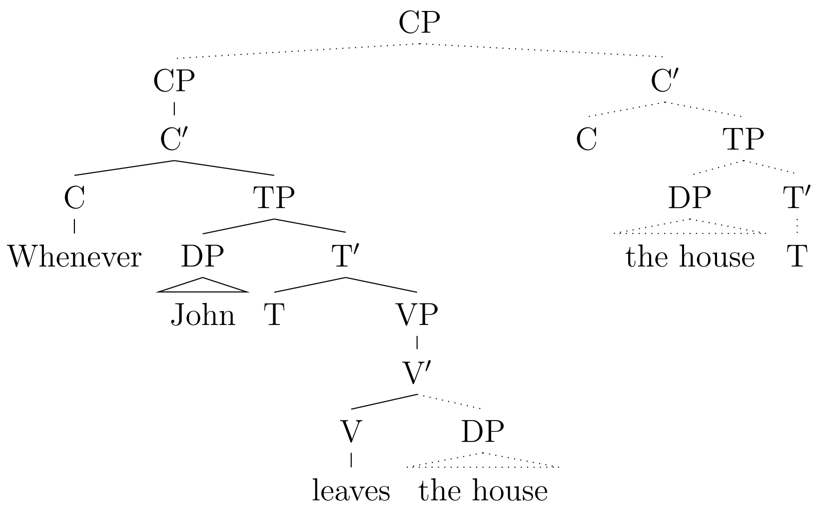
In (18), the identification of Whenever may have activated the structure necessary to build the embedded clause's CP. Since there is no input at this point to facilitate a main clause beside the presence of this embedded CP, it may be the case that the main CP is only weakly activated. I will assume that the absence of a parser means the absence of a predictive mechanism, and therefore I will assume a predominantly input driven (bottom-up) sentence-processing mechanism. This is consistent with some of the assumptions in MacDonald et al. (1994) about lexical processing and thus about their formulation of syntactic processing. From this, it may be the case that (18)-a. is the representation of (18) at the point where only *Whenever* has been encountered.

(18) a.



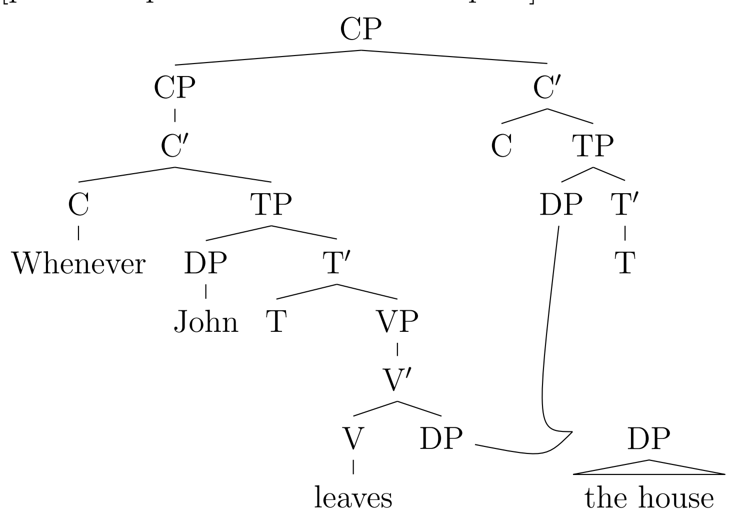
I will assume, next, that the input *John* does not add much to the ambiguity resolution of this sentence, since it is unambiguously a proper name, and very likely to be the subject of the embedded clause. Although this may more highly activate items within the grammatical category of VERB, the internal argument structure of those verbs is completely unknown and unlike to have a notable disparity in activation levels. Thus, the next relevant representation to this discussion is for (18)-b, once the ambiguous verb *leaves* is encountered.

(18) b. [parallel representations – neither adopted]

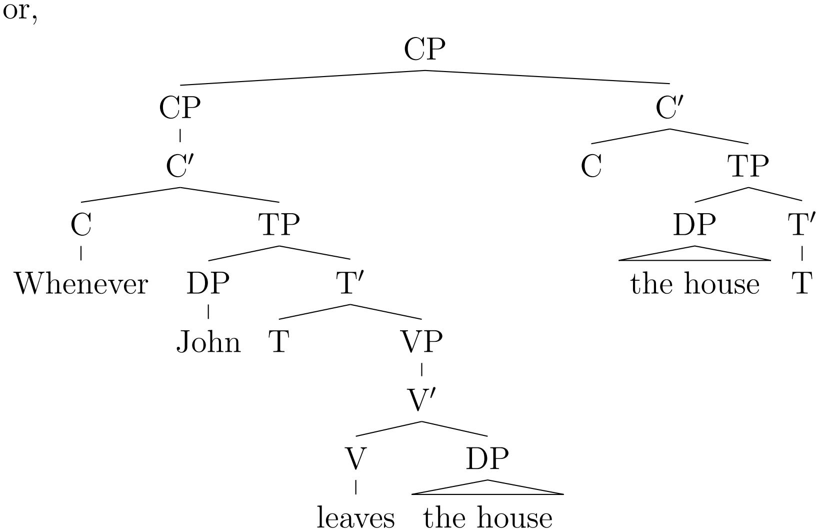


At this point, the frequency bias from *leaves* is no help to the sentence processing mechanism because it is equally likely to be a monotransitive structure as an intransitive structure. Thus, these two structures may be activated simultaneously (i.e., in parallel). Since there is competition between these two representations, it is possible that neither one will reach the threshold activation level to be adopted in the representation. Alternatively, one of the two may reach threshold and be adopted based on environmental factors (e.g., the transitivity of the previous verb heard, the individual experience of the listener, or noise in the lexicon). The former case is of more interest, as it differs more distinctly from the traditional garden path models of the parser. When neither representation reaches threshold, and both are equally activated, this may manifest as a delay in "choosing" a representation. That is, if no representation is adopted, the processing mechanism may have to wait for further input until one of the two representations is sufficiently activated to reach threshold. This is inconsistent with the eagerness observed in long-distance dependency formation (e.g., Omaki et al., 2015; Stowe, 1986), which leads me to suggest another possible course of events.

(18) c. [parallel representations both adopted]



or,



If, somehow, the activation of two argument structures were to reach the threshold for adoption at the same time, the sentence processing mechanism may be able to simultaneously adopt both, either in parallel or as a sort of hybrid (18)-b. Without sorting out what implications this would have more generally, I will step through what this could mean for the specific ambiguity displayed here. If the processing mechanism adopts both, it may be that the two representations are stored overlapping, since there has been up until now only one representation adopted. In this case, it could be that the DP *the house* is represented only once, since it only occurs once, and two different connections are established: one as the verbal complement and one as the specifier of the matrix TP (18)-c. This structure would be ungrammatical if it were stored as a single representation because the DP the house would be dominated by two unrelated projections (the embedded VP and the matrix TP). However, if these two representations are stored as two branches in an abstract parse space, then they may exist separately. In either case, both argument structures have been sufficiently activated, so the next item in the input will have to interface with both parses or, one of the two parses must be discarded.

If the monotransitive verb frame reaches threshold and is adopted, then presumably the intransitive representation will decay over time or be discarded. If the intransitive verb frame is adopted, then the monotransitive representation will be lost in the same way. This model predicts that whichever frame is adopted, especially if the other frame does not reach the threshold activation level, will be the predominant output. This type of output has the potential to still influence activation and behavior of input that follows shortly afterward, since decay is not an abrupt process. Nevertheless, MacDonald and colleagues do not discuss how their formulation would deal with partially activated representations, so it is speculation on my part that their model has such a potential.

The main takeaway from MacDonald et al. (1994) is that the preference for transitive or intransitive verb frames could conceivably be encoded in the lexicon along with the frame or other structural configurations. If this is the case, these structures and associated preferences could be accessed through a similar (or identical) mechanism to lexical retrieval. As in lexical retrieval, this would mean that multiple possible frames (or structures) could be activated and could be in competition during lexical retrieval, and only the most highly activated structure is ultimately adopted. This entails a type of parallel parsing mechanism whereby one structure is built, but many other alternative structures are activated in the process. By extension, this could mean that reanalysis could take advantage of the structures that were activated but did not reach whatever threshold the parser sets for adoption into the structure. As this applies to garden path sentences, the structures that are so unlikely as to be barely activated (if at all) would be difficult to identify during reanalysis since their activation levels are so low, thus the parser might be unable to detect them as viable alternatives.

2.2.3. Trueswell and Kim (1998)

The hypothesis that a word with two possible structures associated with it may activate both options in parallel, thus (temporarily) building two alternate representations in parallel until one reaches an activation threshold whereby it is adopted as the representations going forward. This could mimic the process of lexical retrieval and thus also account for the observations that the parser is sensitive to frequency effects in a similar way to the lexical retrieval mechanism (e.g., facilitatory effects of priming) (MacDonald et al., 1994). To test this hypothesis, Trueswell and Kim (1998) ask whether argument structures associated with specific verbs can be primed in the same way that lexical processes can be primed. Their study follows Garnsey et al. (1997), which examined transitively biased and equibiased verb frames. Garnsey found that transitively biased verbs displayed an implausibility effect when the available fillers were implausible given the thematic roles associated with the argument structure of the verb. In contrast, equibiased verbs showed no such effect, leading her to conclude that the alternative frame for the equibiased verbs in which there was no following gap was made available during structure building. Trueswell and Kim demonstrate that brief exposure to a prime that is associated with a particular argument structure (either a verb followed by a direct object or by a sentence complement) influences ambiguity resolution by increasing the initial adoption of the primed argument structure. In other words, the subcategorization information of a verb is accessed by the parser early during lexical recognition and is influential in ambiguity resolution during the initial structure building process. These results are consistent with the hypothesis that lexical frames are stored in a similar way as other lexical properties, and can be activated and manipulated by priming paradigms.

2.2.4. van Gompel et al. (2012)

In a more recent study, van Gompel et al. (2012) use a priming paradigm to compare the ambiguity introduced between monotransitive and intransitive argument structures. This ambiguity differs from Trueswell and Kim (1998) in that intransitive verbs do not have any complement, so rather than alternating a noun phrase complement with a sentence complement, they alternate the presence or absence of a complement. Since Trueswell and Kim compare the influence two different types of complements have on ambiguity resolution, both frames were providing some information about the complement of the verb. However, van Gompel and Pickering compare two frames in which one frame supplies no information about what might follow the verb (other than whatever follows is not the complement of the verb). Thus their question might be phrased as whether the absence of information (an intransitive verb frame) exerts as much influence on ambiguity resolution as the presence of information (a monotransitive verb frame). They find, generally speaking, that intransitive verbs were stronger priming influences on ambiguity resolution than monotransitive verbs.

This is compatible with Trueswell and Kim's conclusions since they found that verb frames were influential in ambiguity resolution in a priming paradigm, thus replicating their results. However, they additionally conclude that the intransitive verb frame is more marked than the monotransitive frame, which would explain why the parser was more strongly influenced by intransitive primes than monotransitive primes. That is, an intransitive verb frame is a more salient property than a monotransitive frame, giving a greater boost in activation to the intransitive argument structure in the following experimental trial. Van Gompel and Pickering suggest that this observation is consistent with the monotransitive frame being a category-general default for verbs, which alternative structures only available if marked specifically in the lexicon. This builds upon work MacDonald et al. (1994), in which it is suggested that lexical frames are encoded and stored in the lexicon, and are activated and adopted by a similar (or identical) mechanism to that which activates and retrieves lexical properties used in other aspects of sentence processing and comprehension.

The studies discussed in this section agree that the parser has access to the frequency of potential verb frames or argument structures of ambiguous verbs, and this information can influence the parser's behavior during disambiguation. Although it is still contentious how the linguistic processing mechanism uses this information in order to disambiguate between a transitive and intransitive frame. The processes by which the parser (or other processing mechanism) uses this information to resolve an ambiguity once it is identified, however, will be discussed in more detail in Section 2.4.

2.3. Local attachment

Another possible influence on the parser that must be considered besides frequency (or properties stored in the lexicon) is the bias the parser might hold as default, all other things being equal. That is, it is possible that the parser has some underlying biases that guide its choices when other sources of information are insufficient, either because they are absent or because they are balanced between two or more alternatives. A potential "default" bias that has been investigated throughout the psycholinguistics literature is that the parser prefers to attach incoming material to the most recently built structure, which by virtue of being local, is also the most highly activated, most salient, or most recently built structure before the next incoming lexical item is encountered. This bias has been described in a number of ways (depending on the theoretical framework and details of the instantiation), including Right Association (Kimball, 1973; Phillips, 1995), Late Closure (Frazier, 1979), Attach Low (Abney, 1989), Recency (Gibson et al., 1996;

Gibson, 1991), and Minimal Connections (Fodor and Frazier, 1983). Although these hypotheses differ in some ways, the basic premise shared by all is that incoming input is preferentially associated with the lowest and most recent non-terminal node in the syntactic representation. Although the exact mechanism varies between formulations, there does seem to be a consensus in the literature that such a property could be a native strategy of the parser (Kimball, 1973; Phillips and Gibson, 1997, a.o.). This section will explore what such a property would mean for the resolution of a Matrix Subject / Embedded Object ambiguity.

2.3.1. Minimal Attachment

Before exploring the EO/MS ambiguity in detail, I will address a notable principle that decisively does not fall into the same class as the principles listed above. Frazier and Fodor (1978) described above in Section 2.1, suggest a property of the parser called Minimal Attachment. The common interpretation of Minimal Attachment differs substantially from the other properties in the class I will hereby refer to as local attachment type principles, but the motivation for describing this property was fundamentally the same. That is, the class of properties subsumed under my label of *local attachment type* (in the spirit of Phillips and Gibson, 1997), but in additional Minimal Attachment, describe the conservative structure-building behavior of the parser in which it apparently prefers to attach incoming structure in such a way as to build the least costly representation. The method by which the parser chooses the location to attach the incoming materials substantially differs between Minimal Attachment and the local attachment type property of the parser. The difference is motivated by the challenge of describing sentences like

(19), in which the parser does not preferentially associated *for Susan* with *the book*, but rather with *bought*, which is higher in the tree.

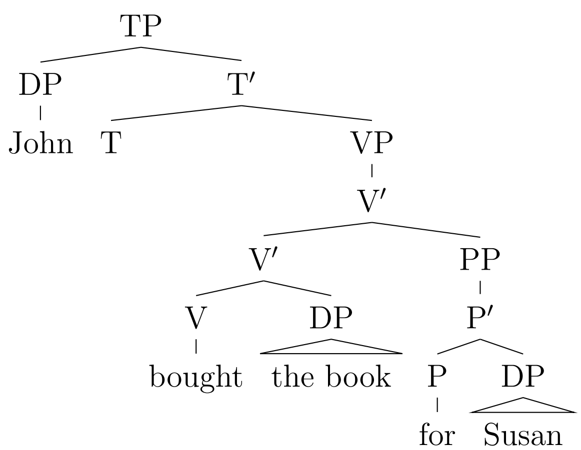
(19) John bought the book for Susan.

This type of sentence seemed to (Frazier and Fodor, 1978) to be accounted for in an unsatisfactory manner by a property of the parser called Right Association by Kimball (1973). Although Kimball directly addresses this type of sentence, he stipulates that the principle of Closure would account for why *for Susan* is preferentially interpreted as an argument of the verb instead of a modifier of *the book*. Right Association, alone, would prima facie predict that *for Susan* would associate with *the book*, which is not the observed behavior. Frazier and Fodor view Closure as an ad hoc solution to this issue, thus formulate Minimal Attachment to account for the preferences observed by the parser without additional stipulations. Minimal Attachment states that the parser will attach incoming material to the parse tree using the fewest possible non-terminal nodes to link the existing tree to the new material. This stipulation differs from the formulations of local attachment type principles in that it does not make reference to the position (either low in the tree or recently built) at which the attachment site is located, thus placing it outside the class of local attachment type principles. Since the formulation of Minimal Attachment, several other local attachment type principles have been described that more closely resemble Right Association, but find other ways of overcoming the issue illustrated by (19).

2.3.2. Phillips (1995)

Phillips (1995) specifically addresses how Right Association could be modified to operate in a situation where all other potentially competing forces are balanced, and Minimal Attachment is not needed to account for sentences like (19). The basic mechanism behind his version, Right Association is a “step”-counting method, in which the parser prefers the syntactic structure that has the fewest number of steps, or arcs, between the most recently attached maximal projection and the incoming lexical item.

(20)

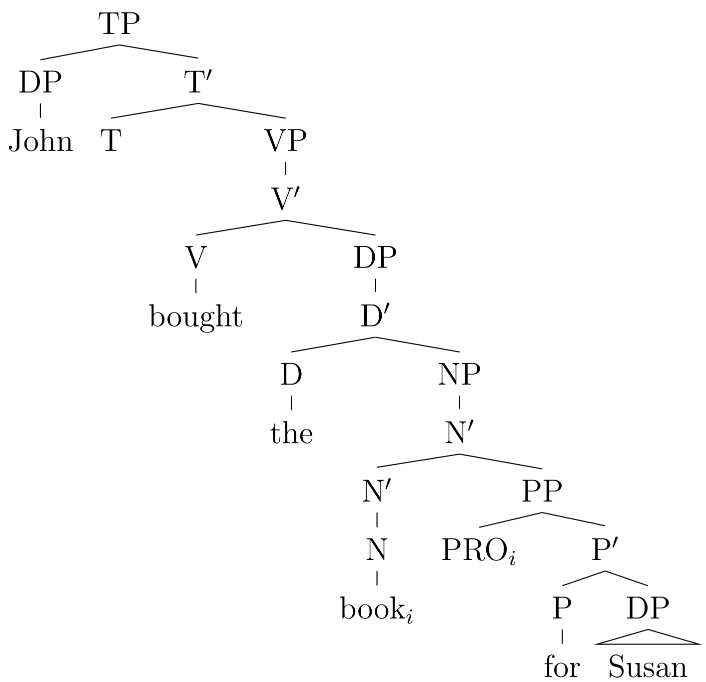


For instance, given the option of associating *for Susan* into the phrase structure for the sentence in (19), there are two steps up and three steps down between for and the previous DP the book when *for Susan* is an argument of the verb (20), and two steps up and four steps down when it is modifying *the book* (21)[2]. This accounts for the preference for interpreting (19) as having the structure in (20), without the additional stipulation of Closure or the other pitfalls of Minimal Attachment. Thus a local attachment type

[2]Phillips assumes that the PP serving as a nominal modifier has a null anaphoric element (PRO) in the specifier position of PP that corefers with the head of the NP within which it is attached. This parallels relative clause constructions, which are similarly served by Phillips’ modified principle of Right Association.

property like Right Association can account for (slightly) non-local attachments without resorting to extra stipulations.

(21)

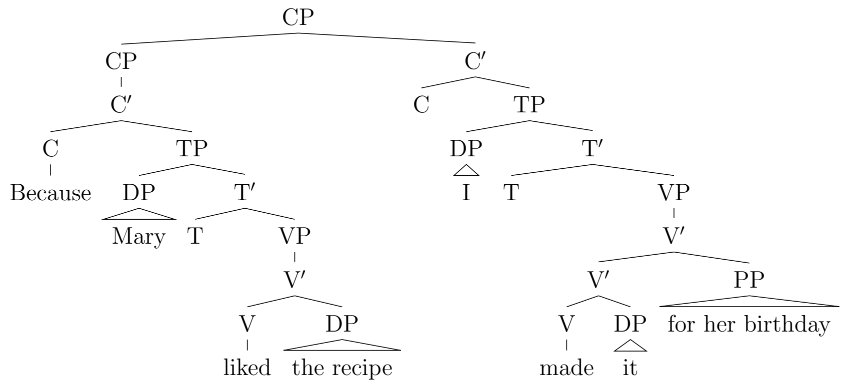


2.3.3. Phillips and Gibson (1997)

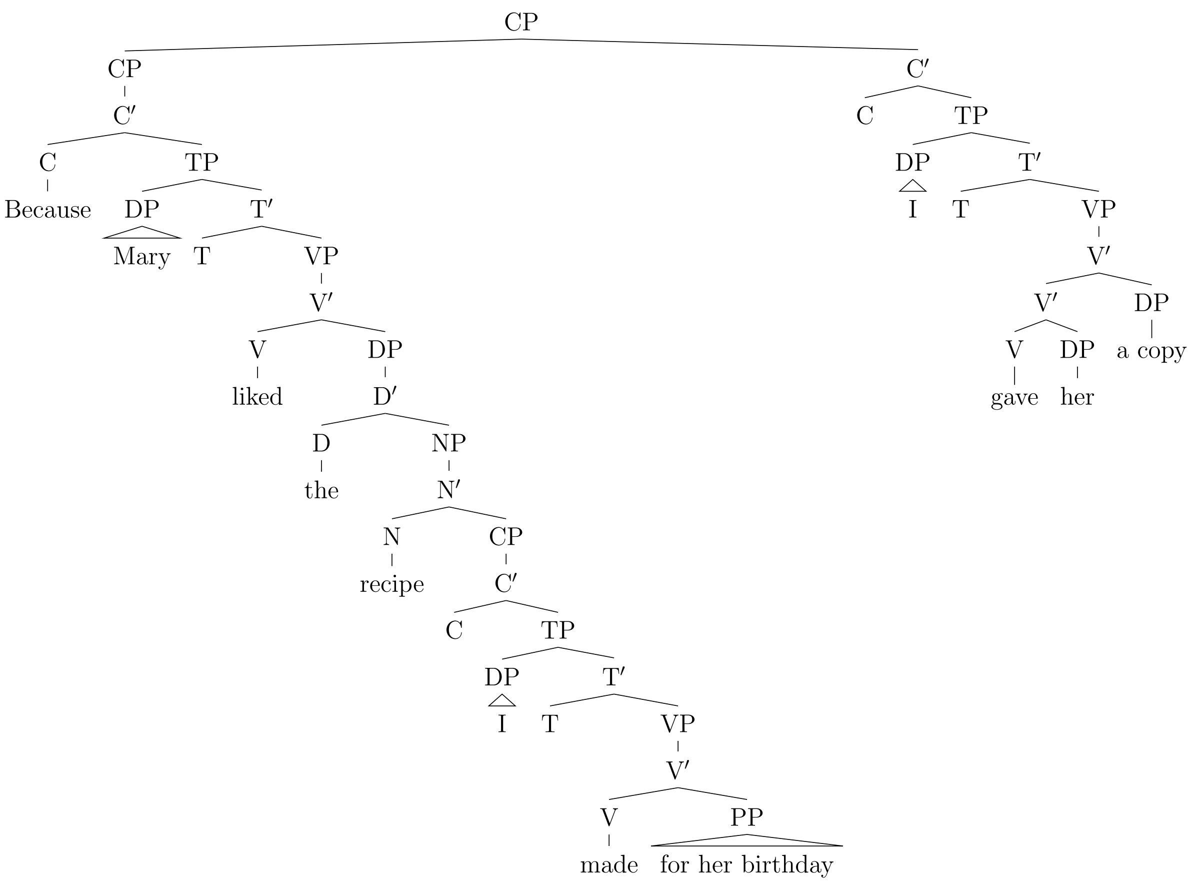
Following this paper, Phillips and Gibson (1997) provide further evidence that Right Association (or a similar local attachment type property) is in operation, and more importantly, it asserts a stronger influence on the parser's behavior than had been previously thought. To do so, they examined an ambiguity that differs only slightly from the MS/EO ambiguity examined in this dissertation; they use a Matrix Subject / Relative Clause (RC) ambiguity, as shown in (22).

(22) Because Mary liked the recipe I made...

a. it for her birthday.

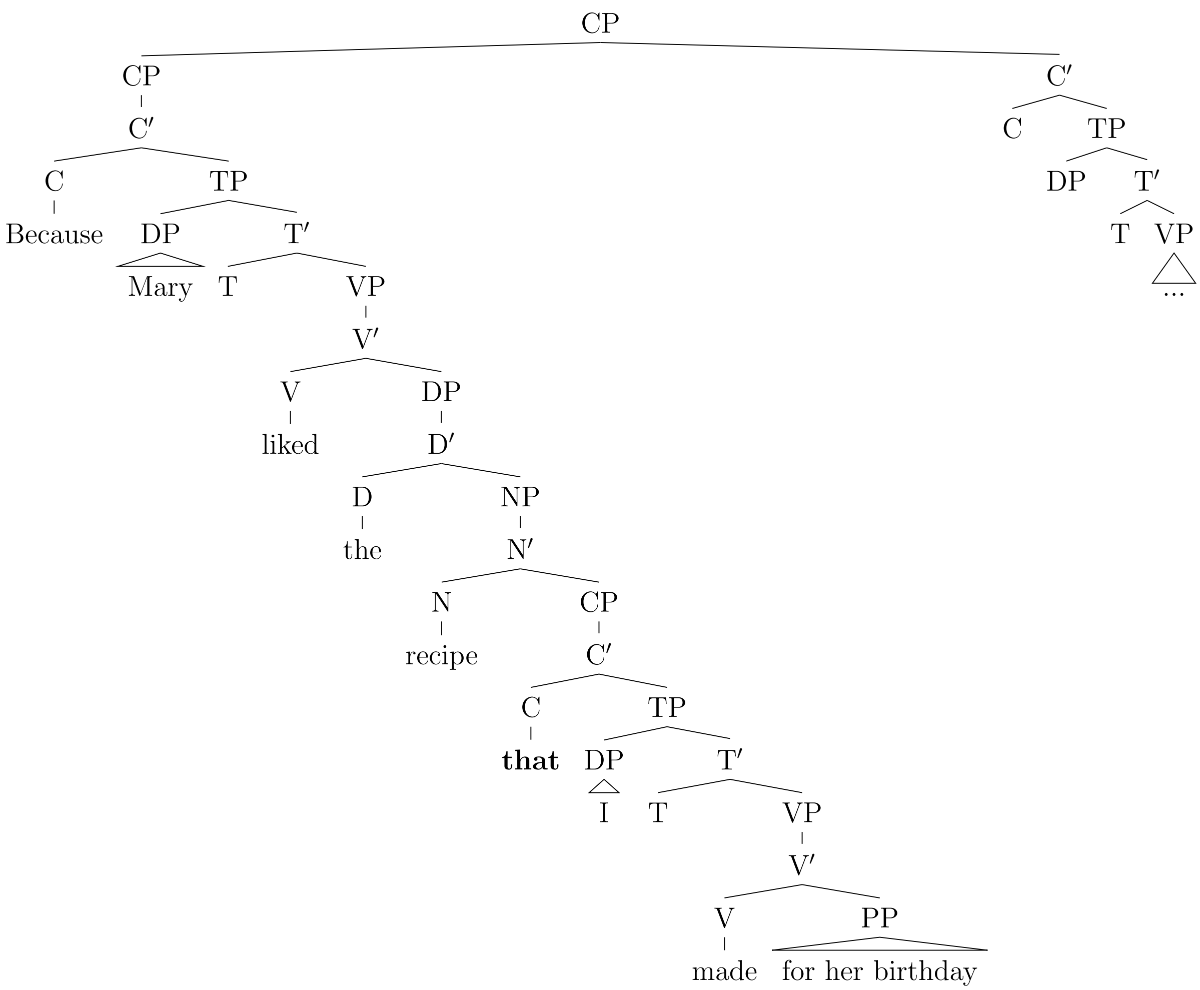


b. for her birthday, I gave her a copy.



Here, the tree in (22-a) illustrates that the recipe is the terminus of the embedded clause, while I made comprises the matrix subject and verb. This contrasts with (22-b), in which I made comprises the subject and verb of a relative clause modifying the recipe. This is unambiguous in a sentence like (22), in which an overt complementizer initiates the relative clause.

(22) b′. Because Mary liked the recipe that I made for her birthday ...



In sentences like (22), the parser seems to prefer the less local attachment illustrated by (22-b). In order to account for this preference, Phillips and Gibson perform an experiment to test the strength of their local attachment strategy for relative clauses in English. Since principles like Minimal Attachment (Frazier and Fodor, 1978; Frazier and Rayner, 1987) predict that a temporarily ambiguous string like (22) would be preferentially disambiguated with *I made* in the main clause because this configuration requires the least extra construction and fulfills obligatory structural constraints without building optional structure. In contrast, Phillips and Gibson's local attachment principles (including Recency and Right Association) predict that the string *I made* would be preferentially attached as a relative clause modifying *the recipe*. They observed that their stimuli containing a non-temporal[3] complementizer at the beginning of the first (embedded) clause (e.g., *Because* in (22)) exhibited a strong preference for the relative clause parse. This suggests that in at least some conditions, the strength of the local attachment bias is stronger than the bias to complete all obligatory arguments as soon as the input would allow, and the bias to build only the minimal structure necessary to account for the (current) input.

It is important to note that Phillips and Gibson assert that this preference is fundamental to the parser, but not strictly ranked with regard to other preferences or biases the parser may have. They are clear in their position that multiple other influences can override an underlying preference such as Right Association when added together. In other words, the biases that influence the parser's behavior are additive and can thus several

[3]The distinction between $\pm$temporal features in this complementizer is only relevant in that the $+$temporal stimuli (which were confounded with another factor) displayed the reverse preference, that is, the matrix clause preference. This is ascribed to the "ganging up" of multiple biases against the local attachment bias, including the bias (or requirement) for tenses to match between clauses when a temporal complementizer is used. Since this was not controlled for, the $-$temporal condition is sufficient to display the strong preference for local attachment.

weaker biases can "gang up" on a stronger bias to override it in certain circumstances. If Right Association is as strong a bias as Phillips and Gibson suggest, then this could explain why previous studies had not found it to be as influential as this study did.

2.3.4. Gibson (1998); Gibson and Pearlmutter (2000); Gibson et al. (1996)

Gibson describes another account of this type of local account in his 1996, 1998, and 2000 papers. These papers discuss how the memory load required to process of different types of ambiguities might impact the parser's behavior. In particular, Gibson's theory of parsing called Dependency Locality Theory (DLT) is of relevance to the current discussion. In DLT, both long and short dependencies are formed based on principles of distance rather than hierarchical structure. Ignoring for the time being the conflicting evidence from Spanish and Dutch[4], the DLT and its predecessor (SPLT) are based on the premise that the parser prefers to associate things that are closer together in time (and the theoretical space of the parse tree) because this requires less effort from the working memory. In other words, the parser prefers to build structures that are easier to remember and maintain in memory. A dependency built between two positions that are closer together, i.e., a shorter

[4]There has been noted a disparity between the preference for attaching, say, a relative clause to the lowest noun phrase in a parse tree and a higher position. Specifically, English, Romanian, and Basque show a consistent preference to attach incoming structures such as relative clauses low in the tree, whereas languages like Spanish, Japanese and Dutch display a consistent preference for attaching relative clauses higher in the tree (Grillo, 2012). Grillo notes that this difference between the two classes of languages could be attributable to the presence of a pseudo-relative construction. That is, languages like Spanish and Dutch that display a high attachment preference also have a construction similar to a relative clause, but in the ambiguous cases in which the high attachment preference is observed, the ambiguity stems from a configuration that would be typically used in the pseudo-relative construction. However, since it is not a pseudo-relative, the parser is presented with a less probable set of alternatives, of which the high attachment option is more likely. Thus, as described by Phillips and Gibson (1997), the principle of Right Association (or, in this case, the preference for Low Attachment) can be overridden if there is enough of a pull from other biases. Otherwise, in more balanced environments, "High Attachment" languages (e.g., Spanish) and "Low Attachment" languages (e.g., English) should still exhibit the same local attachment bias according to Grillo (2012).

dependency, should take up less space in working memory than a longer dependency (Gibson, 1998). Thus a preference to form dependencies as locally as possible could fall out from properties of working memory.

The property I have been referring to as a local attachment type preference may conflict with a verb's frequency information encoded in the lexicon. Thus, even if we examine idealized equibiased verbs that are found exactly half the time with an intransitive frame and the other half of the time with a monotransitive frame, the parser may still be more likely to interpret a string containing this verb as being transitive, due to the strength of the local attachment bias. On the other hand, one may speculate that this influential property of the parser may already be affecting the content and balance of the use of different frames in well-designed corpora. In other words, the calculation of which verbs are equibiased verbs in a corpus may already take into account the parser's preference, since these verbs are produced and comprehended by the linguistic processing mechanism (and it's not certain that the linguistic processing mechanism isn't also influenced by a similar property). Thus it may be the case that given a parser without any such local attachment preference would otherwise actually show a preference for the intransitive form of these ambiguous verbs, and it is only because of the strength of the local attachment preference that they appear as equibiased in the corpora. These two influences are difficult to tease apart, and since this is not the focus or intention of my dissertation, I will leave it to future work to determine whether the parser's biases and the verb frame frequency information are factors that mutually influence each other. What can be taken away from this discussion is that there is a detectable influence of verb frame frequency that can influence the parser's choices in disambiguating a Matrix Subject / Embedded Object ambiguity, but there seems to also be an underlying bias of the parser to build transitive structures over intransitive structures, given the option.

2.4. Ambiguity resolution

The parser's priorities, biases, and preferences are meaningless without an environment within which to demonstrate their implications. The environment most suited to the present research and, in general, the question of how the parser reacts to different sources of information is in syntactic ambiguities. Specifically, temporary ambiguities are useful because they are disambiguated by information later on, so that the parser has a finite period of dealing with the ambiguity before it either gets confirmation that it made good choices or it realizes it made a mistake and must reanalyze the representation(s) it has constructed. In this type of ambiguity, the parser has impetus make "wise" choices, i.e., to avoid making costly mistakes that could cause the representation(s) it builds to be discarded. In serial models of the parser, a representation that is ultimately incoherent would be costly because it would require the parser to revise or reanalyze a structure that has been built and potentially shunted off to another stage of processing (Crain and Steedman, 1985; Frazier and Fodor, 1978). In parallel models, this type of error is only marginally less costly (if at all), especially if the parser has committed to the wrong representation or has inaccurately de-prioritized or discarded the (ultimately) correct parse. In this section, I will review how ambiguities have been discussed and approached by the literature on models of the human parser.

2.4.1. Frazier and Rayner (1982)

In the classic literature, Frazier and Fodor (1978) put forth a two-staged serial model of the parser, which is discussed in detail in Section 2.1. Following on the predictions of this model, Frazier and Rayner (1982) elaborate on how mistakes in parsing and subsequent reanalysis might unfold using selective reanalysis. The set of hypotheses that fall under the category Selective Reanalysis describe how a serial parser may target the specific regions that contain errors. The purpose of such hypotheses is to describe how the parser is able to handle making mistakes (an uncontroversial observation), without becoming overwhelmed by the process of revision or reanalysis. The parser could be prone to being overwhelmed or the memory of the input could decay too far to be re-analyzable if the parser does not have a method for selectively finding its errors and correcting them in real time during processing. The two hypotheses explicated below follow from these limitations, followed by a hypothesis that makes use of selectivity to help the parser make "wiser" choices.

Forward reanalysis is the process by which a parser returns to the beginning of a tree and progresses from left to right through the tree until it detects a relevant "choice point". A choice point is effectively a place in the representation that could have been analyzed in an alternative manner, and specifically that the alternative analysis might provide a better fit for the input that is inconsistent with the original representation. If the parser finds a choice point in its forward progression through its previously built structure, it may prune the right-branching section of the tree following this choice point and rebuild a new structure that can accommodate the input that was inconsistent with the original representation. Some of the issues this hypothesis encounters make it an unlikely choice. For instance, the memory storage needed to maintain the entire tree in an accessible form for an indefinite period of time may be too massive for the human linguistic processing system and memory capacity. Moreover, the parser must have some way of identifying where the choice points in the tree are, and which alternative structure would better accommodate the new input. The only way for it to have that foresight is to have built the alternative structure in comparison. In order to build this structure within the constraints of a serial model of the parser, either the parser must guess what new direction to take at each choice point (potentially overwhelming itself with revisions) or it must build the structure backwards based on the latest input. This is not forward reanalysis, but rather the next hypothesis in this class.

Backward reanalysis, as hinted at, is the process by which the parser reverses direction and looks backward (leftward) through the tree for potential choice points that would accommodate a better fitting structure. In this hypothesis, one would expect reanalysis to preferentially happen lower in the tree, closer to where the parser starts its search for alternative analyses. This would happen because the parser would be less likely to go all the way back to the beginning of the tree before it finds a choice point at which to attempt revision. This contrasts with the Forward Reanalysis hypothesis, which should show more frequent revisions at high positions in the tree, since this is where the parser is most likely to encounter a choice point. Because these two hypotheses rely on some amount of uninformed guesswork or trial-and-error reanalysis, there is no good way to determine whether one would be more successful without modifying it to be more similar to the Selective Reanalysis hypothesis.

The Selective Reanalysis hypothesis (originally put forward by Winograd (1972)) includes the ability for the parser to “flag” choice points it recognizes during its initial structure building process. By flagging choice points, the parser should be able to identify and return to them more quickly if and when it discovers it has made a mistake in the adopted representation. It is uncertain how the parser may choose which points are relevant or useful choice points, since ambiguity comes in many different forms and thus may increase drastically as the length of the sentence increases. But, if the parser has access to some sources of information besides broad syntactic category, then it is not impossible to formulate a way in which it could flag places where the serial analysis of an ambiguity is more likely to cause problems down the line. For instance, if the parser encounters a verb that is equally likely to be monotransitive as intransitive, then it might use this frequency information to flag the choice point at this verb in case whichever structure it adopts ends up being the other (equally likely) structure. This may appear as illustrated in the trees in (23).

(23) Whenever John leaves the house is dark. (Early Closure / MS)

Cf. (23)′ Whenever John leaves the house it is dark. (Late Closure / EO)

a. Whenever John leaves (transitive – FLAGGED)

b. Whenever John leaves the house (EO)

c. Whenever John leaves the house is (ALERT!)

d. Whenever John leaves the house is (RETURN TO FLAG)

e. Whenever John leaves (MS) the house is

f. Whenever John leaves (MS) the house is dark (SUCCESS)

The EO/MS ambiguity that trees depict can also be called an Early/Late Closure ambiguity, where closure refers to the point in (23) at which the embedded clause is complete and thus the following material is associated with the matrix clause. In the Early Closure version of this sentence, the embedded clause is "closed" after *leaves*, which is earlier than the Late Closure version (closed after *house*). In this example, then, Early/Late Closure ambiguity is synonymous with the EO/MS ambiguity I reference throughout this dissertation. By looking at how the parser treats the resolution of sentences like this, one may surmise which structure or structures the parser has built (MS/Early closure or EO/Late closure).

Frazier and Rayner (1982) explore another possible behavior of the parser, which they call the Minimal Commitment Hypothesis. In this model of the parser, the parser avoids making mistakes by waiting for the disambiguating information before choosing which structure to build. However, there is substantial evidence that the parser does not wait for disambiguating input before committing to or adopting a representation (e.g., Stowe, 1986). Rather, the parser consistently displays an eagerness whereby it adopts representations preemptively, apparently committing to structure well before evidence of that structure has been confirmed. This is observed by studies that examine where the parser posits gaps (e.g., Omaki et al., 2015; Stowe, 1986), in languages where the argument structure the verb is ambiguous because the verb is encountered late (i.e., SOV languages like Japanese; Yoshida, 2006), and in other "ACB" constructions, in which the parser reaches a point of A, apparently anticipates the position of an item B, whereby it infers the presence (and position) of item C (Phillips, 2006). In other words, the parser does not wait for bottom-up information before establishing some syntactic structure that allows it to predict some basic, required aspects of the upcoming material, and by extension, it may be able to anticipate optional material as well.

Frazier and Rayner also discuss how a parallel parser might be able to cope with an MS / EO (Early closure/Late closure) ambiguity. They find that a simple, unconstrained parallel parser is unlikely because of the evidence that suggests the parser does not experience difficulty in all sentences equally. That is, the parser experiences more difficulty in Early closure sentences more than in Late closure sentences, which should only be the case if the parser is biased, either by some source of information or from some built-in feature of the parsing mechanism. The parser cannot have equal access to all possible structures, otherwise it would be as easy to adopt the Early closure sentences (or any strong garden path sentence) as it apparently is to adopt Late closure sentences (or to deal with other types of milder ambiguities). Moreover, the parser does not show the same kind of slowdown in sentences where there is no ambiguity, which would indicate that any observed slowdown was not due to ambiguity resolution but to some other aspect of sentence processing. Thus, they conclude that the parser is experiencing a garden path effect in ambiguous sentences with Early closure resolutions, which supports their garden path model of the parser's method of resolving syntactic ambiguity.

2.4.2. Ferreira and Henderson (1990)

As an extension of this work, Ferreira and Henderson (1990) investigate whether the parser is sensitive to the information encoded in verbs, such as subcategorization properties. They also ask whether this information is used for initial structure building or only for reanalysis of a structure that is determined to be built incorrectly, as might happen in a serial deterministic parser described in Section 2.1. Their study makes use of a self-paced reading task which manipulated the target verbs' subcategorization properties and the frequency with which the verbs occur with those associated frames. For example, (24) illustrates the two manipulations used in this study. The absence of the complementizer that makes the sentence ambiguous between the sentence complement and direct object interpretations, but wished is highly unlikely to have a direct object, whereas with forgot, it is reasonable to have one. Thus, if the parser is immediately sensitive to the subcategorization preference of the verb, it will not experience a garden path in (24-a) when the complementizer is absent.

(24) a. He wished (that) Pam needed a ride with him.
  b. He forgot (that) Pam needed a ride with him.

They find that the parser does not seem to be initially sensitive to the verb's subcategorization preferences, because early reading times did not differ between the ambiguous versions of (24-a) and (24-b). However, they do see garden path effects in later reading times, which lead them to conclude that the lexical frame information encoded in the verb is used by the parser during reanalysis. This accounts for the observation that verb frame biases do not seem to prevent garden path effects, but do seem to help the parser recover once the garden path has been recognized.

2.4.3. Staub (2007)

Another way to examine ambiguity surrounding verb subcategorization preferences is through focusing on intransitive verbs. Since there is conflicting evidence that the parser ignores the argument structure of ambiguous verbs (Blodgett, 2004; Ferreira and Henderson, 1990, and references therein), Staub (2007) investigates whether the parser will ever posit a gap after an obligatorily intransitive verb, as it seems to consistently do after transitive verbs (or ambiguous verbs with the potential of being transitive) when engaged in the search for a gap. To do so, he compares three categories of verbs: transitive, unaccusative (intransitive), and unergative (intransitive). Unaccusative verbs are verbs in which the accusative case cannot be assigned, so in English there cannot be a direct object (e.g., *arrive*). Unergative verbs cannot assign ergative case, so typically they do not have a direct object. In rare circumstances, they may appear with an object (or object-like) noun phrase, as illustrated in (25). For the comparison in Staub's study, this is crucial because it means that only unaccusative verbs are truly obligatorily intransitive, whereas unergative verbs are not even though they are heavily biased toward the intransitive argument structure.

(25) a. She slept a dreamless sleep.

b. He sneezed the paper off the table.

c. They laughed their way to the bank.

Thus, by comparing unaccusative verbs to unergative verbs, Staub investigates whether even the remote possibility of being followed by an object will influence the parser's behavior upon encountering a verb. In fact, he finds that the parser does respect the intransitive status of unaccusative verbs, while unergative verbs seem to be slightly less likely to be initially treated as intransitive during initial structure building. From this, he concludes that the parser has access to the obligatory subcategorization properties of verbs immediately upon encountering them. However, even if it is an extremely infrequent configuration, the parser is aware that unergative verbs are able to appear in frames that have an object position. This suggests that the source of the conflicting evidence described in previous work, at least in part, may be the parser's sensitivity to the possibility of even unlikely structures, rather than the parser's disregard for obligatory subcategorization.

2.4.4. Ferreira and Henderson (1991)

Returning to the tests of how the parser does and does not avoid falling victim to garden path sentences, Ferreira and Henderson (1991) examine what process the parser undergoes in order to recover once it finds itself having been led down a garden path. To do so, they make use of the MS / EO ambiguity, additionally manipulating the complexity of the ambiguous region. In (26) and (27), the capitalized words indicate the manipulated region as adapted from one of their experiments. If the parser is more easily able to recover after a short or less complex ambiguous region, then the garden path effects should be less strong in (26-a) as compared to (26-c). In addition, if the parser is influenced by Minimal Attachment (or, in the case of (26)/(27), other formulations of local attachment hypotheses which are compatible with Minimal Attachment here), then (26-a),(26-c) should show higher processing cost at the disambiguating region than (27-a),(27-c).

(26) Early closure versions

a. *Short region*:

After the Martians invaded THE TOWN was evacuated.

b. *Long, unambiguous region*:

After the Martians invaded THE TOWN THAT THE CITY BORDERED was evacuated.

c. *Long, ambiguous region*:

After the Martians invaded THE TOWN THE CITY BORDERED was evacuated.

(27) Late closure versions

a. *Short region*:

After the Martians invaded THE TOWN the people were evacuated.

b. *Long, unambiguous region*:

After the Martians invaded THE TOWN THAT THE CITY BORDERED the people were evacuated.

c. *Long, ambiguous region*:

After the Martians invaded THE TOWN THE CITY BORDERED the people were evacuated.

They find that the parser experiences more difficulty recovering from a garden path if the distance from the head of the ambiguous region to the disambiguating word is longer ((26-a)/(27-a) compared to (26-c)/(27-c)). The measure of distance for this observation is not the linear distance between two words (i.e., the number of intervening words) or the syntactic complexity of the intervening material (not depicted in (26)/(27); manipulating whether the ambiguous region contained a relative clause or a prepositional phrase for length). Instead, it seems to be related to the linear distance between the head of the ambiguous phrase and the disambiguating region. This is consistent with a two-stage parser in the tradition of Frazier and Fodor (1978) model, in which the first stage chunks the input and analyzes the syntactic content of the chunk, and the second stage analyzes the thematic and semantic information. In this way, the longer time that the thematic role of the head of the ambiguous region must be stored in memory, the more ingrained it becomes and the harder it becomes to revise. After a longer intervening region between the head and the disambiguating material, the parser finds it more difficult to reanalyze the thematic role assigned to the head and experiences a stronger garden path effect.

This so-called “digging-in” effect of syntactic (or semantic) content of a garden path has spawned a large and fruitful like of research into how strict the syntactic (and semantic) sentential representations are during and after reanalysis. In other words, does the parser fully discard the structures it builds that turn out to be erroneous, or are those structures still present in some fashion once the parser has completed building a coherent representation of the entire sentence? Numerous researchers have noted the perseverance of globally ungrammatical parses, but the research examining this phenomenon has taken a stance decidedly in favor of a serial model of the parser. In the following subsections, I will briefly discuss what types of approaches have been attempted along the lines of the “Good Enough” theories of processing, which suggest that the parser does not build a rich and detailed representation during processing thereby allowing for multiple interpretations to exist simultaneously for a single parse. I will additionally speculate on how parallel models of the parser may or may not account for the findings discussed therein.

2.4.5. Tabor and Hutchins (2004)

Tabor and Hutchins (2004) compare two selection mechanisms whereby the parser could choose one representation over another during the processing of the temporarily ambiguous region of a garden path sentence. One class of mechanisms they examined is a top-down selection mechanism (TDSM), which makes use of two (independent) stages: one that constructs an initial representation and one that makes repairs (or reanalysis) as necessary. Included in this class are the two-stage serial parsers discussed in depth in Section 2.1. This class is compared to a dynamic self-organizing (DSO) system, in which has a single mechanism whereby the representation is built and modified as support for different possibilities becomes stronger or weaker. Included in this category are constraint-based models, and some parallel models such as an unrestricted race model. By this comparison, they aim to identify the mechanism behind the digging-in effects in which properties of an original interpretation or representation of a substring perseverate after the sentence has been reanalyzed (after the parser corrects a mistake in its structural or thematic analysis). To do this analysis, they examine offline and online behavior in reading tasks. In the offline tasks, distance from the start of the ambiguous region to the disambiguating word and the bias toward transitivity interact robustly, with long distances and intransitively biased verbs reported as being less acceptable than the other conditions. This pattern of interactivity is also observed in online experiments, with the condition that contained an intransitive verb and a long ambiguous region was read slower than the other conditions. This indicates that the transitive interpretation was harder to discard during reanalysis (and subsequently adopting the intransitive analysis was also difficult) due to the increased length of the ambiguous region.

(28) As the author wrote

a. (monotonic) ... the essay, the book that she envisioned grew rapidly in her mind.

b. (one revision) ... the book that she envisioned grew rapidly in her mind.

c. (two revisions) ... the book she envisioned grew rapidly in her mind.

In following up on their findings, Tabor and Hutchins compare sentences that could contain a single ambiguous region to those that contain two ambiguous regions. In sentences like (28), the number of possible revisions increases with the number of ambiguous regions. (28-a) is unambiguous. (28-b) is ambiguous at *the book*, which could be either the object of *wrote* or the subject of the following clause. (28-c) is ambiguous at *the book*, as well, and also at *she*, which does not disambiguate between the object/subject ambiguity of *the book*, but also adds a subject/relative clause ambiguity, depending on the position *the book* has been placed in. Both types of ambiguous sentences are found to give rise to garden path effects, but the length of the disambiguating region influences the ease or difficulty with which the parser completes reanalysis. Specifically, one-word disambiguating regions seem to increase the intensity or concentration of the garden path effects there is less time (or space) for the parser to recognize the disambiguating material in a short region.

From these experiments, Tabor and Hutchins conclude that their observations are more compatible with the DSO models than the TDSM models. Since DSO models allow for compounding strength of reanalysis effects, which is the type of effect observed in two-revision sentences as compared with one-revision sentences. Moreover, TDSM models cannot allow such a gradient effect; either an effect is observed or it is not, so there is no way to accommodate the two distinct effect strengths observed. One drawback of the DSO model is that the constraints that it uses are difficult to falsify, making it potentially susceptible to over-generation. In other words, it may be able to mimic many other types of models (and over-fit the data) rather than providing clear, testable, and falsifiable predictions.

2.4.6. Christianson et al. (2001)

Another study that examines digging-in effects is by Christianson et al. (2001). The aim of this study is to determine whether readers always fully reanalyze sentences in which they experience a garden path, or whether there is the possibility of partial reanalysis. In the first part of this study, the influence of plausibility and length of the ambiguous regions are manipulated. They replicate the findings of Tabor and Hutchins (2004) in that they find that the parser has more trouble recovering from garden paths caused by longer ambiguous regions than by shorter ambiguous regions. They also make a more fine-grained distinction in the properties of the ambiguous region, comparing the distance of the head of the ambiguous NP to the disambiguating word, rather than the total length of the disambiguating region (as was done by Ferreira and Henderson, 1991, and Tabor and Hutchins, 2004). Again, they find that an increased distance from the head of the ambiguous NP to the disambiguating region was associated with an increase in the processing cost of reanalysis that the parser exhibited.

(29) As the hunter shot the deer ran into the woods.

Q: Did the hunter shoot the deer?

At this point, it is of some interest to note that even when participants answered comprehension questions incorrectly, they reported that they were highly confident in their answers (in the plausible conditions). This is taken to mean that whatever parse is settled upon, it is not necessarily the rich and detailed structure that is required to determine the correct parse[5].

Finally, Christianson and colleagues examined the influence of semireflexive verbs like *bathe*, *scratch*, and *dress*, in which the verb assigns theta roles twice to the subject, effectively creating a reflexive relation without any overt grammatical markers. Below in (30) are three examples of semireflexive verbs in their bare, intransitive frames and in their reflexive frames. The sentences on either side of the table describe the same event; in all cases the reflexive pronoun provides redundant information. In contrast, when the verb occurs with a lexical object (i.e., in a transitive, non-reflexive frame), the reflexive interpretation is not available. This indicates that the theta role assigned to the lexical objects (in (30)-a″, b″, c″) is being assigned to the subjects in (30)-a, b, c and the reflexive pronouns in (30)-a′, b′, c′.

(30) a. Mary bathes in the morning

[5]I speculate that another interpretation is that the pragmatic content of these sentences allows for the interpretation that is not entailed by the syntactic structure. In other words, if the intransitive version of the verb *shot* is plausibly referring to the activity of shooting deer, then there is no reason to revise the initial interpretation that the *unmentioned* object of the shooting was, in fact, the same deer that did the running as well. In this case, it is not that the parse is lacking richness or detail, but that the participants' world experience allows them to imagine a situation in which the unmentioned object is also the subject of the main clause.

a′. Mary bathes herself in the morning

a″. Mary bathes the baby in the morning.

b. The dog scratched all day.

b′. The dog scratched itself all day.

b″. The dog scratched the door all day.

c. John dressed in a hurry.

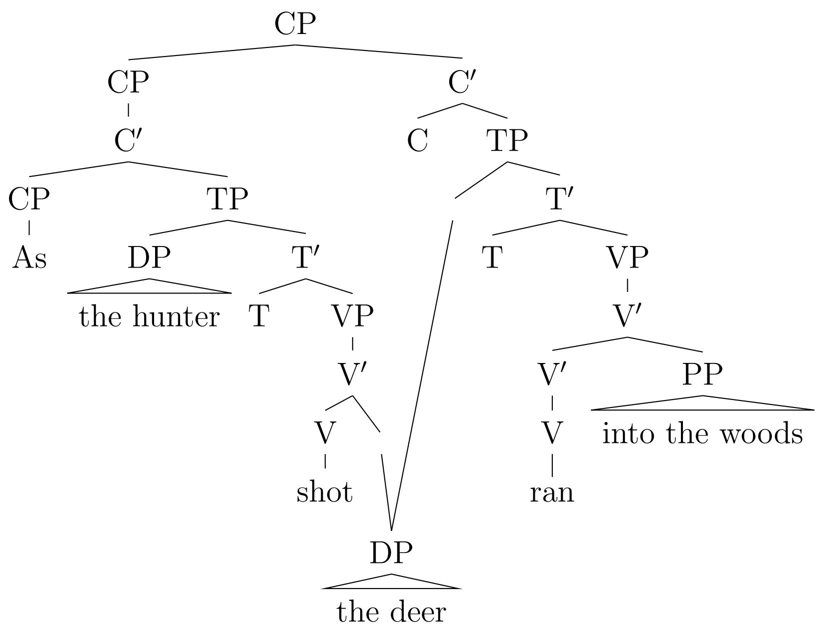
c′. John dressed himself in a hurry.

c″. John dressed his son in a hurry.

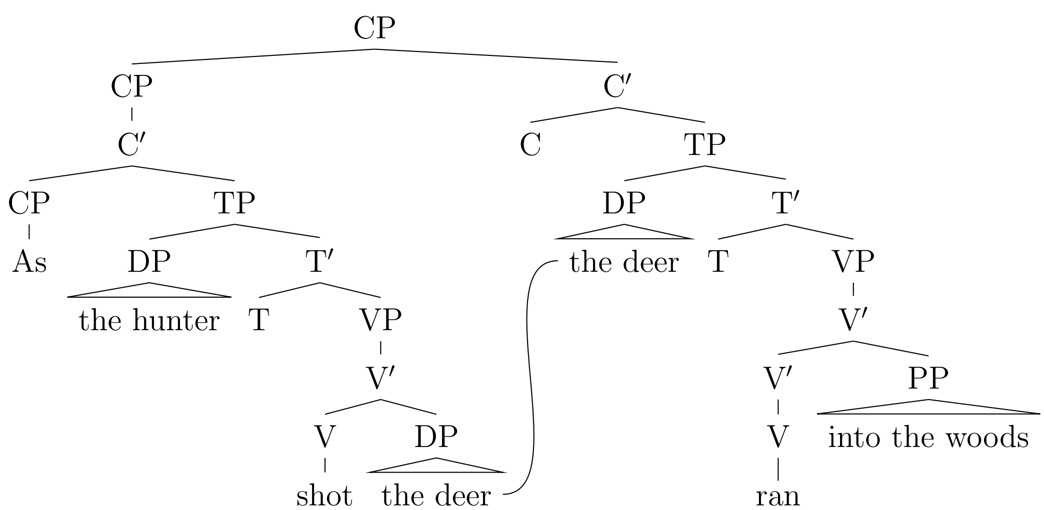
Christianson and colleagues find that the semireflexive verbs display similar effects to the ambiguous verbs used in the previous two experiments, but to a lesser extent. This is attributed to the semireflexive frame retaining a transitive structure even when no object is present. Since the over, transitive frame (not semireflexive) has a lower probability, a garden path is less likely to occur when the parser is more likely to presume a reflexive interpretation.

From this, they conclude that the structure built in the initial embedded object analysis (which is globally ungrammatical) is never fully pruned or removed from memory during reanalysis. Rather, the parser somehow "steals" the phrase from the object position and connects it to the matrix subject position, illustrated in (31). Thus, the original structure is still present to some degree, and still receives some amount of interpretation. One way that this might be feasible is if the parser tries to conserve energy by doing the minimum work necessary to build the final parse, thus it does not expend energy for pruning original analyses (especially if the original interpretations are consistent with the final parse).

(31)



(32)



Another way is if the parser fails to delete the erroneous copy of the phrase in the embedded object, but instead duplicates it and places the duplicate in the matrix subject position (32). However, there is substantial evidence that the parser is highly sensitive to subtle syntactic relations between words in unambiguous constructions, so it seems that the parser is capable of building a rich and detailed representation online (Kazanina et al., 2007; Stowe, 1986; van Gompel and Liversedge, 2003). Yet it is unspecified how either of the possibilities in (31)/(32) would be interpreted after construction, or why such duplication or double association would be tolerated.

What is made clear from this study is that the initial garden path is consistent with the final interpretation, despite the parser presumably attempting to reanalyze the sentence to make it globally grammatical. This suggests that the initial parse lingers in some way, whether in memory or in the semantic or pragmatic representations. Whichever way this comes about, the effect is present long enough and strong enough to influence the way in which participants responded to the subsequent offline comprehension question. Furthermore, some comprehension questions are able to detect the lingering (mis)interpretation, while others are not. In order to distinguish when a reader has completely discarded the original analysis in favor of the revised, grammatically correct analysis, the design of the comprehension questions is something future work should be careful to consider.

Christianson and colleagues end by speculating whether syntactic representation might not be as rich and detailed as previously thought. This speculation is not consistent with some of the other literature on how the parser seems to respond to constraints based on rich representations of hierarchical relations, such as those discussed in Section 2.5. But more importantly, I suggest that a serial model of the parser that is influenced by two interpretations simultaneously is better modeled by a parallel parser, thus eliminating the conflict between the lack of sensitivity to detailed structure as asserted in "Good

Enough" hypotheses and the apparent grammatical sensitivity observed in Stowe (1986), van Gompel and Liversedge (2003), and Sturt (2007), among others. In other words, if the parser constructs a rich and detailed representation of the sentence and is sensitive to grammatical constraints that rely on the detailed representations, the interference of the locally grammatical (but globally ungrammatical) parse with the globally grammatical parse is a prediction of some parallel models of the parser. Rather than weaken the sensitivity of a serial parser to grammatical relations, it seems more prudent to adopt a parallel model that does not strictly rank alternate parses and allows two potential parses to be highly activated and influence comprehension.

2.4.7. Sturt (2007)

Sturt (2007) takes a more optimistic view of the sensitivity of the parser to structural constraints and limitations. In this study, he investigates how the semantic interpretations of an ambiguous sentence change in response to reanalysis. This is based primarily on Christianson et al. (2001), but uses eye tracking in order to get a finer-grained picture of the time-course of reanalysis and interpretation. In order to approach this task, the stimuli in this experiment consist of a context sentence (33-a) followed by an experimental trial sentence (33-b) and a final sentence (33-c):

(33) a. The Antarctic expedition had been going on for months.

b. *Ambiguous match*:

The explorers found the South Pole was actually right at their feet.

Unambiguous match:

The explorers found that the South Pole was actually right at their feet.

Ambiguous mismatch:

The explorers found the South Pole was actually impossible to reach.

Unambiguous mismatch:

The explorers found that the South Pole was actually impossible to reach.

c. The temperature was 30 below zero.

Sturt finds that there is a robust effect of ambiguity in regression path reading times (alternately called go-past times), with ambiguous regions exhibiting longer reading times than unambiguous ones. The presence of such an effect when the semantic congruity mismatched was clearly seen in second-pass times, in particular. This leads Sturt to suggest that semantic persistence is the cause, since the syntactically ambiguous region was so short in length. That is, due to the short length of the syntactic ambiguity *found the South Pole*, the semantic interpretation is not discarded before the syntactic reanalysis is complete. This results in the observed perseverance of the interpretation associated with the original representation. Further, he speculates that semantic and syntactic processing might be more independent than is often assumed in the literature.

This is an intriguing possibility, which could explain some of the plausibility effects observed in previous literature in which interpretation of the thematic role (or other semantic content) may be somehow distinct from the syntactic constraints on their roles (Tabor and Hutchins, 2004; Tanenhaus et al., 1989; Trueswell and Tanenhaus, 1994). That is, if a relation between entities is described in a way that is initially interpreted as one configuration, but is later reanalyzed as being another configuration, *and* if the first configuration isn't directly in conflict with the newly constructed syntactic representation, then the semantic processing mechanism is not under sufficient pressure from the syntactic processing mechanism to revise its representation. For example, in (34), there is nothing about the syntax of the sentence that would prevent Mary from also dressing the baby, as well as herself.[6]

(34) While Mary dressed the baby cried in the other room.

In contrast, there is no reported semantic persistence of which I am aware for sentences like (35), in which the parser constructs a representation where the tail of the wh-dependency triggered by *who* is initially expected after *bring*.

(35) I wonder who Ruth will bring us home to ____ at Christmas.

The parser experiences a slowdown at *us*, which is the word that prevents the tail of the wh-dependency from being located after *bring*. Yet, I have not found reports that the semantic interpretation of this sentence is ever inconsistent with the syntactic representation after analysis is complete. It seems plausible, therefore, that the syntax prevents the possible configuration in which the entity indexed by *who* is both brought home by Ruth and being presented something or someone by Ruth once at home. Put another way, I speculate that who does not semantically occupy both positions, and a potential explanation for this is that the syntax prohibits such a configuration because *us* blocks

[6]The world knowledge of how one can occupy space in one room but not also a different room simultaneously (as indicated by *While*) does prevent this configuration from being plausible.

it. In this case, the difference between local ambiguities in sentences like (35) where perseverance is not reported (or intuited) and (34) where perseverance is quite noticeable is that the initial interpretation of (35) is inconsistent with the final interpretation, whereas the initial interpretation of (34) is compatible with the final interpretation.

2.4.8. Slattery et al. (2013)

One final paper investigating the properties of the Good Enough approach to parsing is by Slattery et al. (2013). In this paper, they ask whether syntactic reanalysis is actually completed in a rich and detailed manner in those sentences that are reported to exhibit a lingering effect of the initial garden path parse. To do so, they manipulate the presence of an ambiguity (with or without a comma after the ambiguous verb *telephoned*), and the gender match between a noun phrase with definitional gender (e.g., *mother* or *father*) that falls within a temporarily ambiguous region and a reflexive pronoun (*himself*) later in the sentence, e.g. (36).

(36) a/b. After the bank manager telephoned(,) David's father/mother grew worried and gave himself approximately five days to reply.

If the parser is building a detailed representation of the sentence, then there should be a slowdown observed at the disambiguating word (*grew*) in the ambiguous conditions and not the unambiguous conditions. Moreover, there should be a slowdown at the reflexive pronoun (*himself*) in the gender-mismatched conditions (*mother*) and not the gender-matched conditions (*father*). This should only happen if the parser is accurately reanalyzing the ambiguous noun phrase *David's father/mother* as the matrix subject (i.e., the subject of *grew*), which would make this noun phrase accessible as an antecedent for *himself*, via constraints on reflexive binding as described in Binding Condition A (Chomsky, 1981). If the parser is not attempting to reanalyze the sentence, then the ambiguous noun phrase is in the embedded object position and is not an accessible antecedent for *himself*. Therefore there should be a slowdown at the reflexive pronoun regardless of gender, as the parser fails to find an accessible antecedent. Furthermore, if the parser is not building a rich and detailed representation of the sentence, either in the first or second attempt at an analysis, then there is no reason that *David* or *father/mother* should be distinguished between during an antecedent search triggered by *himself*. In this case, as *David* matches in gender with *himself*, there should be no slowdown at the reflexive pronoun, since there is always a potential (albeit ungrammatical) antecedent available.

Slattery and colleagues find that the parser builds a sufficiently detailed representation of the sentence to support reflexive binding constraints, with a slowdown observed in gender-mismatched conditions. This suggests that the parser is not relying on an underspecified "Good Enough" representation in this circumstance. However, they do observe semantic lingering effects which are consistent with Sturt (2007), where the semantic interpretation lingers long enough to influence the interpretation of the sentence following the online portion of the trial.

In order to account for their observations within the assumptions of previous Good Enough hypotheses, Slattery and colleagues suggest that rather than building underspecified parses, the parser is failing to prune old, erroneous structure from the parse tree. Thus, when the parse is interpreted, there remains some representation in memory from the original garden path parse that has not been removed despite being overwritten, and this remnant parse can receive interpretation. One way this could happen is that the parser may be operating with a fallible working memory store that is unable to distinguish between the original, erroneous parse and the correct reanalyzed parse when selecting a representation to be interpreted. This possibility requires that multiple distinct representations be stored in working memory, whether intentionally or not, thus a parallel parser might be better suited for such a model, rather than a serial parser (which is assumed). Alternatively, the parser may overlay the revised structure on top of the original structure, without fully deleting or discarding the original structure. However, this possibility is not clearly distinguishable from a multiple-output model, which falls into the parallel model category. Such a parallel model of the parser (as I have described above) is a prudent and efficient way of describing the behavior of the parser in these studies, without resorting to weakening the parser's ability to retain rich and detailed representations.

In whichever way Slattery and colleagues would decide that their observations reflect on the properties of the parser, they do make one strong conclusion: the parser does construct a detailed syntactic representation of the incoming information after it recovers from the initial garden path. This is clear because the reflexive coreference dependencies are able to accurately form between the reflexive pronoun and its (grammatically accessible) antecedent, which is only possible with a rich syntactic representation. Thus, lingering effects that are observed may be attributable to the semantic interpretation mechanism or to the parser's failure to discard previously built structure rather than a parser that underspecified the syntactic representation that it built. This is further supported by ERP evidence from Qian and Garnsey (2015), who showed that any mistakes made by participants in answering questions were related to plausibility measures and not due to a lack of reanalysis after the disambiguating region was encountered.

It is important to emphasize that aspects of models assumed in each of the studies are stipulations when discussing the behavior of a serial parser, but are theoretically motivated and built into parallel models, and specifically multiple-output models of the parser. Since Slattery et al. (2013) does demonstrate that the parser must be constructing a sufficiently detailed representation during the reanalysis of the sentence, any other model that attempts to account for this data must have a detailed representation available. In multiple-output models, it is possible that both the garden path parse and the ultimately correct parse built after disambiguation are stored and accessed simultaneously during interpretation. However, the garden path structure is preferred (or more highly activated or ranked) initially, which could account for the garden path effects observed. Then, once the parser encounters the disambiguating region, the alternative (globally grammatical) parse rises above the garden path parse (in rank or activation) and is eventually adopted. In order to account for the apparent lingering semantic effects that have been consistently observed, one may look to the discussion in Sturt (2007) that suggests distinct syntactic and semantic processing mechanisms. Whatever the case may be, it seems that the digging-in effects from semantic reinterpretation do not indicate that the parser is building an underspecified syntactic representation, either in its first attempt that results in a garden path, or after the ambiguity has been resolved.

2.5. Long-distance dependency formation

A common tool for examining the parser's preferences concerning verbal argument structure is long distance dependency formation. Long distance dependency formation may introduce additional biases that are not present in the absence of this operation. Since many studies regarding verb frame biases make use of long distance dependencies to test the parser's preference (e.g., the search for the associated tail of a moved wh-phrase as in Boland et al., 1995; Stowe, 1986; Traxler and Pickering, 1996, a.o.), the discussion of these studies on verb frames is complete without a discussion of the influence exerted by the parser's search for the tail of a long distance dependency.

2.5.1. Wh-dependency formation

To begin, I will define a dependency as a syntactically constrained link between terminal nodes in a syntactic tree. Much of the following discussion of dependencies will involve links between wh-phrases and positions in trees that are not pronounced, but are associated with their respective wh-phrase. While remaining theoretically agnostic to the content and creation of these dependencies, I will henceforth refer to the tail of the dependency either as such or as the gap position (when discussion wh-phrases). In contrast, the focus of this dissertation and the experiments described in the following chapters examine coreference dependencies, which differ from wh-dependencies in at least one crucial way. Rather than connecting a filler to a gap (e.g., a wh-phrase to its gap), coreference dependencies connect an anaphoric element like a pronoun to an expression with which it corefers. This coreferring expression could be another anaphoric element, or more relevantly, an R-expression (e.g., a proper name or definite noun phrase). Because there is no gap in this type of long distance dependency, the two linked positions in a coreference dependency have pronounced or written realizations and can thus form critical regions for experimental examination. However, coreference dependencies are not syntactically obligatory as wh-dependencies are. This means that the parser might expect a tail or antecedent for a coreference dependency, but would not violate any grammatical rules of no such dependency can be formed. In contrast, wh-dependencies that are incomplete at the end of a sentence are ungrammatical. For instance, compare (37) and (38).

(37) Did he eat an apple? (No sentence-internal coreference, no coreference dependency formed)

(38) *What did John eat an apple? (Wh-dependency triggered but no gap site found, ungrammatical)

In this way, wh-dependencies and coreference dependencies differ and are not always necessarily comparable. In the following section, the implications of dependency formation are discussed in detail, including what can be learned from both wh-dependencies and coreference dependencies.

2.5.1.1. Stowe (1986). Starting with wh-dependencies, which have been paid much attention in the literature, Stowe (1986) investigates how the parser searches for the gap or tail of an open dependency. It is possible that the parser waits for confirmation that it has found a gap before forming a dependency, in which case it should make relatively few mistakes regarding closing dependencies. On the other hand, if the parser is eager (or "active"), then it may attempt to close a dependency as soon as a potential position for the tail is identified. In this case, the parser may be surprised to find a word in the position of the tail which conflicts with the representation it has built. That is, the position it may have chosen to be a gap linked with a wh-word is, in fact, filled and thus cannot be the location of the gap. These are the predictions made by Stowe (1986), and her results support the second hypothesis: the parser is eager. This means that in sentences such as (39), restated from (35), the parser attempts to close the dependency after the verb, but is thwarted when it encounters the word us, resulting in a reading time slowdown.

(39) I wonder who Ruth will bring us home to ____ at Christmas.

Moreover, the parser only seems to posit gaps where they would be grammatically licit and not where they would require the dependency to violate syntactic constraints. For instance, (40) illustrates how the tail of a wh-dependency cannot be located in a prepositional complement of the subject noun phrase (40-b), but when there is no wh-dependency, the corresponding position can be occupied by a noun phrase (40-a) (Stowe, 1986).

(40) a. The story about Susan annoyed her boyfriend.

b. *Who did the story about ____ annoy her boyfriend?

In other words, wh-dependencies cannot link the wh-phrase at the beginning of a sentence (presumably located in the specifier of CP) with a position within an island. The parser respects this constraint, and does not show any slowdown in the grammatical island. This consistency between the syntactic constraints and the lack of processing difficulty that the parser has while reading islands (as compared to non-island candidate for gap locations) suggests that the parser is not positing an illicit gap in an island. Stowe suggests that that the parser could still be positing a gap in this location, where it cannot grammatically occur, but this possibility is filtered out of the possible options quickly enough to show no increased processing cost. But, if the possible illicit gap is posited and filtered quickly enough to be undetectable, it may not be possible to distinguish these two possibilities. To conclude, Stowe suggests that the wh-dependencies she examined were completed eagerly, with the parser anticipating structure in a top-down manner, and respecting syntactic constraints.

2.5.1.2. Traxler and Pickering (1996). Traxler and Pickering (1996) also examine the behavior of the parser during long distance dependency formation. They ask whether the parser waits to find a gap to form the dependency or instead whether the parser anticipates the location of the gap site as soon as it has enough other syntactic information to make a prediction. To examine this question, they manipulate the plausibility of the filler and the first potential gap site. That is, as in (41), if the parser predicts the gap site well in advance, it should exhibit plausibility-related effects before the gap site is reached. Otherwise, it should only exhibit such effects when the parser reaches the gap and subsequently realizes that the relation between the filler (the word that triggered the search for the tail of a dependency) and the argument position in which the gap is located.

(41) a. That's the pistol with which the heartless killer shot the man ____...

b. That's the garage with which the heartless killer shot the man ____...

They find an increase in reading time at the implausible verb, before the gap site is reached, thus before it is known whether a gap site exists in that location. From their results, Traxler and Pickering conclude that, independent of the theoretical details of the parser, the search for the creation of a long distance dependency between a filler and a gap has predictive qualities. That is, the parser seems to anticipate the position of the tail of the dependency in advance of encountering the semantic and plausibility information supplied by the verb that dominates the first potential position for the tail of the dependency. In the theoretical framework that I am using to describe the parser, this could also be stated as the parser predicting the argument structure of the verb in order to find a potential gap site before the verb has been encountered.[7] The First Resort strategy or predictive nature of this parser may also explain how it does not seem to generate dependencies that will violate grammatical island constraints. Since the construction of an island should alert the parser to the impenetrability of the island, the parser will not predict a gap site in the island if it respects grammatical constraints.

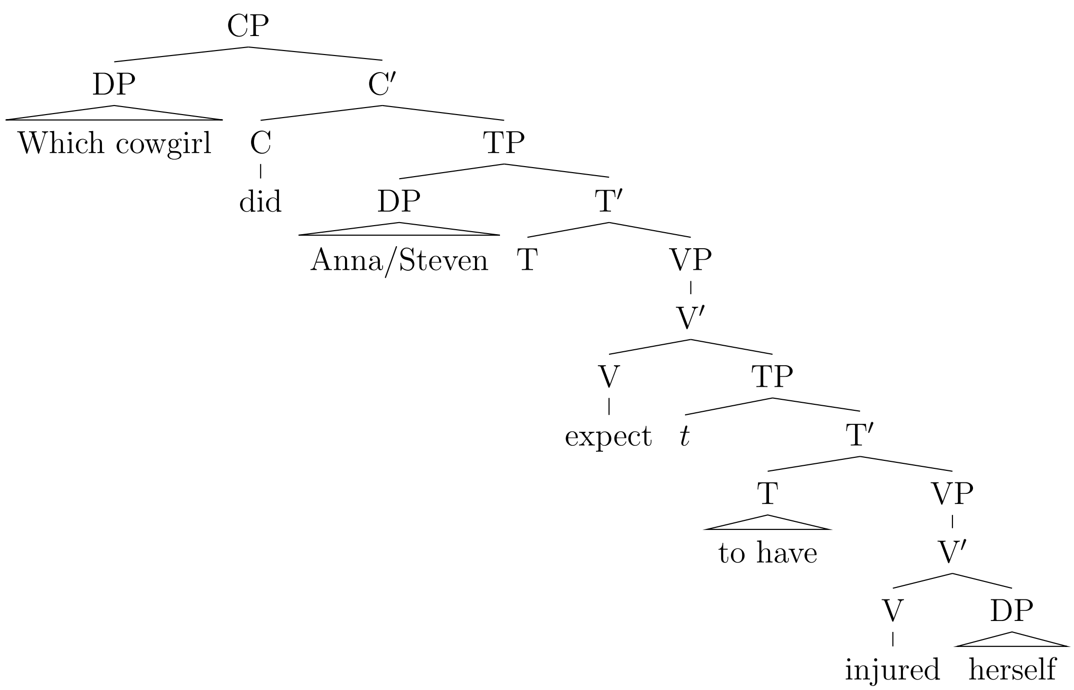
2.5.1.3. Frazier et al. (2015). If the parser is able to represent the parser tree in rich detail both initially and after reanalysis (Slattery et al., 2013), and the parser eagerly constructs wh-dependencies during this structure building process, one may wonder how the dependency formation process interacts with other competing processes. Frazier et al. (2015) examine the way in which the parser navigates a type of local ambiguity during dependency formation. We make use of the phenomenon of local coherence to determine how strongly the representation of a wh-dependency is stored in memory during competition of other dependency formations. Local coherence, briefly stated, is a phenomenon in which the parser begins to construct a representation for a sentence but apparently loses

[7] I make no claim to the validity of this Government and Binding type of framework over others, rather I am using the terms and structures supplied by it for convenience and consistency. I believe that the majority of the results and phenomena discussed in this dissertation can easily fit with different frameworks, thus I leave making explicit reference to the distinctions between frameworks to future work.

track of the global structure and builds a locally coherent representation for a non-initial substring of a sentence, which does not form a coherent syntactic subgroup with the global representation. For instance, (43) is a substring of (42) that can form an independently coherent syntactic structure, but in the context of (42) is not consistent with the structure required for the wh-question.

(42) a. Which cowgirl did Anna expect to have injured herself?

b. Which cowgirl did Steven expect to have injured herself?

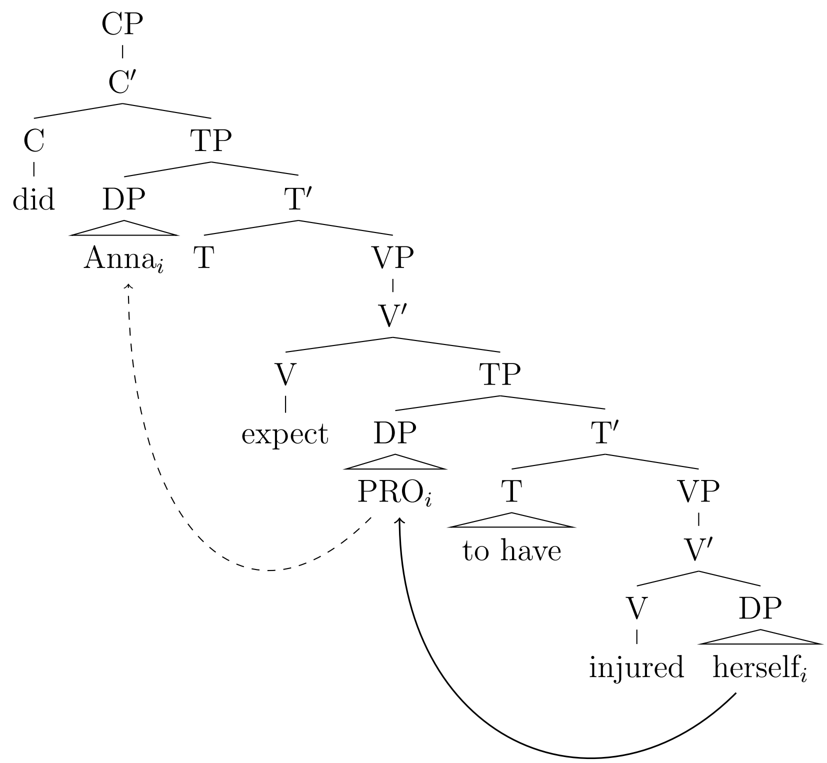


In these sentences, a substring like (43) could be analyzed independently as a coherent parse. This is locally coherent and a grammatical analysis, but globally it is ungrammatical. The local coherence phenomenon is comparable to the garden path phenomenon in that both involve the parser making local attachments that are likely given information from some of the most recent previous input but ultimately are globally ungrammatical attachments (Konieczny, 2005).

(43) ...[did Anna expect to have injured herself]?

a. Did Anna$_i$ expect PRO$_i$ to have injured herself$_i$?

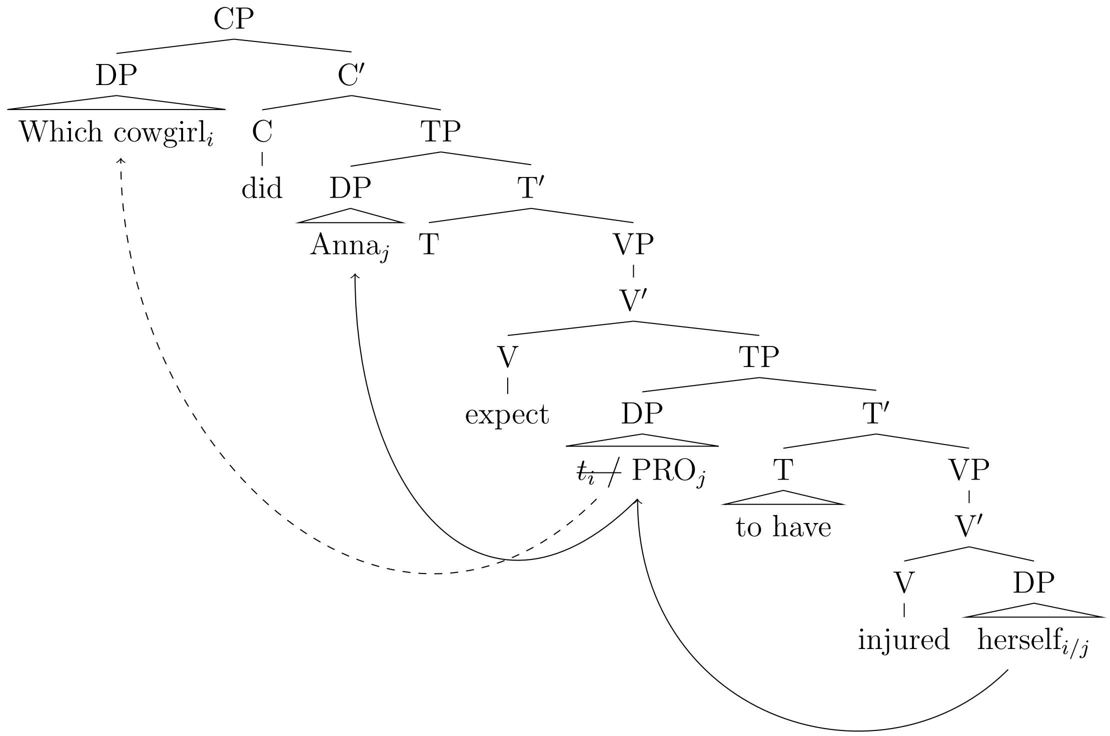
b.



The locally coherent string in (42-a) and (42-b) that has the potential to mislead the parser is delimited and restated in (43). (43) is grammatical by itself, because the string *expect to have injured* contains the phonologically null element PRO (43-a), which can corefer with both *Anna* and *herself*. This allows PRO in (43) to form a link between *Anna* and *herself*, which are dependencies the parser may attempt to form despite their non-obligatory nature (van Gompel and Liversedge, 2003). This is in contrast to the globally grammatical parse in which PRO links *which cowgirl* and *herself*, an obligatory set of dependencies. The anaphor *herself* must corefer with an antecedent that c-commands it, since, by Binding Condition A[8], it must be bound by its antecedent Chomsky (1981). In both (42) and (43), PRO binds (and corefers) with *herself*. But since this position is phonologically null, *Anna* is the linearly closest candidate antecedent. So, the question Frazier et al. (2015) asks is whether the active search for a gap triggered by the wh-phrase (44) will interfere with the locally coherent analysis in (43-b).

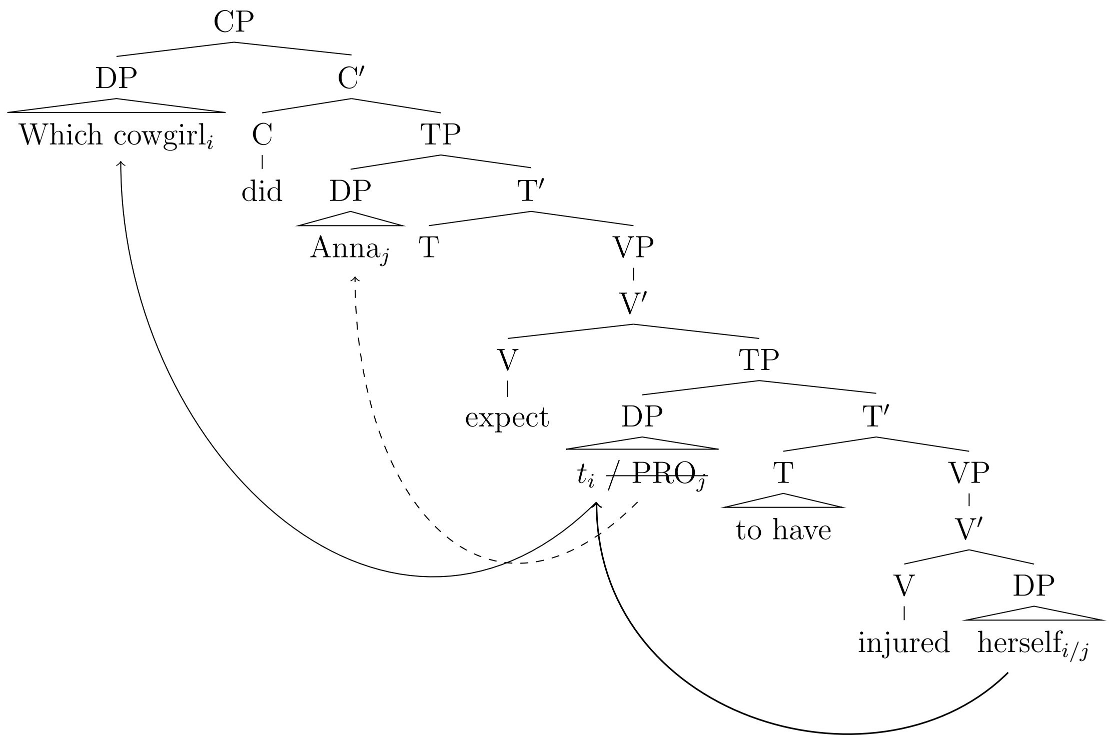
[8]Binding Condition A, briefly, states that a reflexive pronoun, like *herself*, must be bound by its antecedent. Typically, this means the antecedent is found linearly close to the reflexive pronoun, though there is no requirement outside of c-command relations for it to be so.

(44) Locally coherent but globally ungrammatical mid-parse state of (42):



Here, the search for the antecedent of *herself* looks through the locally coherent (but globally ungrammatical) tree in which the position after *expect* is a PRO rather than a trace of the wh-phrase. This PRO is controlled by *Anna* in the locally coherent parse. Thus, the parser interprets the coreference relation in (44) as being between *herself* and *Anna*, instead of between *herself* and *Which cowgirl*, as it should in the globally grammatical parse (45). In the case of (44), interference effects from the locally coherent parse are expected to obtain.

(45) Globally grammatical mid-parse state of (42):



The competition between dependency formation arises when the parser reaches *herself* at the end of the sentence in (42). At this point, the parser must form a coreference dependency between *herself* (a reflexive pronoun) and some antecedent that c-commands it. Once the parser reaches the reflexive pronoun, it must look leftward through the tree to find a candidate antecedent. The grammatical constraint that limits at which positions a candidate antecedent can be located is Binding Condition A (Chomsky, 1981), which briefly states that a reflexive pronoun must be c-commanded (i.e., bound) by its antecedent. Often, this means that the reflexive antecedent is linearly close to the reflexive pronoun. However, in stimuli used in this experiment, the closest (overt) candidate antecedent is not the globally coherent one. The closest antecedent is the gap locations that serves as the subject of the embedded clause. If the parser has constructed the locally coherent (43) representation, then this gap is already linked to *Adam*, and *herself* should be interpreted as referring to *Adam* (at least at stages of processing before semantic consistency is evaluated). This would result in a slowdown associated with a gender mismatch. However, if the parser has maintained the global representation because the tail of the wh-dependency is strongly represented and held in memory as being in the gap position, the *Which cowgirl* is linked with *herself.* Because these two phrases are gender congruent, no slowdown is predicted. We found that this locally coherent substring does not influence the parser during dependency formation, despite previous work showing that this type of sentence is frequently subject to local coherence effects (Konieczny, 2005). Thus, we concluded that the influence exerted by the formation of the wh-dependency can prevent the parser from considering the locally coherent (but globally incoherent) representation during the formation of the later coreference dependency.

2.5.2. Coreference dependencies

Coreference dependencies, as mentioned above, seem to operate in similar ways as wh-dependencies. Although there is no "gap" or unfilled argument position for the parser to consider as the tail of the dependency, the parser still performs a search for candidates for the antecedent or dependency tail position. The following literature discusses how the parser behaves during a search for a coreference dependency, and what types of information (lexical, morphological, syntactic, and semantic) may constrain or influence the search.

2.5.2.1. Cowart and Cairns (1987). To begin, Cowart and Cairns (1987) examine how the parser reacts to the morphological feature of number between the pronoun they and an ambiguous verb in a subsequent clause. In their manipulation, they was contrasted with a lexical noun phrase, and the main clause contained either the verb is (singular) or are (plural). The subject of the main clause was in the form VERB+ing NOUN+s, in which the verb was ambiguous between a nominalization of a verb followed by an object noun phrase and a modifier gerund followed by a noun, as in (46) *packing cases*.

(46) a. While the boxes usually come with several internal partitions, packing cases are/is...
(the action of packing something into cases)

b. While they usually come with several internal partitions, packing cases are/is...
(cases that are of the variety called packing cases)

They find that the presence of the pronoun *they* in the first clause influences how the second clause is read, with sentences containing the pronoun exhibited faster reading times for *are* and slower reading times for *is*, while the reverse effect is found in sentences containing the lexical noun. Moreover, in their subsequent experiments, they observe that this effect is mediated by the syntactic relation between the pronoun and the VERB+ing NOUN+s phrase, indicating that this is a syntactic process. Semantic relations do not seem to have an effect, providing further evidence that these results are due to the syntactic and not semantic structure of the sentences. From their observations, Cowart and Cairns come to the conclusions that the parser is not strongly interactive because such theories would predict that non-syntactic (i.e., semantic) information is integral to interpretation.

Rather, the parser must be at least somewhat modular, such as a two-stage approach akin to Frazier and Fodor (1978). I would like to add that the strongest theories of interactive parsing may make this prediction, but they are a bit of a straw man. Since one is able to produce and accept sentences that are semantically bizarre but grammatical, it seems that any theory that would prevent the parser from parsing a bizarre but grammatical sentence would massively undergenerate possible outcomes. Rather than preventing the parser from building semantically implausible (but grammatical) sentences, a reasonable model of interactivity would include plausibility as a biasing force instead of a constraint.

2.5.2.2. van Gompel and Liversedge (2003). van Gompel and Liversedge (2003) extend this work done by Cowart and Cairns (1987) to look at more specific non-syntactic properties that might influence coreference dependency formation. Specifically, they focus on gender and number, while ensuring that any effects they observe are not due to the introduction of a new discourse entity. One thing to note here is that number in English is a morphological property of the lexical item, whereas gender is a more complex type of information, and is not purely morphologically encoded. That is, while *he* and *she* denote male and female genders, even canonically gendered names such as *John* and *Mary* are not morphologically marked for gender. Thus, any gender effects observed in these sentences are not due to syntactic or morphological encoding of gender, but by the encoding or association of gender through other means.[9]

[9]A tangential but important note, since the use of canonically gender-specific names can change over even short periods of time. Even definitional gender nouns, such as *cowgirl*, can be associated with male pronouns in certain, limited contexts. For example, in the context of a costume party, it would be interpretable (and possibly also acceptable) to say "The cowgirl left his lasso in the kitchen," if the referent of *cowgirl* was a man dressed as a cowgirl. Thus, one must be careful when discussing the property of gender in the context of syntactic and morphological constraints.

When the parser encounters a pronoun, the dependency search is triggered, much as it would be upon considering a wh-phrase that triggers a search for its associated gap. If the anaphoric element is a pronoun that is encountered early on in the sentence, then the parser may not look backward through the previously encountered material (within the sentence) for a coreferent. This type of pronoun might be a cataphor, or backwards anaphor, which look for their "antecedent" rightward in the tree rather than leftward. I will continue to use the term "antecedent" to refer to the candidate coreferring expression at the tail of the coreference dependency for the sake of consistency. As the parser builds new structure and receives new input, it will need to determine when and where to form a coreference dependency. The method by which the parser actually forms a dependency is discussed in the following few papers. What is known, though, is that coreference dependencies are not grammatically obligatory. Pronouns (excluding reflexive pronouns) may corefer with expressions in other sentences, or they may remain unresolved or arbitrary. Based on the findings discussed below, the parser seems to prefer to search for a coreferring expression within the sentence which it is located, despite its nonobligatory status.

In their first experiment, van Gompel and Liversedge observe that a pronoun that is later on followed by a gender-congruent lexical noun phrase was read more easily than when it was followed by a gender-incongruent lexical noun phrase. The pattern of reading times they observed indicated that the structural relation between the pronoun and the candidate antecedent is computed before realizing that the gender incongruent noun phrase makes a poor antecedent. In addition, they find that the parser preferentially forms the dependency at the first candidate noun phrase it encounters, even when there is a better, gender-congruent noun phrase following shortly after the first candidate. This is consistent with an active dependency search and an eager parser that predictively forms a dependency before it has computed the non-syntactic information.

2.5.2.3. Kazanina et al. (2007). To further investigate the eagerness of the parser in forming coreference dependencies, Kazanina et al. (2007) ask at what point syntactic constraints influence the coreference dependency formation process. In the formation of wh-dependencies, syntactic constraints are known to influence the locations that the parser can initially posit a gap. However, since coreference dependencies do differ from wh-dependencies in some key ways, it cannot be taken for granted that they operate under the same constraints. Although van Gompel and Liversedge (2003) observed a slowdown associated with incongruent gender, Kazanina is interested in whether this is mediated by the syntactic accessibility of the antecedent. To test this question, she compares gender congruent and incongruent pairs of pronouns and canonically gendered names. These pairs are either in a configuration that would grammatically violate a syntactic constraint if a dependency were formed between the two words, or in a configuration that would allow such a dependency, as illustrated in (47).

(47) a. Because s/he was taking classes while Kathryn was working two jobs Erica felt guilty.

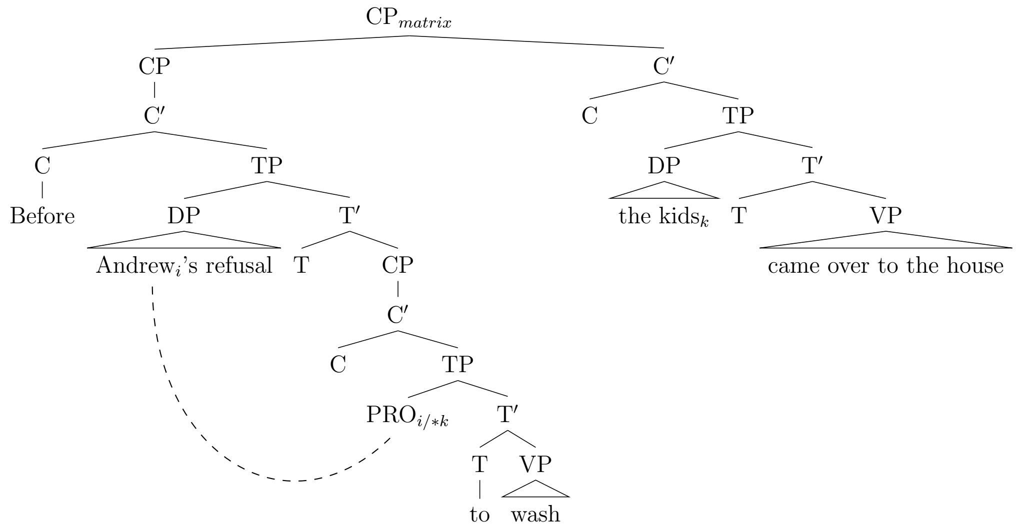
b. Because while s/he was taking classes Kathryn was working two jobs Russell was lonely.

The constraint that would prohibit (47-a) is Binding Condition C, which is a constraint on the hierarchical relation between an R-expression (referring expression, such as a name or lexical noun) and a pronoun. Briefly put, Binding Condition C (BCC) states that an R-expression cannot be c-commanded by a pronoun with which it corefers (Chomsky, 1981). This means that in (47-a), where *she* c-commands *Kathryn*, *Kathryn* is bound by *she*, thus they cannot corefer. However, in contrast in (47-b), *Kathryn* is not c-commanded by *she*, thus a coreference dependency between these two positions would be licit.

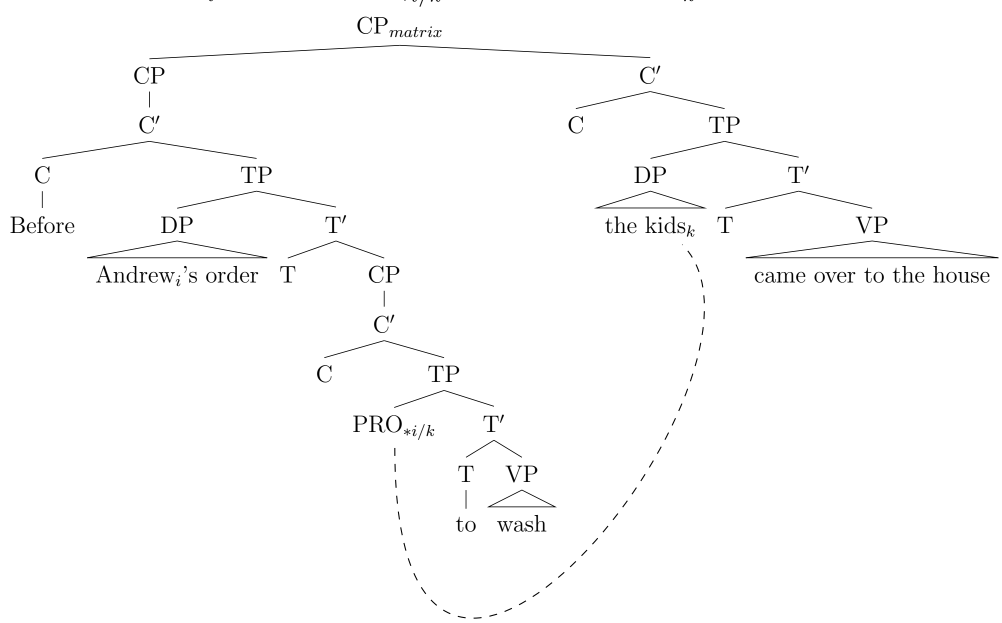
Kazanina and colleagues find that the grammatical constraint immediately influences the parser's construction of dependencies. In cases like (47-a), no effect of gender mismatch is observed, whereas in cases like (47-b), there is a significant slowdown in the mismatch condition. These observations are replicated in a number of related conditions, indicating that the parser is acutely influenced by grammatical constraints such as BCC when anticipating a coreference dependency. Moreover, they suggest that coreference dependencies are built as soon as the parser can anticipate a location that could contain a candidate antecedent, well before the content of the lexical items is processed, or even encountered. Then, once the bottom-up lexical information is available, the dependency (which has already been built) is evaluated for consistency given what information is encountered. If the information is inconsistent (such as incongruent gender, or presumably an implausible relation due to real-world knowledge, as seen in van Gompel and Liversedge (2003)), then the dependency is discarded. There is still a possibility that the constraints are respected early on, during initial structure building, but are violated later on as the parser revisits or reconsiders previously built representations (Sturt, 2003). However, this is not pursued further.

2.5.2.4. Kwon and Sturt (2014). In their 2014 paper, Kwon and Sturt test whether the predictive or anticipatory nature of dependency formation is strong enough to influence ambiguity resolution. They probe the parser's behavior during ambiguity resolution using a long-distance dependency triggered by an anaphor. The anaphor in question in this case is PRO, which is a phonologically unpronounced element, in the case described herein, the subject of a non-finite clause that is assigned coreference by a controller (Chomsky, 1981; Kwon and Sturt, 2014). Like overt pronouns, arbitrary PRO does not require an antecedent to be present within the same sentence. If an antecedent exists, the positions in which it must be located in order to form licit dependencies are restricted by the PRO Theorem (Chomsky, 1981), whereby PRO is ungoverned. Because PRO is silent, it does not carry indicators of gender or animacy. This is useful for testing the parser's behavior in structurally ambiguous strings where potential antecedents could be probed using gender or plausibility information because without indication of these properties, the ambiguity is not resolvable at this point and the parser will continue to build structures that can be probed later on. The structurally ambiguous strings used in this study are based on control relations in sentences like (48). In control relations such as these, an argument of a verb, such as Andrew in (48-a) and the kids in (48-b), determines the antecedent of PRO. In particular, they examined whether the parser's search for the antecedent of PRO would bias the parser against the transitive analysis of the verb in the adverbial adjunct clause in order to more quickly resolve the open dependency. If the parser exhibits strongly active processing in which the bias to complete the dependency early surpasses all other influences on the parser, then they expect that the dependency should be completed before the disambiguating region.

(48) a. Before Andrew$_i$'s refusal PRO$_{i/*k}$ to wash the kids$_k$ came over to the house.



b. Before Andrew$_i$'s order PRO$_{*i/k}$ to wash the kids$_k$ came over to the house.



In (48), there is a local attachment ambiguity in which the kids could be either the object of *to wash* or the matrix subject. In (48-a), PRO (the person washing) cannot corefer with the matrix subject *the kids* because this relation is determined by the semantics of the control nominal *refusal*. In its verb form, *(to) refuse* is a subject control verb, whereas *(to) order* is an object control verb. The properties of their verbal forms are passed on to their nominal forms. Thus, the specifier of *refusal* (i.e., *Andrew*) will control PRO in (48-a). In (48-b), this PRO is arbitrary, but could corefer with *the kids* and cannot corefer with *Andrew*, due to the constraints on object control. Thus at the point during structure building when the parser posits a PRO in (48-a), PRO can then be assigned a controller (*Andrew*), since *Andrew* is in the proper structural relation to control PRO and assign coreference.

In contrast, in (48-b), PRO cannot corefer with *Andrew* at the point PRO is posited because *order* is an object control nominal and thus even if PRO is arbitrary, *Andrew* is not an accessible antecedent. Furthermore, this PRO must be linked to a DP that comes in the downstream, which is not c-commanded by PRO (Binding Condition C blocks such link with lexical noun phrases). Now the only possible antecedent for PRO is the matrix subject. Although the input *the kids* has not yet been encountered at this point, there is motivation for the parser to start a search for the antecedent of PRO, the tail of the coreference dependency, because the parser has built the skeleton of the matrix clause which contains a matrix subject position. When the embedded verb is encountered, the parser recognizes that the verb is ambiguous. Thus the incoming NP, *the kids*, can be either the embedded object or the matrix subject.

One possibility is that the parser prioritizes the formation of the dependency triggered by PRO, which is actively searching for a candidate antecedent, over forming a local attachment, which would determine where the parser could licitly find an antecedent. If so, then it may analyze *the kids* as being the matrix subject, thus assigning an object controller to PRO quickly. If local attachment exerts a stronger influence on the parser than completing the control coreference dependency, then the parser may instead analyze *the kids* as the object of *to wash*, making it an illicit antecedent for PRO, and causing a garden path effect when the parser reaches the disambiguating information.

In the case where the ambiguous region contained a control noun with giver control, a strong influence of eager dependency formation should result in the parser preferentially building the late closure (embedded object) parse. In this parse, *the kids* are the object of *wash*, which is not consistent with the continuation of the sentence, thus should result in a slowdown. In the case of recipient control, a strong influence of dependency formation would then result in the early closure (matrix subject) parse, in anticipation of the main clause containing a suitable subject to serve as the coreferent of PRO which is not c-commanded by PRO (by Binding Condition C). Thus, this situation should result in no slowdown at the main verb since the dependency formation already biased the parser toward the correct (late closure) parse.

However, they find that there is an effect of the garden path in both conditions, albeit slightly smaller in in the Recipient condition, which is evidence against the strongly predictive account.[10] Since the coreference in these stimuli is optional, they execute a

[10]At this point, I would like to note a few things about their first experiment that weaken their claim. The stimuli they used in their first experiment have the potential for another interpretation in which the control noun is the final word in the initial embedded clause. For example, **??** could be initially parsed as "After Andrew's order, to wash the kids... seemed like a bad idea." Thus, even in the giver control

subsequent experiment that makes use of adjunct control (which has a stronger obligation for intrasentential resolution). In this second experiment, they make use of plausibility to test whether the parser predicts a dependency based on the control information. Again, they find garden path effects in both conditions (Giver control and Recipient control) and do not find sufficient evidence of a reading time benefit of Recipient control to conclude that the parser predicts early closure and the matrix clause based on the bias supplied by the control dependency.

They conclude that top-down PRO-dependency formation triggered by the control item does not interfere with the initial process of structure building because the garden path effect was not completely avoided in early reading time measures. Even so, they found that sentences like (48-b) with control information, which could bias the parser toward a matrix subject analysis, did not. Rather, they observed less reading time slowdown during regression path reading times, which are presumed to reflect reanalysis (after the initial garden path effects). This suggests that the parser is only able to make use of the control information after initial structure building.

From this study, it is clear that ambiguity resolution is not overwhelmingly influenced by the parser's eager search for the tail of a dependency. However, based on the comparable performance of coreference dependencies to wh-dependencies and the potentially

condition, PRO has the option of not coreferring with *Andrew*. Furthermore, it seems that Kwon and Sturt (2014) assume a serial parser or a parser that could only be influenced by one parse at a time. However, in the case that the parser maintains both the early and late closure parses simultaneously, then it is not clear that their results can be attributed to the parser experiencing a garden path because it only constructed the late closure parse. Since there is no position in the ambiguous region that can be tested for dependency formation (since PRO is silent and orthographically null), it is impossible to know whether the parser has or has not formed a dependency between *John* and PRO in either sentence. The only region that can be tested is the matrix verb, which means their results can only attest to whether the parser has constructed the late closure parse (seen in a slowdown), but not whether the parser has constructed the early closure parse. This is one way in which my dissertation builds upon this study, which will be discussed in more detail in Chapter 3.

predictive and eager nature of dependency formation, making use of cataphoric triggers for dependency search may clarify what Kwon and Sturt (2014) attempted.

In the following chapters, I will lay out my assumptions about the parser based on the literature described above, and I will make and test predictions based on the aspects of the parser I have considered here: the influence of verb frame frequency and transitivity, the underlying or default biases of the parser, and the influence exerted by the parser's eager dependency formation operation.

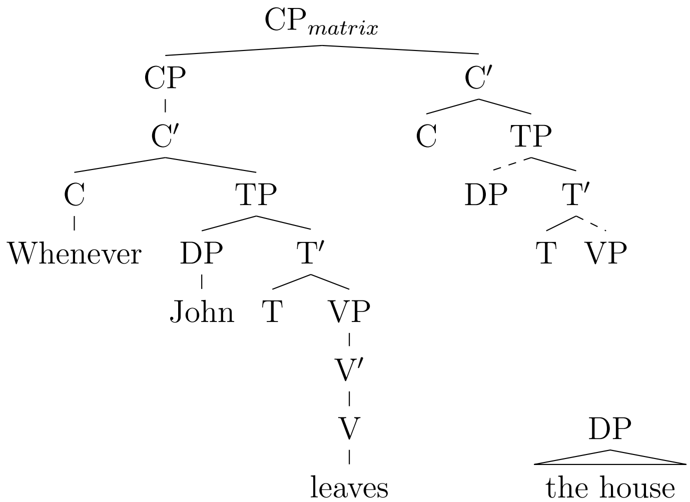
CHAPTER 3

Influence of equibiased verbs on cataphoric dependency formation

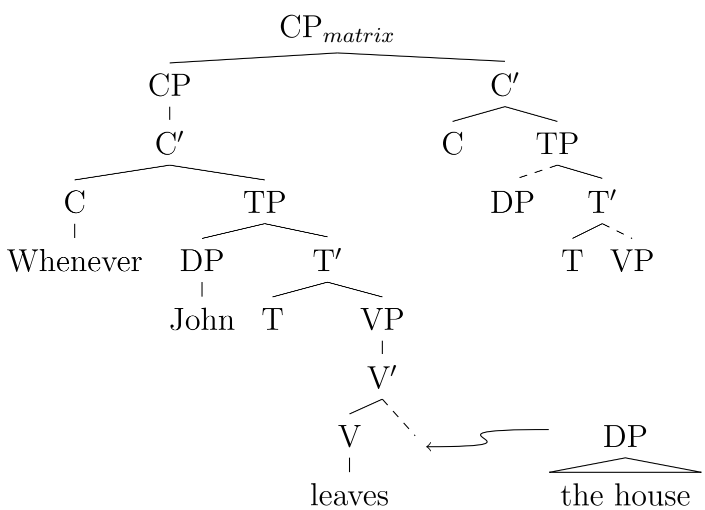
3.1. Introduction

As described in the previous chapter, the parser seems to have a preference toward building syntactic structures that is minimally costly in a local domain (Frazier and Rayner, 1982, 8; Gibson et al., 1996; Kimball, 1973; Phillips, 1995; Phillips and Gibson, 1997, a.o.). Given these observations, one may then expect that the parser will depend on local, bottom-up information such as lexical frequency and verb frame frequency (MacDonald et al., 1994; Staub, 2007) to resolve local ambiguity. However, it is unclear how the parser chooses to prioritize these sources of information. Below, I explore the ways in which the different types of information may affect the parser's structure building using an embedded object (EO)/matrix subject (MS) ambiguity, as in (1). In this example, the DP the house can be analyzed as the direct object of the verb leave (1-a) or the matrix subject located in the specifier of the matrix clause's TP (1-b).

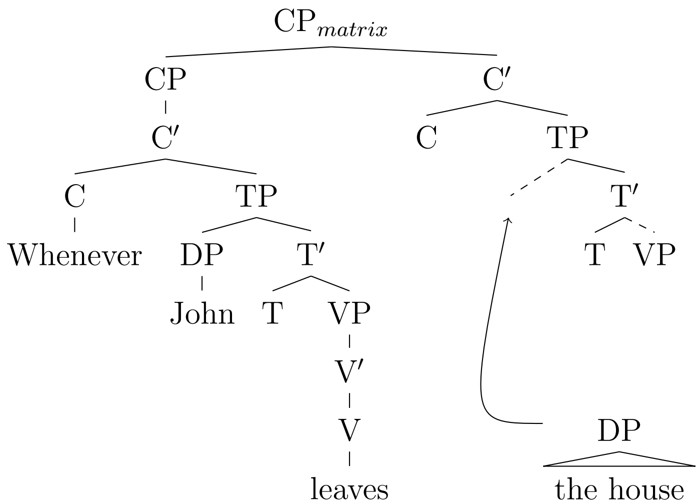
(1) Whenever John leaves the house...



a.



b.



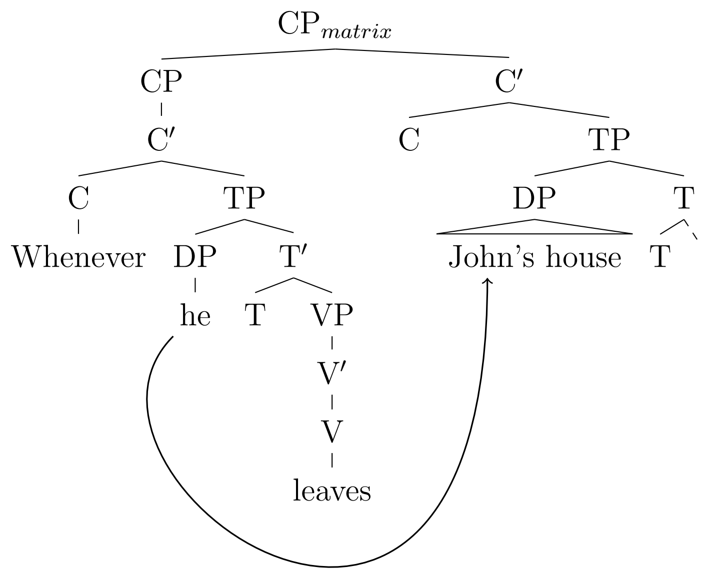
One way the parser may behave upon encountering an ambiguous string is to default to a minimally costly structure, in the case of (1), by making a local attachment during initial structure building (Phillips and Gibson, 1997). The strategy of local attachment illustrated in (1-a), which is also consistent with the strategies of Minimal Attachment+Late Closure (Frazier and Fodor, 1978; Frazier and Rayner, 1982), Properties of Recency Preference Gibson et al. (1996); Gibson (1991), and Right Association (Kimball, 1973; Phillips, 1995) in this example, will not always lead to a globally grammatical analysis. For instance, in a sentence like (1-b), making a local attachment (as the direct object of the embedded VP) will create globally ungrammatical analysis. Then, the parser must recover from this mistake in order to find a globally grammatical analysis when *is dark* is encountered. In (1-b), as shown by the arrow, a grammatical parse must attach *the house* as the matrix subject, rather than the embedded object.

One possible source of conflicting biases is the interaction of long-distance dependency formation, an eager process, and the parser's other biases, such as verb frame frequency information (Frazier and Clifton, 1989; Kwon and Sturt, 2014; Stowe, 1986; Tanenhaus et al., 1989, and discussion therein). If these two biases come into conflict, then the parser may be pulled in two directions. The verb frame information may bias the parser toward building an embedded object parse, as shown in (1-a). However, if the search for the tail of a long-distance dependency is a strong influence and if the embedded object parse is in conflict with the goal of quickly closing the dependency, the parser may prefer a different construction. In order to explore how such a competition may appear, (2) demonstrates how an active dependency search triggered by a pronoun may come into conflict with a bias toward local attachment.

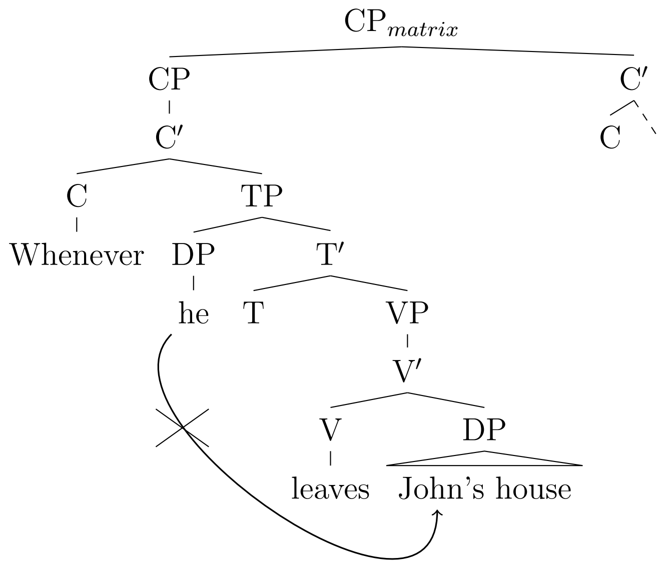
When the parser encounters an anaphor (i.e., a pronoun), a dependency search for its antecedent is triggered. In cases like (1) and (2), there is no candidate antecedent leftward in the tree, so the parser searches rightward. This "backward" dependency search for an antecedent after an anaphor (i.e., rightwards in the tree) will henceforth be referred to as a cataphoric dependency search. This search for the tail (in this case, the antecedent) of a long-distance dependency could influence the parser's preference for local attachment because the parser is known to actively look for (and possibly anticipate) the location of the tail of a dependency (Cowart and Cairns, 1987; Frazier and Clifton, 1989; Stowe, 1986). However, local attachment has also been shown to be a strong bias (Phillips and Gibson, 1997, a.o.). In the case of sentences like (2), the parser encounters the pronoun *he*, triggering the search to form a dependency between it and a candidate antecedent (Cowart and Cairns, 1987; Kazanina et al., 2007; van Gompel and Liversedge, 2003). This search is triggered before the parser encounters the ambiguous region. If dependency formation is a strong enough biasing factor to prevent the parser from building a local attachment, the parser might therefore build a parse that allows for the dependency to be completed sooner than what would have been possible with the local attachment. That is, a search for an antecedent of the pronoun may bias the parser toward an analysis that allows for the early completion of the coreference dependency by minimizing the distance between the pronoun and its antecedent.

(2) Whenever he leaves John's house…

Matrix subject:



Embedded object:



In (2), if *he* triggers an active search for an antecedent, and the parser is indeed influenced to build a structure in which the domain c-commanded by *he* contains as little material as possible, then it should be biased toward building an intransitive structure for the embedded verb leaves because this would minimize the material c-commanded by the pronoun, thus ineligible for coreference. This bias may be coupled with the bias of the adjunct complementizer *Whenever*, which should influence the parser to build the skeleton of the matrix clause. These two biases may conspire to influence the parser to predict that the next position in which a candidate antecedent could occur is within the matrix subject. This contrasts with a candidate antecedent in the embedded object position, which would violate the grammatical constraint called Binding Condition C (Chomsky, 1981). Binding Condition C asserts that a referring expression (in this case, *John*) cannot corefer with a pronoun that c-commands it (*he*, in the embedded object parse of (2)). If the parser builds a structure such that *John* occurs in the embedded object position or any other position that is c-commanded by *he*, it will not be available for coreference, thus the search for the antecedent of *he* must continue on for a longer duration than in the matrix subject parse. The parser must keep the pronoun that triggered the search in working memory for the duration of the search, which is a costly operation. Thus, by minimizing the distance between the pronoun and its antecedent, the parser minimizes the processing cost of the dependency. In this case, the parser may be motivated to build the intransitive version of the embedded VP because this could potentially allow it to hold the pronoun in working memory for a shorter duration.

Another potentially conflicting bias that the parser may be influenced by during ambiguity resolution is the frequency or likelihood of a particular argument structure occurring. Probability of a particular argument structure or verb frame has been shown to affect the parser's dependency formation behavior. In this case, the parser is biased toward building an argument structure that is likely to occur with the particular verb in question, rather than building a default structure determined by broad syntactic category alone (e.g., Crain and Steedman, 1985; MacDonald et al., 1994; Staub, 2007; Cf. Frazier and Fodor, 1978; van Gompel et al., 2012). That is, upon encountering an ambiguous verb that frequently occurs with a particular argument structure, the parser will be more likely to build that frequent argument structure even when other frames are possible (Lapata et al., 2001; Trueswell, 1996; Trueswell and Kim, 1998; Cf. Staub, 2007; van Gompel et al., 2012). Given this influence on the parser's behavior during ambiguity resolution, it may be expected that the frequency of the possible verb frames might supply an additional, independent bias toward which structure the parser choses to build. A verb with two equally likely frames, henceforth an equibiased verb, thus may have minimal influence on the parse that is ultimately preferred by the parser, through whatever mechanism builds and evaluates parses (discussed further below). For instance, in sentences like (2), if the bias supplied by the verb frame frequency of leaves is a strong influence on the parser and if the direction of the bias is to build a transitive frame, then the parser may prefer to build the embedded object parse even though this does not minimize the distance between the pronoun and the first possible candidate antecedent. On the other hand, if the bias supplied by the embedded verb does not have a direction (transitive or intransitive) because the verb is equibiased, then independent of whether the verb frame frequency supplies a strong bias or not, the preferred structure will not be dictated by the verb frame. This leaves open the possibility that the dependency search could supply a bias to the parser, and the parser will construct the matrix subject parse in (2).

The experiments described in this chapter use the (minimal) influence of equibiased verbs to ask how the parser behaves in a structurally ambiguous region while it is also engaged in an active search to form a dependency between a cataphoric pronoun and a (potential) coreference antecedent. Does the parser prioritize building structures that would allow for the early formation of the coreference dependency, thus choosing an intransitive verb frame? Or does it prefer to construct a locally attached structure, using a transitive verb frame? Or, maybe, does it struggle to balance the two biases?

In this chapter, I will address some of the biases present during incremental structure building, such as local attachment and the parser's eager search for a coreference dependency between a cataphoric pronoun and its candidate antecedent, i.e., Active Search (e.g. Kazanina et al., 2007; Stowe, 1986; van Gompel and Liversedge, 2003). In Section 3.2, I will review the literature describing these biases and the parser's ability to cope with them in real time. In Section 3.3, I discuss possible outcomes of structural representations based on the biases toward forming a local attachment and completing a dependency, and specifically when these biases are in conflict with one another. In Sections 3.4, 3.5, and 3.6, I discuss pilot experiments that lay the groundwork for Section 3.7: an online experiment that tracks subjects' eye-movements while reading in order to determine the locations and degree of processing slowdown associated with structural revision. In Section 3.8, I analyze the results of the online experiment and in Section 3.9, I discuss the wider implications of the results of the pilot and main experiments. Finally, I conclude with Section 3.10, which suggests that the two biases examined in this study may be simultaneously considered by the parser, thus a parallel processor may best account for our observations.

3.2. Background

One of the prime environments for observing the parser's behavior is in sentences with a local ambiguity that is later resolved, i.e., garden path sentences (Bever, 1970; Crain and Steedman, 1985; Frazier and Fodor, 1978; Frazier and Rayner, 1982, etc.). Garden path sentences are sentences in which an ambiguous region is potentially easily misanalyzed, thus leading the parser "down a garden path", after which it must[1] select or build a viable alternative analysis in order to construct a globally grammatical parse of the sentence (Bever, 1970; Ferreira and Henderson, 1991; Frazier and Fodor, 1978;

[1] In canonical garden path cases, the parser is unable to find or construct the alternative analysis that is ultimately the correct analysis, in which case the derivation crashes and the sentence is rendered uninterpretable (Bever, 1970; Frazier and Rayner, 1987).

Cf. Altmann and Steedman, 1988). This misanalysis occurs when a string internal to the sentence underlined in (3) forms a coherent and apparently viable parse, which could be continued as it is in parentheses in (3-a), but cannot ultimately fit into the global structure of the sentence[2], shown in (3-b).

(3) a. <u>[Whenever John leaves the house]</u>... [he forgets to lock the door].
    b. <u>[Whenever John leaves] [the house</u> is always left unlocked].

The parser's limitations and biases, such as the available processing capacity (Gibson, 1991; Kimball, 1973; Miller and Chomsky, 1963) and the mechanism for structure building (Crain and Steedman, 1985; Frazier and Fodor, 1978; Gibson and Pearlmutter, 2000; Gibson et al., 1996; Lewis, 2000), may be the source of the garden path phenomenon. For instance, as the parser encounters new material, it may attempt to incorporate the material by locally attaching the new material onto the current structure, when in fact the grammatical analysis will ultimately require a more distant attachment (Phillips and Gibson, 1997). This mechanism has been discussed throughout the literature under a number of different schemas, all of which share a basic property of locality in some form, and will be elaborated upon in the following sections.

3.2.1. Garden paths

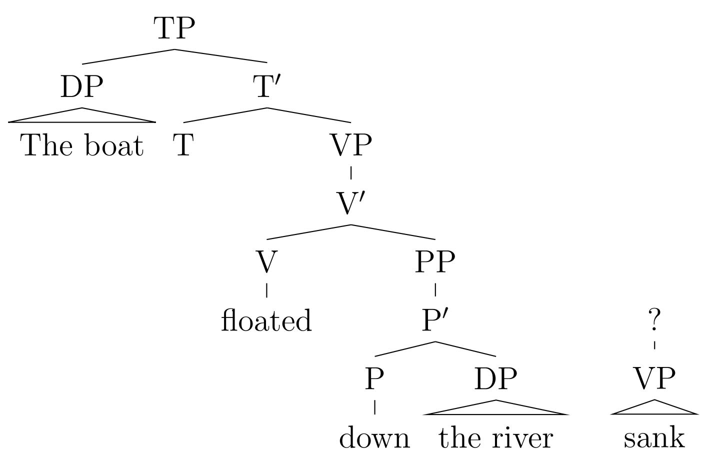
In garden path sentences, the parser is likely to initially (begin to) build a parse that is globally ungrammatical. Typically, this ungrammaticality is due to an analysis of a

[2] It might be relevant to note that when this string occurs as an initial substring of a sentence, it is considered a garden path, but when it occurs non-initially, it is considered a local coherence (Christianson et al., 2001; Konieczny, 2005).

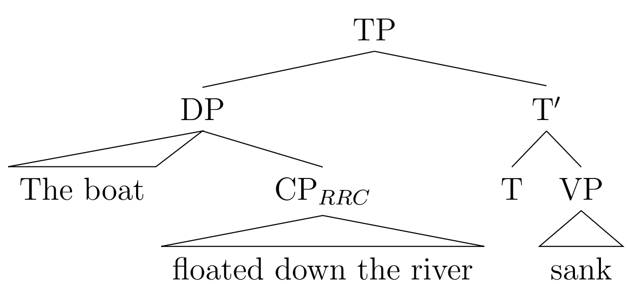
locally ambiguous substring that is potentially grammatical, but turns out to be globally ungrammatical when the parser encounters disambiguating information after the local ambiguity. This can especially happen when the parser encounters a lexical item that biases the parser to build a high frequency structure that is ultimately ungrammatical, rather than a lower frequency structure that would ultimately allow for a globally grammatical parse. A classic example of such a garden path sentence that makes use of a reduced relative clause (RRC) is illustrated (4), originally from Bever (1970):

(4) The boat floated down the river sank.

a.



b.



In (4-a), a typical analysis of *floated* is as a tensed verb and *the boat* makes a good subject of *floated* since boats are things that can float down rivers, thus floated is likely to be the main verb. So the parser might build a structure like (4-a) that treats the string *the boat* *floated down the river* as a complete TP. The parser would then not find a position that the next word, *sank*, could occupy. This may presumably indicate to the parser that the sentence is ill formed or that the parser had misanalyzed the input string, because there may be an expectation for input to be analyzed as grammatical. In fact, *floated* is the beginning of a RRC in this example, and is not the main verb but rather a past participle that has the same surface form as a main verb, as illustrated in (4-b). An unreduced relative clause with a similar interpretation as (4) is (5)[3]:

(5) The boat which was floated down the river sank.

or: The boat ridden down the river sank.

In sentences like (4), which are notoriously difficult to interpret even when the grammatical parse is known, there is no overt indication of a subordinate clause. A sentence that has some indication of the presence of a subordinate clause might make a garden path easier to recover from, or might make it easier for the parser to choose the correct structure initially (Ferreira and Henderson, 1990, 9). Other sources of information or biases of the parser, such as the frequency of a verb frame or the process of dependency formation, may

[3]Another way to illustrate the difficulty of reduced relative clauses is to choose a past participle that is not ambiguous with a main (tensed) verb, as in (i-a) and (i-b), respectively (adapted from Tabor et al. (2004)):

(i) a. The athlete tossed the ball smiled.
b. The athlete thrown the ball smiled.

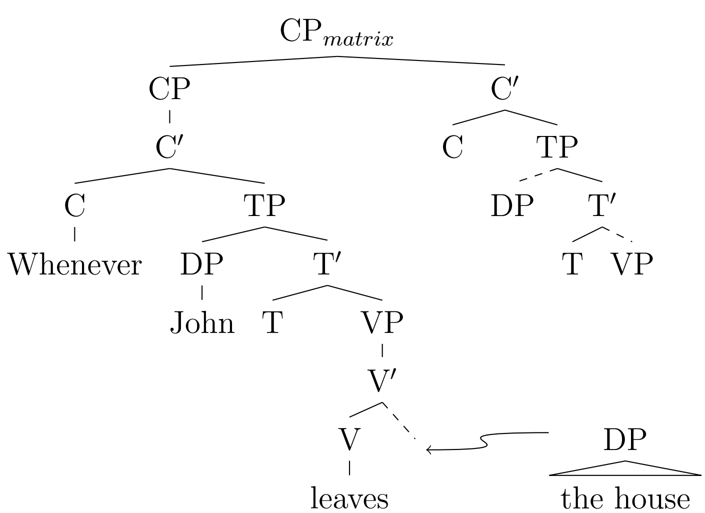
The two sentences in (i) have functionally identical structures, but (i-a) is more likely to trigger a garden path effect than (i-b), due to the ambiguity and relatively probability of tossed being analyzed as a main verb (Tabor et al., 2004).

also influence the parser's behavior to make a garden path more or less likely to occur, and will be discussed further below.

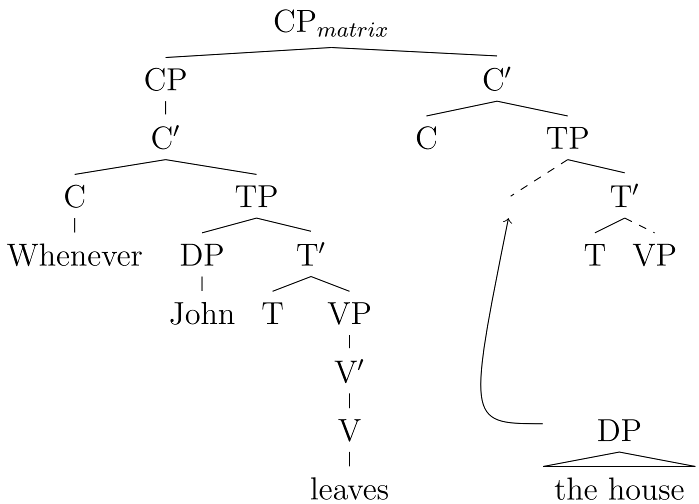
3.2.2. Argument structure ambiguity

Verbal argument structure is a source of ambiguity because a single form of a verb may be associated with multiple distinct argument structures. Each distinct argument structure requires a distinct structural analysis, which leads the parser in different directions. This is the case in sentences with sentence-initial adjunct clauses such as (1), restated below as (6), in which the verb in the embedded clause can have multiple distinct argument structures, leading to distinct possible continuations:

(6) a. Whenever John leaves the house. . . it's dark.



b. Whenever John leaves the house. . . is dark.



In (6), the verb *leaves* could be transitive (6-a) or intransitive (6-b). In both cases, *leaves* is followed by a noun phrase (*the house*), which does not yet disambiguate between the two possible structures. Thus, given that the parser does not wait to build a structure until after the disambiguating information is reached (Stowe, 1986; Tabor and Hutchins, 2004; Traxler and Pickering, 1996), it must either choose between the two options or construct the two representations in parallel. In either case, the parser may have a preference for one of the two possibilities, based on the frequency or likelihood of each argument structure, or other information outside of broad syntactic categorization.

The lexical frame of a verb is the configuration of a verb's arguments, primarily referring to the order of the complements a verb may take (Connine et al., 1984; Lapata et al., 2001). The frequency with which a particular verb frame is used affects the parser's preference for building that verb frame, with high-frequency verb frames more preferred (or assigned greater probability) than low-frequency verb frames. This frequency is not necessarily consistent across the population of speakers, but can be calculated with fairly high accuracy from large-scale corpus studies (Arppe and Järvikivi, 2007; Gahl et al., 2004). Thus, it is possible to estimate what verbs have multiple possible verb frames with frequencies that are approximately equal, i.e., equibiased verbs (MacDonald et al., 1994; Staub, 2007; Tanenhaus et al., 1989; Trueswell and Kim, 1998).

Such a measurement is important to calculate because there seems to be a bias toward transitive argument structure formation, which is part of the general explanation for garden path phenomena. But even when a verb is biased toward an intransitive analysis based on corpus frequency, the parser may still prioritize the less frequent (but locally grammatical) transitive analysis (Christianson et al., 2001; Frazier and Rayner, 1982). Yet, the parser does respect at least some of the constraints on verb frame content. Staub (2007) demonstrated that the parser does not posit a gap after obligatorily intransitive verbs, but does seem to posit such a gap after preferentially intransitive but potentially transitive verbs. Furthermore, Blodgett (2004) observed that intransitive verbs were significantly less likely than equibiased verbs to elicit a garden path effect. This study used auditory stimuli, so it is not necessarily comparable due to the presence of other information in the speech stream, such as prosodic cues. The relevant observation from Staub (2007) and Blodgett (2004) is that the argument structure (and possibly frequency of that structure) does influence the parser's behavior in otherwise structurally ambiguous strings.

3.2.3. Models of the Parser

The literature proposes that models of the parser could be serial (only constructing a single possible representation at a time) or parallel (having the ability to build and maintain multiple representations simultaneously). In the reviews by Lewis (2000) and Gibson and Pearlmutter (2000), they compare the broad categories of these models. The general conclusions they agree on are that strictly autonomous serial and parallel parsers are unviable models of the human sentence processor because there is too much evidence that the parser is influenced by factors besides syntactic category. The early serial models of the parser were strongly syntax-first and strictly serial (Frazier and Fodor, 1978, a.o.), and did not allow for the parser to take into account information beyond broad syntactic categories in initial structure building. This type of model is inconsistent with observations that the parser can be biased by contextual or lexical information toward a grammatical but less likely parse. That is, the difficulty associated with reading a sentence containing an ambiguous or improbable structure is mediated by the frequency with which the specific lexical items in that sentence occur in such structures (Trueswell, 1996; Trueswell and Kim, 1998; Cf. Gibson, 2006). In comparison, strictly autonomous, unconstrained parallel parsers have no available mechanism for limiting the number of parses constructed simultaneously, since only broad syntactic information is available to them initially, and no probabilistic frequency information can be taken into account until later stages. This unbounded use of resources does not seem to reflect the parser's behavior, even in the hypothetical case where resources available for building parallel structures are unlimited (Crain and Steedman, 1985). Thus, the primary debate between parsing models can be simplified to interactive models of probabilistic serial parsers and interactive models of less-strict parallel parsers, in which interactivity allows probabilistic characteristics of the input to influence the parser. Lewis (2000) and Gibson and Pearlmutter (2000) disagree on which of these two classes best accounts for the data available in the literature, so the following subsections lay out the arguments for and against both possibilities.

3.2.3.1. Serial Models. In interactive serial models, the parser is able to take into account contextual and lexical information during its initial structure building process. This means that the parser can make use of verb-frame frequency information to choose the most likely parse, thus it will preclude analyses that could likely require (costly) revisions later on. One way to model the behavior of the parser within the umbrella of serial parsers is to adopt a probabilistic decision mechanism for determining what structure the parser will build when it encounters an ambiguous string. In this probabilistic serial parsing model, the parser will pursue structural options with probabilities proportional to their likelihoods. This mechanism allows for the parser to build both the preferred sentential complement parse and the dispreferred object relative clause parse in sentences like (13), across trials. According to Lewis, this type of parser is difficult to falsify, and the attempts by Pearlmutter and Mendelsohn (1998, 9) to do so are based on a null finding, thus is not strong evidence against the probabilistic serial model. I will leave the question open as to whether the probabilistic serial parsing model accurately accounts for the observed behavior of the parser, and return to it in Chapter 5.

3.2.3.2. Parallel Models. Another class of models allows the parser to build multiple parses simultaneously. This class includes models that allow for temporary representation of multiple parses (until disambiguation resolves the structure) and models that can simultaneously incrementally build parse trees throughout the process. The literature in favor of a parallel model of the parser has converged on a subset of characteristics that such a model must have. In order for the parser to be parallel and feasible given cognitive resources, it must be limited in what representations it can maintain at a given time. Two of the typical ways to model such a limitation are by ranking potential parses such that the highest ranked parse is ultimately adopted, and through a "race" in which parses "compete" to reach some threshold of activation first, in order to be adopted.

One criticism of parallel parsing models is that an increased level of processing difficulty should be observed at the point in which the greatest number of competing parses are being maintained. This processing difficulty would thus occur during an ambiguous region in which multiple ambiguities are present. But this is not necessarily the case, as discussed by Gibson and Pearlmutter (2000). Ambiguous regions — even highly ambiguous ones — are shown to be read quickly, whereas disambiguating regions display the slowdown associated with processing difficulty. One way parallel models of the parser can account for this fact is through competition. Even if there are multiple parses being represented at the point of ambiguity, there is not necessarily much competition between them. If, for instance, the syntactic representations can share storage space where they overlap in structure, they may not tax the parser during the ambiguous region because they do not take up much extra "space" in working memory. Another possibility is that the competition is mainly between interpretations, i.e., within the semantic comprehension mechanism. In this case, the difficulty would not be expected within the ambiguous region, without making further assumptions about how the input is processed and how the representations are stored (Gibson and Pearlmutter, 2000). An alternative interpretation of a parallel parsing mechanism, which is consistent with the reading times in ambiguous regions and corresponding disambiguations, is that the point at which the parser must discard or re-rank parses that are not consistent with the current input would show an increase in processing difficulty. This, too, requires further assumptions about how the input is processed and stored. That a superficial reconfiguration of the parallel parsing model can offer a solution to this criticism suggests that the criticism is not broadly applicable. In any case, there is no strong evidence to distinguish between (limited) parallel and (probabilistic) serial models as of yet, so both must be considered when interpreting studies of the parser's online behavior.

3.2.4. Conflicting Evidence for the Influence of Dependency Formation

While it is clear that lexical properties can affect the parser's online structure-building decisions (MacDonald et al., 1994; Trueswell, 1996), it is possible that the parser also takes into account what types of elements must be encountered to satisfy various grammatical constraints or principles. For instance, if the parser encounters a word that opens a long-distance dependency, such as a wh-word or anaphoric element, the parser may prioritize structures that lead toward closing that open dependency. By closing an open dependency, the parser no longer needs to maintain that operation in working memory, so by closing the dependency as soon as possible, the parser ensures that the processing cost is lower. This is an intriguing possibility because the information which influences the parser during structure building is uncertain and it supplies a clearly defined ground for examining the parser's behavior. As discussed in the previous section, the parser is able to access information about lexical frequency that can preclude it from building structures that could cause costly reanalysis processes. Can the search for a long-distance dependency also aid the parser in ambiguity resolution?

As explicated in Section 2.5.2.4, Kwon and Sturt (2014) test the influence of dependency formation between the anaphor PRO and an antecedent. In their study, they use the ambiguity provided by a control noun, which allows the ambiguous region to extend past the location of PRO without providing the parser with morphological information that would aid in an early reanalysis. From their observations, they conclude that the noun phrase in the ambiguous region was not initially analyzed as a matrix subject because of the processing difficulty experienced at the disambiguating verb. This would mean that the parser was only using the control information after it initially chooses how to build the structure of the ambiguous region.

Their conclusion is of particular relevance to this dissertation because it suggests that the coreference dependency relation may not be "strong" enough to supersede the parser's bias toward locally attaching incoming input. On the other hand, there is evidence that the parser uses dependency information online, even if not during initial structure building (Stowe, 1986; Traxler and Pickering, 1996). Furthermore, since PRO is phonologically and orthographically null, it is not possible to see online effects at the point where PRO may be in the sentence using an eye-tracking while reading paradigm. This point will be discussed further in Section 3.9.

However, Frazier et al. (2015) draw a conclusion that conflicts with Kwon and Sturt (2014), in their study on wh-dependency formation described in detail in Section 2.5.1.3. They examine the influence of dependency formation on structure building using wh-filler-gap dependencies and potential local coherence constructions. The dependency formed between a wh-word and its gap is obligatory, unlike dependencies triggered by pronouns. This makes the wh-dependency a less variable relation, which the parser may be able take advantage of by making predictions based on the presence of the dependency (Altmann and Kamide, 1999; Kwon and Sturt, 2014). If the presence of a wh-dependency always indicates that a gap will occur later in the parser (in a grammatically licit location), then the parser may be able to anticipate this relation.

If the active search for the wh-dependency does not interfere with the parser's local behavior, Frazier and colleagues expected that the conditions in which the linearly closest (but globally ungrammatical) antecedent mismatches in gender with the reflexive anaphor, as in (42-b), should show a reading time slowdown associated with the processing difficulty of gender incongruity (the Gender Mismatch Effect, Kazanina et al., 2007; van Gompel and Liversedge, 2003). If the formation of the wh-dependency interferes with the locally coherent analysis and prevents it from forming, they should see no such slowdown. In fact, they find no influence of a local gender mismatch in the conditions in where the wh-phrase and reflexive antecedent match in gender. This indicates that, indeed, the search triggered by the wh-phrase for the tail of its dependency prevents the parser from adopting the locally coherent analysis (43-b). Thus, this study provides evidence to suggest that an active dependency search mechanism can prevent a parser from making mistakes during (local) ambiguity resolution. In other words, the presence of an active dependency search may limit what representations the parser is willing to build, if those representations would interfere with the process of dependency resolution.

3.2.5. Cataphoric coreference dependencies

The type of long-distance dependency which will be tested in this dissertation is the cataphoric dependency. In a cataphoric dependency, an anaphor such as a pronoun triggers an antecedent search rightwards through the (incoming) string[4] and this search is crucially constrained by the grammatical constraints on coreference and binding (Chomsky, 1981). This makes the formation of cataphoric dependencies an excellent probe for structure building. The distinction between anaphoric and cataphoric dependencies is illustrated below in (7-a), in which the pronoun *he* triggers an anaphoric antecedent search leftwards in (7-a) and a cataphoric antecedent search rightwards in (7-b). In both (7-a) and (7-b), a candidate antecedent for *he* is *John*, which occurs in (7-a) to the left of the anaphor *he*, and in (7-b) to the right of the cataphor *he*.

(7) a. Anaphor: Whenever John$_i$ left the house, $\underline{\text{he}_i}$ remembered to lock the door.
b. Cataphor: Whenever $\underline{\text{he}_i}$ left the house, John$_i$ remembered to lock the door.

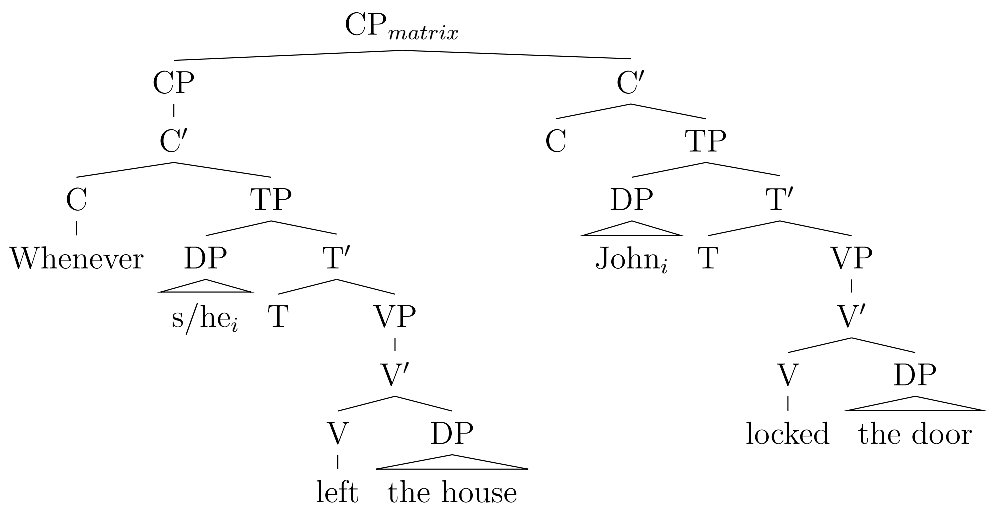
What makes the cataphoric dependency of particular interest to this study is not only that the candidate antecedents for a cataphor are grammatically constrained by the Binding Conditions, but that pronouns in English have definitional gender, which must be checked and matched before a dependency between a pronoun and its antecedent is successfully established (van Gompel and Liversedge, 2003). This is a useful property because any attempts at dependency formation between a pronoun and name (or other R-expression) that do not match in gender can result in a processing slowdown, known as a Gender Mismatch Effect (GMME) (Kazanina et al., 2007; van Gompel and Liversedge, 2003).

[4]This is "backwards" from the typical description of pronominal antecedents, which search for an antecedent leftwards through the tree, hence the alternative name "backwards anaphora".

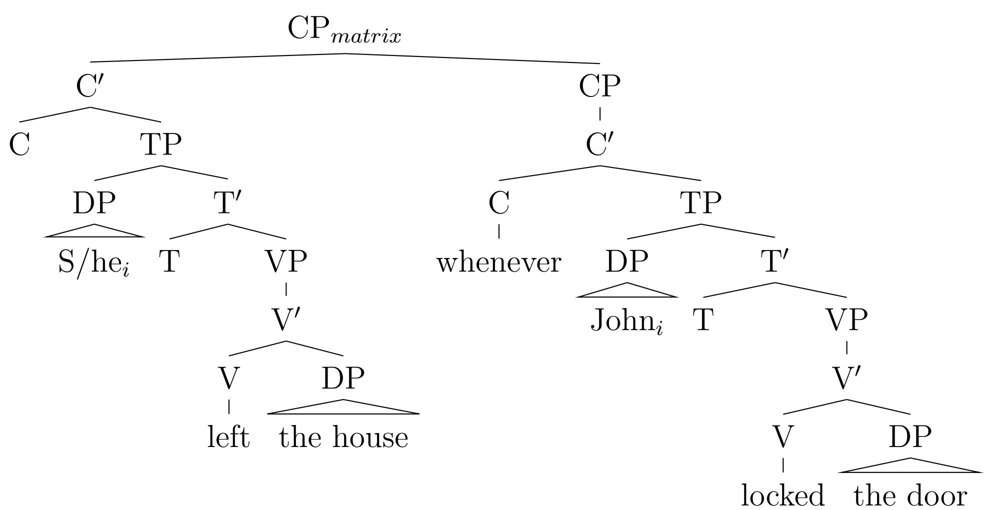
This effect occurs when the parser attempts to form a dependency between a cataphor and a candidate antecedent, but does so before the gender features are confirmed to match. The slowdown presumably occurs as the parser revises or deletes the dependency, which is grammatically licit, but is otherwise prohibited because of the gender features do not match between the cataphor and the candidate antecedent. The process undergone to undo the dependency formation is sufficiently costly for the processing mechanism to manifest as a reading time slowdown.

As discussed in Section 2.5.2.3, Kazanina et al. (2007) concludes that the parser represents the position of the incoming material before gender information (specifically) is integrated or interpreted. After the parser has created a structure for an incoming R-expression, it can determine whether or not the R-expression can licitly corefer with a preceding cataphor. If the cataphor does not c-command the R-expression, then dependency formation is attempted, as shown in (8)-b. On the other hand, if the cataphor does c-command the R-expression, then dependency formation is not attempted, (8)-d.

(8) a/b. Whenever [he$_i$ / she$_i$] left the house, John$_i$ locked the door...



c/d. [He$_i$ / She$_i$] left the house whenever John$_i$ locked the door...



This is evidenced by a GMME at the region analogous to *John* in sentences like (8)-b, where *John* is not c-commanded by *she*, but in contrast, there is no GMME at *John* in sentences like (8)-d where *John* is c-commanded by *she*. Moreover, this type of effect was not apparently due to the introduction of a new discourse referent, and it persisted when potential antecedent positions were controlled for predictability. All together, this suggests that the cataphoric dependency search is a useful probe into the parser's structure building behavior because it is a consistent and early effect of the structural representation of the sentence.

3.2.6. Interim summary

Given this background, this dissertation asks whether another long-distance dependency, namely cataphoric dependency, can influence structure building in the case of local attachment ambiguity by giving the parser an additional bias that overrides or supersedes its other structure-building biases. There is evidence that an active dependency search can interrupt a locally grammatical (but globally ungrammatical) parse, as well as evidence that lexical properties of the verb can strengthen the pressures of locally coherent parses. However, there is also evidence that certain types of information, like verb frame frequency, do not prevent the parser from being led down a garden path. So which biases influence the parser? The parser is influenced by multiple competing pressures, but ultimately it can only test a limited number of parses online before the processing cost becomes too high. By pitting the pressures exerted by a search for a dependency completion and the preference for locally attached parses, I aim to explore the parser's prioritization method and behavioral choices surrounding structurally complex ambiguities.

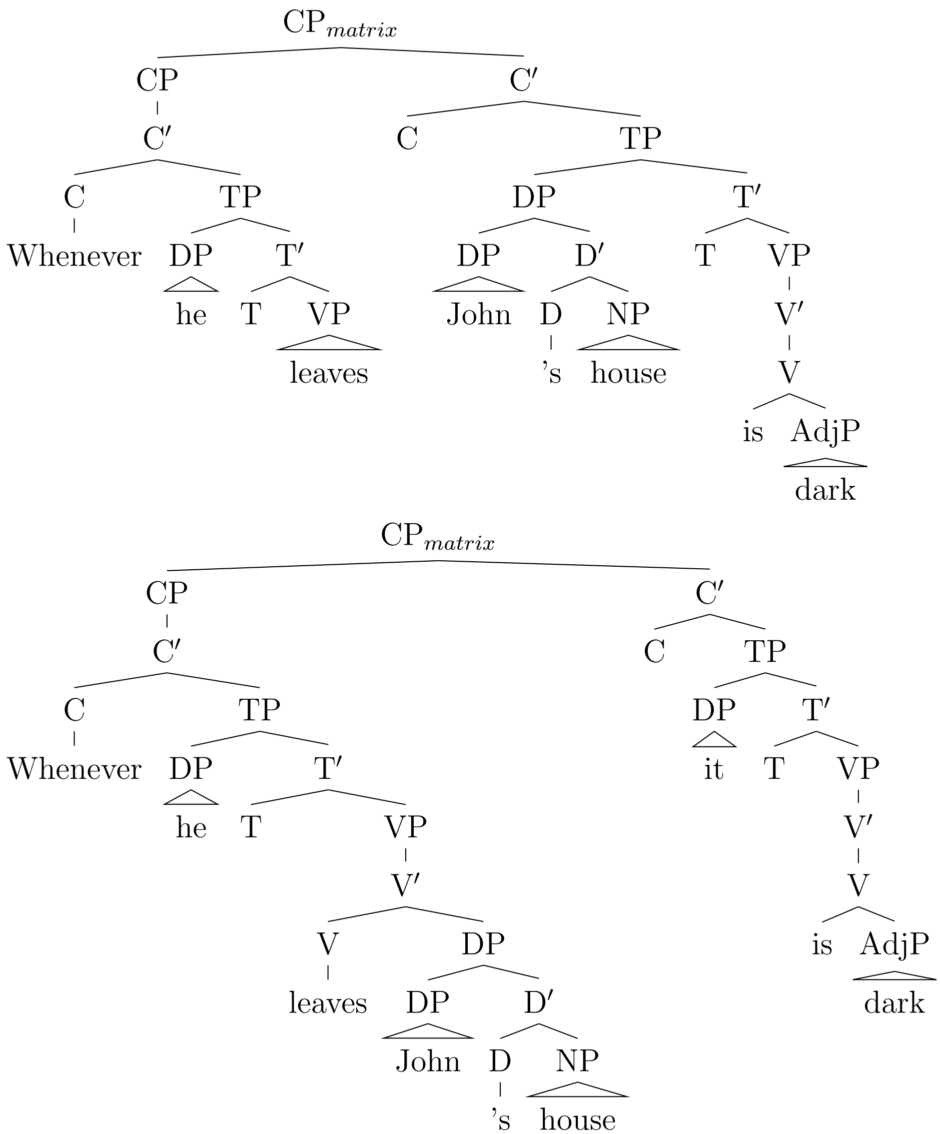
3.3. Local Attachment and Cataphoric Dependency Formation: Predictions

The experiments discussed in this chapter focus on the parser's structure-building biases: a bias for local attachment and the potentially conflicting goal of eagerly forming a cataphoric dependency. As I have reviewed in the previous section, the parser has a strong bias toward local attachment during online structure building, which presumably takes the form of an embedded object analysis in an embedded-object/matrix-subject ambiguous string, as illustrated in (9), adapted from (2).

This bias can still be superseded by the parser's desire to form dependencies quickly based on what it can predict about the upcoming content of the sentence — at least for wh-dependency formation, as Frazier and Clifton (1989) and Frazier et al. (2015) show. Furthermore, Kazanina et al. (2007) and van Gompel and Liversedge (2003) demonstrate that cataphoric dependency formation is similarly eager to wh-dependency formation, and Kazanina demonstrates that the parser respects grammatical constraints in a comparable way. Assuming, then, that cataphoric dependency formation triggers similar parsing processes to wh-dependency formation, what might be expected from the parser's behavior in situation where cataphoric dependency formation could interfere with the bias for local attachment?

An environment that highlights how these biases compete is in sentences such as (9), where a dependency search for a cataphoric antecedent is triggered by an early pronoun, and subsequently encounters a verb that is locally ambiguous between a transitive and intransitive verb frame. This will be the frame within which this dissertation examines competition between biases on the parser.

(9) Whenever he leaves John's house ... [is / it is] dark.

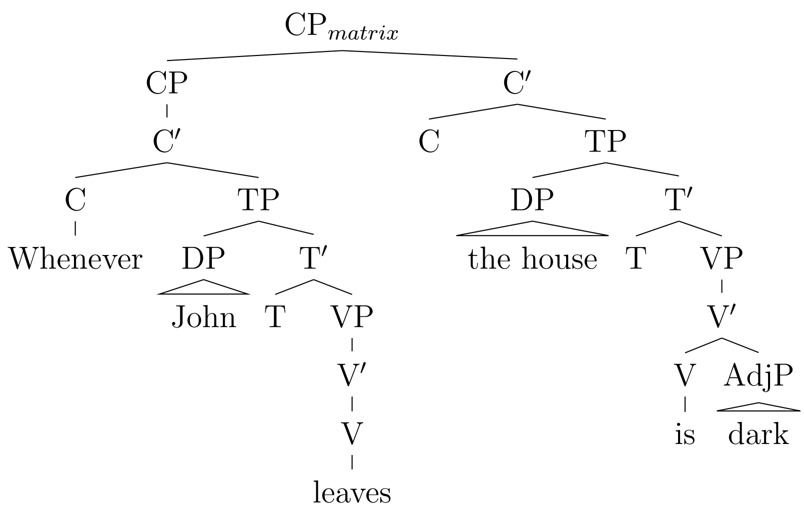


In such circumstances, the open cataphoric dependency could be completed more quickly if the intransitive verb frame is selected, since this allows coreference between the cataphoric pronoun, which triggered the active antecedent search, and *John*. Otherwise, if the transitive verb frame is selected, the cataphoric pronoun c-commands *John* and thus cannot corefer with *John* (Chomsky, 1981, BCC). These hypotheses are based on the distinct parses that could be built, but there is also a possibility that a multiple-output parser would display a hybrid set of effects whereby the parser reacts to both the matrix subject and the embedded object parses independently and simultaneously. The parser's prioritization of local attachment and early dependency formation will be tested in the experiments presented in the remainder of this chapter. In the following subsections, I will describe different scenarios and step through what my hypotheses predict for each. Note also that I will start off discussing how the parser could construct a parse based on the probabilistic or interactive serial models of the parser. The initial structure building behavior of an interactive or probabilistic serial parser is compatible with ways in which a parallel or multiple-output parser would build one of its representations. Therefore, I will focus my discussion on interactive and probabilistic serial parsers until Section 3.3.3, in which I integrate the processes of building individual parses into a discussion of how a parallel parser might behave.

3.3.1. Embedded Object/Matrix Subject Ambiguity with No Cataphor

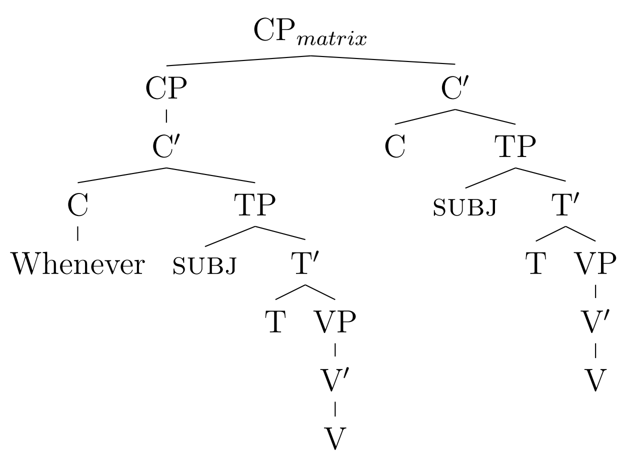
In the first scenario I shall step through, the sentence has a local embedded-object/matrix-subject ambiguity, and contains no cataphoric pronoun. This allows a baseline to be established against which the focus of this study can be compared. Given a sentence such as (10), what steps could the parser take to form a coherent and grammatical analysis?

(10) Whenever John leaves the house is dark.



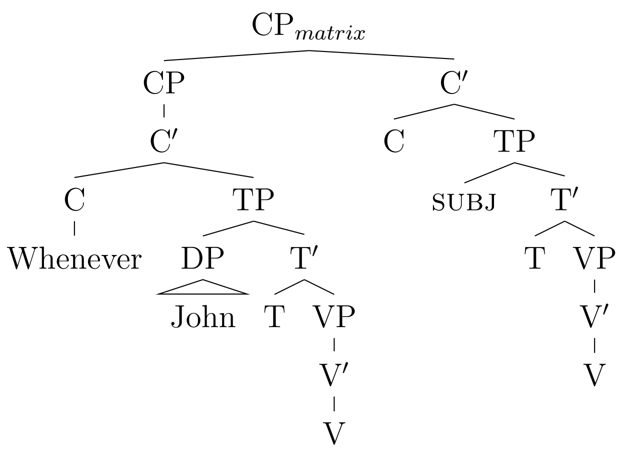
The first word, *Whenever*, may indicate to the parser that the utterance begins with an adverbial subordinate clause that will somehow later on be related to a different clause, the matrix clause. At this point, if one assumes that the parser constructs the syntactic representation as far rightward as it can, based on category-selectional restrictions alone, one may expect the parser to build the structure in (10) at the point of the first word.

(10) a. Whenever



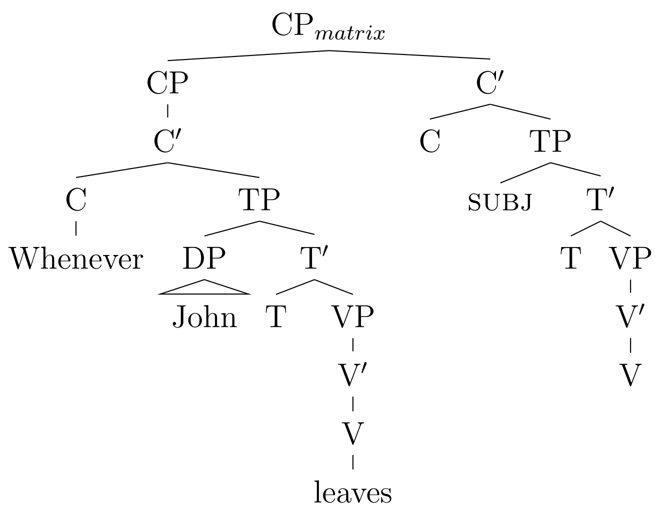
This structure consists of whatever projections are necessary to extend the terminal node of *Whenever* up to the CP of the matrix clause and rightward as far as is minimally required by English grammar. I will assume that the parser will represent CP_{matrix} at this point in the analysis because the parser should recognize that *Whenever* is the beginning of an adjunct clause embedded within the matrix CP. Based on the selectional restrictions of English, the parser must then build a TP for each clause, because TP is unambiguously the complement of C. Furthermore, the projection of each TP entails the projection of a subject position (SUBJ), by means of the Extended Projection Principle (EPP) (Chomsky, 1981). Finally, TP selects a verb phrase complement, VP, which projects down to V in both clauses. Although this type of parser is not one that is typically described in the literature, it follows from the literature on the hyper-active nature of the parser (Omaki et al., 2015; Traxler and Pickering, 1996; Wagers and Phillips, 2009) that structure is anticipated whenever possible, but that the parser does not build structure infinitely far in advance. Thus, the parser as described here, builds as much structure as it is obligated to and does not engage in building structure for which it must make assumptions about the bottom-up content until that content is identified.

(10) b. Whenever John



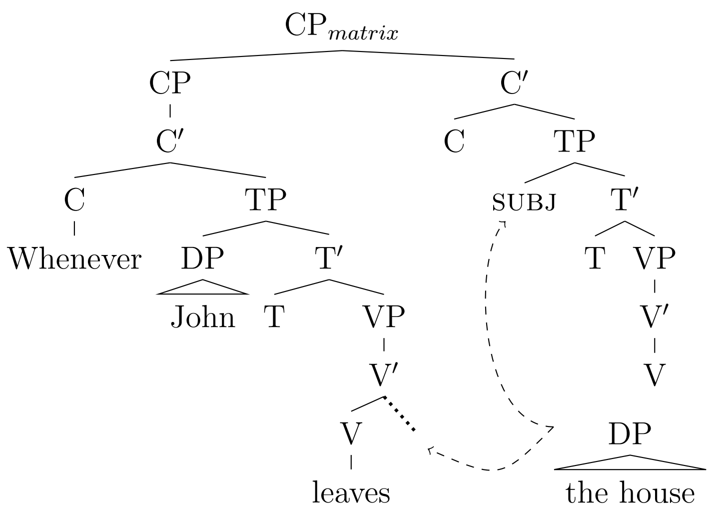
The next word, *John*, may be analyzed as the embedded subject since the next position in the tree is the embedded subject, and therefore this parse would require minimal processing cost to construct. Thus, if we assume the parser prefers to locally attach incoming structure (at least in the absence of other influences), this is both the easiest and most likely structure it will build. At this point, the parser has constructed the tree as shown in (10), which consists of nodes based on the syntactic categories that are minimally required by English, but no content beyond what can be predicted based on the unambiguous grammatical principles.

(10) c. Whenever John leaves



At the embedded verb *leaves*, the parser is now faced with a dilemma. Since *leaves* is ambiguous between a transitive and an intransitive analysis, there is no guarantee that the parser can anticipate what type of input will be encountered next. In serial models of the parser, this means the parser will have to choose between (at least two) potential verb frames as it continues forward. In parallel models of the parser, this could indicate that the parser will construct the (two) divergent parses concurrently. This only affects the predictions of theories that assume a multiple-output parser which experiences increased cost associated with competition between alternate parses when ambiguity is initially encountered (e.g., Christianson et al., 2001; Lewis, 2000; Slattery et al., 2013; Cf. van Gompel et al., 2000) and theories that require the retrieval of a lexical entry that specifies argument structure (e.g., MacDonald et al., 1994; Trueswell and Kim, 1998; van Gompel et al., 2012). In the case of the theories which predict increased cost associated with competition of representations maintained in parallel, the competing parses could increase the processing difficulty at this point, which would result in a reading time slow-down as the parser attempts to represent more than one derivation at once. In contrast, some previous work indicates that ambiguous regions like this correlate to faster reading times, potentially due to structural underspecification (Christianson et al., 2001). Although it is up for debate whether fast reading times at the ambiguous verb indicate that the parser is or is not representing multiple parses in the region of ambiguity, it does show that the ambiguity of encountering an equibiased verb does not slow down sentence processing. In the theories that assume lexical frames are stored as properties of lexical items, it is debatable whether a lexical frame is selected immediately upon recognizing the word based on what is available, or whether competing lexical representations lead to competing syntactic representations (by means of competition between the verb frames associated with the lexical items in competition). Again, this may not matter for our argument at this stage in the parse, but it is a notable feature of (10). This stage in the parse marks the beginning of an ambiguous region, so however the parser anticipates dealing with such a situation, it may become more apparent by the way in which it deals with subsequent input.

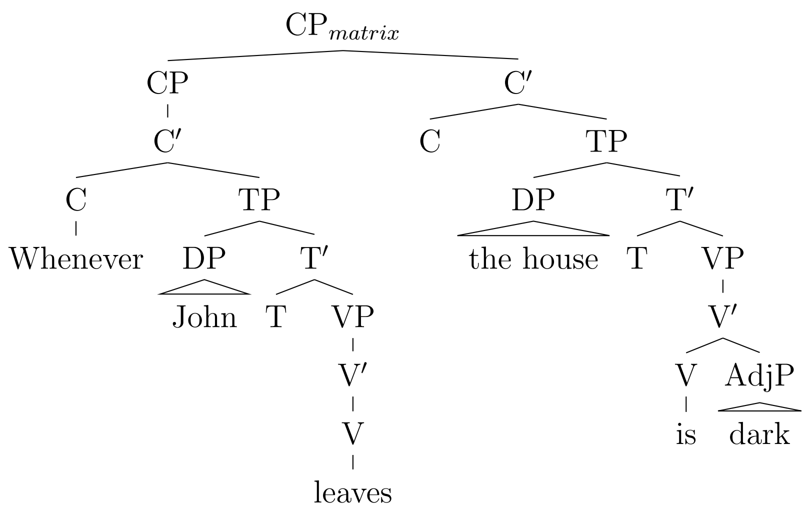
(10) d. Whenever John leaves the house



After the ambiguous verb, the parser must contend with representation of the ambiguous string. If the parser is serial in nature, then it builds a single syntactic representation. To do so, it may take into account such biases as lexical frequency (and, specific to this case, verb frame frequency) and local attachment. These biases may be additive or categorical, meaning that they may build upon one another when in the same direction, or pull the parser in different directions thereby increasing ambiguity. If these biases are in collusion toward a transitive structure, in which case the parser may easily select the transitive verb frame since this allows for local attachment (and would be consistent with a transitively biased verb). Alternatively, these biases may be in competition, in which case the parser may do one of several things. The parser may prioritize lexical frequency, in which case it would build whichever structure is more frequent. It may prioritize local attachment and select the transitive verb frame. Or it may depend on how strong the bias from the lexical frequency is; in which case, we may expect the parser to be more heavily influenced by verbs with highly likely intransitive verb frames than by verbs that are equally likely to have transitive frames as intransitive frames.

In (10), the verb frame frequency for *leaves* is of this final type because both transitive and intransitive argument structures are approximately equally likely. Thus whether or not the parser makes use of this information, it is not a decisive factor that weights the likelihood of a particular resolution over the alternate. On the other hand, the bias toward local attachment is more likely to influence the parser's behavior (Phillips and Gibson, 1997). Thus, one might expect the parser to be more likely to analyze *the house* as the embedded object of *leaves*, even though the parser is (presumably) aware that the embedded clause may end at any moment and return to the matrix clause. This assumes that the parser represents only a single derivation or that it favors one derivation sufficiently more than others at any given point to block effects from alternate parses. A parallel parser that activates competing representations gradually, similar to a Neighborhood Activation Model (Luce and Pisoni, 1998) of lexical access, may not necessarily have "chosen" between multiple possible structures at this point (MacDonald et al., 1994). The effects predicted by such a parser will be discussed below in Section 3.3.3.

(10) e. Whenever John leaves the house is dark

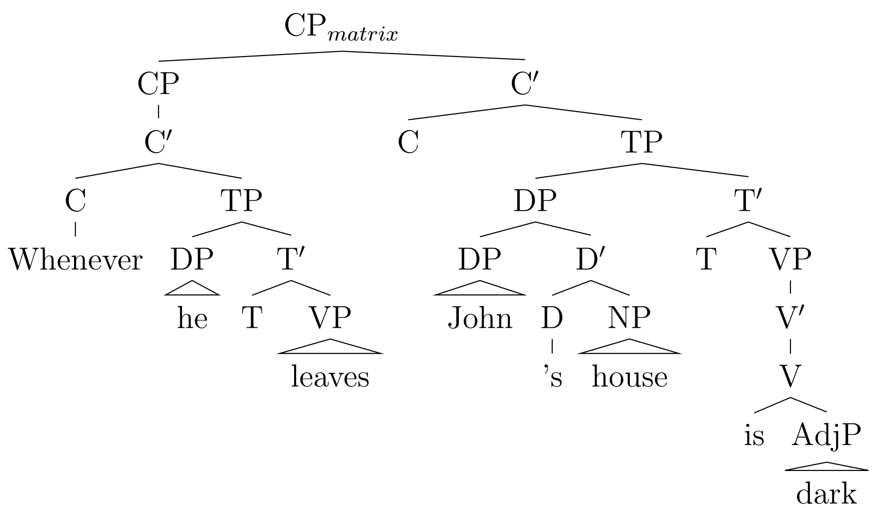


Finally, the parser encounters the disambiguating predicate *is dark*. At this point, the parser may need to undergo reanalysis if it has represented the alternative parse(s), whether by backtracking and rebuilding structure (as in some serial models) or by some re-ranking or activation process (as in some parallel models). Crucially, in the relevant cases of either serial or parallel models of the parser, there is an increase in difficulty at the point of disambiguation if the previously favored parse cannot accommodate the new input (Gibson and Pearlmutter, 2000; van Gompel et al., 2000).

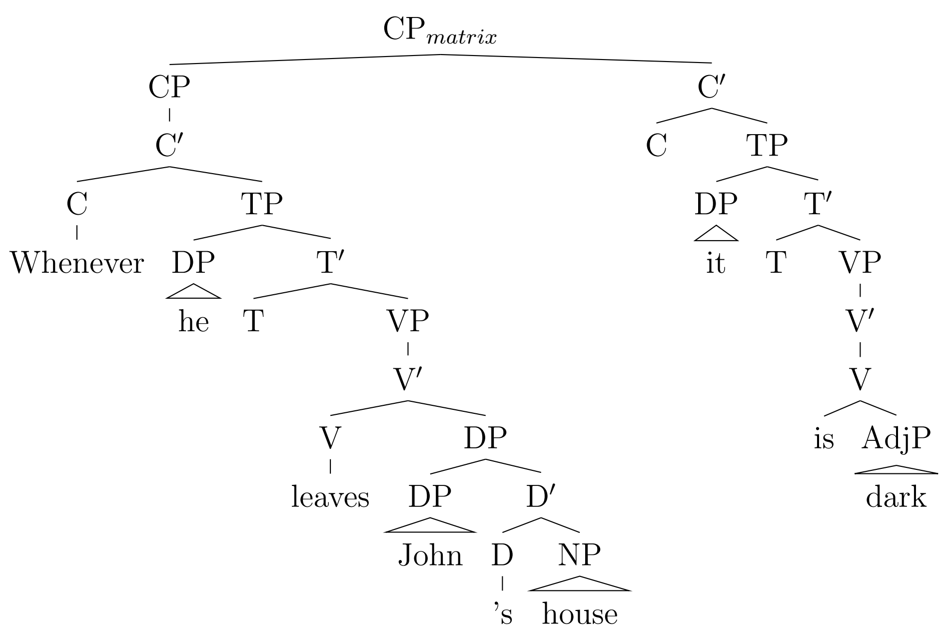
3.3.2. Embedded Object/Matrix Subject Ambiguity with Cataphor

3.3.2.1. Before reaching the ambiguous region. This section examines a sentence that follows the same syntactic structure as the examples in (10), with one critical difference. In (11) and (11)′, there is an additional and competing pressure on the parser in the form of a cataphoric dependency. These two sentences differ in the disambiguating material, with (11) depicting the matrix subject parse and (11)′ depicting the embedded object parse.

(11) Whenever he leaves John's house is dark.

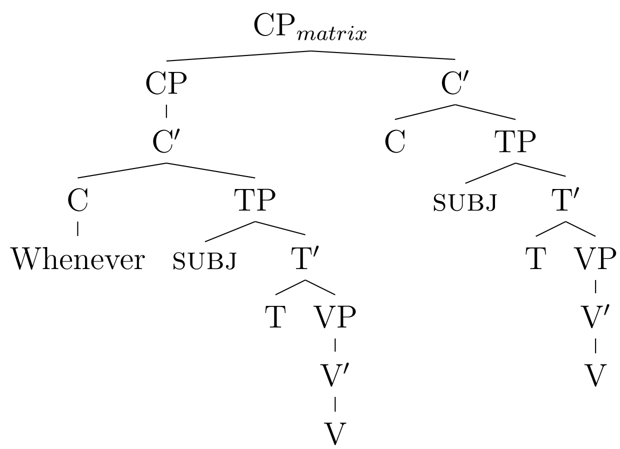


(11)′ Whenever he leaves John's house is dark.



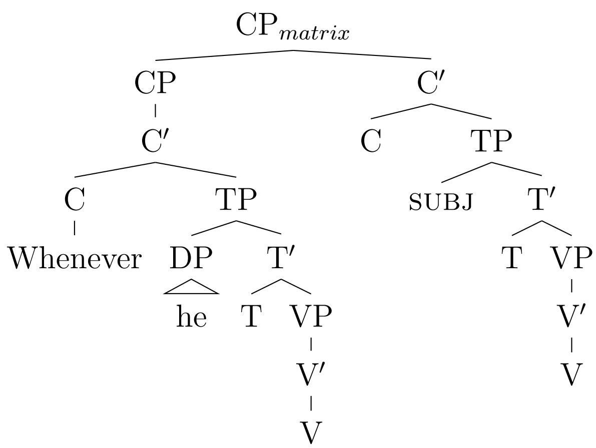
Since the experiments described in this study make use of sentences like (11), (11)′ serves to demonstrate the structural differences between the two representations, but the following discussion will only refer to the construction of (11).

(11) a. Whenever



As in Section 3.1, *Whenever* initializes the sentence, and since the content is the same as in (10), the parser has access to the same information and the same options for representation. It is at the second word that the scenario below diverges from the previous one.

(11) b. Whenever he



In (11)-b, the second word is not a proper name (i.e., *John*, in (10)), but rather a pronoun (i.e., *he*). This pronoun does not yet have an antecedent within the input thus far encountered, but it can still corefer with a (yet unknown) antecedent from lower in the parse tree. Based on Kazanina (2005) and Kazanina et al. (2007), there is strong evidence that such a cataphoric pronoun triggers a search for a candidate antecedent. At the point where (11)-b is the input, the parse tree may consist of two CPs: the embedded clause and the matrix clause (assuming that the parser can extrapolate from the CP established by *Whenever* as an embedded clause that is dominated by a matrix CP). In addition, this entails that each clause also has a TP and thus also a VP. *He* occupies the position of embedded subject, while the position of matrix subject is yet to be filled although it is predicted based on the projection of CP$_{matrix}$.

Given the available structure that the parser is able to predict, the process of active search may anticipate potential locations for the tail of the coreference dependency, shown in (11)-b′. Based on the previous discussion of the influence of Binding Condition C on active search, our understanding of the parser suggests that the parser will not attempt to build a dependency between *he* and the (potential) complement of the embedded verb[5], which is suggested by the work of Kazanina et al. (2007), among others. This is because *he* c-commands the other arguments of *leaves*, and by BCC, R-expressions in these positions cannot corefer with anything in this structural relation.

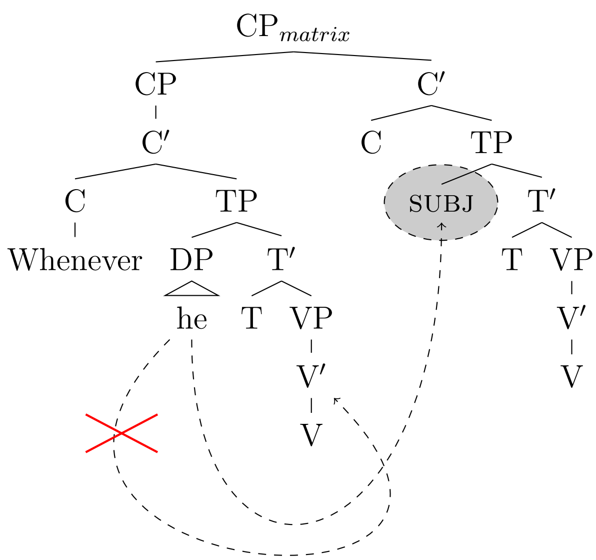
The next location that is available as the tail of the coreference dependency is the matrix subject. This position has been predicted and the parser has not encountered any other information about it that would prevent the dependency from being viable (as plausibility information might be able to do). The parser may then eagerly establish the link between *he* and the matrix subject position, if *he* triggers the parser to "look" for a non-c-commanded noun (or noun-like) antecedent.

[5]It is conceivable, however unlikely, that the parser attempts to form a dependency between *he* and the pronoun his in a genitive DP (i-a), which parallels the construction used in this study in (i-b):

(i) a. Whenever he leaves *his* house, it is dark...
    b. Whenever he leaves *John's* house, it is dark...

While he and his may ultimately corefer in (i-a), a dependency formed between them does not necessarily require the parser to have searched out that position. If the active search for a coreference antecedent skips over the embedded object, it is possible that the dependency search is put on hold during the parsing of illicit positions, but is restarted when the anaphoric item his is identified, since coreference is permitted. It is also possible that the application of Binding Condition B (roughly: a pronoun must remain free within its binding domain) is what allows such a dependency to form. If it begins with a backwards search through the parse, the antecedent search triggered by *he* would not attempt to form a dependency at *his* in (i-a). Rather, *his* would trigger a backward search for an antecedent outside of its binding domain. *He* satisfies this search, permitting the coreference dependency. Additionally, the formation of coreference dependencies is optional, which may allow the antecedent search to be unresolved over the course of (i-a), if no other candidate antecedent is found. In other words, although the differences between (i-a) and (i-b) may superficially suggest the antecedent search triggered by *he* would check within the genitive DP, there are alternative explanations for why this is unnecessary. Coupled with evidence from previous work on coreference dependencies, there is no evidence that the parser attempts to resolve coreference dependencies at positions with the potential to have a genitive DP structure. Since this is beyond the scope of this dissertation, I direct the reader to Badecker and Straub (2002) for discussion of Principle B effects in processing.

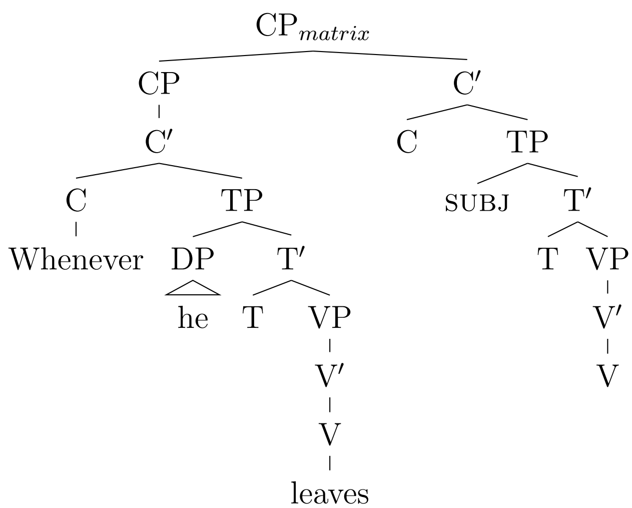
(11) b′.



At this point, there is no way for the parser to tell what the structure of the matrix subject will be or whether it will contain a simple noun phrase, or some sort of proform, etc. If the parser builds the structure that is required by grammatical constraints and selectional restrictions, as mentioned above in Section 3.3.1, but does not go further than it is obligated, then it will not guess as to the internal structure of the matrix subject. The presence of the matrix subject, in whatever form, is required due to the Extended Projection Principe (EPP, Chomsky, 1981). In this case, one might conceive of the dependency tail being located in an underspecified position, visualized by the gray oval in (11)-b′.

Here, there is no evidence that the matrix subject will be an NP dominated by a DP, for instance, if the framework labels pronouns as DPs. Furthermore, even if the subject is assumed to contain an NP dominated by a DP, the specifier position of the DP is not predictable. Thus, the parser does not know what the location of the first candidate antecedent will be and cannot make a more specific prediction than "within the matrix subject".

(11) c. Whenever he leaves

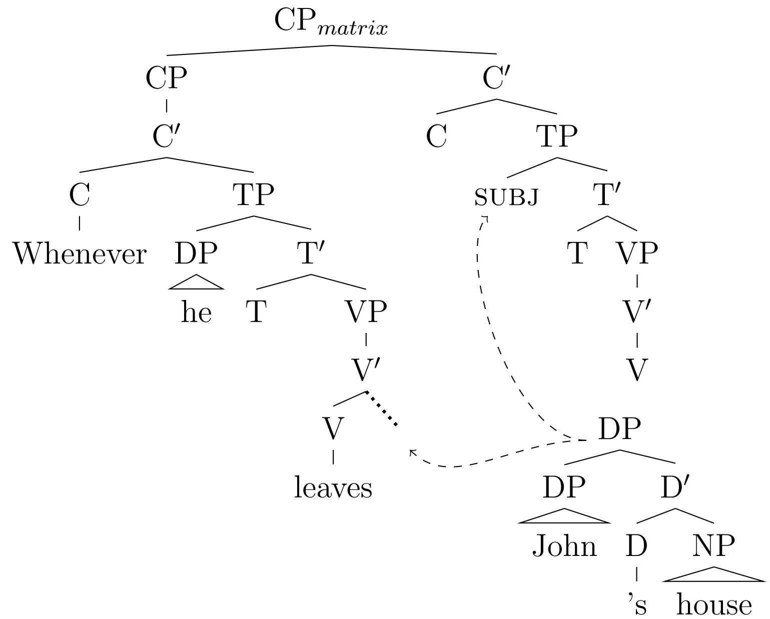


When the parser encounters the verb, it may try to resolve the ambiguity based on frequency information. However, *leave* may be equally likely to have a transitive frame as an intransitive frame, as mentioned before, providing the parser with no additional help regarding disambiguation. It may be of interest to note here that a comma placed after this ambiguous verb could disambiguate it as having an intransitive frame, but in contrast, the lack of comma is not necessarily an indicator of a transitive frame (Baldwin and Coady, 1978; Niikuni and Muramoto, 2014). The ambiguity established by these two equally likely verb frames in the absence of a comma or other disambiguating information creates the particular type of local ambiguity with which this study is concerned. That is, it is unknown at this point whether a DP following *leaves* would form an embedded object for the transitive verb frame or the matrix subject following an intransitive verb frame,[6] as demonstrated in (11)-d.

As a brief aside, I will assume that when parser encounters *John's*, it has not built the internal structure of the matrix subject yet. If it has built anything, it will have only built what projects from the head D of the dominating DP, as mentioned above. Then, when *John's* is encountered, it is recognized to have the genitive *'s* marker, which is inserted as the head D and *John* is inserted in the specifier position of the DP. At this point, the complement of D is still unknown and unpredictable (especially since a phrase like "John's was dark" is grammatical and lacks that overt NP content). Thus, the parser does not construct the internal structure of the complement NP until *house* is identified. This order of operations will serve an important purpose below, in Section 3.3.2.2.

[6] Although the matrix TP is predicted, there is substantial evidence that the parser prefers to build local attachments anyway. These observations have contributed to the hypotheses termed Late Closure (Frazier and Rayner, 1987), Recency (Gibson et al., 1996), and Right Association (Phillips, 1995; Phillips and Gibson, 1997).

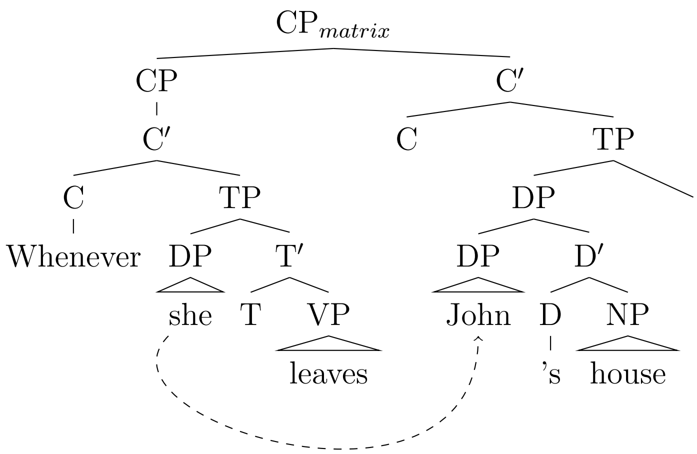
(11) d. Whenever he leaves John's house



As before, the ambiguity is not guaranteed to be correctly resolved at the point where the DP is encountered. Even though the parser may be anticipating a dependency linking *he* and the matrix subject, it does not yet have confirmation whether *John's house* is located in the matrix subject or not. This presents the parser with an immediate problem. *John's house* is a candidate antecedent for *he* (i.e., it is available to evaluate for coreference in terms of gender, number, and other lexical and pragmatic properties) if and only if this new input is in the matrix subject position. Crucially, it is not candidate antecedent if it is located in the embedded object position. Since the search for the antecedent of cataphoric pronouns has been shown to be an eager, active process, the parser may not want to delay dependency formation until after the ambiguity is resolved. If it does not delay dependency formation, then it is faced with a choice. The parser can prioritize the representation in which *John's house* is attached low (as the embedded object), which blocks *John* from being a candidate antecedent and delays evaluating the feasibility of the dependency between *he* and whatever is in the matrix subject. Alternatively, it can favor the parse where *John's house* is attached high (as the matrix subject), which allows the active search to evaluate the dependency immediately but could then also result in a high processing cost associated with reanalysis, if the disambiguating region indicates that *John's house* is actually the embedded object. This conflict is the crux of this study.

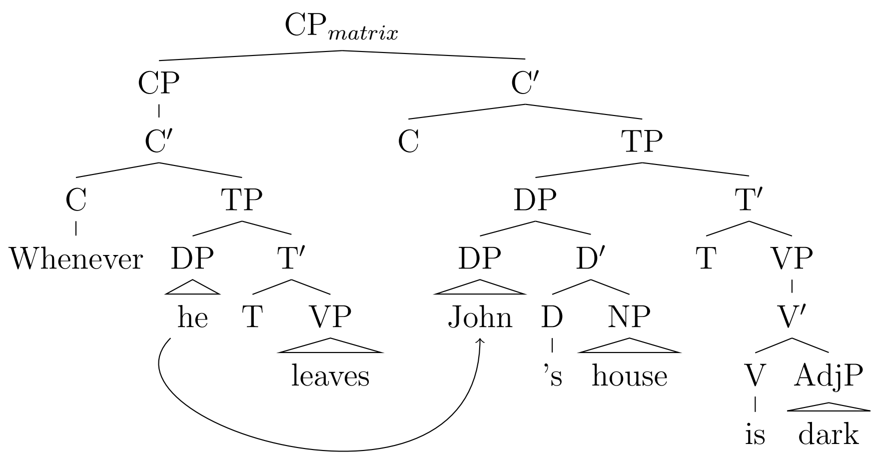
3.3.2.2. Matrix subject parse. In the case where the constructed parse is the one in which *John's house* is in the matrix subject position (henceforth the matrix subject parse), the formation of a coreference dependency between *he* and *John* is possible. If this dependency is attempted, the parser will check the lexical properties of *John* to see if such a coreference is feasible. For instance, previous studies have demonstrated that both gender and number are properties that, when incongruent, create a detectable reading time slowdown at the region of the R-expression (Kazanina et al., 2007; van Gompel and Liversedge, 2003). To extend this observation to the present sentence, in the case where *he* is replaced by *she*, as in (11)-d′, the parser may experience a gender mismatch effect (GMME) upon attempting to form a structurally licit dependency between *she* and *John*.

(11) d.′ Whenever she leaves John's house



In other words, the congruency of gender information between the pronoun and a candidate R-expression probes the parser's actions in terms of dependency formation. The presence of a GMME at the R-expression indicates that the parser has attempted to form a coreference dependency between the R-expression and another phrase (the cataphoric pronoun) that does not match in gender.

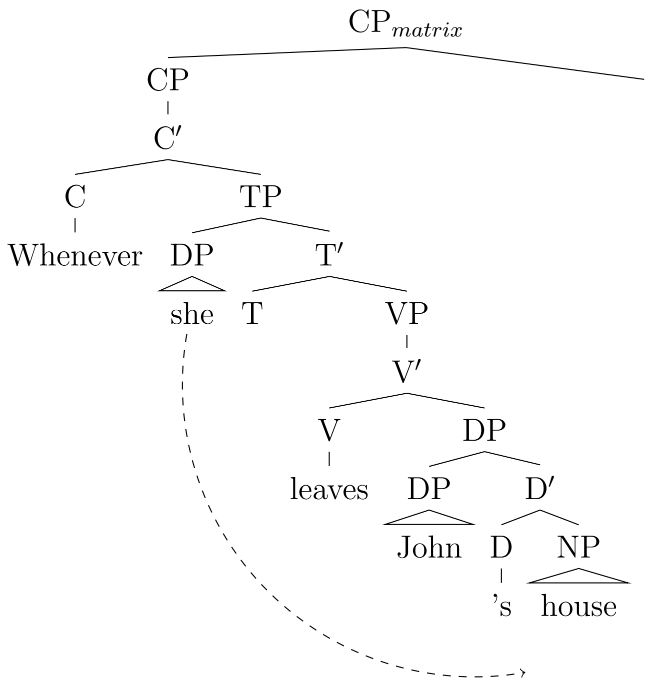
(11) e. Whenever he leaves John's house is dark.



In addition, the parser will be unsurprised by encountering the matrix verb in this parse because it analyzed the ambiguous DP as the matrix subject, so it may have been anticipating the matrix verb would follow next. These two possibilities are distinguishable by the parser's behavior at and following the matrix verb. The lack of a reading time slowdown in these regions would support the case where the parser prioritizes the ambiguous DP as the matrix subject.

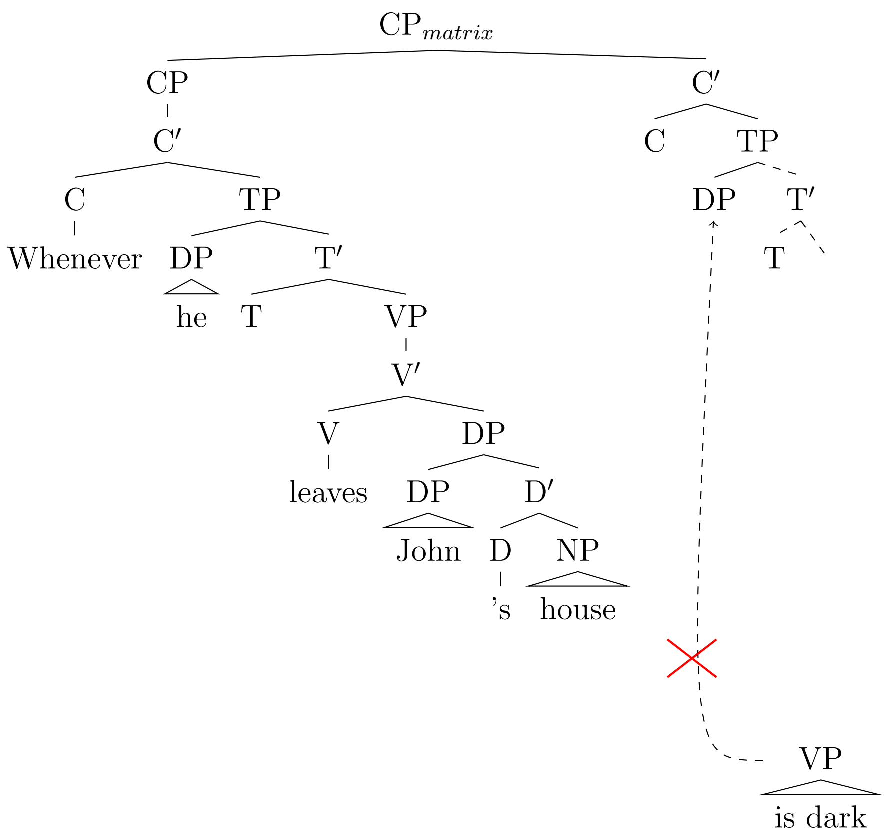
3.3.2.3. Embedded object parse. In a representation where the parser attaches *John's house* low into the parse tree, i.e., as the object of *leaves*, it would be unexpected to observe a GMME at *John*.

(11) d.″ Whenever she leaves John's house



The reason a GMME would be unexpected is because of the constraints placed on coreference dependency formation by Binding Condition C. In other words, if the parser prioritizes a structure in which the R-expression is c-commanded by the cataphor (henceforth the embedded object parse), there should be no difference in reading times regardless of gender congruency because there is no attempt at forming a coreference dependency.

(11) e.′ Whenever he leaves John’s house is dark.



The resolution of the local ambiguity is also an important and telling region. After dealing with the ambiguous region, the parser may have committed to particular a syntactic structure in order to quickly insert the DP *John’s house* into the parse tree and apply (or not apply) any relevant operations. The necessity of placing the DP in the parse tree is increased because of the active search. If the parser has constructed the embedded object parse, then the phrase following the ambiguous DP could be the matrix subject. In this case, the parser would be surprised to encounter the matrix verb immediately following the ambiguous DP. An increased processing time at this point would be an indication of the parser’s surprise at encountering the matrix verb, thus supporting the case where the parser commits to or favors the representation where the ambiguous DP is the embedded object.

The distinction between these two outcomes is mediated by the assumption that the parser has constructed only one representation of the input. In the case where the parser constructs multiple analyses in parallel for ambiguous inputs, there may be one favored representation (ranked higher or more highly activated, for instance). Even if one representation is favored, this does not exclude other representations from influencing or affecting the behavior of the parser. The details of such a situation are described in the next section.

3.3.3. Parallel parsing, or Multiple Outputs Model

It is a matter of simplicity that effects expected from only individual parses have been discussed thus far. It is possible that the parser can represent more than one individual parse at a time during structure building. That is, the parser may build alternative structures in parallel when it encounters the point where an ambiguous region begins, potentially producing multiple outputs at a given stage in parsing. These multiple-output models are defined "...as those in which a single processing stage passes more than one alternative on to a later stage where further decisions are made" (Boland and Cutler, 1996). This accounts for most of the models of the parser relevant to this section. While it is not necessary for a non-autonomous, non-serial parser to be a multiple-output model, those few versions here will not be considered here. The remainder of this section will consider the case where the parser builds two or more structures in parallel until a resolution is reached and one output overtakes the alternates, either as a determinately selected winner or as the most highly ranked or activated parse. This selection of a preferred parse does not preclude the possibility that the parser still maintains other parses or that, once selected, other potential interpretations are removed from storage. Instead, this section focuses on the route taken by the winning parse only.

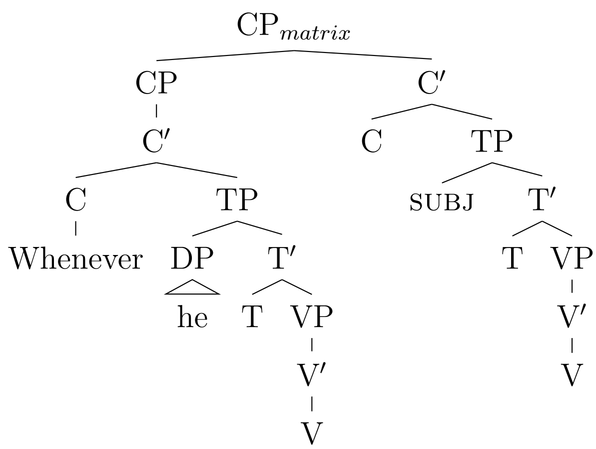
As in Sections 3.1 and 3.2, the multiple-output parser can build each structure in the same step-by-step procedures as a serial model of the parser. The question is at which points does the parser output a single representation or multiple representations, and, more importantly, how the active search for the tail of a dependency influences the way the parser navigates these representations. In the case where the parser's structure building is motivated by lexical information, the retrieval and recognition of lexical features could influence the parser to build simultaneous, competing structures for some lexical items, but not others. In a more "syntax-first" model of the parser, the building of a node with a label that could potentially have multiple argument structures might be a trigger for building simultaneous, competing structures. In this latter case, the lexical information (such as the argument structure of a verb) may later serve to filter out structures that are incompatible, rather than affect the initial constraints on structure building.

The parser could be triggered to build parallel structures in order to favor each bias it encounters. There is evidence suggesting that the parser does not wait to encounter disambiguating or clarifying bottom-up information before constructing a syntactic structure (Omaki et al., 2015; Phillips, 2006; Stowe, 1986; Tabor and Hutchins, 2004; Traxler and Pickering, 1996). This is the source of garden path effects in the classical serial model, in which the parser builds a structure that accounts for the information in an initial substring, but ends up constructing a parse that cannot be made to grammatically accommodate the new incoming material. In this serial model of the parser, it must then return to a previous location in the parse and revise the structure to account for the current input. The mechanism for this type of reanalysis would have to include a way of storing the structure of the parse for an indefinite amount of time, or risk decay. This model would also have to include a method of backtracking or searching the parse for a location where an alternative structure could be substituted.

Another possibility is that the parser does not need to backtrack to change which parse it is (most actively) pursuing. In this scenario outlined by Lewis (1998) the parser can "jump" from one parse to another within the parse space. However, this still requires that the parser maintain the intermediate stages of alternative parses in order to have constructed such an alternative parse to jump to. The parser's preference for a representation may be determined in a number of ways, but I leave the details of that calculation to future research. The parser's preference for a particular representation that is relevant to this dissertation may be calculated by lexical properties such as verb frame frequency and structural biases, such as the eagerness of the parser to complete an open dependency early and the general preference of the parser for Local Attachment.

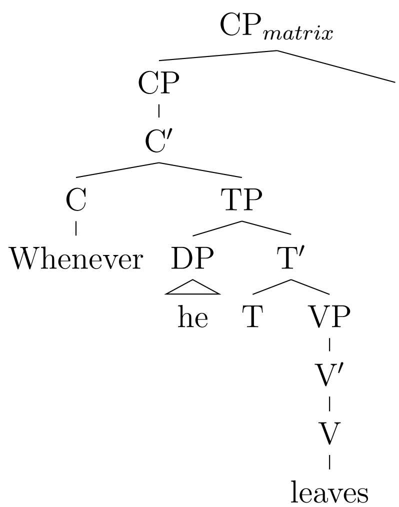
In multiple-output type models, the serial parsing procedures described above may be used to lay out step-by-step what the parser would be building in each of the parallel representations. To illustrate this, the trees in (12) are restated from the previous sections.

(12) a/b. Whenever he



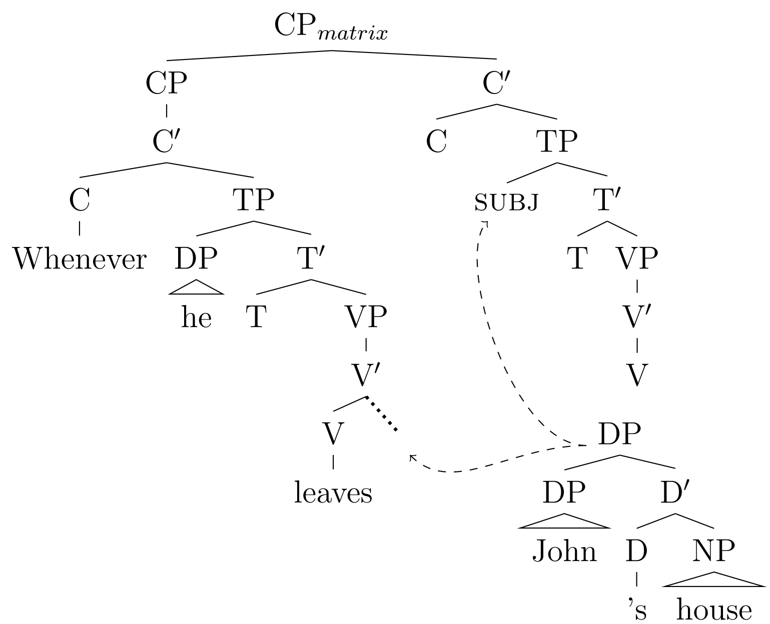
There is no ambiguity relevant to the discussion at the point where all the parser has encountered is *Whenever*, so the parser will proceed as it did before. Next, at the point where *he* is encountered, the search for the antecedent of the coreference dependency is triggered. Since the parser will have built the skeleton of a matrix clause, it may also anticipate that the first accessible candidate antecedent is located in the matrix subject position.

(12) c. Whenever he leaves



The addition of *leaves* is the point at which ambiguity can first be recognized. Although the structure built up to present is still unambiguous for our purposes, the parser may detect that *leaves* has two potential argument structures which would create two incompatible parses that the following word will have to deal with. That is, the parser may anticipate that the verb *leaves* could be either intransitive, thus being the final word in the embedded clause, or transitive, thus having a following object position.

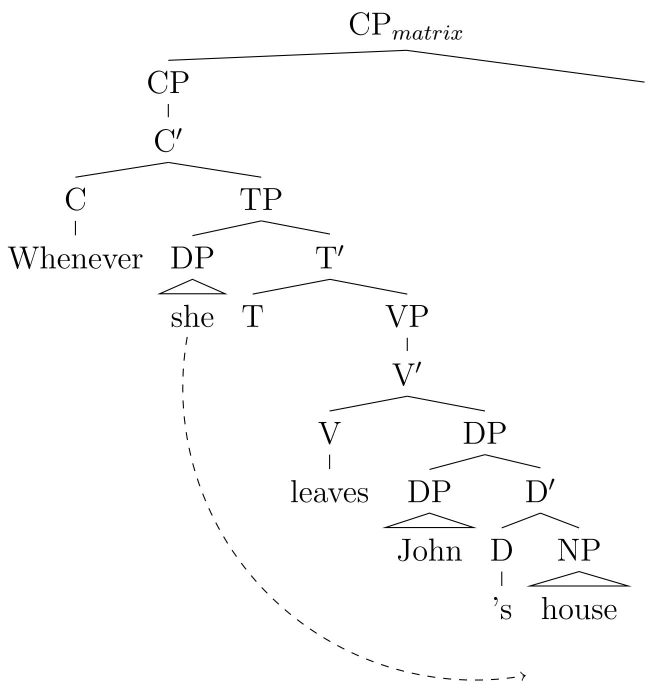
(12) d. Whenever he leaves John's house



This ambiguity is strengthened when the following input *John's house* does not disambiguate between the two possible parses because the longer an ambiguity is held, the harder it is for the parser to reanalyze it (Ferreira and Henderson, 1991). So, the parser generates the two alternatives. In the matrix subject parse, the parser is able to find a suitable antecedent at *John* and attempt to complete the dependency. This might terminate the antecedent search. It is conjectural whether the parser will continue to search for potential antecedents in the same type of active search once a successful candidate antecedent is found. It is the case that there may be multiple potential (gender-matched) antecedents, which would complicate matters. Additionally, there may be other coreferring phrases or names within the sentence that are not relevant to the current discussion. But, the relation between these other positions and the cataphoric pronoun are not relevant because they are not structurally constrained the way the R-expression is here.

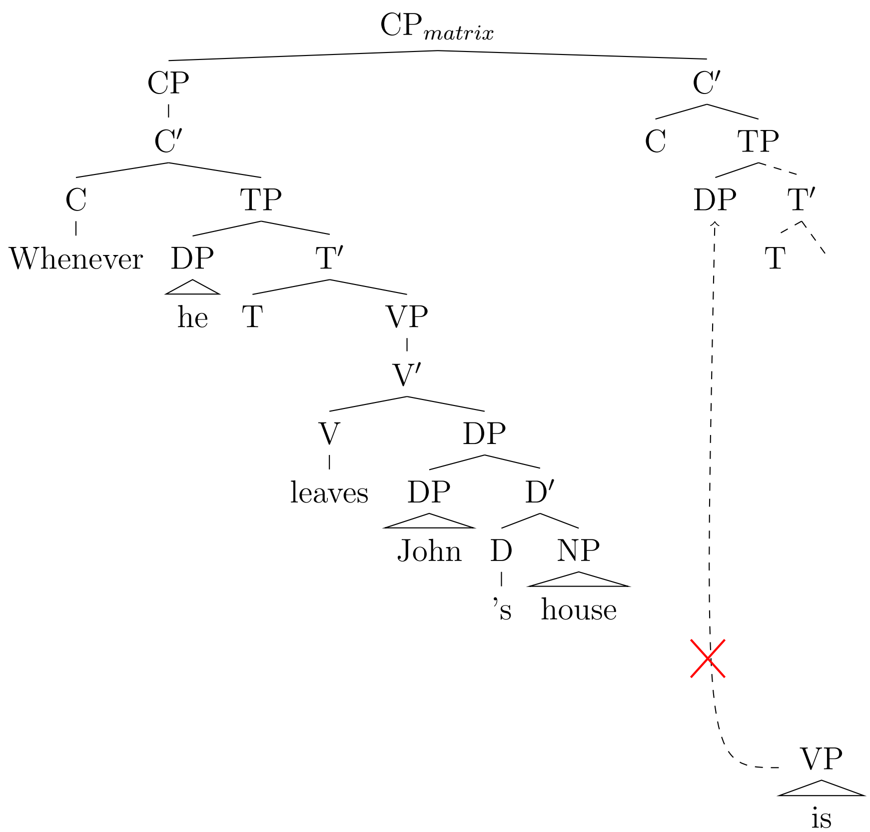
Moreover, they are more distant thus increasing the possibility that they will fall in other clauses and not be c-commanded by the pronoun (thus they would be accessible for coreference). Therefore, the current predictions focus only on the search for the first candidate antecedent and leave second candidates for future research. In the embedded object parse, the parser cannot legally form a dependency between *he* and *John*, so it continues on as described above. An alternative example, (12) highlights a potential way to probe the structure that the parser builds.

(12) d.′ Whenever she leaves John's house



In the parse where the cataphoric pronoun and the R-expression do not match in gender, the parser will attempt to form the dependency before the gender information is available, leading to a gender mismatch effect at the position of the R-expression when the R-expression is placed in an accessible position (Kazanina et al., 2007, a.o.). Moreover, in any parse where the cataphoric pronoun and the proper name in the ambiguous DP mismatch in gender, the parser will not yet have found a suitable candidate antecedent for the pronoun. Due to the active and eager nature of this search (Kazanina et al., 2007; Stowe, 1986; van Gompel and Liversedge, 2003), the parser may anticipate that another position will be available to contain a candidate antecedent. In the embedded object parse, the next available location is the matrix subject, which is not c-commanded by the pronoun, and is already anticipated as being the tail of the coreference dependency. In the matrix subject parse, the matrix subject position has already been located and discarded as an unsuitable candidate due to the gender mismatch, so the next potentially available location is the matrix object. In order to anticipate the possibility of a matrix object, the parser will consider the structure dominated by the matrix VP. In other words, in the condition where the parser fails to find an antecedent in the matrix subject, it may predict a DP complement of the matrix verb in order to expedite a new dependency formation, which would target the next location where a candidate antecedent could be. This type of behavior from the parser would be consistent with work done in head-final languages, such as Japanese, which shows prediction of the verb's structure in a rich syntactic representation before bottom-up information in encountered (Yoshida, 2006).

(12) e. Whenever he leaves John's house is ...



There is still the question of how two alternative representations being maintained by the parser affect each other. This becomes a crucial question at the point of resolution, in (12)-e. At this point, there is a matrix subject representation (that either has or has not completed the cataphoric dependency, depending on the gender congruency) and an embedded object representation (which has not completed the dependency, independent of the gender of the R-expression, due to Binding Condition C).

On one hand, it is possible that only the leading parse — the parse that is most highly ranked or activated, or the parse that the parser is currently pursuing — will display effects in reading times. In this case, there is no way to distinguish between the serial and multiple-output models in the paradigm used for this experiment. On the other hand, if the alternative parses can influence reading time measures despite being dispreferred (à la Fodor and Inoue, 1994; Lewis, 2000), this is a point at which one may be able to detect these effects. That is, if the parser has constructed a matrix subject representation, one may expect to see a gender mismatch effect in sentences like (12)-d′ (with mismatched gender) as compared to (12)-d (with matched gender). And, if the parser maintains both a matrix subject representation and an embedded object representation in parallel, one would also expect to see a surprisal effect at the matrix verb, from the embedded object parse's anticipation of a matrix subject following *John's house*. In the case where both effects are detected, one would have strong evidence that both parses are pursued in parallel.

It is possible that a slowdown in reading time at the matrix verb *is*, instead of a surprisal effect, a slowdown associated with pruning the inconsistent parse from the parse space. In this case one would only expect to see the slowdown in sentences where the parser has maintained multiple representations and where the parser has not had reason to anticipate or predict the internal structure of the matrix verb. This means that sentences where the gender congruency between the cataphor and R-expression is mismatched (both in ambiguous and unambiguous sentences) may actually show an increased reading speed at the matrix verb, due to the parser anticipating a lower candidate antecedent after its first attempt to complete the dependency fails. In contrast, one would expect to see slower reading times at the matrix verb for both ambiguous and unambiguous sentences in which the parser has been able to complete the cataphoric dependency. That is, if the embedded object parse is maintained long enough to influence the reading time at the matrix verb, the strongest effect should be from sentences like (12)-e, in which the parser has found a candidate antecedent and is still maintaining multiple representations.

3.4. Pilot 1: Offline Coreference Acceptability

In order to determine the ability of a reader to establish coreference between the initial cataphoric pronoun and the R-expression that follows the verb, first an offline forced-coreference acceptability-rating task was conducted. This task was designed to explicitly test the extent to which the global interpretation of the sentence affected the reader's ability to form a coreference dependency between the cataphor and the R-expression.

This experiment had a 2×2 factorial design, contrasting the factors of Gender (Match versus Mismatch) and the Location of the R-expression (Matrix Subject versus Embedded Object). None of the sentences had internal punctuation, although they all ended with a period ("."). The pronoun and R-expression were both underlined in all conditions, illustrated below in (13), and the participants were instructed to specifically rate how natural it sounds for the two underlined words to refer to the same person (i.e., corefer). Details of how the stimuli were constructed are described in Section 3.7.2. A complete list of stimuli can be found in the Appendix.

(13) MS+Match: Whenever <u>she</u> was trying to casually hum <u>Annie</u>'s favorite melody was hauntingly beautiful but Karen couldn't do it justice all alone.

MS+Mismatch: Whenever <u>he</u> was trying to casually hum <u>Annie</u>'s favorite melody was hauntingly beautiful but Jimmy couldn't do it justice all alone.

EO+Match: Whenever <u>she</u> was trying to casually hum <u>Annie</u>'s favorite melody it was hauntingly beautiful but Karen couldn't do it justice all alone.

EO+Mismatch: Whenever he was trying to casually hum Annie's favorite melody it was hauntingly beautiful but Jimmy couldn't do it justice all alone.

This experiment was run using Amazon Mechanical Turk (MTurk) and stimuli were randomized into 40 counterbalanced orders, one for each participant, using the Turktools toolbox (Erlewine and Kotek, 2015) and presented in vertical lists without breaks between stimuli, meaning that participants could scroll through the trials freely and see multiple trials at once. Each list contained one token of each item, with a total of 104 items. Twenty-four of these items were target stimuli, 24 were fillers which were especially designed to distract from the target items and use the full range of the Likert scale, and the remaining 56 were stimuli from three unrelated and non-conflicting, concurrent experiments. None of the filler experiments contained the type of verbal ambiguity manipulated for this study.

Forty participants were recruited from the United States who self-identified as having English as their native language. Their self-reported ages ranged from 23 years old to 60 years old, with a mean of 37.7 years and a standard deviation of 10.6 years. Participants were instructed to rate how normal or natural it would sound for the two underlined words (the cataphor and the R-expression) to refer to the same person. There were ten practice trials in which one instructed the participants to rate it a "7", one instructed the participants to rate it a "1", and one instructed the participants to rate it something between "1" and "7" in order to give them practice using the middle range of the scale. The remaining practice sentences did not give instructions. Experimental sentences were each presented above a seven-point Likert scale with (1) labeled "Bad" and (7) labeled "Good".

If the participants are sensitive to Gender, the factor intended to be use as a probe in the online experiments, one may expect then Gender Match condition to be rated significantly higher than the Gender Mismatched condition.

If the participants are using the syntactic resolution of the sentence as part of their evaluation of coreference, and if they have a single, clear representation of the structural relation between the cataphor and the R-expression, it is predicted that the Matrix Subject condition will be rated higher than the Embedded Object condition. This would be consistent with predictions that fall out from Binding Condition C. Given that our participants are sensitive to the factor of Gender, one may also predict a significant interaction between Gender and Location, with Gender Match + Matrix Subject rated higher than Gender Mismatch + Matrix Subject.

Depending on the strength of the effects, one may also observe a difference between Gender Match and Mismatch in the Embedded Object condition. Since participants are asked to actively evaluate potential coreference between words, even in the case of ungrammatical coreference, one may find that the Gender Matched condition is rated higher than the Gender Mismatched condition. Alternatively, the absence of such an observation may simply be a floor effect, thus still consistent with the previous predictions.

If participants do not use the global syntactic structure in their evaluations of coreference, one should find no difference between Matrix Subject and Embedded Object conditions. This would be extremely unexpected, but it would also indicate that the task of evaluating forced coreference may not be relevant to online coreference formation.

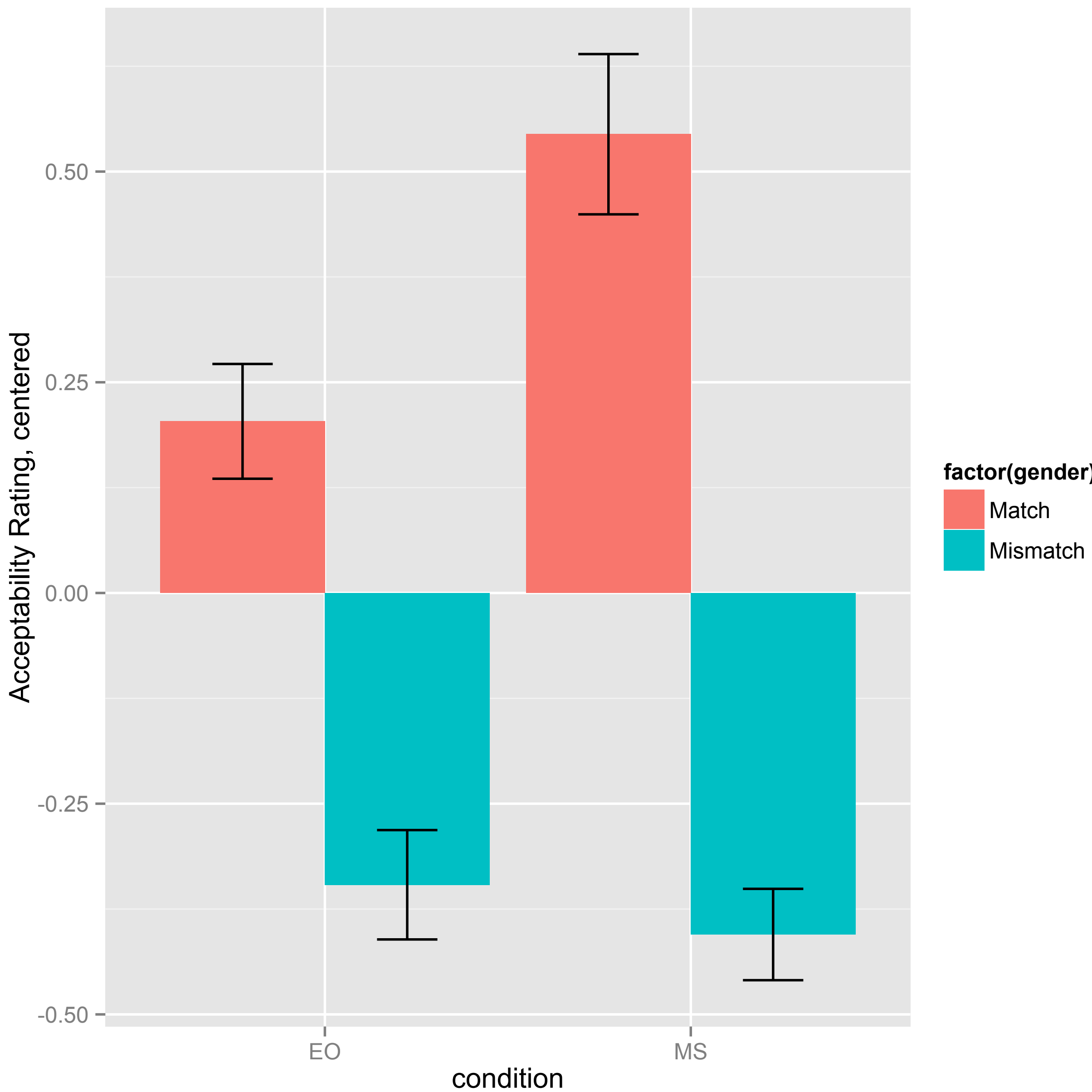


Figure 3.1. Scaled Likert Ratings for Pilot 1: Forced Coreference Acceptability

Results were analyzed using a linear mixed effects model (lmer), available in the *languageR* package (Baayen, 2011), with fixed effect and interaction of Location of the R-expression and Gender, and with random intercepts and correlated random slopes for participants and items,[7] since this was the maximal random effects structure that would allow the model to converge. An "item" henceforth refers to the set of four sentences that share content external to the manipulated factors, e.g. the four sentences in (13). The two fixed effects were contrast coded, in which each level was assigned either the value 0.5 or -0.5. There was found to be a significant main effect of Gender, with the Matched condition rated higher than the Mismatched condition (β = -0.547, S.E. = 0.093, t = -5.86, $p < 0.0001$). This unsurprising result indicates that participants are paying attention to the canonical genders of the pronoun and R-expression, so this manipulation can be used as a probe for dependency formation in future experiments.

There was also a significant main effect of Location, with R-expressions in the Matrix Subject position rated higher than in the Embedded Object position (β = 0.335, S.E. = 0.093, t = 3.58, $p < 0.001$). This indicates that participants are also paying attention to syntactic structural constraints such as Binding Condition C. This result is consistent with Binding Condition C constraining coreference in this manner.

Finally, there was a significant interaction between Gender and Location (β = -0.392, S.E. = 0.132, t = -2.96, $p < 0.005$). Visually, it appears that this interaction is driven by the Matrix Subject + Gender Matched condition, which has a broader distribution of data that extends to higher values than any other condition. However, there is a striking similarity in mean ratings between the two Gender Matched sub-conditions. This may indicate that, as described above, a forced-coreference rating task is not necessarily an accurate measure of how readers will interpret a sentence. Since, grammatically speaking, coreference is constrained by the Binding Conditions, thus the embedded object

[7] Model: $lmer(rate \sim disambiguation * gender + (1|subj) + (1|item), data = data)$

R-expression could not (grammatically) corefer with the subject that c-commands it, it seems that the higher ratings given to the Embedded Object + Matching condition may actually be an artifact of the task. Yet, the statistical significance of the interaction is sufficient to conclude that gender-matched R-expressions in the Matrix Subject position are, in fact, more acceptable and available for coreference than in the other conditions.

We may take from this experiment that gender is a useful probe for coreference dependency formation. One may also assume that readers may pay attention to the global structure of a sentence in in offline evaluations of coreference.

3.5. Pilot 2: Offline Forced-Choice

One of the shortcomings of rating-style tasks is that subjects are forced to evaluate garden path sentences with global structures that they may not be able to accommodate. One way to evaluate what kinds of structures the parser prefers to build initially, after encountering an ambiguous string, is to allow readers to complete the sentence in a way that makes sense to them. Given the variability in responses obtained through freeform sentence completion tasks (like production tasks in general), I used a forced-choice sentence completion paradigm that limited participants' responses to those relevant to our question.

In the previous pilot experiment, readers were presented with sentences that manipulated gender matching between the cataphor and first R-expression, as well as ultimate syntactic location of the R-expression, while punctuation, especially a comma that would have disambiguated the sentence before the matrix verb, was omitted in order to eliminate disambiguating cues. The current experiment is designed to test the disambiguation strength of the presence of a comma after the (otherwise) ambiguous verb. By determining how likely a subject is to prefer a matrix subject or embedded object interpretation of the relevant R-expression in the presence and absence of a comma, some insight into the effect of punctuation on the structure building process should be gleaned. By doing so, there will be an established basis for evaluating in future experiments the time-course of processing strings with and without commas during online tasks, as well.

This experiment had a 1 × 2 factorial design, contrasting a single factor: Punctuation (Comma versus No Comma). Stimuli were directly adapted from the Pilot 1, and modified as described. Sentences were presented with an underscored blank in the position after the first R-expression, illustrated in (14). Beneath the frame sentence, two options that could fill in the blank were presented. The presentation consisted of 144 items in total, with three sets of 24 fillers (= 72) especially designed to distract from the 24 target items, and the remaining 48 items were stimuli from two unrelated and non-conflicting, concurrent experiments. These filler experiments did not manipulate verbal ambiguity, and neither contained cataphoric pronouns. The full set of 144 stimuli were randomized into 40 counterbalanced orders using the Turktools toolbox (Erlewine and Kotek, 2015). For the stimuli of interest to this study, these two options were always *was* and *it was*. A choice of *was* would indicate that the reader built a structure in which the R-expression was in the Matrix Subject position. A choice of *it was* would indicate that the R-expression was in the Embedded Object position. All pronouns and R-expressions matched in gender within a sentence.

(14) Comma: Whenever she was trying to casually hum, Annie's favorite melody ___ hauntingly beautiful but Karen couldn't do it justice all alone.

◯ it was ◯ was

No Comma: Whenever she was trying to casually hum Annie's favorite melody ___ hauntingly beautiful but Karen couldn't do it justice all alone.

◯ it was ◯ was

If the presence of a comma after an ambiguously transitive verb is a strong indicator of upcoming verbal argument structure, specifically if it indicates that the verb is intransitive, one may expect that the Comma condition will show a strong preference for interpreting the R-expression as a Matrix subject, thus a high proportion of was choices. Alternatively, if commas are not strong indicators of intransitive verbal argument structure, one may expect the options to be chosen at chance levels.

There is also the possibility that the absence of a comma is a strong indicator of transitive verbal argument structure. If this is the case, one may expect that the No Comma condition will show a preference for the Embedded Object interpretation, and there will be a high proportion of *it was* selections.

Importantly, this result could be independent from the effect of the *presence* of a comma. For instance, if a comma indicates a stylistic pause (which may or may not correspond to the end of a major relevant syntactic constituent), there may be no strong motive to assign a syntactic structural property to the punctuation. It may conversely be that the absence of a comma is a strong *negative* indicator of the end of a syntactic constituent, which could bias the parser to construct a structure in which the verb phrase has continuing arguments. This leads to a second way to phrase the predictions:

If the presence of a comma is a strong indicator of syntactic constituency, one may expect that matrix subject interpretation will be indicated at a much higher rate than Embedded Object interpretations in the Comma condition. Alternatively, if the presence of a comma is a weak indicator of syntactic constituency, one should see a pattern of selection that is influenced by another factor (such as verb frame frequency).

If the absence of a comma is a strong indicator of syntactic constituency, one may expect that Embedded Object interpretation will be indicated at a much higher rate than Matrix Subject interpretations in the No Comma condition. Alternatively, if the absence of a comma is a weak indicator of syntactic constituency, one should see a pattern of selection that is controlled by another factor.

In any of these cases, it is possible that there is some external priority of the parser that is not currently accounted for. To that end, this study serves primarily as an exploration of what kinds of factors go into determining parser preference during offline structure building. From the previous literature, one may expect to see an overall bias toward transitive interpretations of these verbs, resulting in an overall bias toward the Embedded Object interpretation, regardless of the manipulation of Punctuation. If this is the case, given that the verbs in our target stimuli were chosen to be equibiased between transitive and intransitive readings, one may find that what constitutes an "equibiased" verb needs further exploration. Furthermore, it may also turn out that the cataphoric dependency search does affect the parser's behavior in this offline task, resulting in readers wanting to produce a sentence that allows for early dependency formation (thus giving the R-expression a Matrix Subject location).

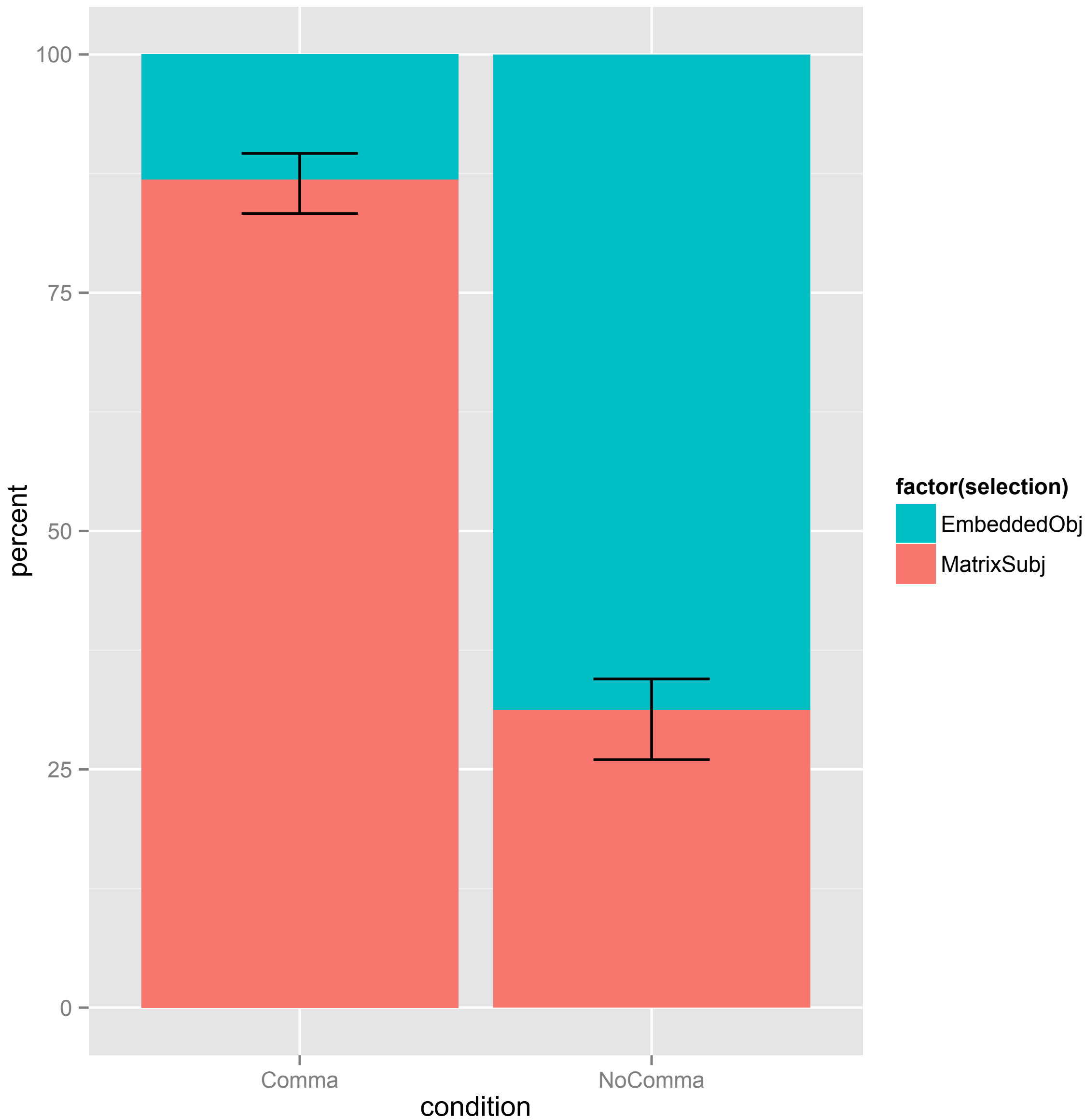


Figure 3.2. Proportion of Responses in Pilot 2: Matrix Subject versus Embedded Object. Error bars represent 95% Confidence Intervals.

In sum, the possible outcomes of this experiment are myriad and have many different explanations. Thus, one must be careful in attributing cause to whatever outcome is observed. Rather, one may use this experiment as an exploration of what kinds of effects are likely or even available to explore in online studies.

Selections of *it was* were coded with a“1” and selections of *was* were coded with a “0”. Proportions were calculated by summing the responses by condition and dividing by total number of responses. Confidence intervals were calculated using a test of equal or given proportions and are indicated by error bars in Figure 3.3. This 1-sample proportions test with continuity correction provides a Pearson's chi-square test statistic based on the binary data gathered, and indicates that the proportion of *was* selections in the Comma condition are significantly higher than *it was* selections (β = 86.76, $\chi^2(1)$ = 255.88, p < 0.0001, 95% C.I. = 83.31–89.61). This trend is reversed in the No Comma condition, in which *was* selections were significantly more frequent than *it was* (β = 30.08, $\chi^2(1)$ = 74.09, p < 0.0001, 95% C.I. = 26.02–34.48). The results were also analyzed using a logistic linear mixed effects regression model (lmer) with fixed effect of Punctuation, with random intercepts for participants and items, which was the maximal random effects structure that would allow the model to converge[8]. This analysis was performed using the *languageR* package (Baayen, 2011). The fixed effect was contrast coded, in which each level was assigned either the value 0.5 or -0.5. A significant main effect of Punctuation obtained, with the Comma condition having a higher proportion of *was* responses than the No Comma condition (β = 3.23, S.E. = 0.203, z = 15.93, p < 0.0001). Although this does not strongly indicate any particular mechanism is at work or that any particular influence

[8]Model: $lmer(response \sim condition + (1|subj) + (1|item), family = \text{“}binomial\text{”}, data = data)$

on structure building is prioritized by the parser, it is consistent with our hypothesis that the presence of a comma is an indicator of the right edge of a syntactic constituent. In other words, the presence of the comma after the (otherwise) ambiguous verb successfully disambiguated the sentence toward the matrix subject configuration, thereby completing the embedded VP constituent, and by extension the embedded CP constituent.

3.6. Pilot 3: Offline Acceptability with Line-Breaks

Before running an online eye-tracking study, it was also important to look for any effects the formatting requirements of the eye-tracker presentation interface would have on the comprehension of the stimuli. Since the eye-tracking while reading paradigm requires a fairly sensitive detector to determine which character on the screen the participant is fixed on, there are a limited number of characters available per display line. Moreover, if one considers that line breaks, like commas, could indicate some syntactic or prosodic feature of the sentence (illustrated by the comical example in (15)), it may be crucial to determine what kinds of effects line breaks required by the eye-tracker introduce.

(15) a. please eat
grandma

b. please
eat grandma

Thus, in order to determine whether the sentences formatted for the eye-tracking experiment are treated in a comparable way to those in Pilot 1, this experiment comprises the same task as in Pilot 1, but with the additional manipulation of line break location. In this way, the experiment had a 2 × 2 × 2 factorial design that manipulated the factors of Punctuation (Comma versus No Comma), Gender (Matched versus Mismatched), and Line-break (Present versus Absent). Briefly, the manipulation of Line-Break is illustrated in (16), within the Match + Comma condition. To avoid any obvious confounds introduced by line breaks between critical regions or near the potential location of the comma, line breaks were placed within syntactic constituents that were unlikely to be split by prosodic boundaries and in more natural locations that were outside the critical regions. For instance, in the Present Line-Break condition in (16-a), line breaks occur within a complex verb phrase, but nine characters before the ambiguous verb. This is between the matrix verb's spillover region and the continuing material. This final material (including the name *Karen* in (16)) is designed to provide an alternative coreferent for the pronoun, so it always matched in gender with the pronoun. It is important to note that, as below, the Absent Line-Break condition may still occupy two lines of text. This would depend on the size of a participant's browser window, the size of the font in their browser, and the size of the screen on which the window was displayed. Since these three factors are not within the experimenter's control through MTurk (the platform for this experiment), the difference between the Present and Absent Line-Break conditions may also be described as Controlled Line-Break and Uncontrolled Line-Break, respectively.

(16) a. Whenever <u>she</u> was trying to
casually hum, <u>Annie's</u> favorite melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

b. Whenever she was trying to casually hum, Annie's favorite melody was hauntingly beautiful but Karen couldn't do it justice all alone.

The presentation format was identical to Pilot 1, including forced coreference between the pronoun and the first candidate antecedent. In addition, all target sentences ultimately had the first R-expression in the Matrix Subject position. This design is directly based on the stimuli design used in the eye-tracking experiment, in order to give us basis for comparison between the offline results and online results. This format was followed for the 24 target stimuli. There were also 24 fillers designed especially to distract from the target stimuli. The remaining 32 stimuli were from two unrelated and non-conflicting, concurrent experiments. These filler experiments contained no manipulations of verbal ambiguity. The experiment was run using MTurk, with 40 participants in the United States who self-identified as having English as a native language. Their ages ranged from 21 to 67, with a mean of 39.8 and a standard deviation of 13.

The results were analyzed using linear mixed effects regression (lmer) with fixed effect of Location of the R-expression, Gender and Line-Break, with random intercepts for participants and items, which was the maximal random effects structure that would allow the model to converge[9]. This analysis was performed using the *languageR* package (Baayen, 2011). The three fixed effects were contrast coded, in which each level was assigned either the value 0.5 or -0.5. There was a main effect of Gender, with gender-matched coreference rated higher than gender-mismatched (β = -2.43, S.E. = 0.17, t = -14.1, p < 0.0001). A floor effect for gender-mismatched sentences is visibly evident, which is similar to the results found in Pilot 1. There was also a main effect of Punctuation, in which sentences

[9]Model: $lmer(rate \sim gender * punctuation * linebreak + (1|subj) + (1|item), data = data)$

with a Comma were rated higher than sentences No Comma sentences (β = -0.99, S.E. = 0.17, t = -5.75, $p < 0.0001$). There was an interaction between Gender and Punctuation, in which gender-matched sentences with a comma were rated the highest overall (β = 0.99, S.E. = 0.24, t = 4.07, $p < 0.0001$). Finally, although there was no main effect of Line-break, there was an interaction between Line-break and Punctuation, in which the No Comma + Absent Line-break condition was rated lower than the No Comma + Line-break condition (β = -0.64, S.E. = 0.24, t = -2.63, p = 0.0059). This effect persisted in several different analyses of these data, which are not reported.

The results of this pilot experiment show that the primary effects observed in Pilot 1 are still present in the eye-tracking stimuli. This means that any coreference effects observed in the online eye-tracking experiment can be discussed in terms of the participants' ability to force coreference. Yet the presence of an effect of line-break, especially in such a limited case, is strange. The persistence of this effect through removal of marginal participants and different data transformations (scaled based on the range of the Likert scale the individual used, both over all the stimuli and just the target stimuli) suggests that this effect is strong. The strength of the effect does not rule out the possibility that it is a fluke and would not be replicable, but it also encourages us to consider line break as a potential confound in following experiment.

On the other hand, it appears that the effect which a line-break has on the ability to force a coreference between target words is that it strengthens the effect of punctuation. A possible explanation for this involves implicit prosody, which is fundamentally beyond the scope of this experiment, but relevant to the study as a whole. A simplified version of the Implicit Prosody Hypothesis (Fodor, 2002) states that the task of silent reading

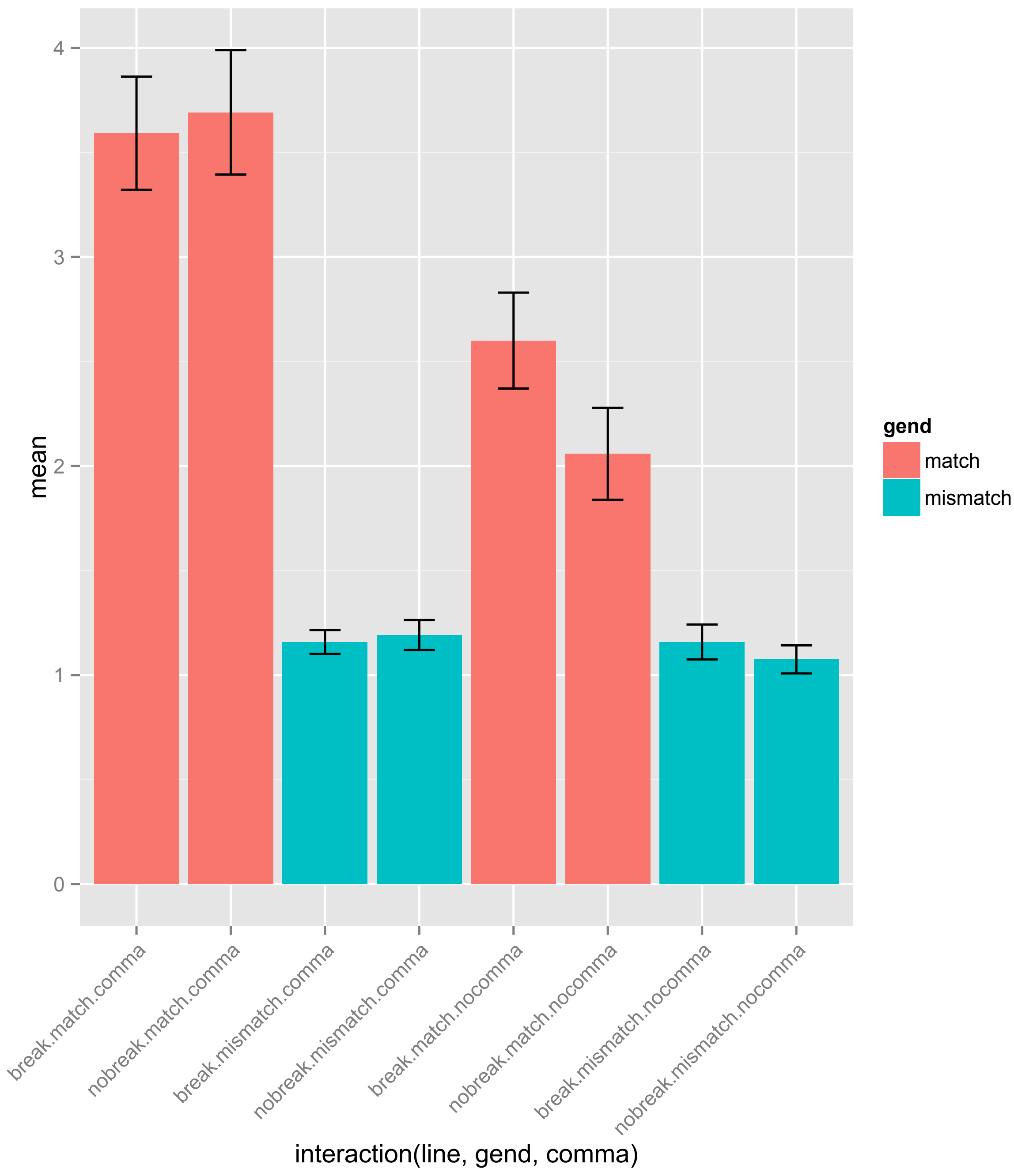


Figure 3.3. Likert Ratings for Pilot 3: Forced Coreference Acceptability with and additional manipulation of Line-Breaks.

is not exempt from prosodic effects because the reader "creates" some silent correlate to prosody internally. If line breaks imply prosodic boundaries, as intuitively demonstrated in (15), above, then it is possible that the line break locations used in this experiment induce prosodic phrasing of smaller constituents. This type of prosody is not necessarily unnatural, but would only occur in speech in limited environments modulated by paralinguistic information: for instance, shouting, speaking when out-of-breath, or speaking very slowly and/or carefully as if to someone who is hard-of-hearing or otherwise struggling to hear (Fujisaki, 1997). Ultimately, it is questionable why such a prosodic structure would affect the readers' ability to force an offline coreference in a sentence where that coreference is grammatical. So although this pilot demonstrates that line breaks do not diminish the effects observed in Pilot 1, there may be unintended confounds to consider in analyzing the eye-tracking results.

3.7. Eye tracking task

3.7.1. Eye-tracking methodology

The purpose of using the eye-tracking methodology here is to gather online measures of structure building during reading. Specifically, eye-tracking has advantages over other online methodologies because it allows the participant to engage in a fairly natural (common) task, keeps track of several types of measures, and importantly, both allows for the participant to look at previously read words and records those actions. This regression path measurement is of particular interest to garden path sentences and ambiguity studies because it can track the reader's attempts to reanalyze the sentence and rebuild or reconstruct the syntactic structures they ostensibly built on their first read-through.

Fundamentally, eye-tracking is a useful tool for studies of sentence processing because there is a correlation between location of attention and direction of gaze. Simply put, people look at what they are thinking about. Thus, the eye movements and fixations recorded during reading are informative as to in what order words are being attended to, and for what durations of time. From these measures, one can then assume that the readers' behavior is an arguably direct measure of the parser's behavior as it builds and revises syntactic structures.

Moreover, the sensitive measurements gathered by modern eye-trackers allow us to discriminate small-scale time course effects such as duration of first fixation and first-pass reading time on down to the scale of individual characters. In contrast to self-paced moving window tasks, eye-tracking does not rely on hand-eye coordination or cumulative processing effects. Although some measures similar to those gathered by self-paced moving window tasks can be useful, these measures are gathered in conjunction with the more fine-grained ones.

3.7.2. Materials

We recorded eye movements using an EyeLink 1000 tower with chin rest. The chin rest / tower was positioned 30 inches from a CRT monitor, ensuring that the display monitor was fully occupied within the tracking range of 55° horizontal and 45° vertical, as specified in the equipment manual. The experiment was presented on the monitor using the software EyeLink 0.07m, in a 12-point font size for the experimental stimuli (Figure 3.4).

This experiment had a 2 × 2 factorial design that manipulated the factors of Punctuation (Comma versus No Comma) and Gender (Matched versus Mismatched). There were a total of 96 items. Of these, twenty four were stimuli especially crafted to distract from the target stimuli and forty eight were adapted from unrelated experiments to provide a range of sentence lengths. None of the fillers manipulated verbal ambiguity or coreference, although some contained pronouns. The remaining sentences were twenty four target stimuli, listed in full in the Appendix and illustrated in (17), which consisted of a sentence-initial adverb clause beginning with either *Whenever* or *Any time*. This was followed by the cataphoric pronoun, which was followed by a five-word long past progressive verb phrase containing an adverb. A line break always occurred before the adverb, for the reasons described in Pilot 3. The verbs in this location were selected from the corpus designed by Gahl et al. (2004). Verb frame frequency was determined by calculating the proportion of attestations with a transitive frame (of any type) and an intransitive frame (of any type). Proportion of intransitive frames was subtracted from proportion of transitive frames to find the bias. Positive values were transitive-biased (with a value of 1 hypothetically indicating 100% of attestations have simple transitive frames), negative values were intransitive-biased (with a value of -1 hypothetically indicating that 100% of attestations have a simple intransitive frame), and a value of 0 represented perfect equibias. Bias was used rather than raw counts due to the variable number of total attestations in the corpus. Verbs with a bias of as close to 0 as possible that could also be followed by a concrete, possessable noun were selected. Due to the difficulty of constructing plausibly ambiguous stimuli, the ambiguous verbs were skewed toward having a higher proportion of intransitive frames, with biases ranging from +0.26 to -0.65. Based on Sturt (2007), however, it is reasonable to assume that intransitive-biased verbs that can take an object may still be treated as transitive verbs during dependency formation.

Following the head of this verb phrase, a comma was either present or absent, depending on the condition of the token. The ambiguous DP consisted of a genitive proper name (the "ambiguous NP") possessing the head noun, which was modified by an adjective (*Annie's favorite melody*). Since all items (thus all tokens within an item) ultimately have Matrix Subject configurations, the words following the ambiguous DP were always the matrix predicate, consisting of *was* and two words describing the head noun of the ambiguous DP. The final content of the sentence was less tightly controlled, but always started with the word *but* and a proper name that matched in gender with the cataphoric pronoun. Three items were excluded post hoc for a total of 21 items used in the analysis. These items were identified as contributing to longer reading times in all conditions due to the fact they contained low-frequency words, as calculated from the Corpus of Contemporary American English (COCA) (Davies, 2008). These three words were in item 5.5: *rambunctious*(=370), item 5.5 *racquet*(=390), and item 5.5 *dinghy*(=325), which were the only words in any of the stimuli with a frequency below 450 (out of 450 million words, as of 2015).

(17) Comma+Match: Whenever she was trying to
casually hum, Annie's favorite melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

Comma+Mismatch: Whenever he was trying to
casually hum, Annie's favorite melody was hauntingly beautiful
but Jimmy couldn't do it justice all alone.

No Comma+Match: Whenever she was trying to
casually hum Annie's favorite melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

No Comma+Mismatch: Whenever he was trying to
casually hum Annie's favorite melody was hauntingly beautiful
but Jimmy couldn't do it justice all alone.

Thirty eight participants were recruited from Northwestern University's introductory linguistics classes and received credit for participation. Three participants were excluded, for a total of N = 35, for the following reasons. One self-reported dyslexia, which may effect reading time and comprehension; one fell asleep during the experiment; and one had long, thick eyelashes that interfered with the calibration procedure (although this last participant was able to complete the experiment, the data gathered was unreliable and noisy). Participants were calibrated at the beginning of the experiment, and then periodically throughout the experiment as needed. Stimuli were presented after fixation on a black rectangle (in the location of the first character) was identified. Participants were instructed to press a button on a hand-held controller when they had read and understood the sentence, at which point they were presented with a comprehension question. Comprehension questions were answered with 'yes' or 'no', indicated by buttons on the hand-held controller.

Data gathered during recording was saved as an .EDF file on a secure external server. Data files were then processed and converted to .ASC files using the EyeDoctor software available through the UMass Eye-tracking website (available at:

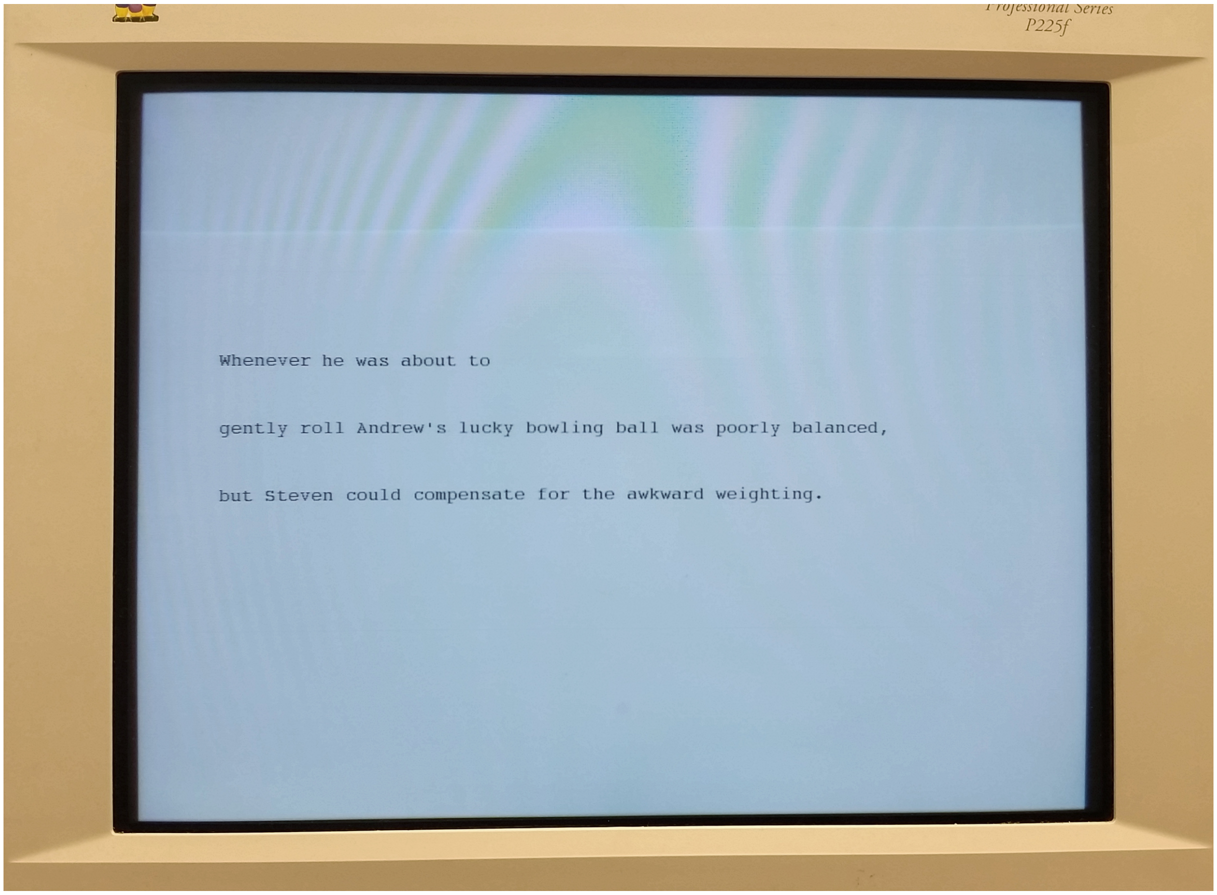


Figure 3.4. Sample presentation screen for eye-tracking task.

http://blogs.umass.edu/eyelab/software). During this process, data files were hand-aligned to correct for any vertical drift in the fixation locations. This process ensured that fixation locations that clearly indicated the participant was reading a word were saved within the boundaries of that word, thus could be used in statistical analysis. After hand-alignment, files were saved in .DA1 format and processed into a statistically analyzable dataset using Python scripts written by Shevaun Lewis (UMD wiki, archived at http://archive.languagescience.umd.edu/wiki/Eyelink%201000%20Eyetracker).

3.7.3. Predictions

Predictions, as generally described in Section 3.3, are as follows. If the parser prioritizes the cataphoric dependency search, one expects to see a gender mismatch effect (GMME) at the location of the first R-expression, due to the parser's preference for completing the cataphoric dependency early. This would be present regardless of punctuation, since the ambiguous verb allows the parser to "choose" to build an intransitive verbal argument structure in order to complete the dependency sooner. In addition, one would not expect to see a surprisal effect at the matrix verb, because the parser will have already placed the R-expression in a matrix subject position, thus may be expecting the matrix verb to follow.

If the parser prioritizes building locally coherent strings, one would expect to see a GMME only in the presence of a comma, which serves as direct evidence for an intransitive structure. In the absence of a comma, then, one does not expect a GMME, thus indicating that the parser did not attempt to form a coreference dependency at that location because it initially builds a structure with the R-expression in the embedded object position. Moreover, one would expect a surprisal effect at the location of the matrix verb because the parser will have predicted the presence of a possible matrix subject following the R-expression. Upon encountering no such subject, the parser will attempt to revise the built structure and recover from the garden path.

3.8. Results

For the following analysis of the eye-tracking data, a linear mixed effects model was used with fixed effect of gender and punctuation, their interaction, as well as presentation order and length of the region in characters, with random intercepts for participants and items along with an uncorrelated random slope for each of the fixed effects by participant and by item, which was the maximal random effects structure that would allow the model to converge.[10] This analysis was performed using the *lme4* package (Bates et al., 2015b). The factors of Gender and Punctuation were contrast coded, in which each level was assigned either the value 0.5 or -0.5. Correlations were removed from the random effects in order to allow model convergence while maintaining a maximal effect structure (Barr et al., 2013; Bates et al., 2015a). After computing this maximal LMER model, an effect of interest was removed. This diminished model was then compared to the maximal model using an ANOVA. This comparison calculated the contribution of the effect of interest to the maximal model. The statistics produced by this procedure are listed in Table 3.2.

The regions of interest examined by each comparison were the R-expression in the ambiguous DP (*Annie's*), the modifier in the ambiguous DP (*favorite*), the head of the ambiguous DP (*melody*), the matrix verb (*was*), and the region following the matrix verb (*hauntingly*) (which may show spillover effects, due to the short length of the matrix verb itself). The region following *hauntingly* was not analyzed because it contained a line-break, thus spanned two lines and had a negative length (in characters). The following sections report the analysis of the ambiguous name, and the matrix verb with its spillover region, which were the only three regions of those listed to show significant effects.

It is notable that no effects obtained in the region corresponding to *melody*, which contains the head noun of the matrix subject. I do not have a strong argument for why this should be the case, but I suggest the following. In the cases where the dependency

[10] Model: $lmer(value \sim gender * punctuation + order + length + (1 + gender + punctuation + order + length||subj) + (1 + gender + punctuation + order + length||item), data = data)$

successfully identifies *John* as the antecedent of the cataphoric pronoun (i.e., of *he*), there is no need to consider the head noun as an alternative because an earlier candidate has been identified. In the cases where the gender of *she* prevents *John* from being an acceptable antecedent, the short amount of time it takes the parser to reevaluate or otherwise "undo" the dependency from this position is enough such that *melody* has already been identified as a poor candidate antecedent (being inanimate, for instance), and thus has already been excluded from being a potential antecedent. However, this is not something these data can clarify and must be left to future work.

We examined two types of eye-tracking measures, which I will call early and late, for the sake of convenience. Early measures include first fixation (the duration of the first fixation within a region) and first pass reading time (the duration of time spent within a word from the first time the eye enters the region until the first time the eye leaves the region). Late measures include regression path duration (the duration of time spent from the first time the eye enters the region until the last time the eye exits the region to the right) and total time (total time spent within a region). The means and standard errors in milliseconds for each of these regions by condition are listed in Table 3.1.

3.8.1. Ambiguous NP critical region (R-expression)

In the region corresponding to *Annie's* in the example sentence (17), there were significant effects found in first pass reading times, reread duration and in total time. No other measures reached significance. The estimates (β) and standard errors calculated by the maximal LMER, plus the chi-squared statistic and p-value from the comparison of the

Table 3.1. Means (and Standard Errors) for Eye-tracking Experiment (Section 3.7).

Region		*Annie's*	*favorite*	*melody*	*was*	*hauntingly*
Punctuation	Gender					
First fixation						
Comma	Match	210.0 (6)	226.7 (8)	214.1 (8)	190.7 (11)	216.6 (6)
Comma	Mismatch	208.3 (6)	237.6 (7)	209.7 (6)	167.5 (9)	226.2 (7)
No Comma	Match	200.7 (6)	226.9 (6)	200.7 (6)	190.1 (13)	217.7 (6)
No Comma	Mismatch	211.9 (7)	222.8 (6)	210.9 (8)	184.3 (7)	221.5 (7)
First pass						
Comma	Match	230.5 (9)	263.3 (12)	241.7 (11)	190.7 (11)	274.7 (13)
Comma	Mismatch	261.0 (14)	266.6 (11)	233.7 (10)	167.5 (9)	266.4 (11)
No Comma	Match	224.9 (9)	256.5 (9)	218.5 (8)	199.0 (14)	297.3 (13)
No Comma	Mismatch	251.5 (11)	261.2 (12)	225.0 (9)	187.9 (8)	270.8 (11)
Regression path						
Comma	Match	317.0 (25)	339.6 (24)	288.5 (18)	254.4 (26)	319.1 (24)
Comma	Mismatch	362.4 (28)	349.5 (27)	303.3 (25)	175.4 (11)	297.4 (19)
No Comma	Match	357.3 (28)	348.0 (26)	254.3 (15)	270.8 (35)	404.7 (33)
No Comma	Mismatch	379.2 (30)	303.1 (16)	271.1 (18)	213.3 (16)	412.4 (36)
Re-read time						
Comma	Match	327.2 (33)	384.1 (45)	266.0 (31)	233.8 (49)	255.7 (21)
Comma	Mismatch	367.3 (43)	375.5 (58)	303.8 (55)	164.6 (7)	273.1 (31)
No Comma	Match	415.9 (42)	369.4 (40)	334.1 (40)	242.6 (57)	348.8 (32)
No Comma	Mismatch	459.3 (44)	340.3 (32)	304.8 (33)	263.9 (57)	365.1 (34)
Total time						
Comma	Match	416.5 (26)	409.3 (26)	315.7 (19)	212.6 (15)	384.1 (22)
Comma	Mismatch	453.7 (33)	401.8 (30)	313.0 (22)	207.9 (12)	370.8 (22)
No Comma	Match	458.2 (29)	435.6 (29)	349.4 (24)	262.3 (26)	462.2 (26)
No Comma	Mismatch	527.9 (34)	446.8 (27)	372.9 (24)	271.4 (18)	468.9 (26)

maximal LMER to the diminished LMER (with critical effect removed) are listed in Table 3.2, by measure, region, and effect.

For first pass reading time, there was a main effect of Gender congruency, with Mismatched gender slower than Matched gender (β = -24.14, S.E. = 9.8, χ^2 = 6.01, p = 0.014). This is consistent with the Gender Mismatch Effect (GMME) described in the

Table 3.2. Combined ANOVA and LME results for Ambiguous NP in Eye-tracking Experiment (Section 3.7).

Region	Effect	Estimate	S.E.	χ^2 (df)	p-value
First fixation					
Annie's	Gender	-4.15	7.2	0.36 (1)	>0.1
	Punctuation	-0.86	5.6	0.02 (1)	>0.1
	Interaction	-9.01	13.7	0.43 (1)	>0.1
First pass					
Annie's	Gender	-24.14	9.8	6.01 (1)	=0.014
	Punctuation	-3.12	9.9	0.11 (1)	>0.1
	Interaction	7.55	19.5	0.15 (1)	>0.1
Regression path					
Annie's	Gender	-26.32	27.7	0.91 (1)	>0.1
	Punctuation	29.7	41.1	0.53 (1)	>0.1
	Interaction	33.2	51.9	0.43 (1)	>0.1
Re-read time					
Annie's	Gender	-37.97	43.6	0.77 (1)	>0.1
	Punctuation	97.25	44.0	4.50 (1)	=0.034
	Interaction	8.06	74.4	0.01 (1)	>0.1
Total time					
Annie's	Gender	-43.11	28.1	2.34 (1)	>0.1
	Punctuation	72.64	27.7	6.13 (1)	=0.013
	Interaction	-24.07	62.4	0.15 (1)	>0.1

literature. Curiously, there is only a main effect in this early measure, suggesting that the presence or absence of a comma does not affect the parser's decision to attempt to form a coreference dependency between the cataphoric pronoun and the ambiguous name. Despite this region visually appearing to be driven by the Match vs Mismatch conditions within the Comma condition, the interaction did not reach even marginal significance. This suggests that the main effect is sufficient present independent of punctuation.

Moreover, to confirm that this effect was present in the No Comma condition independently from the Comma condition (where it was assumed to obtain), a linear mixed effects regression was calculated. In this analysis, the maximal lmer model had fixed

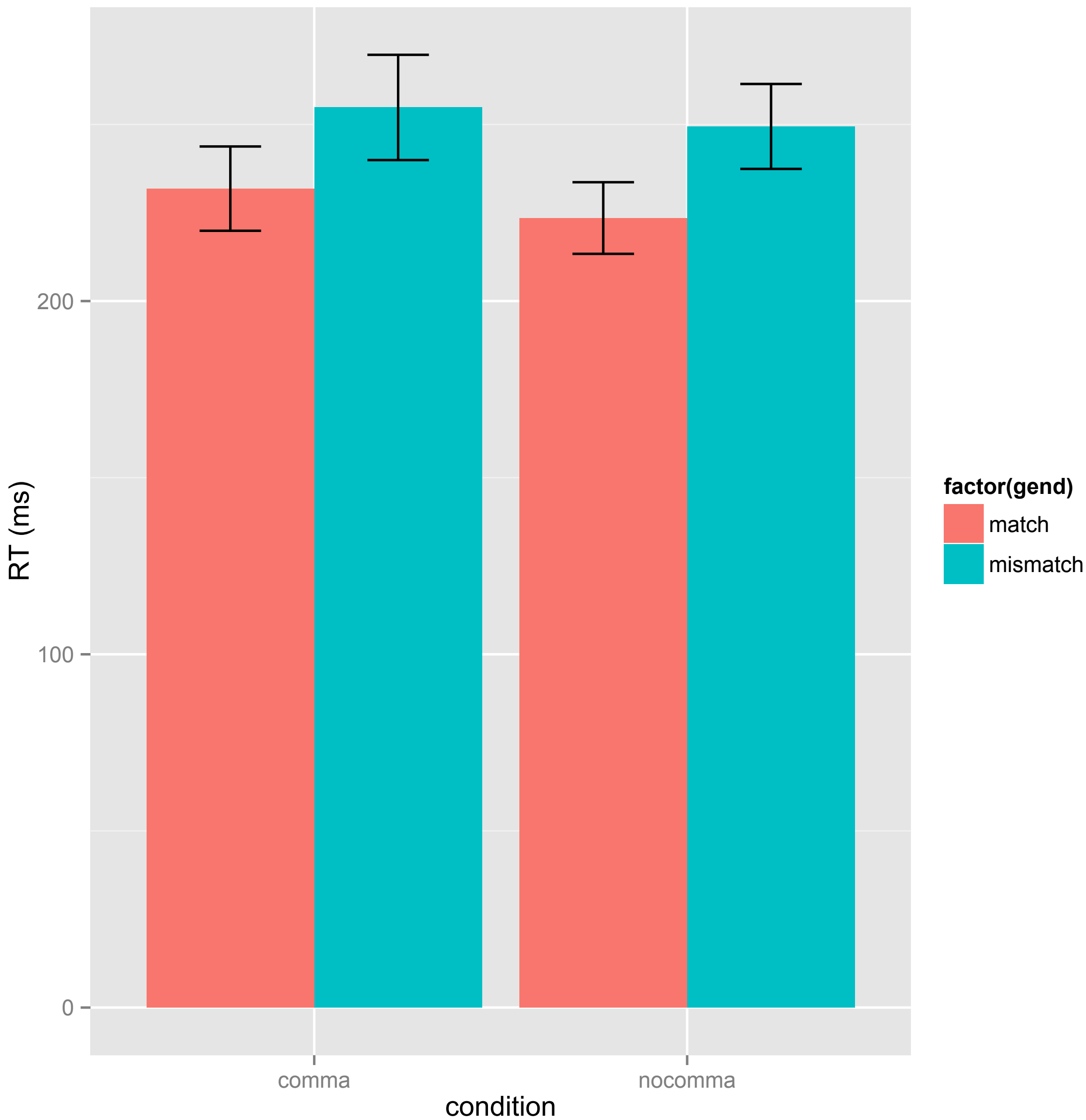


Figure 3.5. First Pass RT (Ambiguous NP Critical region, *Annie's*)

effects of gender and length, with random intercepts for participant and item, along with random slopes for gender and length by both participant and item, which was the maximal random effects structure that would allow the model to converge.[11] Correlations were removed from the random effects in order to allow for convergence. A marginally significant effect of Gender obtained, with Matched gender read faster than Mismatched gender (β = -26.17, S.E. = 13.3, $\chi^2 = 3.77$, p = 0.052).

In re-read duration and total time, a main effect of Punctuation was found, with the No Comma condition read slower than the Comma condition (RR: $\beta = 97.25$, S.E. = 44.0, $\chi^2 = 4.50$, p = 0.034; TT: $\beta = 72.64$, S.E. = 27.7, $\chi^2 = 6.13$, p = 0.013). This indicates that during regressions from later in the sentence back to this critical region (*Annie's*), more time was spent dwelling on these regions in the No Comma condition. A possible explanation for this is that the ambiguity created by the lack of a comma "catches up" with the parser later on in the sentence, after the initial influence from the cataphoric dependency search biases the parser to construct the matrix subject parse. Then, the parser must re-evaluate the R-expression for candidate status. The late influence of punctuation will be considered in more detail in Section 3.9.

3.8.2. Matrix verb critical region

In the region corresponding to the matrix verb *was*, two main effects reached significance. The estimates (β) and standard errors calculated by the maximal LMER, plus the chi-squared statistic and p-value from the comparison of the maximal LMER to the diminished LMER (with critical effect removed) are listed in Table 3.3, by measure, region, and effect. In regression path duration, the only significant effects observed were

[11]Model: $lmer(value \sim gender+length+(1+gender+length||subj)+(1+gender+length||item), data = data[data\$punctuation == \text{“}nocomma\text{”},])$

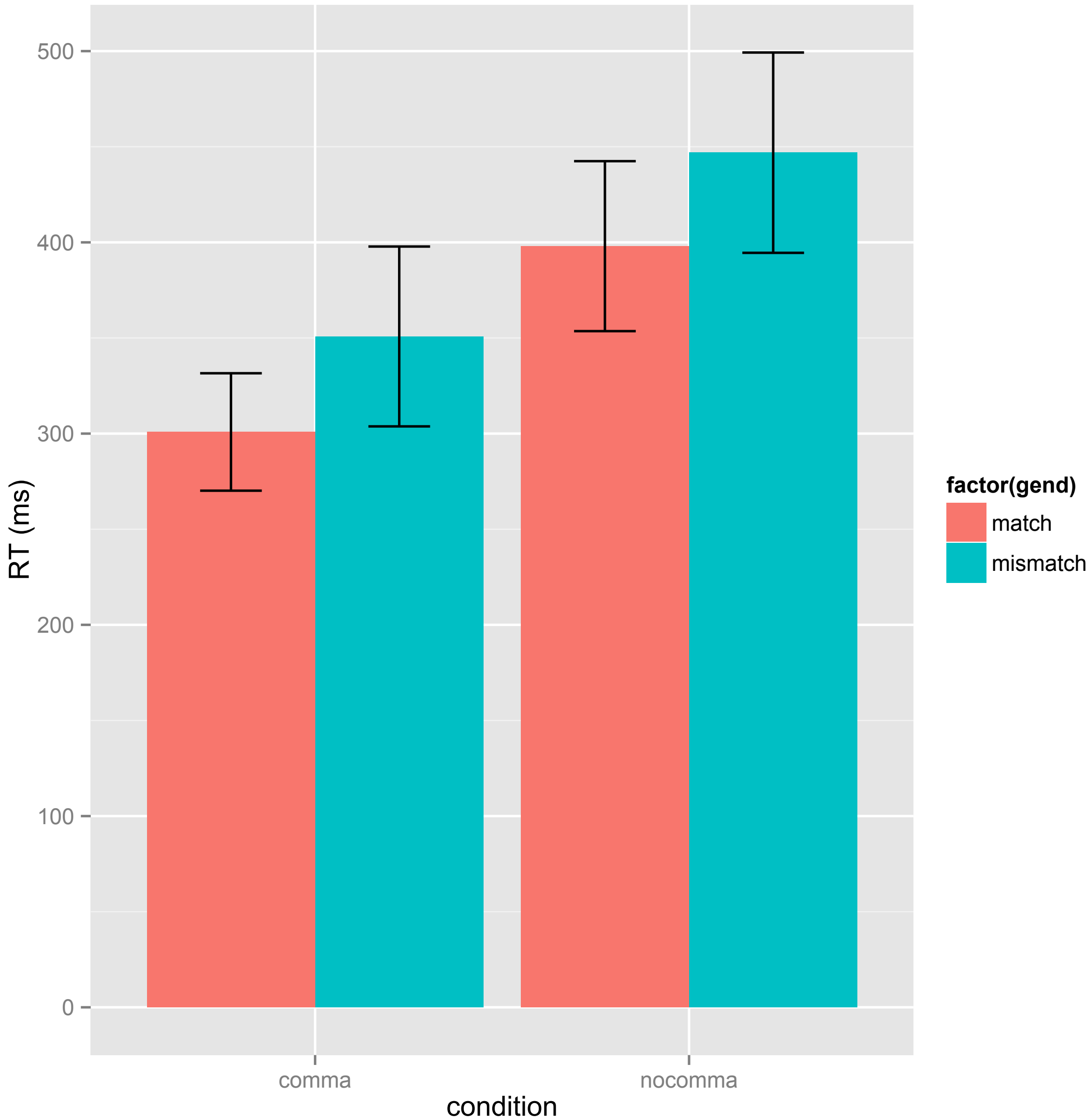


Figure 3.6. Re-read Duration (Ambiguous NP Critical region)

main effects of Gender, with the matrix verb read slower in the Matched conditions than in the Mismatched conditions (β = 64.95, S.E. = 25.4, χ^2 = 6.27, p = 0.012). As mentioned in Section 3.7.2, three subject were excluded post hoc. Before these three subjects

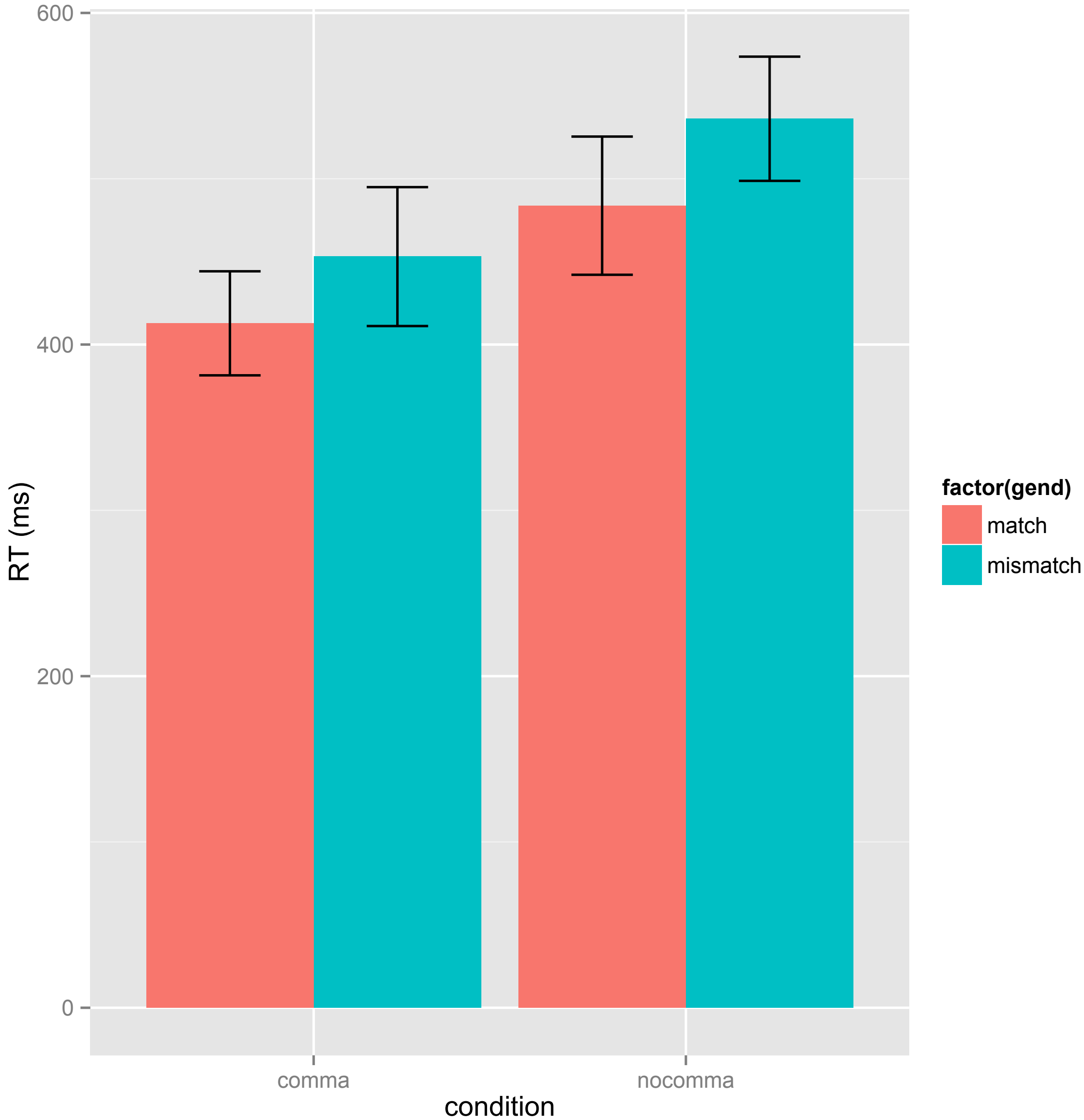


Figure 3.7. Total Time (Ambiguous NP Critical region, *Annie's*)

were identified and removed, however, significant effects of Gender were observed in first fixation duration and first pass reading time. With the problematic subject removed, a visual trend is still apparent (as illustrated by Figure 3.8), but the effect no longer reaches significance. A priori, we have no motivation for predicting such an observation because our hypotheses did not address a mechanism that might allow gender to influence reading time at the verb. Therefore, explanations for this observation will not be addressed currently. For a thorough discussion of what a slowdown at the matrix verb critical region could mean for the parser when it is associated with the Matched gender condition, I direct the reader to Chapter 4, in which this effect is replicated and interpreted in detail.

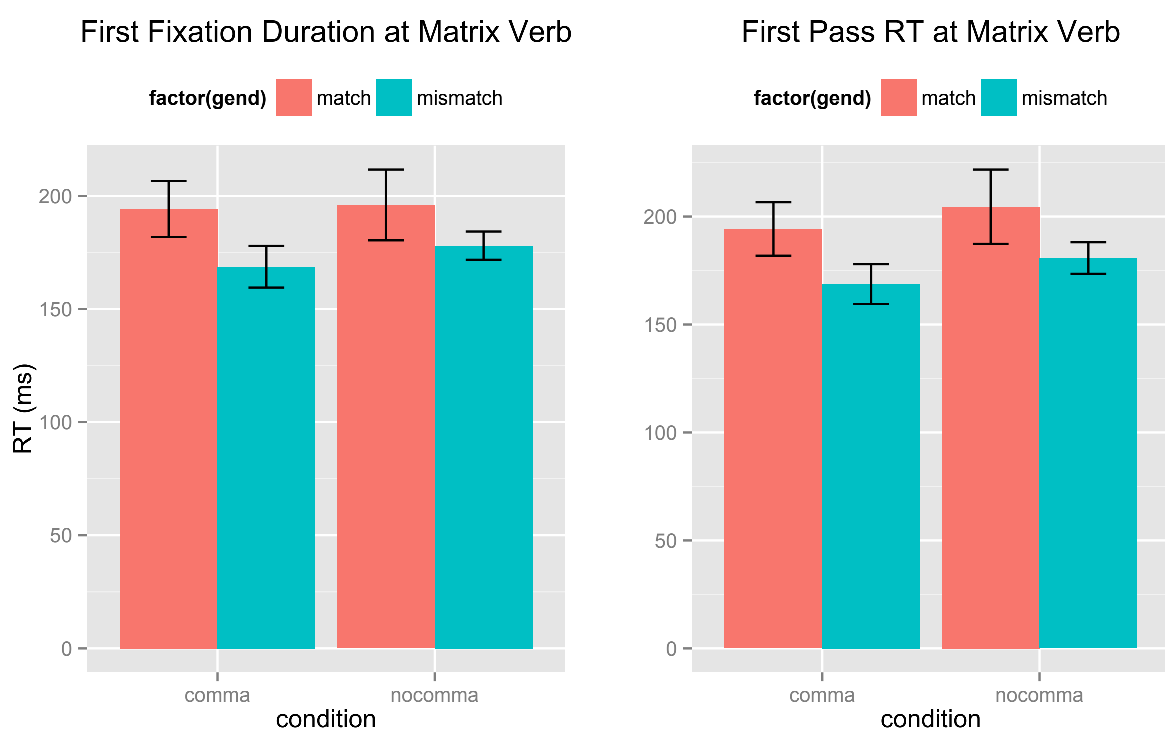


Figure 3.8. First Fixation and First Pass Reading Times (Matrix Verb Critical region, *was*)

On the other hand, there is theoretical motivation for observing a main effect of punctuation, which was found in the total time measure. Significantly more total time was spent in this region in the No Comma condition as compared to the Comma condition

($\beta = 58.76$, S.E. $= 23.7$, $\chi^2 = 5.49$, p $= 0.019$). This suggests that the presence of a comma biases the parser toward building a structure in which the ambiguous DP is the matrix subject. In this case, the parser is not surprised by the matrix verb, since it expects the verb to follow shortly after the subject. In other words, in the absence of a comma, the parser builds a structure in which the ambiguous DP does not serve as the matrix subject, thus it is surprised when it encounters the matrix verb without having assigned an DP to the position of matrix subject yet. However, this effect is only apparent in this cumulative reading time measure, potentially due to the short length of the critical region. For this reason, the spillover region following the matrix verb is examined in the following section.

3.8.3. Matrix verb spillover region

A stronger indication of an increased processing cost at the matrix verb occurs in the following region, which has a more substantial length, thus has a better chance at showing statistical effects. In this spillover region, the region corresponding to *hauntingly*, we observe main effects of Punctuation in regression path duration, re-read time, and total time, all of which show that the No Comma condition was read slower than the Comma condition (RP: $\beta = 112.84$, S.E. $= 34.0$, $\chi^2 = 9.17$, p $= 0.0025$; RR: $\beta = 105.58$, S.E. $= 33.8$, $\chi^2 = 8.81$, p $= 0.003$; TT: $\beta = 93.47$, S.E. $= 26.2$, $\chi^2 = 11.23$, p $= 0.0008$). These effects suggest, as before, that the presence of a comma may override a default embedded object interpretation, thus in the absence of a comma, the parser predicts (or expects) that the ambiguous DP is actually the embedded object Thus upon encountering the matrix verb with no preceding matrix subject, the parser is surprised. It might be noted that there is also a visual trend for the Mismatch conditions to have slower reading

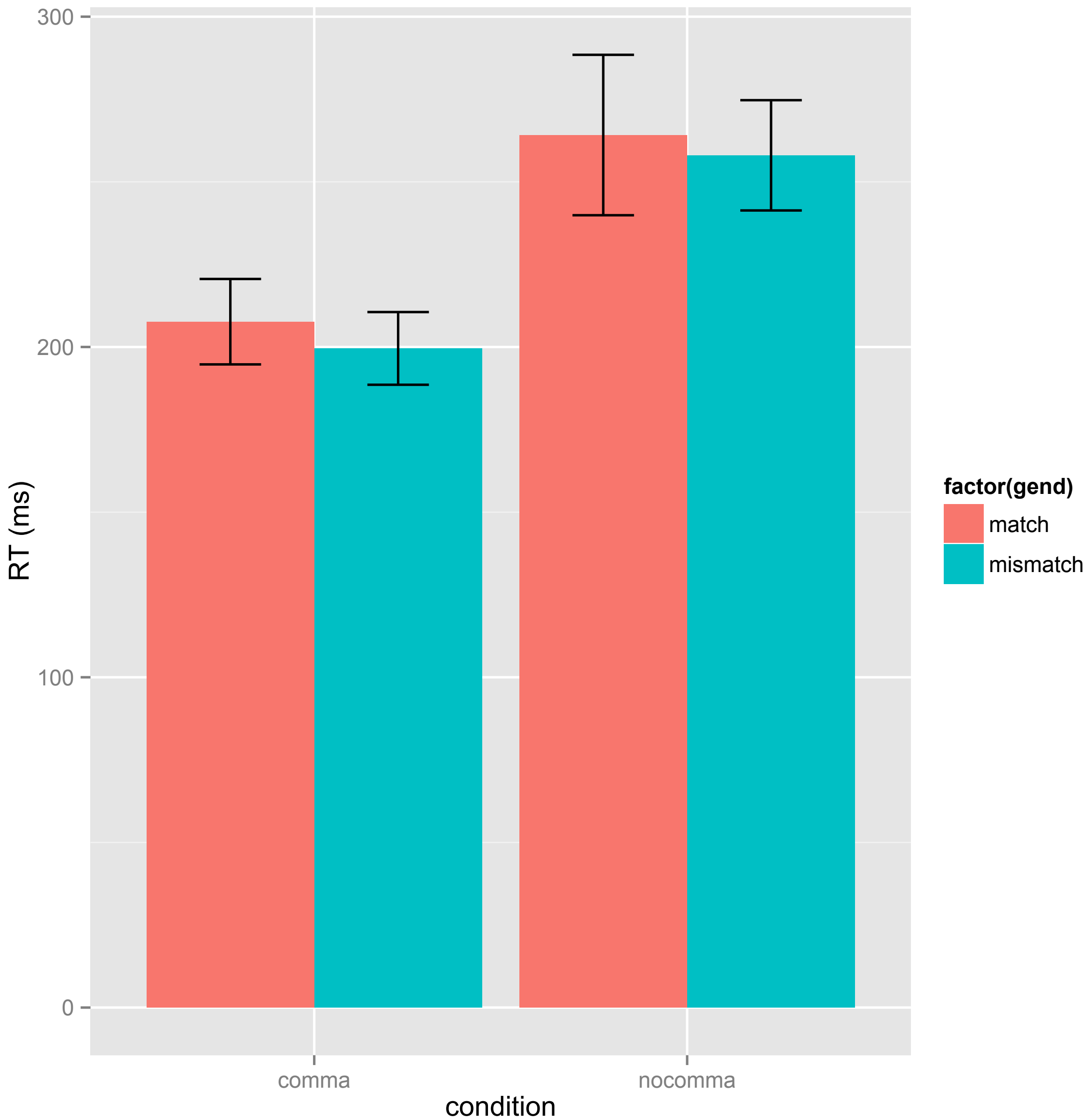


Figure 3.9. Total Time (Matrix Verb Critical region)

times than the Match conditions in total time and reread duration, for instance, but this did not even reach marginal significance.

Table 3.3. Combined ANOVA and LME results for Matrix Verb in Eye-tracking Experiment (Section 3.7).

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First fixation					
was	Gender	17.85	11.9	1.99 (1)	>0.1
	Punctuation	7.38	9.6	0.60 (1)	>0.1
	Interaction	12.16	21.9	0.34 (1)	>0.1
hauntingly	Gender	-5.26	6.91	0.56 (1)	>0.1
	Punctuation	0.41	8.3	0.003 (1)	>0.1
	Interaction	6.53	11.9	0.32 (1)	>0.1
First pass					
was	Gender	17.73	13.4	1.61 (1)	>0.1
	Punctuation	12.15	10.3	1.37 (1)	>0.1
	Interaction	-6.78	21.9	0.11 (1)	>0.1
hauntingly	Gender	17.20	10.7	2.53 (1)	>0.1
	Punctuation	18.85	14.1	1.77 (1)	>0.1
	Interaction	17.28	21.3	0.67 (1)	>0.1
Regression path					
was	Gender	64.95	25.4	6.27 (1)	=0.012
	Punctuation	23.13	27.0	0.79 (1)	>0.1
	Interaction	-13.45	46.5	0.10 (1)	>0.1
hauntingly	Gender	8.97	33.2	0.08 (1)	>0.1
	Punctuation	112.84	34.0	9.17 (1)	=0.0025
	Interaction	28.89	63.3	0.22 (1)	>0.1
Re-read time					
was	Gender	40.90	60.4	0.41 (1)	>0.1
	Punctuation	94.00	61.8	1.83 (1)	>0.1
	Interaction	-110.42	118.6	0.84 (1)	>0.1
hauntingly	Gender	-13.90	30.9	0.23 (1)	>0.1
	Punctuation	105.58	33.8	8.81 (1)	=0.0030
	Interaction	25.19	71.3	0.12 (1)	>0.1
Total time					
was	Gender	1.20	24.7	0.001 (1)	>0.1
	Punctuation	58.76	23.7	5.49 (1)	=0.019
	Interaction	-9.51	42.4	0.05 (1)	>0.1
hauntingly	Gender	12.26	19.6	0.38 (1)	>0.1
	Punctuation	93.47	26.2	11.23 (1)	=0.00080
	Interaction	-17.66	46.2	0.15 (1)	>0.1

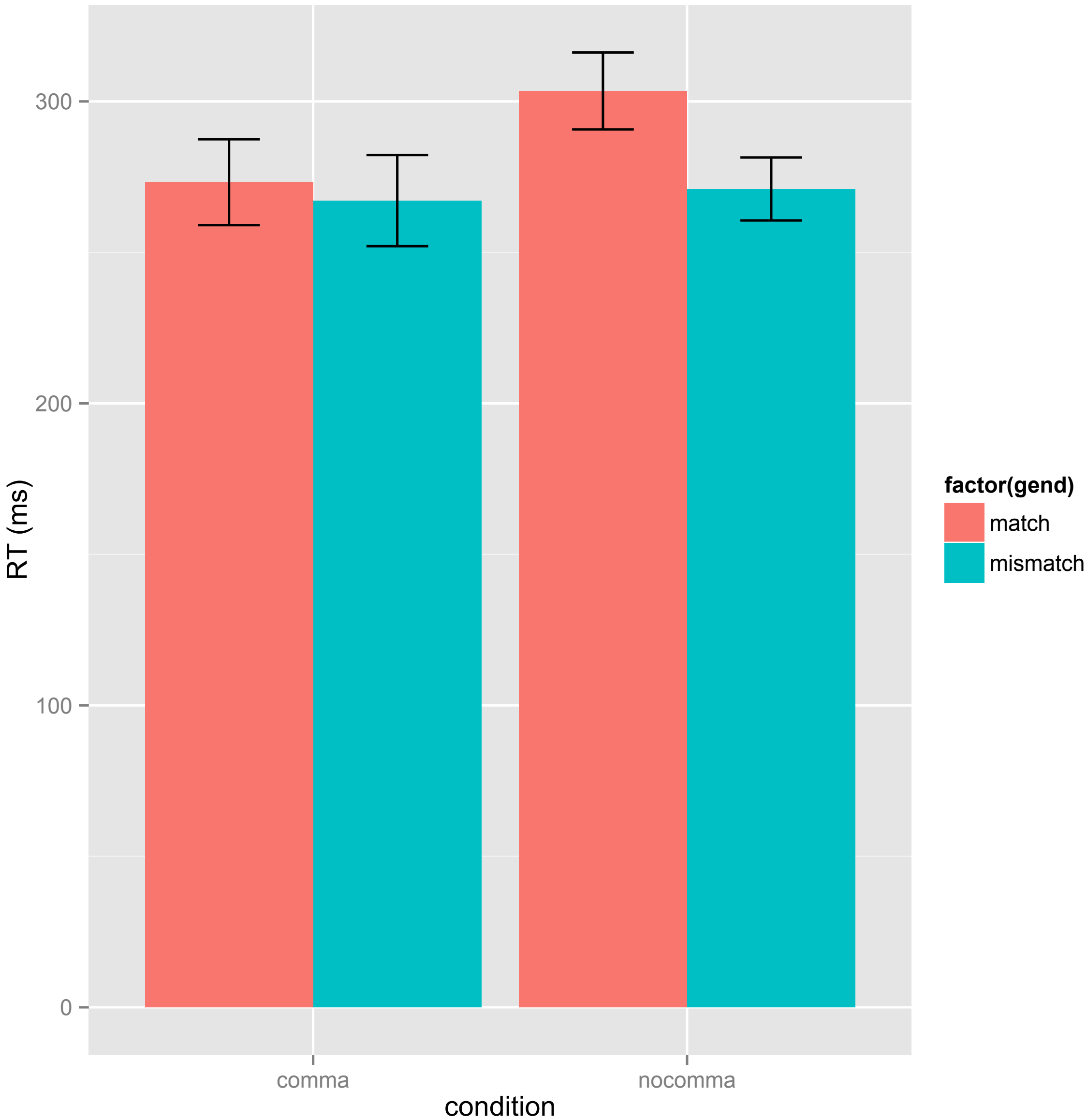


Figure 3.10. First Pass RT (Matrix Verb Spillover region, *hauntingly*)

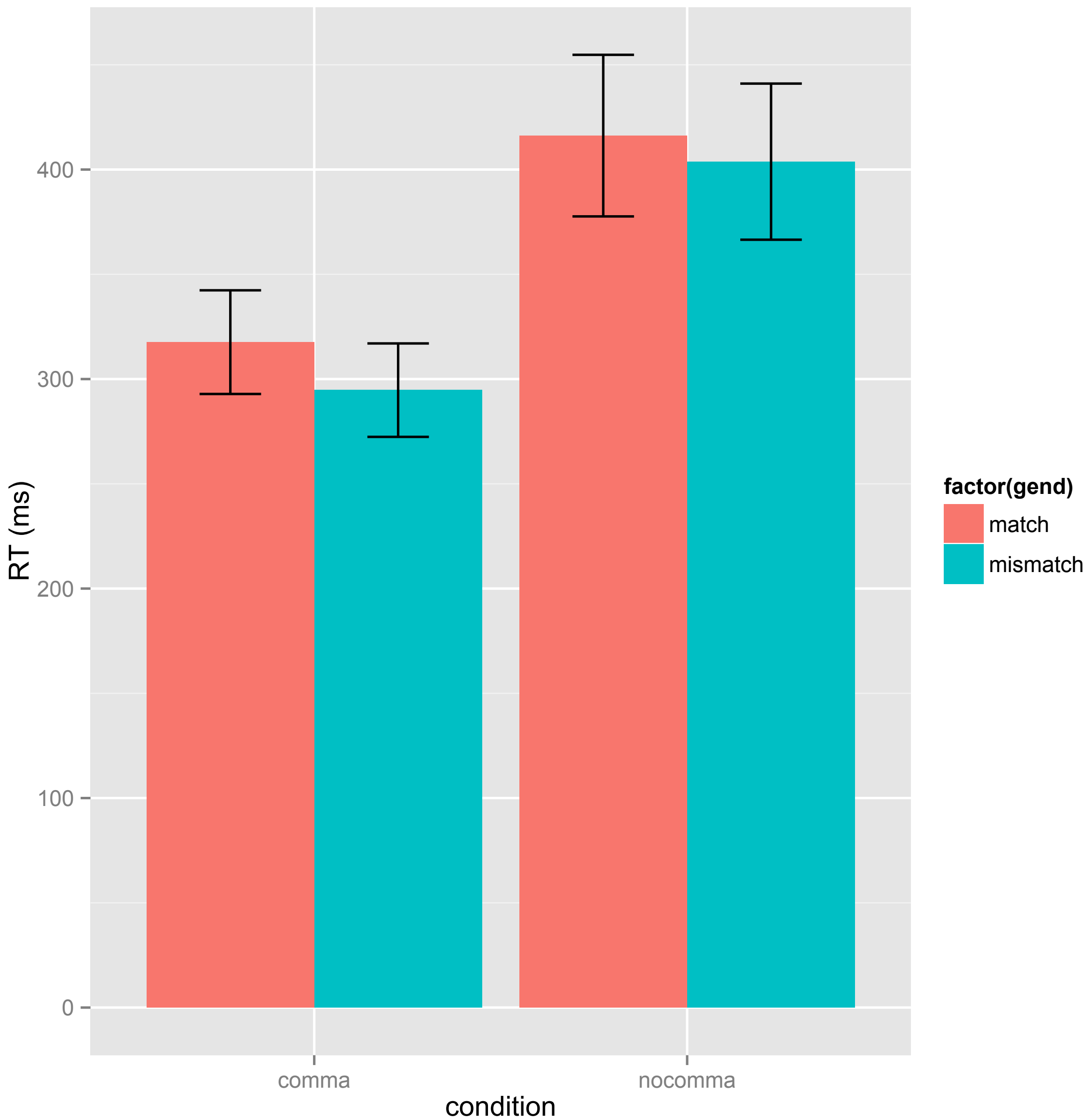


Figure 3.11. Regression Path RT (Matrix Verb Spillover region, *hauntingly*)

3.8.4. Discussion

The observed GMME at the R-expression in the ambiguous NP critical region suggests that the parser does consider a parse in which the NP is not c-commanded by the cataphoric pronoun. This pattern was generally predicted for the unambiguous Comma

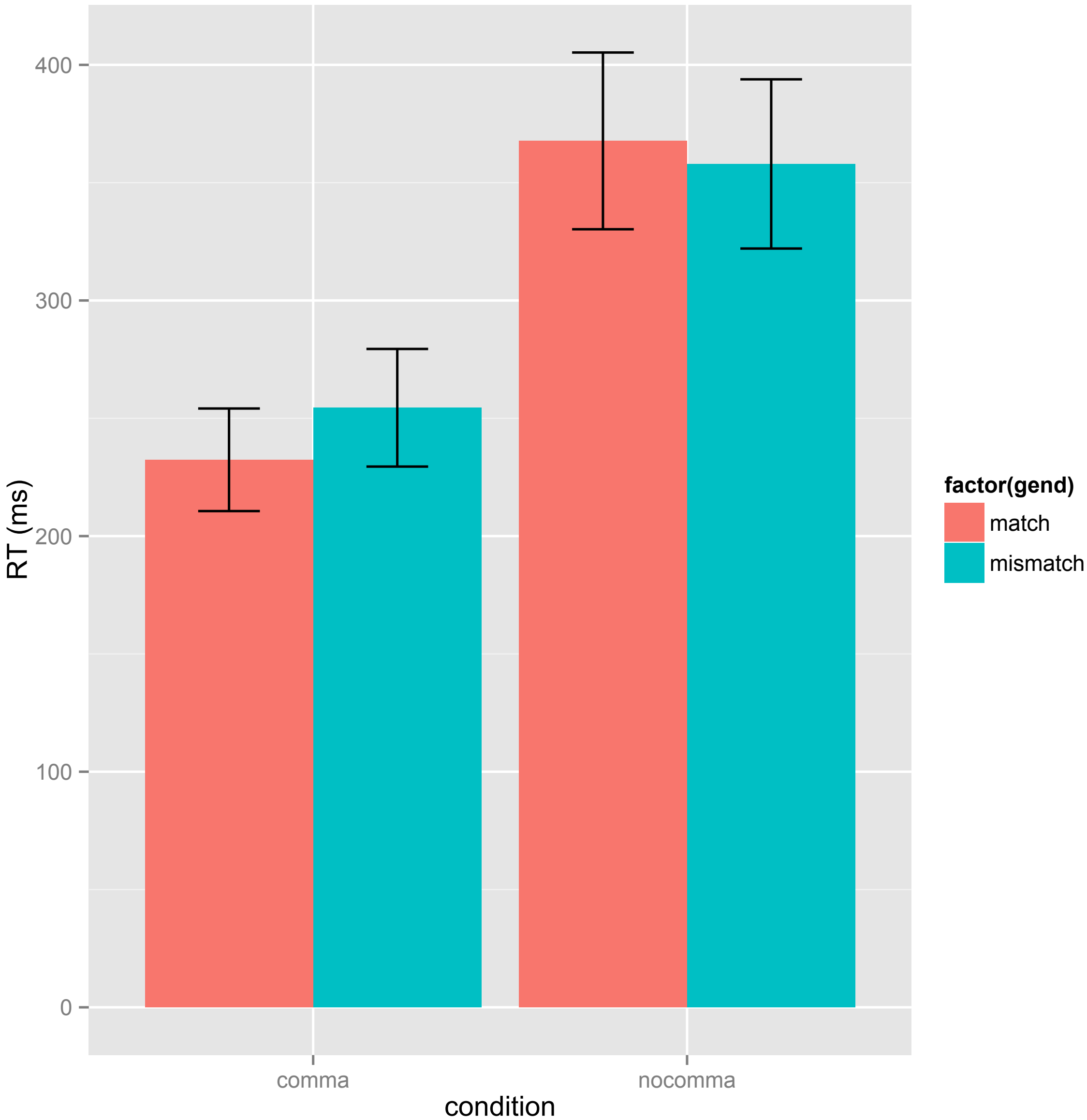


Figure 3.12. Reread Duration RT (Matrix Verb Spillover region, *hauntingly*)

condition, but only predicted in the No Comma condition if Active Search influences the parser's priorities such that it attempts to locate the antecedent of the coreference dependency even in a case where doing so would potentially violate Binding Condition C.

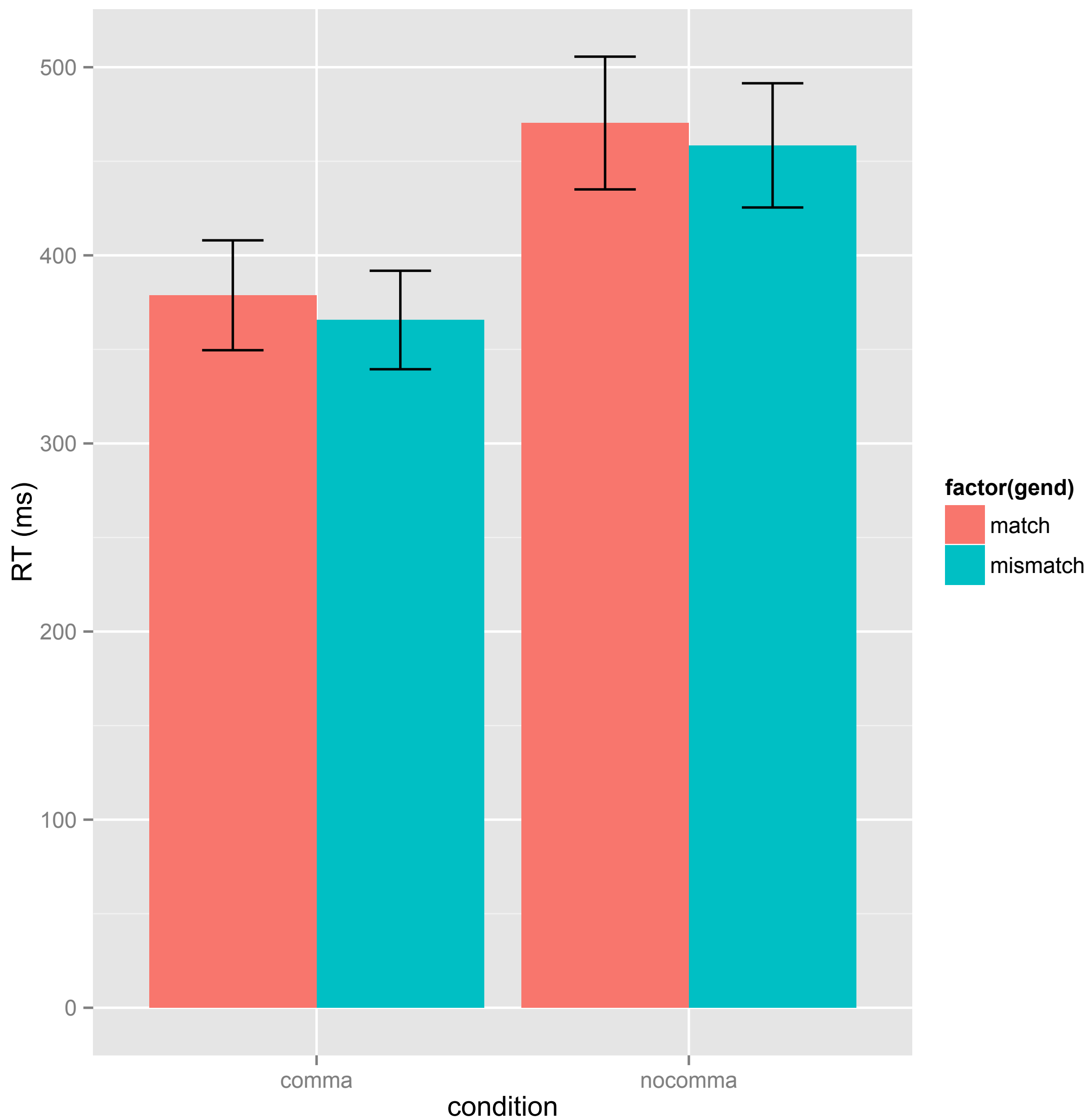


Figure 3.13. Total RT (Matrix Verb Spillover region, *hauntingly*)

However, in later measures at the matrix verb, there is a reading time slow-down in Mismatch as compared to Match conditions consistent with a surprisal effect at the matrix verb (when the parser is expecting a matrix subject. The presence of this effect suggests that the parser is also considering a locally coherent parse despite already finding a globally coherent parse. One potential explanation for this is that each time a sentence is read, the parser constructs either the matrix subject parse or the embedded object parse in proportion to their likelihood as determined by the verb frame frequencies. If this is the case, the parser will sometimes build the matrix subject parse and thus show a relative slowdown at the ambiguous NP critical region (corresponding to the name *Annie's*) and no slowdown at the matrix verb. Then, also, the parser will not show this slowdown at the ambiguous NP when the parser builds the embedded object parse, but it will show a slowdown at the matrix verb. Thus, there will be a correlation between the speed of reading times at the ambiguous NP and the matrix verb. That is, longer reading times at the ambiguous NP (MS parse) would be associated with shorter times at the matrix verb, and shorter reading times at the ambiguous NP (EO parse) would be associated with longer reading times at the matrix verb.

A post hoc analysis of the data was performed to determine whether such a correlation existed. The analysis used a linear mixed effects regression (LMER) in which the dependent variable was the regression path duration at the matrix verb spillover region (*hauntingly*) and the dependent variable was the first pass duration at the ambiguous NP critical region, since these were the two regions and reading times that demonstrated the relevant effects. Specifically, regression path duration was selected over reread time and total time at the matrix verb spillover region because a surprise effect triggered by encountering a verb when the parser expects a subject could conceivably influence the parser to look back (i.e., regress) to earlier points in the sentence, making this measure a key indicator of the embedded object parse. The model also allowed for random intercepts

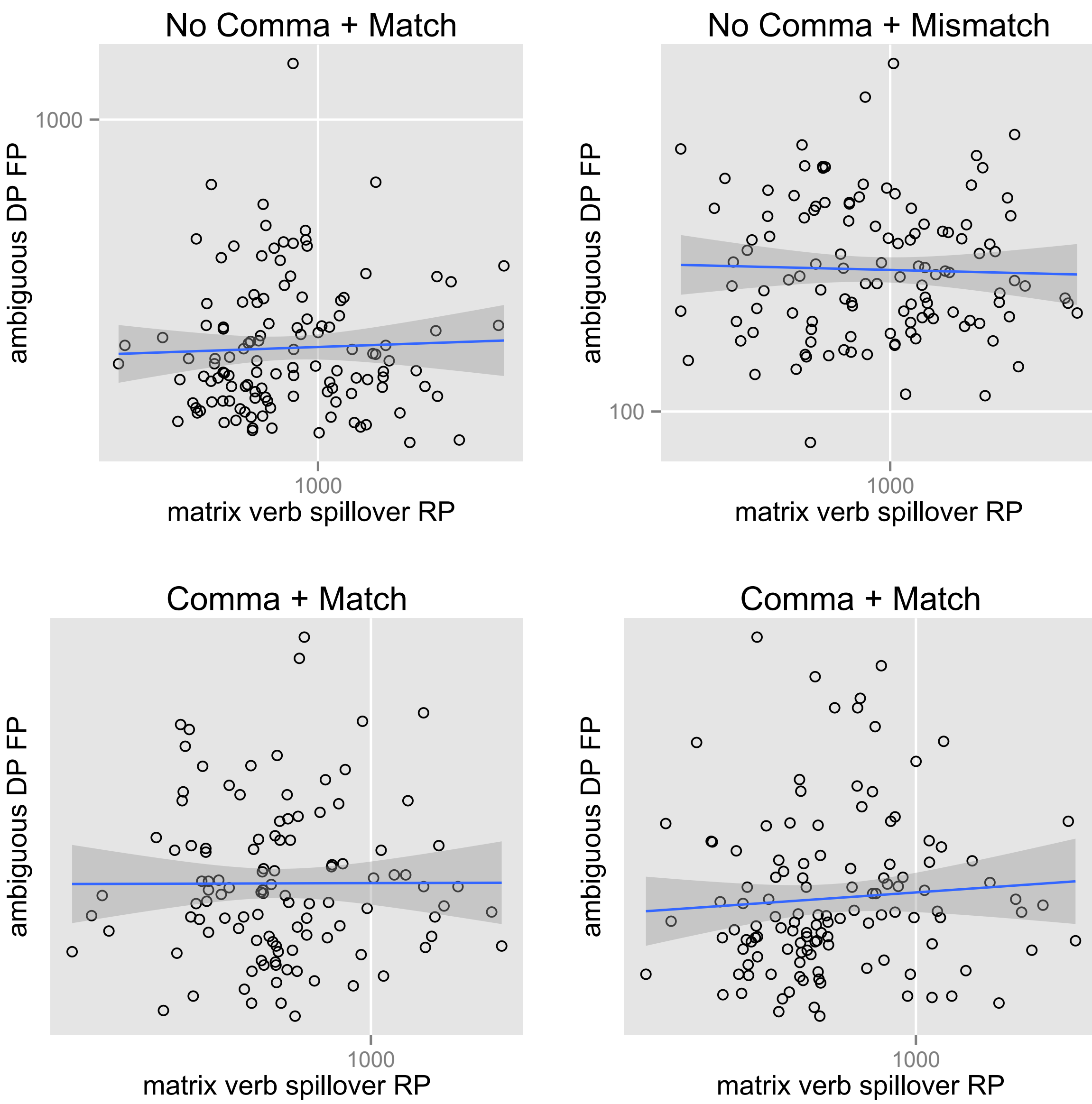


Figure 3.14. Correlation between RT at the ambiguous NP critical and matrix verb spillover regions. Shaded areas represent 95% confidence region.

by participants and items, where an item is the set of four minimally different sentences as created by the manipulation of Gender and Punctuation, which was the maximal random effects structure that would allow the model to converge.[12] The two fixed effects were contrast coded, in which each level was assigned either the value 0.5 or -0.5. Based on this model, no significant results obtained ($\beta = 0.075$, S.E. = 0.28, $\chi^2 = 0.07$). Additionally, a Pearson's product-moment correlation tested the linear correlation between first pass reading times at the ambiguous NP critical region and the matrix subject spillover region. Again, no significant results obtained (t = -0.1, correlation = -0.005). This indicates that it is *not* the case that the parser is choosing the matrix subject or embedded object parse on a by-trial basis, but rather the parser is experiencing the slowdown at both the ambiguous NP and at the matrix verb independently. In other words, the parser is not responding proportionally to the likelihoods of the possible verb frames.

The other possibility is that the initial parse is one in which there is an object DP, and despite an early preference for a structure without an object DP, the influence of the first parse lingers. This will be discussed further in the section below.

3.9. General Discussion

In this study, I examined the parser's behavior while reading locally ambiguous sentences during the search for a coreference antecedent in order to determine how the parser prioritizes concurrently competing pressures. In the first pilot experiments, I found evidence that Binding Condition C has an effect on offline forced coreference judgments, with illicit coreference rated more poorly than licit coreference, even when the gender of the two referents was congruent. Interestingly, the dispreference for mismatched gender coreference is stronger than the dispreference for grammatically illicit coreference. This may be due to the forced nature of the task, since participants were explicitly asked about

[12]Model: $lmer(matrixspilloverRP \sim ambiguousnameFP + (1|subj) + (1|item), data = data)$

coreference. One may speculate that incongruent genders have a non-linguistic basis for the dispreference (in addition to a linguistic gender agreement requirement), whereas grammatical judgments rely on linguistic knowledge alone. It is possible that participants considered the possibility of "speech errors" when evaluating the grammatically illicit sentences, which could induce attempt at reconstructing a sentence where the coreference is grammatical. Therefore, I do not take the higher acceptability of grammatically illicit coreference to indicate that Binding Conditions are not operative in this offline experiment, but rather that the nature of this task might allow participants to identify and focus on the two relevant words (and the genders of those words) without necessarily considering the grammatical context of those words.

Next, I examined readers' preference of global syntactic resolution through a forced-choice task. The two types of forced-choice tasks were designed to provide a choice between two grammatical resolutions to the sentence, either wholly written out in a minimal pair, or as a pair of one- or two-word choices listed below an underscored blank within a frame sentence. In both cases, I found that participants strongly preferred the parse in which the ambiguous DP was analyzed as the embedded object of the equibiased verb, in the absence of a comma. This reflects offline acceptability judgments, from which one may conclude that the locally coherent parse is ultimately "winning out" over the parse that favors early cataphoric dependency completion. Conversely, in the presence of a comma, there is a strong preference for parses that place the ambiguous DP in the matrix subject position. Given that commas are a somewhat reliable indicator of syntactic constituency (e.g., Baldwin and Coady, 1978; Niikuni and Muramoto, 2014), this is unsurprising. Yet the strength of the offline preference for the locally coherent parse in the absence of a comma would suggest that local coherence is a high priority for the parser, giving further credence to our interpretation of the online results, in line with the conclusions of Phillips and Gibson (1997).

The results obtained in the online eye-tracking while reading task are contradictory if one assumes a serial model of the parser. That is, if the parser prioritizes Active Search for the cataphoric antecedent, one would expect its success to preclude any effects that would appear from a failed parse. If the parser prioritizes building a minimally locally coherent structure, one would not expect the parser would check the ambiguous NP for coreference, since it would be a grammatically illicit dependency. However, there are apparent affects from both pressures. This cannot be explained by a serial-type parsing process, but is consistent with the parallel-type parsers (Gibson and Pearlmutter, 2000).

Examining the results more closely, I found a GMME at the R-expression in the ambiguous DP, but no interaction of Gender and Punctuation, which suggests both Comma and No Comma conditions are demonstrating this effect in the same direction. The parser is known to only attempt to form a coreference dependency in syntactic positions that would form grammatically licit dependencies (Kazanina et al., 2007). Thus I conclude that the parser is building a structure in which the ambiguous NP is in a location that is not c-commanded by the cataphoric pronoun.

However, I also found a reading time slowdown at and following the matrix verb in several measures, which suggests that the parser interpreted the ambiguous DP as the embedded object of the preposed adverbial clause, thus building a locally coherent structure. In building a locally coherent structure, the ambiguous DP would be in the position of the embedded object, thus c-commanded by the cataphoric pronoun and unable to form a licit cataphoric dependency.

The stability and persistence of the surprisal effect in later measures at the matrix verb is inconsistent with a parser that deterministically built a structure in which the ambiguous DP was definitively the matrix subject. Yet both effects are present. An explanation that is consistent with the parser constructing two possible syntactic structures, which show effects that emerge at different points in the time-course of processing — and that are consistent with two mutually incompatible parses — is what has been described as lingering garden path effects (Christianson et al., 2001). Christianson et al. (2001) found that the thematic role assignment within certain garden path sentences often appeared to be consistent with the parse that was globally coherent and the parse that was locally coherent (but globally incoherent):

(18) While the mother washed the baby cried in the other room.

Globally coherent: The baby cried in the other room while the mother washed herself.

Locally coherent: The mother washed the baby ...

Resulting interpretation: The mother washed the baby and the baby cried in the other room.

The attempt to repair the original, incorrect parse is incomplete, and the parser creates a so-called "Good Enough" parse of the sentence. In trying to balance accuracy and speed, it sacrifices some of the repair's accuracy for a reduced processing cost from revision. One explanation is that the parser builds a syntactic structure, then realizes the structure will be unable to account for the global structure. At this point, it attempts to revise the previous structure — either by building a parallel repair or by replacing the original structure with a new one. In either case, it seems that the original structure persists and influenced semantic (and structural) interpretation. Slattery et al. (2013) suggest that, of these two possibilities, it is likely that the parser attempts to overwrite the previously built structure but fails to "prune" the globally incoherent parse from its revised structure. Thus, processing effects (semantic or syntactic) from the original parse may still be present after the repair has occurred.

If one were to use this approach for the current study, the parser would build a structure that is initially locally coherent, placing the ambiguous DP in an embedded object position before retrieving the lexical frequency properties of the verb. When the parser retrieves this property and finds that the verb is ambiguously transitive, it builds an additional prioritized parse that places the ambiguous DP in another location, presumably as the matrix subject. Now, with the ambiguous DP in a location that is accessible to the cataphoric antecedent search (ignoring for the moment that it is also in an inaccessible position at the same time), the parser may attempt to form the cataphoric dependency. Simultaneously, the deprioritized locally coherent parse is still exerting its influence on the parser, so despite moving the ambiguous DP to a search-accessible location, the parser still expects to find another DP to serve as the matrix subject of the sentence for the construction of the deprioritized embedded object parse. This expectation manifests as a surprise effect at the matrix verb.

This is problematic because the effects observed from the locally attached (embedded object) parse only show up in later measures. This suggests that the parser is initially affected by the ultimately correct (matrix subject) parse; the effects of which must decay before the influence of the original parse can be detected. In more casual terms, the parser has moved on to complete the task of dependency formation and has built a structure to accommodate that pressure, but the previous task of building a rich argument frame for the ambiguous verb hasn't been abandoned yet[13].

This dichotomous set of effects is intriguing. If the parser fails to prune all discarded parses (or if it builds ungrammatical trees in order to accommodate multiple interpretations from garden paths), and if these discarded parses can linger even after finding strong evidence for a globally coherent parse quite early in the sentence, one might ask what the limit is on the influence of discarded parses. Or, from another perspective, how strong is the influence of Active Search? Since Active Search is ostensibly the pressure that suppresses the parser's bias toward building a locally coherent string, how strong a bias toward local coherence could Active Search overcome? This is addressed in the following chapter, but is relevant to this discussion as well.

The relative strength of the structure building pressures on the parser is not something that has been measured in detail. While working memory and lexical properties evidently affect structure building, it remains uncertain how the parser responds to competition between these pressures. Though there are cognitive limitations on the parser, and these lead to a so-called speed-accuracy trade-off, there is not yet a formulation for how the parser works within its limitations. Rather than asking what happens at the edge of working memory capacity, I ask how the parser behaves within a typical, optimal range of complexity. Simple ambiguity should not tax the parser to its limits, given the complexity

[13]Given the evidence that the parser eagerly and even *hyper*-actively constructs dependencies, this account has more than a few holes. This is discussed in more detail in Section 4.4.3

the parser is able to accommodate. Though I have not explicitly calculated a measure of complexity for our sentences, there is evidence that a single embedded clause does not typically push the parser over its limits (Gibson, 1998). And since participants did not report struggling with interpreting these sentences (unlike canonical garden path sentences, from which recovery is almost impossible), I believe that complexity is not a crucial factor in determining the parser's behavior at the point of ambiguity.

But what if it were? If one instead used sentences with highly transitive verbs (though still with some ambiguity), would the bias for local coherence override the cataphoric antecedent search? Or, in an even more extreme scenario, would the intransitive reading of the verb become inaccessible, and would repairs become too costly to compute online? If one can balance the pressures on the parser to keep processing costs from ambiguity and structure building biases within the cognitive limitations of the parser, one might be able to test the strength of the parser's priorities relative to one another. Specifically, given the pressures from local coherence and Active Search, pushing the parser to the edge of its ability to accommodate different influences, while staying within its strict cognitive limitations would give us insight into how the parser chooses the rankings of its priorities. This is the motivation for the experiment discussed in Chapter 4. In turn, this gives us more information about the parallel nature of the parser, allowing us to formulate a more specific framework for future work. This will be discussed in detail in Chapter 5.

3.10. Conclusion

In this chapter, the parser's behavior is observed as it encounters a locally ambiguous string while balancing pressures from constructing a minimally locally coherent structure, and the eager search for the antecedent to a cataphoric coreference dependency. The parser seems to make an initial prediction for a locally coherent string that is temporarily overridden by the Active Search mechanism. In doing so, it is observed that the parser attempts to complete the cataphoric dependency at the location of the structurally ambiguous DP, putting itself at risk for violating Binding Condition C. However, the parser succeeds in forming this coreference dependency and finds confirmation of the structure in which the ambiguous DP is not c-commanded by the pronoun when it immediately encounters the matrix verb. Yet, the locally coherent parse persists despite evidence against it. The parser apparently reconsiders its original predictions, which manifests in increased reading times at the matrix verb: something only predicted if the parser expected a DP distinct from the ambiguous DP to serve as the matrix subject. In order to more explicitly describe the process the parser undergoes in order to manifest effects of processing from two competing parses, I describe a multiple output model of the parser that accounts for our results.

In the present descriptions of multiple output models of the parser, I must still account for variations in the strength of the initial locally coherent parse. While I can currently account for differences in repairability between sentences whose complexity stems from high frequency lexical properties, new environments must be used to test this hypothesis. For instance, will the processing cost become too high for the parser to accommodate the pressures from an Active Search if the strength of local attachment is increased? Or, instead, will the Active Search be prioritized over the biases from verb frame frequency? Given likelihoods of (probabilistic) outcomes, can one calculate variation in the parser's behavior from dependency type and lexical statistics? Chapter 4 will address the relative strengths of local attachment and Active Search in environments that are less favorable to the pressures to complete a cataphoric dependency quickly.

CHAPTER 4

Influence of transitive-biased verbs on cataphoric dependency formation

4.1. Introduction

Based on the results observed in Chapter 3, I make the claim that the cataphoric dependency search influences the parser (though I do not say to what degree) and this causes the parser to represent both the Matrix Subject and the Embedded Object parses simultaneously. However, the verbs used for this experiment were equibiased — equally frequently observed as monotransitive and intransitive in the Gahl et al. (2004) corpus. Because of this frequency, I am unable to make a claim about the strength of the influence from the cataphoric dependency search, other than to say that there is some influence. Thus, the experiment described in the present chapter aims to identify the degree to which the cataphoric dependency search influences the parser by putting it in competition with a strong monotransitive[1] bias.

There is a long history of examining the influence of a (mono)transitive bias on the parser's online behavior. Much of this literature concludes that the frequency with which a verb is used in a particular form influences the likelihood that the parser will represent that form in its initial structure building stages (Crain and Steedman, 1985; MacDonald

[1]Although the literature alternatively calls the transitive argument structures or frames that include one subject and one direct object position the *simple transitive* form, henceforth, this frame will be called the *monotransitive* form.

et al., 1994; Tanenhaus et al., 1989, a.o.). That is, verbs that are more likely to be used in a transitive frame are more likely to be parsed as transitive, whereas verbs that are more likely to be used in an intransitive frame are more likely to be parsed as intransitive. It has been suggested that this bias is a source for the (conscious and unconscious) garden path effects observed in temporarily ambiguous sentences in which the more likely or more frequent alternative is not the ultimately "correct" form. However, there is evidence that the representation of verbal argument structure is a much more complex process. In the lexicalist literature, it is especially common to describe verbs with multiple argument structures or frames having more than one representation activated during processing, even if the alternative is unlikely (MacDonald et al., 1994; McRae and Matsuki, 2013; Trueswell and Kim, 1998). Just as the retrieval of lexical items might activate orthographic neighbors or infrequent lemmas that share a form (Luce and Pisoni, 1998; van Gompel et al., 2012), the retrieval and representation of verbs might activate multiple structures or other forms that share all or part of the input's form. In this scenario, even an infrequent verb frame might be activated or represented during processing, thus the transitivity bias of the verb may not have as strong influence on what representations are calculated. In either case (whether or not the transitivity bias of the verb influences what representations are available to the parser), the interaction of the bias provided by the cataphoric dependency search and the bias provided by the verbal transitivity information will be informative to how the parser prioritizes these two sources of potentially disambiguating information during parsing.

4.2. Background

4.2.1. Transitive verbs

Much of the classic literature focuses on the influence of transitivity on disambiguation. Tanenhaus et al. (1989) talks about how transitivity is a preference during parsing, so that the other options (i.e., intransitive forms) are overridden in a way. That is, gaps are posited after verbs that are potentially not transitive, thus these verbs have a transitive bias. Boland et al. (1989) concludes that a verb's possible argument structures are made available as soon as the verb has been recognized, which is also supported by evidence in Boland et al. (1995) that more detailed information about the verb frame is available to the parser during ambiguity resolution and long-distance wh-dependency formation. As such, it seems that the dependency formation processes used in these papers to test where gaps might be posited are either actively predicting a transitive interpretation of these verbs since the parser's search for a gap is triggered before reaching (thus before recognizing) the verb, or the parser prefers to use a transitive representation of these words, with at least one internal argument (a direct object, and optionally an indirect object). This might also be supported by free response experiments that show more transitive completions to sentences (van Gompel et al., 2012), but it's certainly not a simple phenomenon. Trueswell and Tanenhaus (1994) demonstrates that a number of different things can influence the apparent transitivity preference of a verb.

During online parsing of filler-gap dependencies, Tanenhaus et al. (1989) assert that there are three types of lexical knowledge that are readily accessible and used in forming syntactic representations: verb transitivity preference, verb argument structure (in terms of plausibility/thematic role), and verb control. Of these, the first type of lexical property is most relevant to the current study. To test the verb transitivity preference, they examined these sentences using an "embedded anomaly technique" in which a critical word (e.g., horse or rock in (1-c) and (1-d)) was either plausible or implausible. Reading times for the whole sentence were recorded as a participant was asked to decide whether the sentence made sense or not.

(1) a. Early gap + Transitive preference:
The district attorney found out which [witness/church] the reporter asked __ about the meeting.

b. Late gap + Transitive preference:
The district attorney found out which [witness/church] the reporter asked anxiously about __.

c. Early gap + Intransitive preference:
The sheriff wasn't sure which [horse/rock] the cowboy raced __ down the hill.

d. Late gap + Intransitive preference:
The sheriff wasn't sure which [horse/rock] the cowboy raced desperately past __.

This was followed by a self-paced reading task of the same sentences. The results show a clear reading time slowdown at the verb in the implausible conditions as compared to the plausible conditions, but only for the transitive verbs. The distinction between transitive and intransitive verbs is taken to indicate that gaps are not posited after verbs that are more likely to be intransitive and are posited after verbs that are likely to be transitive.

If the parser is acutely sensitive to the transitivity preference of a verb during structure building as is suggested by the conclusions drawn by Tanenhaus and colleagues, then we may be able to observe a different pattern of results than those reported in Chapter 3 by making use of this sensitivity.

This type of result indicative of a parser that is sensitive to verb preference is replicated numerous times in the classic literature. In Boland et al. (1995), they conducted several experiments that examined how the syntactic and semantic (i.e., thematic) features of a verb and its potential arguments influenced the filler-gap dependency formation. For instance, they used a "stop making sense" task to determine at what point a sentence such as ((2-a)/(2-b), (3-a)/(3-b)) is deemed implausible.

(2) a. Which film did the salesman show at the medical convention?

b. Which audience did the salesman show at the medical convention?

c. Which audience did the salesman show [the film to __] at the medical convention?

In sentences such as these, the first argument position following the verb would be an implausible position for the tail of the wh-dependency for (2-b) but not (2-a). This is because it is implausible for a salesman to show an audience, but not to show a film. However, in sentences like (2), there is another potential position for the tail of the wh-dependency, illustrated in (2-c). Sentences like (2-c) were not part of the experiment, but rather (2-c) is a hypothetical version in which there is no global implausibility. Boland and colleagues found that in sentences with this indirect object position, the implausibility effect from attempting to associate the wh-phrase with the direct object position was diminished as compared to sentences with no indirect object position. They state that this is because the parser has access to the verb's argument structure and thematic roles immediately upon encountering the verb, thus the parser is not under much pressure to associate the implausible position with the wh-phrase (since there is an alternative). However, this is in stark contrast to the sentences like (3), in which there is no indirect object position associated with the lexical properties of the verb, thus the implausibility of (3-b) is observed as a stronger effect at this position.

(3) a. Which structure did the biologist see through the powerful microscope?

b. Which afternoon did the biologist see through the powerful microscope?

In sentences like (3), Boland and colleagues observed that participants say the sentences "stops making sense" after reaching the position where filler would make an implausible object. This indicates that they participants are aware of or influenced by the implausibility of the tail of wh-dependency landing at the position after this verb. More importantly, the difference between the results for sentences like (2) and (3) suggests that readers quickly have access to the lexical properties of the verb that inform them of the potential argument structure (and thematic roles, which are of less relevance to the current study). From this, we may say that readers will be sensitive to the likelihood of a verb appearing in a mono-transitive form (or not).

The likelihood of a verb's frame (whether defined by argument structure, thematic roles, or other properties) might be stored in a similar way to other lexically specific features. This is what Trueswell examines in his 1996 paper. In particular, he looks at the relative frequency of use of the participle form of transitive verbs in conjunction with relative clauses. For example, sentences like (4) and (5) were presented for subjects to read in a self-paced reading paradigm.

(4) a. HIGH PART + RRC:

The award accepted by the man was very impressive.

b. HIGH PART:

The award that was accepted by the man was very impressive.

(5) a. LOW PART + RRC:

The room searched by the police contained the missing weapon.

b. LOW PART:

The room that was searched by the police contained the missing weapon.

He found that readers struggled more with sentences containing a "LOW PART" verb, whether or not the context of the sentence semantically supported the relative clause interpretation. That is, in his first experiment, the relative clause interpretation was semantically supported, and he found no slowdown in sentences like (4-a) as compared to (5-a), despite the local ambiguity. In contrast, his second experiment had sentences in which the relative clause interpretation was not semantically supported (6).

(6) a. LOW PART + RRC:

The thief searched by the police had the missing weapon.

b. LOW PART:

The thief who was searched by the police had the missing weapon.

While the sentences were eventually interpreted by the readers, Trueswell found that LOW PART sentences caused the readers more trouble than HIGH PART sentences, presumably because of the increased difficulty induced by the infrequent participle structure. This observation suggests that frequency of verb frames do influence the parser's behavior during ambiguity resolution, with more likely frames reducing or eliminating difficulty during reading locally ambiguous regions.

4.2.2. Intransitive verbs

Moreover, more recent work on intransitive-biased verbs suggests that there is important information about the parser available through investigating other frames. Staub (2007) shows that the parser does not ignore intransitivity during online parsing — that is, the parser does not posit gaps after verbs for which a gap would definitely be ungrammatical (e.g., unaccusative verbs like *arrive*). However, one thing he notes is that verbs like *laugh* (which are unergative) do seem to show some amount of transitive-like effects. He suggests that this is because they, in fact, can act as transitive verbs in very rare and specific circumstances. Thus, it seems like intransitivity is a marked condition for which strict conditions are needed constrain the parser's behavior. This is further supported by the results described in van Gompel et al. (2012), in which they assert that mono-transitivity is a category-general property of all verbs, unless specifically marked in their feature set. This is shown by their priming experiments, in which verbs with transitive biases were less influenced by certain primes than verbs with intransitive biases. This basically suggests that verbs that are intransitive have some extra marking, in addition to the (usual) ability to also be monotransitive, so mono-transitivity is somewhat of a default from which other forms diverge. Basically, in the end, this means that transitivity is the baseline, and thus might be "easier" or "more typical". (On the other hand, it might mean that verbs that have any amount of "marking" for being intransitive may in fact have an accessible intransitive representation).

In Staub (2007), he examines whether the parser always posits a gap following a verb, or whether the parser is sensitive not just to the transitivity of a verb, but its intransitivity. To do so, he compares three types of verbs: unaccusative, unergative, and mono-transitive. Mono-transitive, or simple transitive, sentences are used as a baseline for which a gap is assumed to be posited after the verb. The two types of intransitives, therefore, are the experimental conditions. The distinction between unergative and unaccusative verbs is an important component of this study, and is relevant to the current discussion of transitivity.

Unaccusative verbs are intransitive verbs that cannot assign accusative case to their Theme argument which thus has to raise into the subject position, where it is assigned nominative case. This is relevant because the object position of the verb cannot be occupied by another argument, since this is the original location of the now subject. Because the object position is occupied by the trace (or other remnant) of the theme and the subject position is occupied by the overt representation of the theme, all of the argument positions are filled. This makes them unavailable as potential landing sites for the tail of a long-distance dependency.

In contrast, and crucially, unergative verbs have a different pattern of case assignment. Unergative verbs assign the thematic role of Agent to their argument, and in English, this corresponds to the subject position (disregarding where the agent is base generated). Thus, there is still potential (albeit rare) for the verb to assign accusative case to an object, and is much more commonly treated as an intransitive verb. While this type of unergative "transitive" form is uncommon (Staub, 2007, and references therein), it is attested in such forms as (7-a) cognate objects, (7-b) "[X]'s way", and (7-c) resultative objects.

(7) a. They slept a dreamless sleep.
 b. She laughed her way out of the auditorium.
 c. He sneezed the paper off the table.

In these forms, it seems that the unergative verbs may have a few properties common to monotransitives. Despite their canonically intransitive subcategorization, the parser may be aware of these atypical constructions, and therefore have access to the apparently transitive structures that would allow for a gap after the unergative verb, as in (8).

(8) a. [What kind of sleep] had they slept __?
 b. ?[Where] did she laugh __ out of the auditorium?
 c. [Which paper] did he sneeze __ off the table?

In this case, it would not be in violation of the verb subcategorization for the parser to posit a gap after the verb. Thus, in order to determine whether the parser is sensitive to the (in)transitivity of the verb, both unaccusative and unergative verbs must be examined in comparison to (mono)transitive verbs.

In Staub's first experiment, the readers are observed to experience more difficulty via reading time slowdowns after intransitive ((9-a), unaccusative and (9-b), unergative) verbs than after transitive verbs in (9-c) when what follows is the matrix clause (i.e., the verbs in question are in the subordinate, or embedded, clause).

(9) a. When the dog arrived the vet and his new assistant took off the muzzle.
 b. When the dog struggled the vet and his new assistant took off the muzzle.
 c. When the dog scratched the vet and his new assistant took off the muzzle.

This suggests that the parser is sensitive to the intransitivity preference or obligation of the embedded verb in these sentences. However, a more complex picture emerges when Staub compares his results to the previous work of van Gompel et al. (2001). In sum, Staub suggests that the difficulty observed after the embedded clause (i.e., the subject of the matrix clause) may not be due entirely to the parser initially placing the matrix subject in the embedded object position. To test whether the difficulty observed at the matrix subject in both intransitive conditions was due to the parser positing an object position after the verb, Staub compares two types of transitive sentences (deemed (10), NP-preference and (11), PP-preference) to (12) unaccusative verbs. For each verb type, sentences containing a gap shortly after the critical verb (a.) were compared to sentences with no gap (b.). Thus, the reading times measured at (and following) the verb may indicate whether or not the parser has posited an object position.

(10) a. NP-preference gap:
 The gadget that the manager called occasionally about after the accident still didn't work.

b. NP-preference no gap:

The manager called occasionally about the gadget after the accident.

(11) a. PP-preference gap:

The truck that the pilot landed carefully behind in the fog shouldn't have been on the runway.

b. PP-preference no gap:

The pilot landed carefully behind the truck in the fog.

(12) a. Unaccusative gap:

The party that the student arrived promptly for at the fraternity house was late in getting started.

b. Unaccusative no gap:

The student arrived promptly for the party at the fraternity house.

Staub's results suggest that, in fact, gaps were only posited after the transitive verbs, and not after the unaccusative verbs. He concludes that the parser does not disregard obligatory subcategorization restrictions on unaccusative verbs. Since unergative verbs were not examined in this experiment, however, it is impossible to say whether or not the parser would show similar effects to those of the unaccusative condition. Thus it seems that one cannot rule out the parser positing a gap after an unergative verb because of the (remote) possibility of one of the apparently transitive structures depicted in (7), above.

Staub et al. (2006) present evidence that the parser may construct an intransitive reading when presented with the option, given an alternative involving Heavy NP Shift, but it is not a straightforward conclusion to draw that the parser "prefers" the optional intransitive construction over the transitive one. Especially in the face of works concluding the parser will construct an optionally transitive representation given the opportunity (Juliano and Tanenhaus, 1994; Trueswell et al., 1993) and the body of literature suggesting the parser can construct parallel representations (Boland and Cutler, 1996; Pearlmutter and Mendelsohn, 1998, 9, including the current results from Chapter 3), one must be wary of claiming that difficulty observed in one position excludes the possibility that other constructions are considered.

From these studies, one may gather that the parser is sensitive to obligatory intransitivity, but that this is not a straightforward property to test or observe. Moreover, determining whether the parser has a bias, let alone what that bias might favor is a difficult task. Historically, however, there some consensus that the (simple) transitive structure of verbs is somehow more of a default than other possibilities. Indeed, more recent work by van Gompel et al. (2012) suggests that intransitivity is a marked feature, and that simple transitive structures (i.e., mono-transitivity) is "category general" property of the verb label. This suggestion comes from a set of priming experiments in which free responses are tallied by completion type following different types of priming conditions.

In this study, intransitives showed a stronger priming tendency than monotransitives when the specific lexical item was repeated in the free response prompt, while monotransitives only showed priming relative to a no-priming baseline. This suggests that intransitives and monotransitives are represented differently in the lexicon, with the intransitive property encoded as a lexically specific feature. To illustrate how this might be represented in the lexicon, van Gompel makes use of the schematization of lexical representation in Pickering and Branigan (1998), adapted below. In this diagram, all words connected to the VERB category label will activate a monotransitive frame. In contrast, only verbs that are specifically encoded as having an intransitive frame will activate that structure (although this does not seem to allow for verbs with only intransitive forms to be connected to the category of VERB).

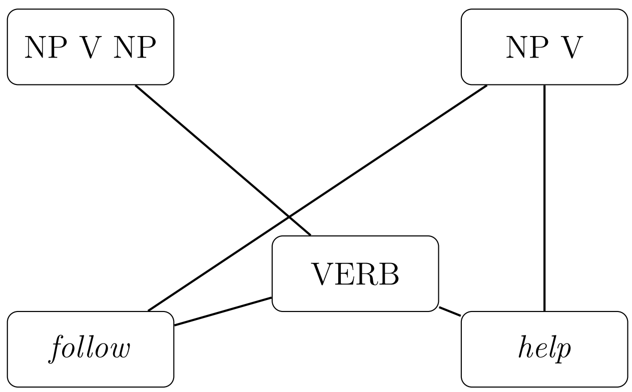


Figure 4.1. Diagram adapted from van Gompel et al. (2012): Category-general representation of monotransitive structures and lexically specific representation of intransitive structures.

Thus, by comparing the relative influence of monotransitive- and intransitive- primes on the types of free-responses given for sentences like (13-c), the repetition of the specific verb used (either shooting or guarding) could strengthen the priming effect. Since the repetition of the specific verb in the prime and prompt only showed a strengthening effect for intransitive primes (13-a), and not for monotransitive primes (13-b), they concluded that the intransitive structure was a more specific, marked property than the monotransitive structure.

(13) a. The marksman and the assassin were shooting/guarding.

b. The marksman was shooting/guarding the assassin.

c. When the gamekeeper was shooting/guarding ...

The relevant aspect of their study is that intransitives are somehow distinct from or a more marked form than monotransitives. They conclude that monotransitive representations are category general (i.e., not attached to the specific verbs but to the lexical category of VERB), and intransitive structures are specific to lexical items. This is also supported, they claim, by language acquisition studies, corpus studies, and their own previous work (van Gompel et al., 2012, and references therein). The point of this is that they suggest that intransitive verbs may easily made to be monotransitive but monotransitive verbs are difficult to make intransitive because it is somehow a marked (or less common) form.

4.2.3. Interim Summary

So far, verb frame preferences do not seem to have a clear-cut influence on the behavior of the parser. Some studies suggest that verbs are treated as transitive when no constraints would be violated in doing so (e.g., van Gompel et al., 2012), while others observe that intransitive interpretations are preferred when that alternative is available (e.g., Staub et al., 2006). Given that the results of the main experiment in Chapter 3 suggest that both the transitive and intransitive verb structures are simultaneously used during structure building, these previous studies are not in conflict. However, this puts us in a unique position to test the relative strength of bias introduced by the transitive verb frame frequency against long-distance dependency formation and compare it to the results obtained from equibiased verbs. In doing so, we will be able to gather evidence that speaks directly to the strength of the verb frame bias in initial structure building.

4.3. Pilot 4: Offline Forced Choice Task

In preparation for the eye-tracking experiment that serves as a follow-up to Chapter 3, this pilot experiment investigates the verb frame transitivity bias of the verbs selected for the adapted stimuli. In order to confirm that the verbs chosen are in fact biased toward the transitive verb frame, a pair of simple forced-choice experiments were carried out on MTurk.

4.3.1. Materials and Procedure

The pilot was conducted on Amazon Mechanical Turk (MTurk) recruiting 80 participants who were self-identified native speakers of English, whose IP addresses were restricted to the United States. Stimuli consisted of 180 items, 24 of which were target stimuli, 72 of which were specially designed fillers to distract from the target stimuli, and the remaining 84 were from four unrelated and non-conflicting, concurrent experiments. These filler experiments contained not manipulation of verbal ambiguity or ambiguity regarding coreference resolution.

This experiment had a 1 × 3 factorial design, with levels of Matched, Mismatched, and No Coreference, e.g., (14) adapted from the experiments in Chapter 3. The Matched condition (14-a) displayed a sentence in which the gender of the pronoun and first candidate antecedent were congruent, whereas the Mismatched condition (14-b) displayed a sentence with incongruent gender between these two positions. The No Coreference condition (14-c) displayed a sentence in which the position occupied by the pronoun in the Matched and Mismatched conditions was filled with a proper name of the same normative gender as the name in the position of the first candidate antecedent. The No Coreference condition serves as a baseline to which the other two conditions can be compared, as it does not contain a long-distance dependency search. Thus, the biases influencing the responses for the No Coreference conditions do not include the active search for the cataphoric antecedent, but it does still include the bias introduced from the transitive-biased verb frame.

(14)
a. Whenever she was trying to gently imitate Annie's dissonant melody [it was/was] hauntingly beautiful but Karen couldn't do it justice all alone.
b. Whenever he was trying to gently imitate Annie's dissonant melody [it was/was] hauntingly beautiful but Jimmy couldn't do it justice all alone.
c. Whenever Mia was trying to gently imitate Annie's dissonant melody [it was/was] hauntingly beautiful but Karen couldn't do it justice all alone.

Two types of tasks were included in this experiment, each with 40 unique participants. Stimuli for each task were presented in a continuous list, which participants could scroll down at their leisure, but each participant only saw one type of task. The first task was a forced choice between two full sentences, or Full Sentence Forced Choice (FS). In this task, the subject was asked to select the sentence that though sounded "better" or more natural.

(15) Example stimulus presentation for Full Sentence Forced Choice:

◯ Whenever he was trying to gently imitate Annie's dissonant melody was hauntingly beautiful but Jimmy couldn't do it justice all alone.

◯ Whenever he was trying to gently imitate Annie's dissonant melody it was hauntingly beautiful but Jimmy couldn't do it justice all alone.

The second type of task was a sentence completion task, or Force Choice Sentence Completion (SC). Participants were presented with each trial sentence in a list, but in the region where the local ambiguity was resolved, they were shown an underscored blank. Beneath this sentence frame, there were two options from which the participant could choose. One option was the matrix verb by itself, and the other option was the matrix verb preceded by *it*. In other words, the two options from which participants could choose were the matrix subject parse and the embedded object parse.

(16) Example stimulus presentation for Sentence Completion Forced Choice:

Match: Whenever she was trying to gently imitate Annie's dissonant melody ____ hauntingly beautiful but Karen couldn't do it justice all alone.

◯ it was ◯ was

Mismatch: Whenever he was trying to gently imitate Annie's dissonant melody ____ hauntingly beautiful but Jimmy couldn't do it justice all alone.

◯ it was ◯ was

No Coreference: Whenever Mia was trying to gently imitate Annie's dissonant melody ____ hauntingly beautiful but Karen couldn't do it justice all alone.

◯ it was ◯ was

These two tasks were designed for an unrelated, non-conflicting experiment in order to gauge bias between two forms. This is also a useful tool for the present study, in which the bias presented by the verb frame frequency information needs to be evaluated.

4.3.2. Predictions

If the ambiguous verbs are transitively biased as we intended, we should find that subjects are more likely than chance to select the embedded object interpretation. In the full sentence forced-choice task, this will manifest as a preference for the full sentence that has *it* in the position of the matrix subject (second option in (15)). In the forced-choice sentence completion task, this will manifest as a preference for the *it was* selection over the *was* selection (second option in (16)). If this pattern of results is observed, we may take these stimuli as representative of transitively biased verbs, thus fit to provide the necessary contrast to the equibiased verbs used in Chapter 3.

If, on the other hand, the verbs are not sufficiently biased toward the transitive structure, we should find that subjects are either selecting between the two types of responses at chance levels or consistently selecting the *was* response ((15)-first, (16)-first). This would indicate that the transitive option was not more natural or attractive to the readers, and thus is not able to supply the necessary bias toward the transitive frame necessary to contrast with the experiment in Chapter 3.

4.3.3. Analysis

In order to understand the pattern of results from the forced choice data, two types of analyses are necessary. The first analysis is a series of logistic linear mixed effects models in order to determine whether any of the conditions (Match, Mismatch, No Coreference) differ from each other. This analysis was conducted by comparinga maximal model to a depleted model with a term of interest removed, in which the term of interest was one of three pairs of conditions within each task type. A generalized linear mixed effect regression (GLMER) was used, and is available in the *lme4* package (Bates et al., 2015b). The random effects structure included random intercepts for subjects and items, which was the maximal random effects structure that would allow the model to converge.[2] The fixed effects were contrast coded, in which each level was assigned either the value 0.5 or -0.5. Since two comparisons were needed for these analyses, corresponding to two separate models, Bonferroni correction was used to modify the confidence intervals ($\alpha = 0.025$). The results of this analysis are reported in Tables 4.1 and 4.3, in which the estimate (β), standard error, chi-squared statistic, and p-value are listed for the three GLMER comparisons.

Table 4.1. GLMER results for Pilot Experiment (Section 4.3).

Effect	**Estimate**	**S.E.**	χ^2	**p-value**
Match vs Mismatch	-0.49	0.22	4.66	=0.062
Match vs No Coreference	0.31	0.23	1.72	>0.1
Mismatch vs No Coreference	0.41	0.20	4.12	=0.085

[2]Model: $glmer(response \sim MatchNoCoref + MismatchNoCoref + (1|subj) + (1|item), data = data, family = \text{“}binomial\text{”})$ and
$glmer(response \sim MatchMismatch + (1|subj) + (1|item), data = data, family = \text{“}binomial\text{”})$

Table 4.2. Within-Condition Analysis for Pilot Experiment (Section 4.3).

Condition	**Estimate (%)**	**95% C.I.**	χ^2 **(df)**	**p-value**
Match	22.45	18.4–27.0	115.14 (1)	<0.0001
Mismatch	26.63	22.3–31.4	82.73 (1)	<0.0001
No Coreference	21.67	17.7–26.2	121.82 (1)	<0.0001

Table 4.3. GLMER results for Pilot Experiment (Section 4.3).

Effect	**Estimate**	**S.E.**	χ^2	**p-value**
Match vs Mismatch	-0.19	0.23	0.66	>0.1
Match vs No Coreference	0.21	0.23	0.78	>0.1
Mismatch vs No Coreference	0.20	0.20	0.20	>0.1

Table 4.4. Within-Condition Analysis for Pilot Experiment (Section 4.3).

Condition	**Estimate (%)**	**95% C.I.**	χ^2 **(df)**	**p-value**
Match	29.66	25.2–34.6	62.25 (1)	<0.0001
Mismatch	31.33	26.8–36.3	52.65 (1)	<0.0001
No Coreference	28.72	24.3–33.6	68.52 (1)	<0.0001

The second analysis was designed to show whether there was a preference for one selection over the other within a condition, independent of how the three conditions compare to each other and computes a χ^2 value. For this analysis, a 1-sample proportions test was performed with the null hypothesis set at $p = 0.5$ (chance-level performance), so a significant result suggests that the participants were not selecting responses randomly. As before, this test is a test of equal or given proportions that provides a Pearson's chi-square test statistic. Tables 4.2 and 4.4 display the results of the within-condition analysis, calculated with a test of equal or given proportions (the same 1-sample proportions test

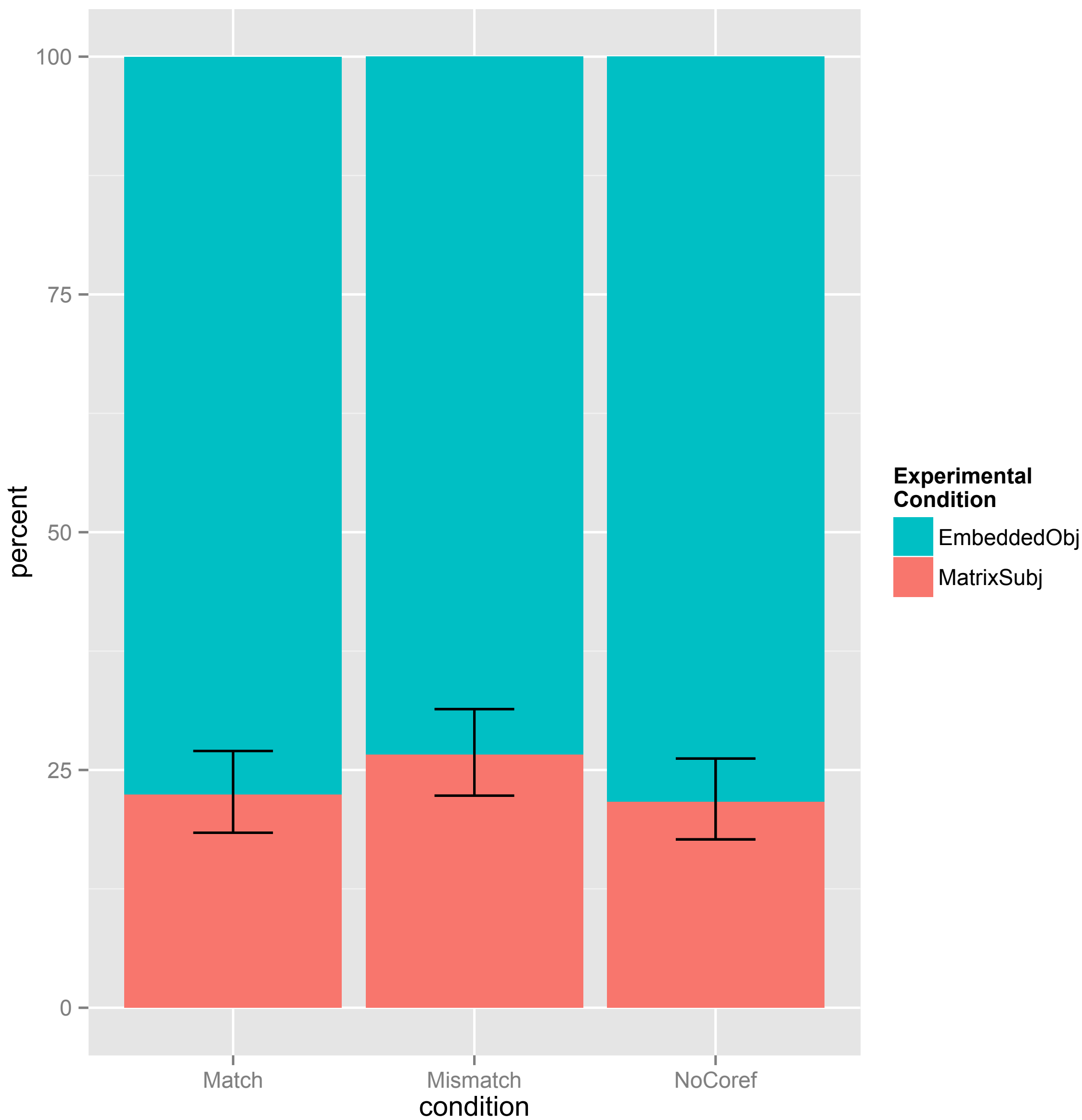


Figure 4.2. Proportion of Full-Sentence Forced Choice Responses in Pilot 4: Matrix Subject vs. Embedded Object. Error bars represent 96% Confidence Intervals.

performed in Section 3.5). This separate analysis is needed because of the binary nature of the data gathered from participants.

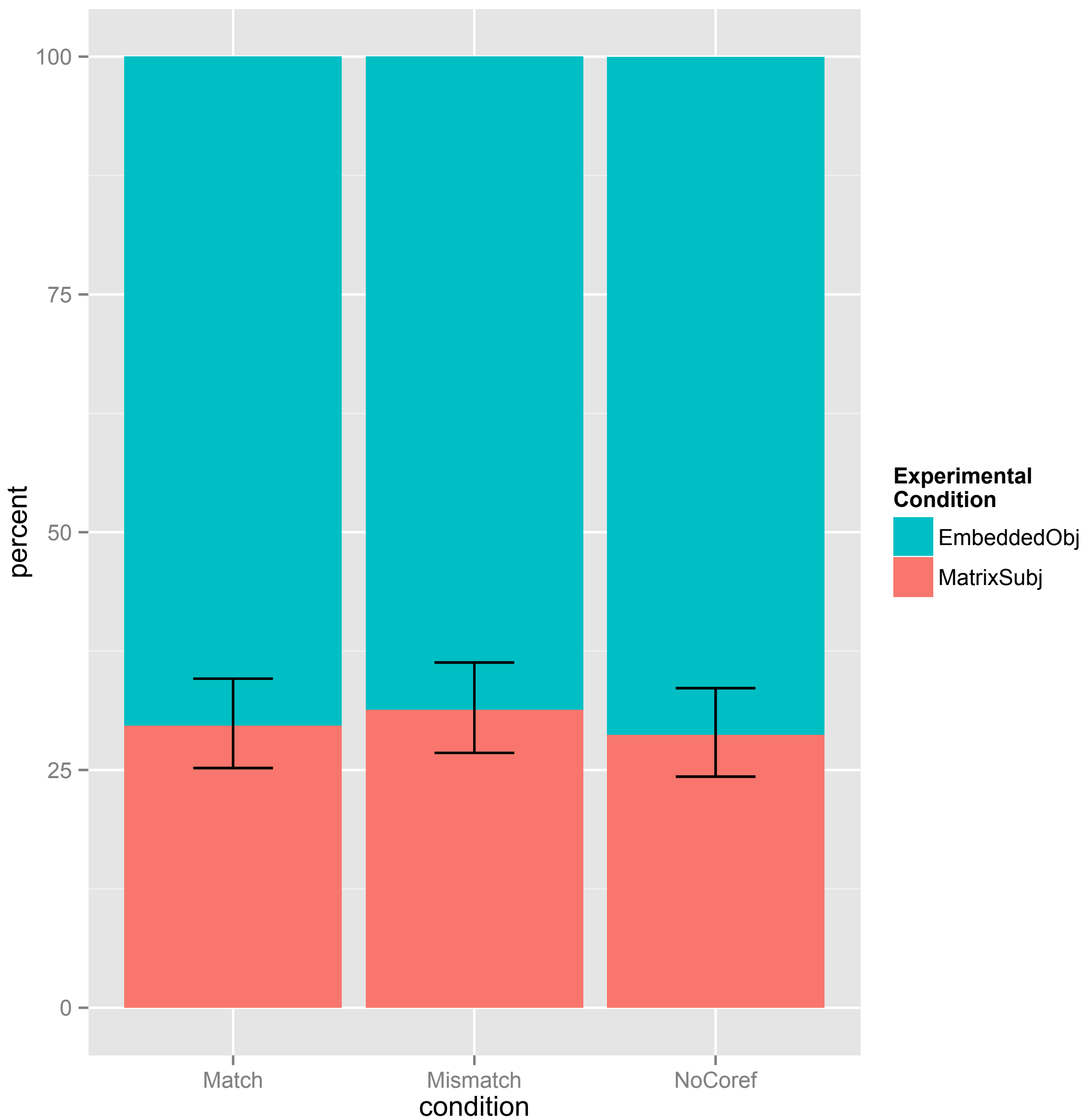


Figure 4.3. Proportion of Forced Choice Sentence Completion Responses in Pilot 4: Matrix Subject vs. Embedded Object. Error bars represent 96% Confidence Intervals.

4.3.4. Results

The logistic linear mixed effects regression produced no significant effects for either task. This suggests that the verbs bias for (or against) the transitive frame was consistent across conditions. There was, however, a main effect of condition before the application of the Bonferroni correction in the comparison between the Mismatched condition and the No Coreference condition in the FS task, with the Mismatched condition having fewer *it was* selections than the Matched condition (β = -0.49, S.E. = 0.22, χ^2 = 4.66, p = 0.031) or the No Coreference condition (β = -0.41, S.E. = 0.20, χ^2 = 4.12, p = 0.042). The Bonferroni correction makes the absence of a significant result somewhat conservative, but the evaluation of these data suggests no interpretation is statistically supported. Although it may behoove to entertain some speculation, I can think of no theoretically valid reason for there to be a consistently lower preference for the transitive verb frame when the gender of the pronoun and first candidate antecedent are incongruent as compared to when the genders are congruent or when there is no cataphoric dependency. Coupled with the lack of any such trend in the SC task, it seems unnecessary to hypothesize about the origin of this non-significant trend.

The 1-sample proportions analysis revealed that there was a significant preference for the embedded object parses across conditions in both tasks. In all comparisons, the p-values were calculated at below the p = 0.0001 level, with percentages of *it was* selections ranging from 73% to 78% in the FS task, as shown in Table 4.2, and from 68% to 71% in the SC task, as shown in Table 4.4. The nature of this analysis cannot describe how strong the preference for the Embedded Object selection (i.e., transitive verb frame) was, but it does confirm that the verbs used for these stimuli are transitively biased.

4.3.5. Discussion

Although the analyses performed for this pilot experiment cannot distinguish the degree of preference for the transitive parse, there is a distinct preference over the intransitive parse in all conditions and both tasks. This is crucial for the interpretation of the main experiment to follow. The transitivity bias supplied by the ambiguous verbs must be present in order to test the relative strength of a cataphoric dependency search against the search-less baseline. So despite the marginal effect of condition observed between the Mismatch and No Coreference conditions in the FS task it is possible to assert that these verbs are consistently treated as transitive, when readers are given the option to choose between transitive and intransitive alternatives. In other words, this pilot experiment confirms that, at least in offline tasks, the transitive parse (i.e., the Embedded Object parse) is preferred for these stimuli.

4.4. Eye-tracking task

Since the transitivity of the verbs has been established by the pilot experiment, the influence of these transitively biased verbs on the parser's disambiguation behavior may now be evaluated. As set out in Chapter 3, the potential for competition between the active search for the antecedent of a cataphoric dependency and whatever bias the parser experiences from the ambiguous verb may influence online reading behavior. In other words, this experiment will test whether the influence from the transitively biased verbs prevents the parser from building the Matrix Subject representation by increasing the likelihood that the ambiguous NP is in the embedded object position.

4.4.1. Materials and Procedure

This experiment was conducted in much the same way as described in Section 3.7.2. We recorded eye movements using an EyeLink 1000 tower with chin rest in the same configuration as before. The experiment was presented on the monitor using the software EyeLink 0.07m, in a 12-point font size for the experimental stimuli. Twenty four target stimuli were the same as used in the pilot experiment described in (15) in Section 4.3.1. Fillers consisted of 144 items from five unrelated and non-conflicting, concurrent experiments. None of the filler experiments manipulated verbal ambiguity or cataphoric coreference resolution. Participants were calibrated at the beginning of the experiment, and then periodically throughout the experiment as needed. Stimuli were presented after fixation on a black rectangle (in the location of the first character) was identified. Participants were instructed to press a button on a hand-held controller when they had read and understood the sentence, at which point they were presented with a comprehension question in approximately 25% of overall trials and 100% of target trials. Comprehension questions were answered with 'yes' or 'no', indicated by buttons on the hand-held controller. Data gathered during recording was saved and processed in the same manner as described in Chapter 3.

The design of the experiment was 1 $\times$ 3, with the three levels being Matched and Mismatched (referring to the gender congruency between a pronoun and the first candidate antecedent) and No Coreference, in which the pronoun was replaced with a proper name, thus no cataphoric dependency search would be triggered, as illustrated in (17). In these stimuli, the general structure of the sentences was taken directly from the stimuli used in Chapter 3, except when the semantic plausibility required slight modifications. The ambiguous verbs in these transitively biased stimuli were taken from the Gahl et al. (2004) corpus. From the same calculation of proportion of transitive to intransitive frames made for the eye-tracking experiment and described in Section 3.7.2, verbs with the highest possible bias (proportion of transitive frames less the proportion of intransitive frames) were selected. All verbs had at least one attestation of an intransitive frame in the corpus, and a calculated bias of at least +0.40. As before, bias was used rather than raw counts due to the highly variable number of total attestations in the corpus.

(17) Match: Whenever she was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

Misatch: Whenever he was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Jimmy couldn't do it justice all alone.

No Coreference: Whenever Mia was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

4.4.2. Predictions

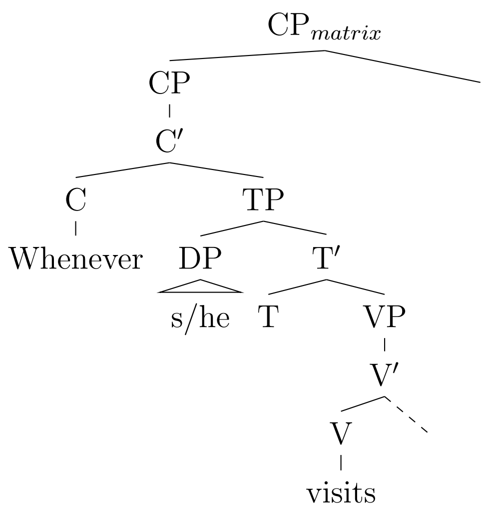
Since the stimuli in this experiment are adapted directly from the previous chapter, with the only relevant change being the verb frame frequency of the ambiguous verb, the step-by-step process which the parser may undergo is much the same as laid out in Chapter 3.3.

The primary difference between the previous experiment and the current one is the number of experimental conditions tested. In the present experiment, the three levels allow for comparison between congruent and incongruent gender for a candidate antecedent (Matched and Mismatched conditions, respectively), with the additional No Coreference condition, which allows the presence of an active search to be contrasted with the parser's behavior in the absence of one.

4.4.2.1. Hypothesis 1: Verb transitivity is strong. If verb frame frequency is a strong influence on the parser's initial structure-building behavior, we predict that the transitively biased stimuli in this experiment will elicit a different pattern of reading times than in the previous experiment in Chapter 3. In order to illustrate how this situation would play out, the following subsection steps through the parser's behavior for sentences like (18) in the case that the transitively biased nature of the ambiguous verbs strongly influences the process of structure building.

(18) Whenever s/he visits John's house is dark, but Joanna/Joe doesn't mind.

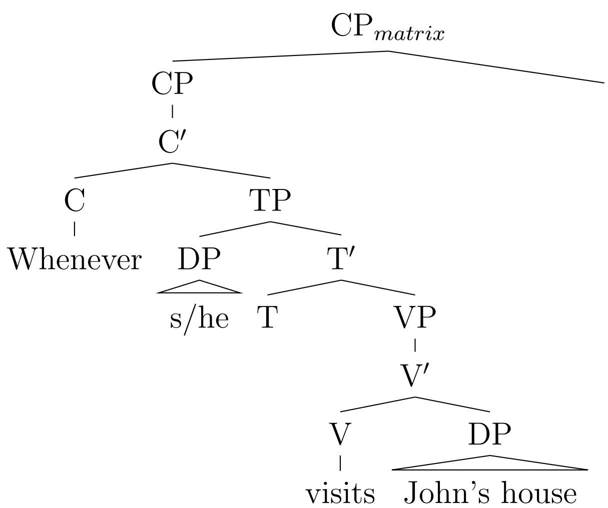
a. Whenever s/he visits



Since so much of the content of this sentence is the same as in the previous experiment, the parser will initially behave in the same way. *Whenever* will start off the structure building process. The parser may at this point recognize that *Whenever* indicates a CP that is dominated by the matrix CP that will eventually contain a matrix clause. As before, the pronoun *s/he* triggers an active search for an antecedent at the tail of the coreference dependency. Next, the parser encounters the transitively biased verb *visits*. It is at this point that the sentences used in this experiment differ from the previous ones and the two hypotheses described in this section diverge.

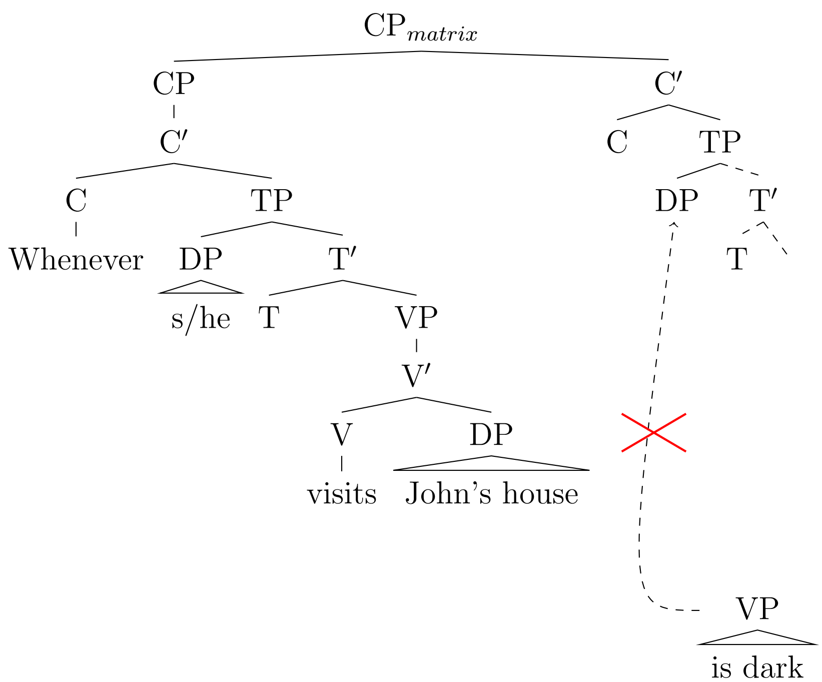
If the parser encounters the transitively biased verb and the verb frame frequency is both immediately accessible and strongly influential on the parser's behavior, then the parser will immediately prioritize the representation in which the ambiguous verb takes an object. It is not necessary for the parser to predict an object, i.e., build a position for the object, before a noun phrase that could serve as an object is encountered in the string. What is necessary is that the parser will prioritize the parse with the embedded object over alternatives. This can lead to the parser to be primarily influenced by the embedded object parse, especially if the additional evidence from the verb frame frequency is a strong influence on the parser.

(18) b. Whenever s/he visits John's house



If the parser is anticipating encountering an object and therefore places the DP *John's house* in the object position of the verb, the R-expression *John* will not be an accessible antecedent for the cataphor. Since in this scenario, the parser has committed to the most likely parse and this parse prohibits coreference between the cataphor and the ambiguous verb's object, there should be no difference in processing time between the Matched and Mismatched conditions. Gender congruency does not come into play in the active search process if the parser is not attempting coreference, which is why we do not predict a difference here. In the No Coreference condition, the active search process was never triggered, so the processing time at the ambiguous NP will be neither sped up by a gender match nor slowed down by a gender mismatch. Since in all three conditions in this scenario the parser is not attempting any coreference dependency formation at this location, we should not see any differences in reading time.

(18) c. Whenever s/he visits John's house is dark



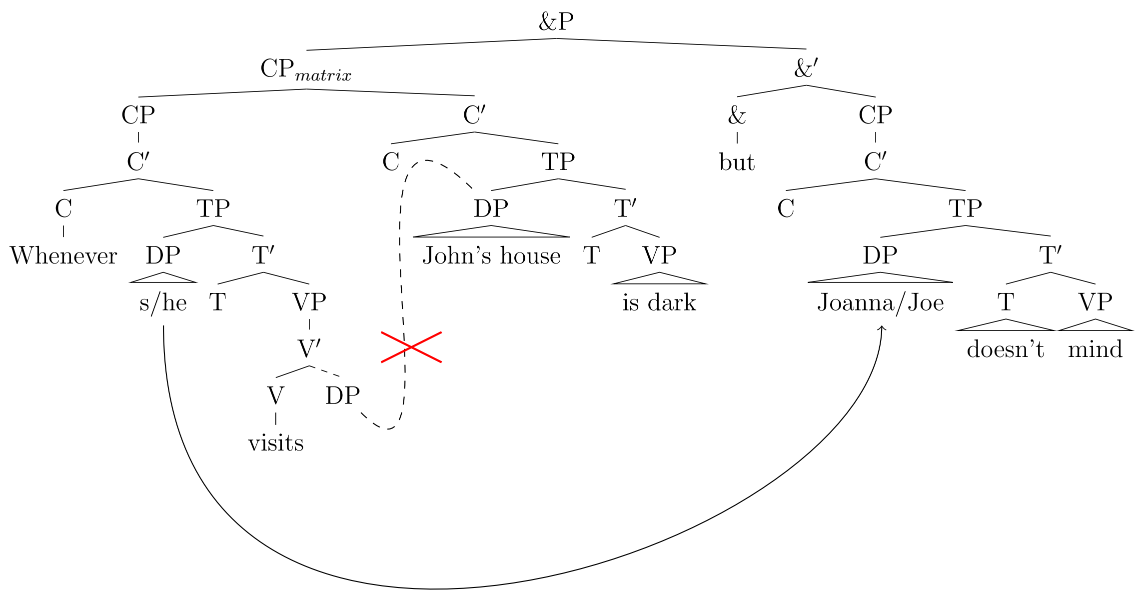
A null result does not form a strong argument for or against the hypothesis, so in order to distinguish what structure the parser has built, it is useful to examine what the parser does when it encounters the matrix verb. All of the sentences used in this experiment are, as before, resolved at the matrix verb, with the ambiguous DP ultimately in the position of the matrix subject. Since these stimuli have transitively biased verbs but ultimately have intransitive argument structures, this should cause a surprisal-related reading time slowdown at the matrix verb.

The parser may exhibit some other behaviors that would indicate that the active search for the cataphoric antecedent was influencing the parser's behavior in a significant way. The experimental design is not set up in a way to clearly identify these possibilities, but speculation on them will provide some basic guidelines for interpreting any results that are obtained in the case where no differences are observed between the conditions at the first candidate antecedent.

If the parser encounters the matrix verb when it expected to find a matrix subject, it may attempt to reanalyze the sentence as a matrix subject parse instead of the embedded object parse before it continues further to the right. In this case, we may see longer regression path times when not only does the reader look back to the first candidate antecedent, but when the gender mismatches as well. In this case, in the re-read times, we may see a gender mismatch effect at the proper name, even though it was originally in a position in which forming the coreference dependency would have violated Binding Condition C. That is, if the parser reanalyzes the string as a matrix subject construction, the proper name is no longer c-commanded by the pronoun and becomes available as an antecedent. Thus, the parser may attempt to form the dependency after the structure of the sentence is resolved.

Alternatively, it may be the case that the parser does not go back to the ambiguous NP and attempt to form a coreference dependency between the pronoun and the proper name because the structure was already built, but so low probability that the parser did not put any resources into searching it for a candidate antecedent. In this case, the parser may "jump" from one parse tree to another in the parse space (as described in Lewis (1998)), but the active search process does not retrace its path in the new tree. That is, the active search process was only active in the highly likely (or high activated) embedded object parse, but since the matrix subject parse was unlikely due to the transitive bias of the ambiguous verb, the active search was not pursued in that parse, as it was with the equibiased verb in Chapter 3. In this case, the parser may start searching for the next candidate antecedent where it left off: at the matrix verb. The next candidate antecedent would then potentially be located in the object position of the matrix verb. There is no candidate antecedent in this location (occupied by *dark*) in (18)-d.

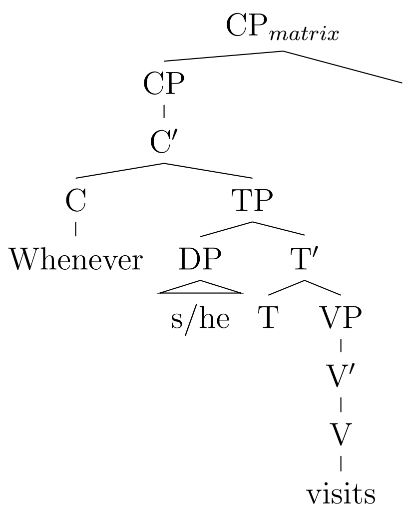
(18) d. Whenever s/he visits John's house is dark, but Joanna/Joe doesn't mind.



In this case, we may observe a gender mismatch effect at this second name. Unfortunately, the design of this experiment is such that there is no mismatched gender condition at the second name, so we may only speculate what the parser might do at this point in the structure-building process under the circumstances put forth in this hypothesis.

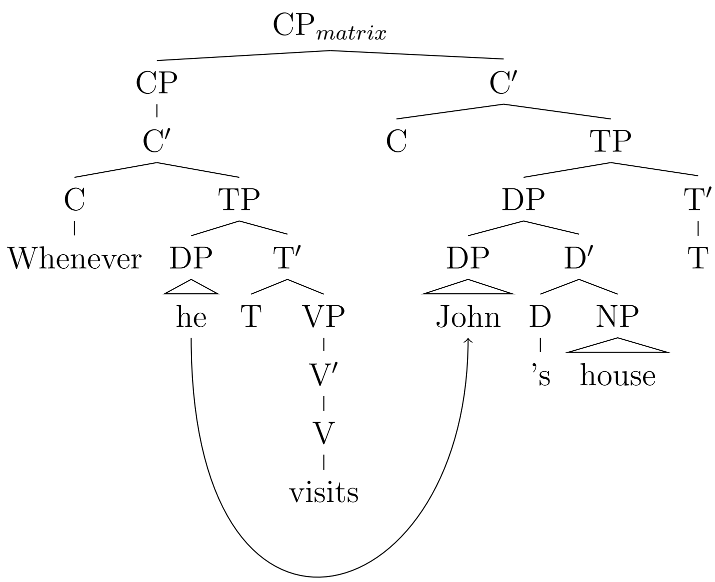
4.4.2.2. Hypothesis 2: Verb transitivity is not strong. In the contrasting hypothesis, that verb transitivity does not contribute a strong influence on the parser's behavior, we will observe a different pattern of reading times. The point where the two hypotheses diverge is at the embedded verb leaves, depicted in (19)-a.

(19) a. Whenever s/he visits



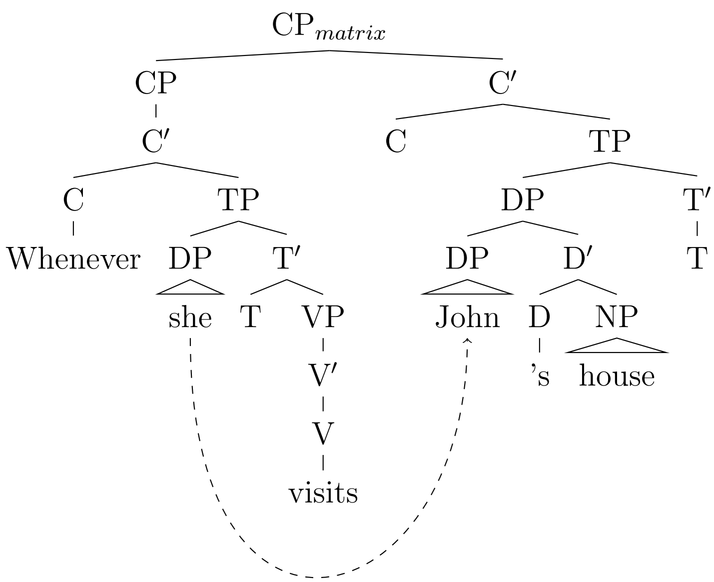
If the verb frame frequency does not influence the parser to commit to the transitive parse, then the parser should behave as it did in Chapter 3. In the Matched condition, shown in (19)-b, the parser finds a candidate antecedent at the proper name *John*, which is accessible in the matrix subject parse. Here, the coreference dependency is formed and the gender feature is checked and everything is consistent. Thus, no slowdown in reading time is expected in this condition.

(19) b. Whenever he visits John's house



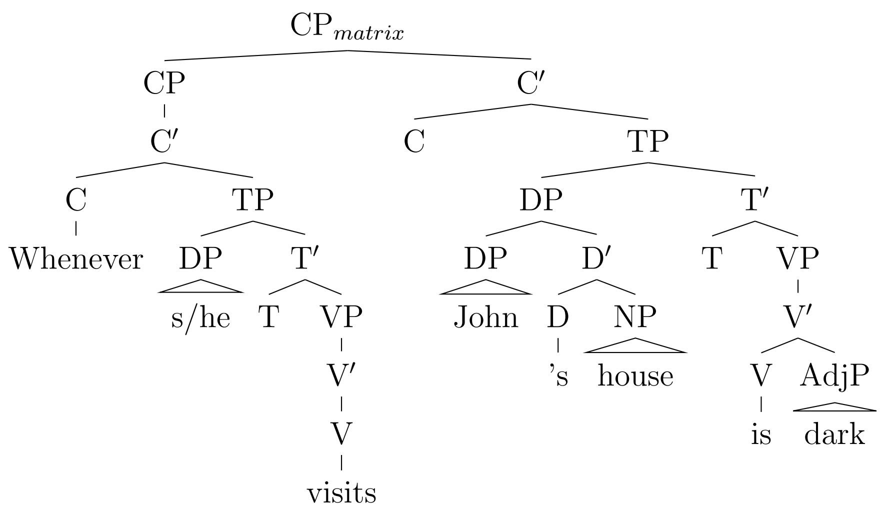
In contrast, the Mismatch condition (19)-b′ should produce a gender mismatch effect at the proper name. Although the matrix subject parse may coexist alongside the embedded object parse, it may also display the effects consistent with building the matrix subject parse, as observed in the previous chapter. Thus, a gender mismatch effect at *John* in (19)-b′ is indicative of the presence of the matrix subject parse, but does not rule out the presence of the embedded object parse.

(19) b.′ Whenever she visits John's house



Next, after the ambiguous noun phrase, the parser encounters the matrix verb, in (19)-c. If the parser is influenced by the active search, as it apparently was in the previous experiment, we would expect the matrix verb to be consistent with an accessible representation and cause no significant surprise to the parser.

(19) c. Whenever s/he visits John's house is dark

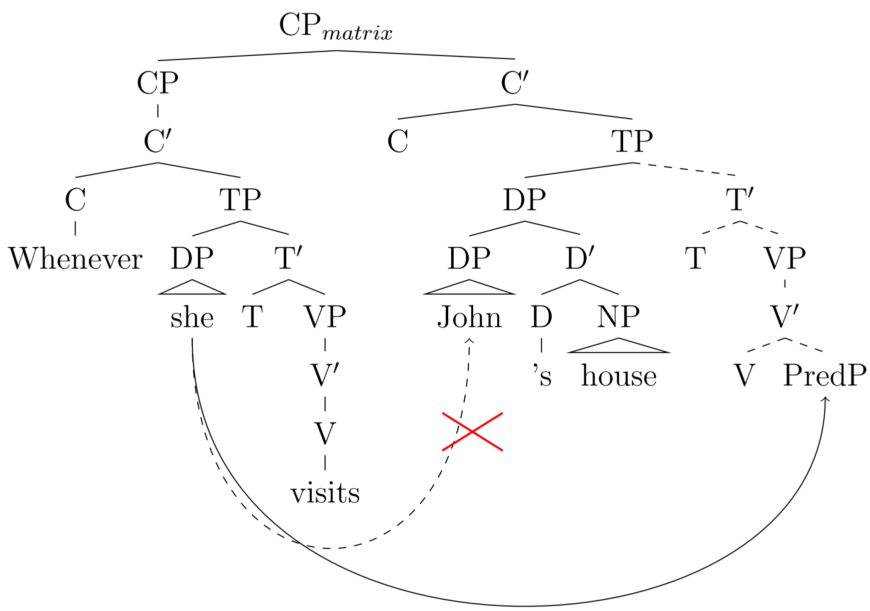


There is no baseline against which to test the relative slowdown of the verb between conditions, since all three conditions contain the same embedded verb and the locally ambiguous DP is resolved to be the matrix subject in each. However, there is another characteristic of the parser's behavior that we can investigate at this point: the parser's anticipation of the matrix verb by way of an active search for a second candidate antecedent.

4.4.2.3. Effects at the Matrix Verb. As mentioned in Chapter 3.8.2 and 3.8.3, there was a main effect of Gender found at the matrix verb, with Mismatched conditions read faster than Matched conditions in first fixation and first pass reading times, as well as in the matrix verb spillover region in first pass reading times. There was no motivation for predicting this in the proposed hypotheses at the time, so it was left for this follow-up to elucidate. The experimental design in this current experiment allows for a more thorough exploration of what the parser may be doing at the point of the matrix verb.

If the parser's active search for a cataphoric antecedent is satisfied when it finds a suitable candidate and is thus terminated or deprioritized, there is no reason for it to make any predictions about the upcoming material. In the Matched condition, we would expect the parser to find a suitable candidate at the location of *John* and be sufficiently satisfied. In contrast, the Mismatched condition does not satisfy the search for a suitable antecedent. So even though the parser initially attempts to form the cataphoric dependency at *John*, it ultimately is unsatisfied, and continues the search. As I concluded in Chapter 3, the parser is influenced by the active search, and if this is the case, the parser will continue to be influenced by the active search while it is ongoing. Thus, the parser may anticipate where the next possible location for a candidate antecedent would be (Omaki et al., 2015). If the parser has placed *John's house* in the position of the matrix subject, the next likely positions for a candidate would be in the matrix object position. In order to calculate the presence of the matrix object, the parser must predict the presence of a matrix verb, at least at the level where its projection in the syntactic tree may accommodate a DP complement.

(20) Whenever she visits John's house



This is in contrast to the Matched condition, in which the parser may not be surprised by a matrix verb, but it has not yet built the syntactic nodes that hold the lexical labels during a process undergone in order to predict the matrix object. In this way, the parser would be expected to have a lower processing cost at the matrix verb in the Mismatched condition than in the Matched condition. In other words, the reading time will be faster at the matrix verb in the Mismatched condition because the parser has a pre-built structure in the Mismatched condition after its attempt to link the pronoun to the matrix subject was foiled, but it does not have any such motivation to predict or pre-build the structure in the Matched condition. Finally, in the No Coreference condition, the parser has no active search to influence its predictions for upcoming material, so the reading time at the matrix verb in this condition should also be slower than in the Mismatched condition.

4.4.3. Analysis and Results

Given the 1×3 design of this experiment, analyses were conducted using two linear mixed effects regressions to compare the three conditions in a pairwise manner. Due to the potential for significance inflation over two model comparisons, alpha levels were modified using a Bonferroni correction ($\alpha = 0.025$). Two models were necessary because one compares the fixed effects of Match and Mismatch to the No Coreference condition, while the other compares the fixed effect of Match to Mismatch directly. All models also had fixed effects of presentation order and region length by character (when length varied across items) random intercepts for participants and items, along with a correlated random slope for the fixed effect(s) of the model and region length by participant, which was the maximal random effects structure that would allow the model to converge.[3] The fixed effects of the Gender comparisons were contrast coded, in which each level was assigned either the value 0.5 or -0.5. The fixed effects of order and length were numeric and therefore did not need to be coded for analysis. Since some of the verbs had the potential for a frame with a sentence complement instead of a direct (DP) object, this likelihood was considered as a potential factor. However, it did not contribute significantly to models and was not included for the analyses presented here. Each of the two models was then considered a maximal model, to which a diminished model (lacking an effect of interest) was compared using an ANOVA to determine the contribution of the effect to the maximal model. This resulted in estimates (β) and standard errors being drawn from the summary of the maximal models, and the chi-squared statistics and p-values being

[3]Model: $lmer(res \sim MatchNoCoref + MismatchNoCoref + order + length + (1 + MatchNoCoref + MismatchNoCoref + length|subj) + (1|item), data = data)$ and
$lmer(res \sim MatchMismatch + order + length + (1 + MatchNoCoref + MismatchNoCoref + length|subj) + (1|item), data = data)$

drawn from the ANOVA model comparisons. These statistics are listed in Tables 4.6 (for the Ambiguous DP regions) and 4.7 (for the Matrix Verb regions), while the means and standard errors by measure, condition, and region are listed in Table 4.5.

Table 4.5. Means (and Standard Errors) for Eye-tracking Experiment (Section 4.4).

Region	*Annie's*	*dissonant melody*	*was*	*haunt... beautiful*	*Karen**
First fixation					
Match	211.6 (4)	242.6 (5)	211.8 (9)	253.9 (5)	199.9 (5)
Mismatch	229.2 (5)	237.5 (5)	228.3 (13)	257.5 (6)	202.3 (5)
No Coref	227.2 (6)	240.7 (5)	213.1 (10)	248.0 (5)	205.6 (6)
First pass					
Match	248.3 (8)	474.8 (15)	226.9 (12)	537.9 (17)	282.2 (9)
Mismatch	290.0 (10)	490.1 (17)	247.7 (16)	549.0 (16)	263.0 (8)
No Coref	281.8 (10)	500.4 (15)	226.8 (14)	511.8 (15)	286.7 (9)
Regression path					
Match	352.3 (18)	613.6 (23)	298.7 (27)	1090.5 (58)	492.1 (50)
Mismatch	385.2 (17)	640.2 (27)	306.6 (45)	956.1 (45)	576.6 (53)
No Coref	387.0 (30)	640.7 (37)	331.7 (59)	1086.6 (58)	445.0 (28)
Re-read time					
Match	465.3 (34)	593.8 (38)	258.3 (27)	609.8 (43)	283.3 (20)
Mismatch	436.8 (31)	522.6 (36)	225.2 (31)	535.6 (34)	298.1 (22)
No Coref	452.4 (38)	610.4 (43)	264.2 (34)	600.6 (37)	245.9 (18)
Total time					
Match	486.7 (23)	789.1 (29)	295.0 (14)	811.4 (29)	340.1 (12)
Mismatch	503.6 (22)	755.1 (28)	287.9 (14)	777.2 (24)	338.1 (12)
No Coref	517.6 (26)	796.8 (29)	292.5 (15)	786.6 (25)	346.6 (11)

*This region was either a male or female name, so the length differed within items and between items.

4.4.3.1. Ambiguous NP region. At the location of the ambiguous noun phrase, there were two comparisons that reached significance. It was found that the Match condition was read significantly faster than the Mismatch condition in first fixation duration and first pass reading time (FF: β = -17.01, S.E. = 6.6, $\chi^2(1)$ = 6.74, p = 0.019; FP: β = -39.97, S.E. = 12.2, $\chi^2(1)$ = 10.73, p = 0.0022). This is a strong indication of a gender

Table 4.6. Combined ANOVA and LME results for Ambiguous DP in Eye-tracking Experiment (Section 4.4).

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First fixation					
Annie's	Match vs Mismatch	-17.01	6.6	6.74 (1)	=0.019
	Match vs No Coref	-21.91	7.6	8.33 (1)	=0.0078
	Mismatch vs No Coref	12.14	7.6	2.60 (1)	>0.1
dissonant	Match vs Mismatch	6.06	6.7	0.82 (1)	>0.1
melody	Match vs No Coref	5.91	7.7	0.66 (1)	>0.1
	Mismatch vs No Coref	-6.20	7.6	0.58 (1)	>0.1
First pass					
Annie's	Match vs Mismatch	-39.97	12.2	10.73 (1)	=0.0022
	Match vs No Coref	-46.43	14.1	10.88 (1)	=0.0019
	Mismatch vs No Coref	33.52	14.1	5.63 (1)	=0.036
dissonant	Match vs Mismatch	-3.38	19.5	0.03 (1)	>0.1
melody	Match vs No Coref	-19.65	22.5	0.77 (1)	>0.1
	Mismatch vs No Coref	-12.28	22.3	0.30 (1)	>0.1
Regression path					
Annie's	Match vs Mismatch	-35.24	30.3	0.59 (1)	>0.1
	Match vs No Coref	-42.88	35.2	1.50 (1)	>0.1
	Mismatch vs No Coref	27.56	35.1	1.33 (1)	>0.1
dissonant	Match vs Mismatch	-3.63	38.2	0.01 (1)	>0.1
melody	Match vs No Coref	-18.72	44.3	0.18 (1)	>0.1
	Mismatch vs No Coref	-10.87	43.9	0.06 (1)	>0.1
Re-read time					
Annie's	Match vs Mismatch	38.71	44.3	0.46 (1)	>0.1
	Match vs No Coref	42.74	50.9	0.68 (1)	>0.1
	Mismatch vs No Coref	-34.66	51.1	0.75 (1)	>0.1
dissonant	Match vs Mismatch	94.68	49.3	5.20 (1)	=0.045
melody	Match vs No Coref	59.40	57.0	1.11 (1)	>0.1
	Mismatch vs No Coref	-130.94	57.6	3.71 (1)	>0.1
Total time					
Annie's	Match vs Mismatch	-13.43	28.2	0.23 (1)	>0.1
	Match vs No Coref	-21.57	32.7	0.45 (1)	>0.1
	Mismatch vs No Coref	5.46	32.5	0.02 (1)	>0.1
dissonant	Match vs Mismatch	50.25	33.1	2.32 (1)	>0.1
melody	Match vs No Coref	24.80	38.4	0.42 (1)	>0.1
	Mismatch vs No Coref	-74.21	37.8	3.87 (1)	=0.099

mismatch effect, due to the parser's attempt to form a dependency between, for example, *she* and *John*. Crucially, this indicates that the parser has constructed a representation of the sentence that allows for *John* to be an accessible antecedent, thus it is not c-commanded by *she*. Furthermore, this supports the hypothesis that the parser is not strongly influenced by the verb frame's probabilistic likelihood, thus maintaining the less likely (but possible) matrix subject parse in a high enough state of activation or high enough rank to influence reading time.

The comparison between the Match condition and the No Coreference condition also reached significance in first fixation duration and first pass duration, with the Match condition read faster than the No Coreference condition (FF: β = -21.91, S.E. = 7.6, $\chi^2(1)$ = 8.33, p = 0.0078; FP: β = -46.43, S.E. = 14.1, $\chi^2(1)$ = 10.88, p = 0.0019). This may indicate some sort of facilitation process or anticipation of the position of the candidate antecedent by the parser in the Match condition, as compared to sentences in which active search is not triggered in the No Coreference condition. If this is so, it is evidence that the parser is being "actively" influenced by the presence of active search, rather than by default maintaining the unlikely parse with the possibility of licit coreference. The implications of this will be discussed further in the General Discussion.

In further support of this apparent gender mismatch effect, the comparison between Mismatch and No Coreference also reach significance in first pass duration, with Mismatch read significantly slower than No Coreference at the ambiguous noun phrase *John* (β = 33.52, S.E. = 14.1, χ^2 = 5.63, p = 0.036). This indicates that not only is there a relatively increased reading time in the Match condition, but also that the Mismatch condition has an additional processing cost that slows down reading time as compared to when

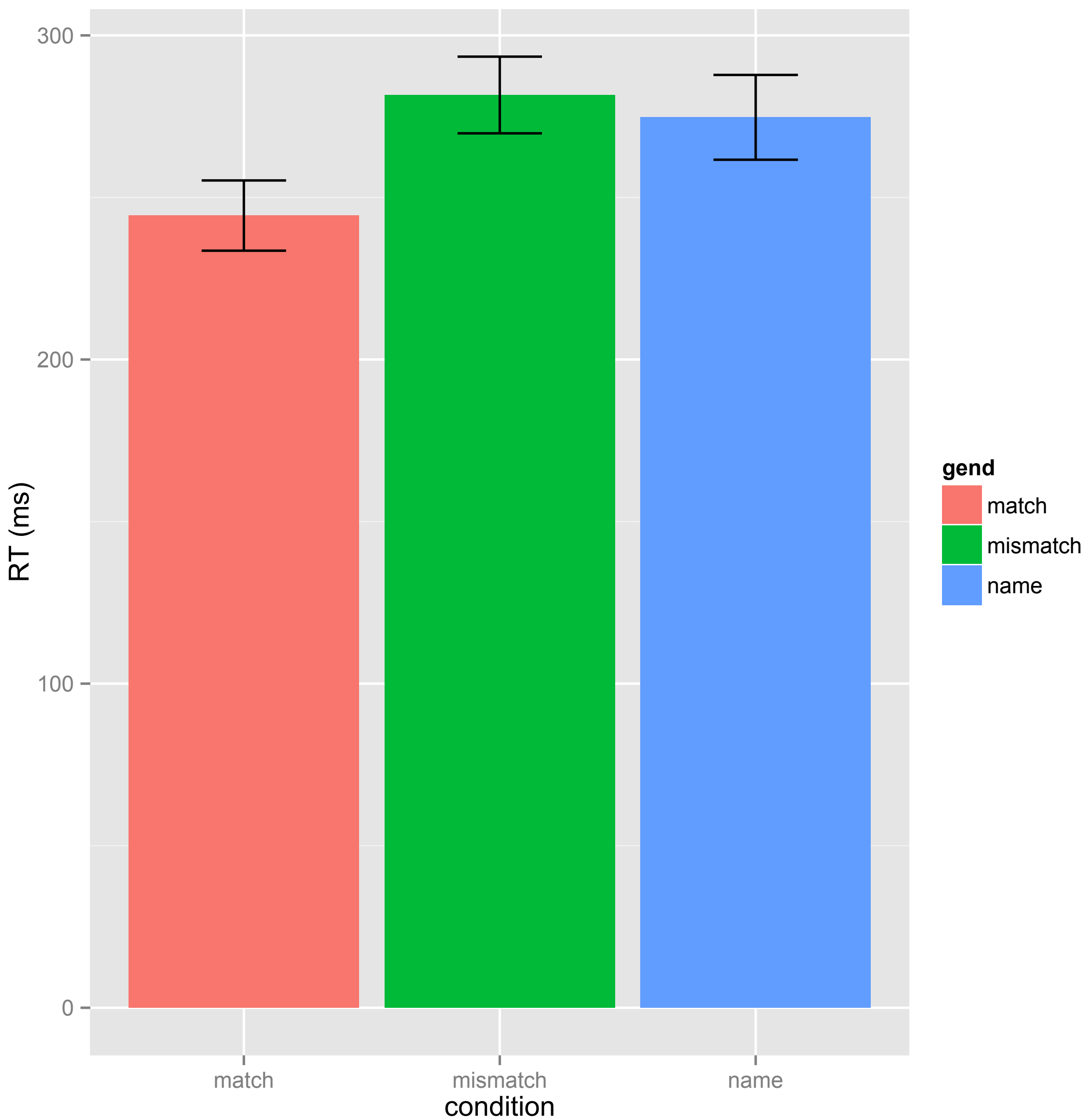


Figure 4.4. First pass reading time at the ambiguous NP *Annie's*

there is no antecedent search. This suggests, more than some facilitation effect of having congruent gender features (Sturt, 2003), that the parser is indeed attempting to form the dependency at this location and is being forced to revise after integrating the necessary lexical properties. The presence of both a speed up in comparing the Match condition to our No Coreference baseline and a slowdown in comparing the Mismatch condition to the baseline is strong evidence that a gender mismatch effect is present, thus that the process of active search has identified *John* as an accessible candidate antecedent. In turn, this picks out the matrix subject parse as being sufficiently prioritized by the parser to exert influence on reading time measures.

4.4.3.2. Matrix Verb region. No significant effects were observed in the matrix verb region. Although some effects were observed in the equivalent region in the previous chapter, it is not surprising that no effects reached significance since this region was only four characters long, which is a particularly short length for fixations to land within. Thus the spillover region following the matrix verb is considered as a substitute for any effects that might be due to the parser encountering the matrix verb.

4.4.3.3. Matrix Verb Spillover region. Following the matrix verb, there is a potential for spillover effects to be observed. In fact, the Mismatch condition was read significantly faster than the No Coreference condition in regression path duration (β = -194.95, S.E. = 76.4, $\chi^2(1)$ = 6.52, p = 0.021). This may indicate that the parser experienced less processing cost at the matrix verb because the syntactic structure had been pre-built to accommodate the search for another candidate antecedent, as described in Section 4.4.2.3.

The Bonferroni correction due to the multiple comparisons means that no other comparisons reached the significance level. However, the comparison between the Match and Mismatch conditions was marginally significant in regression path duration, with Mismatch read faster than Match (β = 145.11, S.E. = 66.7, $\chi^2(1)$ = 4.74, p = 0.059). While

Table 4.7. Combined ANOVA and LME results for Transitive Matrix Verb in Eye-tracking Experiment (Section 4.4).

Region	**Effect**	**Estimate**	**Std error**	χ^2 **(df)**	**p-value**
First fixation					
was	Match vs Mismatch	-14.67	13.4	1.21 (1)	>0.1
	Match vs No Coref	-12.24	15.5	1.24 (1)	>0.1
	Mismatch vs No Coref	16.95	15.3	0.64 (1)	>0.1
hauntingly	Match vs Mismatch	-2.38	7.1	0.11 (1)	>0.1
beautiful	Match vs No Coref	2.70	8.2	0.11 (1)	>0.1
	Mismatch vs No Coref	7.24	8.1	0.81 (1)	>0.1
First pass					
was	Match vs Mismatch	-16.24	16.3	0.97 (1)	>0.1
	Match vs No Coref	-11.20	18.8	0.36 (1)	>0.1
	Mismatch vs No Coref	21.37	18.6	1.28 (1)	>0.1
hauntingly	Match vs Mismatch	-5.75	21.2	0.07 (1)	>0.1
beautiful	Match vs No Coref	16.17	24.6	0.43 (1)	>0.1
	Mismatch vs No Coref	26.74	24.3	1.21 (1)	>0.1
Regression path					
was	Match vs Mismatch	4.91	65.9	0.004 (1)	>0.1
	Match vs No Coref	-13.71	76.1	0.04 (1)	>0.1
	Mismatch vs No Coref	-22.45	75.3	0.08 (1)	>0.1
hauntingly	Match vs Mismatch	145.11	66.7	4.74 (1)	=0.059
beautiful	Match vs No Coref	93.09	77.2	1.46 (1)	>0.1
	Mismatch vs No Coref	-194.95	76.4	6.52 (1)	=0.021
Re-read time					
was	Match vs Mismatch	23.01	44.6	0.28 (1)	>0.1
	Match vs No Coref	10.88	49.5	0.05 (1)	>0.1
	Mismatch vs No Coref	-39.29	52.9	0.58 (1)	>0.1
hauntingly	Match vs Mismatch	79.16	47.3	2.84 (1)	>0.1
beautiful	Match vs No Coref	52.43	53.8	0.99 (1)	>0.1
	Mismatch vs No Coref	-107.68	54.7	3.92 (1)	=0.095
Total time					
was	Match vs Mismatch	1.14	18.4	0.01 (1)	>0.1
	Match vs No Coref	-3.16	21.1	0.02 (1)	>0.1
	Mismatch vs No Coref	-5.81	21.2	0.08 (1)	>0.1
hauntingly	Match vs Mismatch	48.20	32.0	2.27 (1)	>0.1
beautiful	Match vs No Coref	52.33	37.1	1.98 (1)	>0.1
	Mismatch vs No Coref	-44.28	36.6	1.46 (1)	>0.1

this is a statistically tenuous relation given the conservative correction method, there is indeed a visual trend that is consistent with the Mismatch condition being read faster than the Match condition, seen in Figure 4.5. Thus, it is reasonable to speculate that there is a reduced processing cost associated with the Mismatch condition.

4.4.3.4. Second Antecedent region. Although it was not part of the original design of the experiment, analysis of the region corresponding to *Karen* or *Jimmy* in (17) could reveal whether the parser continues to search for an antecedent when it is not able to form one at the first candidate antecedent (i.e., at *Annie*). The means and standard errors for the region corresponding to *Karen* or *Jimmy* are listed in Table 4.8 and the linear mixed effects regressions and the corresponding ANOVA model comparisons, as performed in the previous analyses, are reported in Table 4.5. While nothing reached significance, there was an suggestive trend that was marginally significant before the Bonferroni correction was applied. Once the correction was applied, p-values were all greater than 0.1. While this is a nonconservative metric and should not be considered evidence of an effect, I will speculate about the visual trend that is apparent in Figure 4.6.

If the trends which are visible in Figure 4.6 were strong enough to be indicative of a generalizable pattern, then this pattern would be consistent with what has been observed thus far. In first pass reading times at the second candidate antecedent, the Mismatch condition is (visibly) faster than the No Coreference condition (β = -26.2, S.E. = 13.7, χ^2 = 3.77). Were this significant, it would be consistent with the parser anticipating the subject position in the coordinated clause after encountering the word *but*, which indicates the coordination. In this way, the parser may have been able to predictively

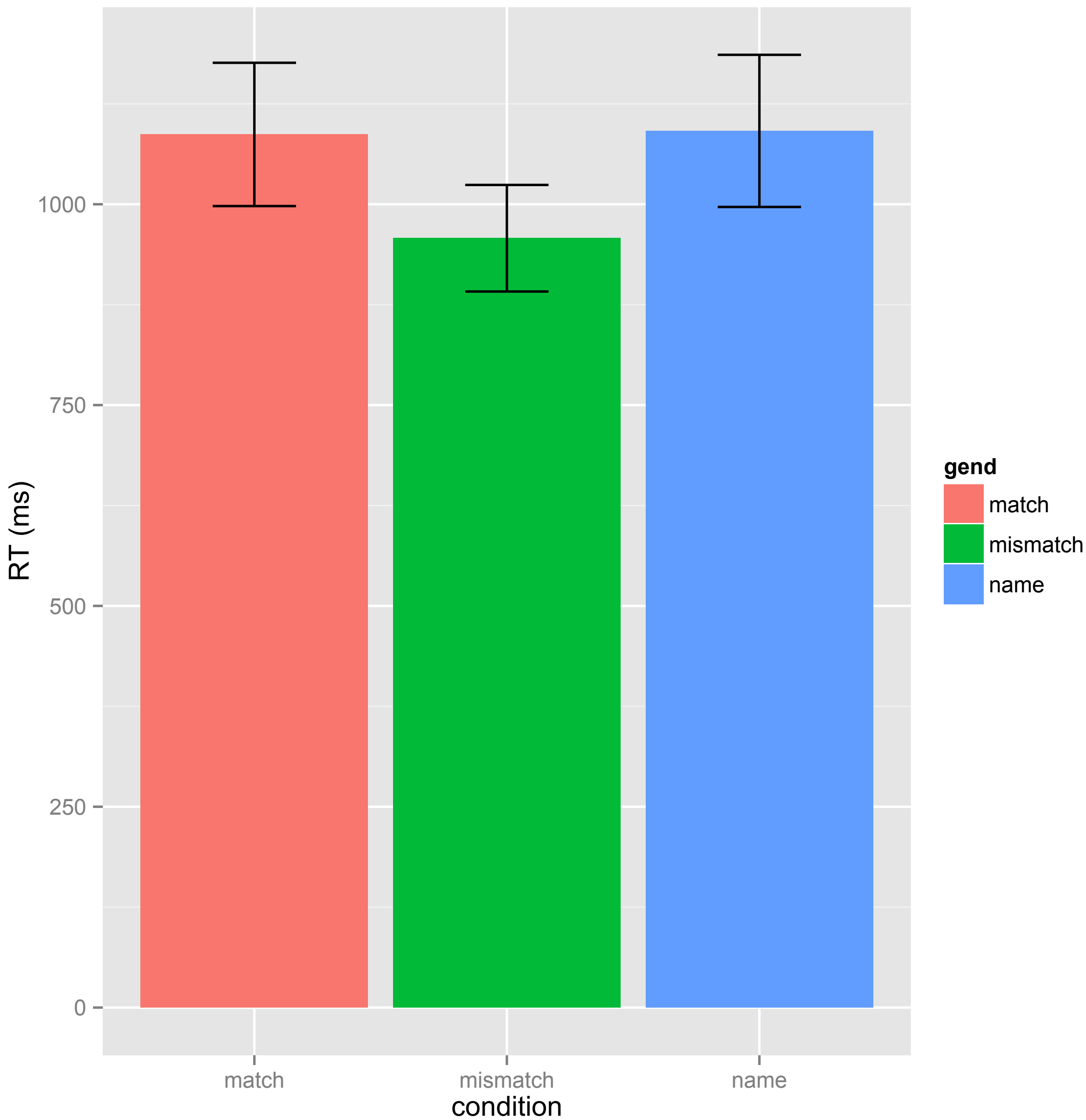


Figure 4.5. RP durations at the matrix verb spillover region

form a dependency between the cataphoric pronoun (e.g., *she*) and this second candidate antecedent (e.g., *Karen*). Furthermore, in regression path reading times, the Mismatch condition is (visibly) slower than the No Coreference condition (β = 137.11, S.E. =

Table 4.8. Combined ANOVA and LME results for Second Antecedent in Eye-tracking Experiment (Section 4.4).

Region	Effect	Estimate	Std error	χ^2 (df)	p-value*
First fixation					
Karen	Match vs Mismatch	-2.54	7.1	0.12 (1)	>0.1
	Match vs No Coref	-6.07	8.2	0.54 (1)	>0.1
	Mismatch vs No Coref	-1.09	8.2	0.02 (1)	>0.1
First pass					
Karen	Match vs Mismatch	17.24	11.9	2.11 (1)	>0.1
	Match vs No Coref	8.26	13.7	0.38 (1)	>0.1
	Mismatch vs No Coref	-26.6	13.7	3.77 (1)	=0.052
Regression path					
Karen	Match vs Mismatch	-84.68	62.6	1.81 (1)	>0.1
	Match vs No Coref	-32.52	71.7	0.20 (1)	>0.1
	Mismatch vs No Coref	137.11	72.0	3.64 (1)	=0.057
Re-read time					
Karen	Match vs Mismatch	-11.72	28.7	0.18 (1)	>0.1
	Match vs No Coref	24.25	33.8	0.50 (1)	>0.1
	Mismatch vs No Coref	40.17	32.0	1.61 (1)	>0.1
Total time					
Karen	Match vs Mismatch	-0.62	15.6	0.002 (1)	>0.1
	Match vs No Coref	-7.51	17.9	0.18 (1)	>0.1
	Mismatch vs No Coref	-6.31	17.9	0.12 (1)	>0.1

*Bonferroni correction was not applied here in order to illustrate nonconservative, marginal effects.

72.0, χ^2 = 3.64). Were this significant, it would be consistent with the parser finally encountering and identifying the second candidate antecedent, which might prompt it to look backward through the sentence to make sure that the gender does indeed match between the cataphoric pronoun and this second candidate, since such a fine-grained feature of the pronoun may have decayed in memory, for instance (Wagers and Phillips, 2014). Such a regression would result in longer reading times in this region, but not in the other regions (where either the dependency has been successfully completed or was never triggered). Although this speculation is based on visual trends that do not reach

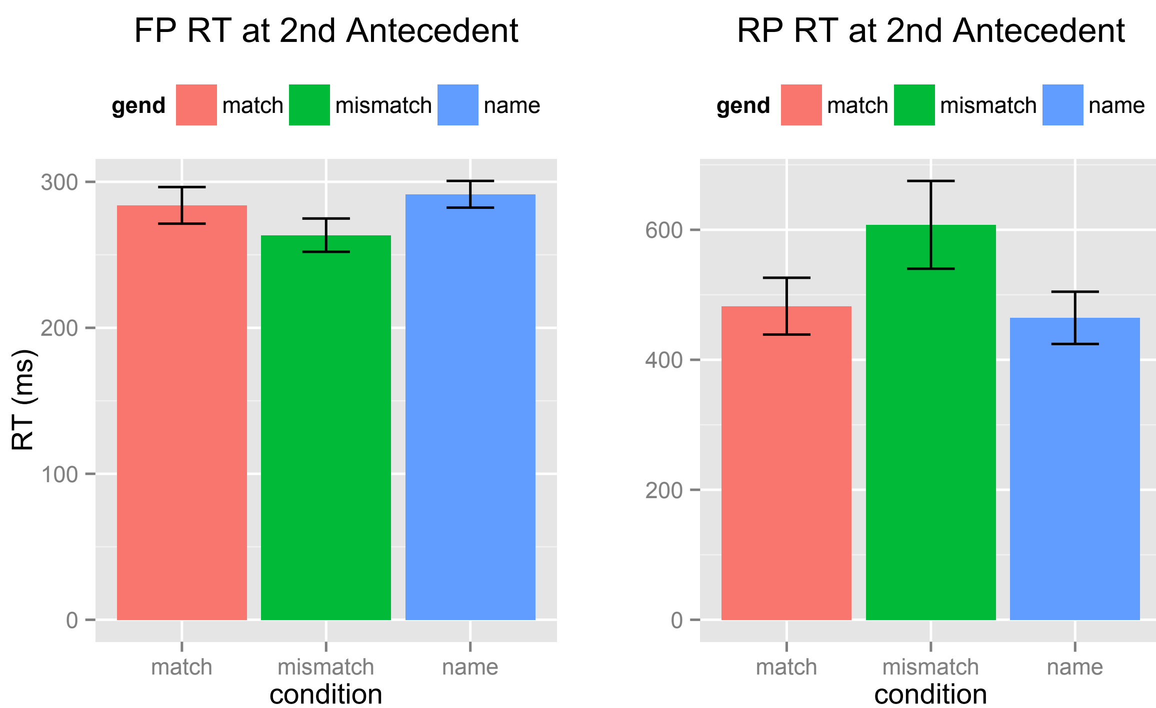


Figure 4.6. FP and RP durations at the second antecedent region

significance, this suggests future work may be successful in more soundly identifying the parser's behavior in situations in which an initial candidate antecedent is rejected, thus the dependency remains open.

4.4.4. Discussion

The presence an effect in which the matrix verb was anticipated during the parser's active search (and prediction) of the locations of potential antecedents was hinted at in Chapter 3. It was not explored because at the time because there was no theoretical motivation for doing so. However, in the current experiment, the three-way comparison allows these results to discussed in a more principled way. In the No Coreference condition, there is no active search triggered, so at no point does the parser have motivation for anticipating the positions in which a noun phrase might appear. Thus this condition forms a baseline that could be said to have a neutral influence, at least relative to the two other conditions. The Match condition does have an active search for an antecedent, but the parser encounters a candidate antecedent and can form a dependency that would satisfy the antecedent search. Thus this condition only has motivation for the parser to anticipate the position of a noun phrase up until but not after the ambiguous DP. In other words, the parser has no motivation to predict structure beyond the matrix subject. In the Mismatch condition, in contrast, the parser may anticipate the position of the ambiguous DP and attempt to form a coreference dependency with the candidate in this position, but the antecedent search ultimately remains unsatisfied. This is due to the gender mismatch between the cataphor and the candidate, neither of which prevents the dependency from initially forming, but may prevent the parser from terminating its active search. In this way, the parser may continue to anticipate the potential positions for a candidate antecedent. Thus, as discussed in Section 4.4.2.3, the process of predicting a structural location for the next candidate incidentally builds the infrastructure for the matrix verb. As a result, at the point when the matrix verb is encountered, the processing cost of integrating it into the parse is lower and it is read faster.

4.5. General Discussion

In this chapter, the influence of verb frame frequency and transitive bias was explored by comparing two conditions in which an active search for a cataphoric antecedent was triggered and a condition in which no search was triggered. Despite the transitive bias supplied by the verbs and their frame frequencies, the parser still attempted to form a cataphoric dependency between the pronoun and the proper name in the locally ambiguous DP. This process replicates the observations described in Chapter 3, which used equibiased verbs rather than transitively biased verbs. Thus it seems that the parser is not strongly influenced by the verb frame frequency when there is still a potential ambiguity, which is consistent with the conclusions of Staub et al. (2006).

In addition, it seems that the gender mismatch effect observed in this experiment somehow facilitated the processing of a gender congruent proper name. This is in contrast to the parser experiencing a large processing cost at a gender incongruent proper name. These two possibilities, while effectively equivalent when comparing structures containing a long-distance dependency, are distinguishable when compared to the third condition that did not contain an active search. In this third condition, the reading times at the proper name were slower than in the gender congruent Matched condition, but not different from the gender incongruent Mismatched condition. This may suggest that either the parser is experiencing some speedup in the Matched condition, contrary to the typical manifestation of a gender mismatch effect, or that the parser is slowing down in the Mismatched condition, but also in the No Coreference condition, due to some other reason. One possible explanation for an unrelated slowdown in the No Coreference condition at the proper name is the addition of a new discourse entity. However, Kazanina et al. tested this possibility in their 2007 study. They found that the addition of a new discourse entity did not increase reading times for their stimuli. Since the stimuli used in this dissertation are based closely on the structure and intent of the stimuli designed for Kazanina's study, I see no reason to expect that the addition of a new discourse entity would slow down the parser an equivalent amount to the observed gender mismatch effect. While it remains a possibility to be explored, the fact that the Matched condition was still read faster than the Mismatched condition is sufficient to conclude that an eager dependency search did attempt to form a coreference between the pronoun and the proper name, and succeeded in the Matched condition.

4.6. Conclusions

This chapter serves as a follow-up to the experiments described in Chapter 3. In Chapter 3, I observed that the parser does attempt to form a coreference dependency between a pronoun and a proper name within an ambiguous DP, despite the potential for such a relation violating the grammatical constraint of Binding Condition C. In the current chapter, the same pattern was observed even though the ambiguous verbs were biased toward a transitive reading, which made the likelihood of violating Binding Condition C higher than with the equibiased verbs used previous. This suggests that the bias contributed by the verb frame frequency is not sufficient to prevent the parser from pursing a parse that would allow for the speedy completion of the cataphoric dependency search.

CHAPTER 5

General Conclusions

5.1. Discussion of results

The experiments in this study were designed to test the relative strength of coreference dependency formation as compared to the influence of the parser's preference for local attachment (Chapter 3) and as compared to verb frame preference (Chapter 4).

5.1.1. Equibiased verbs

In Chapter 3, coreference dependencies were triggered by cataphoric pronouns that occurred before an ambiguous verb. The ambiguous verb was equibiased between a transitive and intransitive frame in order to allow for the distinct possibility of the following noun phrase to be resolved as either the embedded object or matrix subject. Results showed a slowdown in both ambiguous and unambiguous conditions when the gender of the R-expression was incongruent with the cataphoric pronoun. This indicates that the parser formed a coreference dependency between the pronoun and the proper name within the ambiguous noun phrase before gender properties were confirmed to be consistent, replicating van Gompel and Liversedge (2003) and Kazanina et al. (2007). In other words, the parser constructed the matrix subject parse, in which the ambiguous noun phrase was the subject of the main clause, thus available for coreference by Binding Condition C (Chomsky, 1981). However, there was also a reading time slowdown at the matrix verb and subsequent region in the ambiguous conditions, indicating that the parser was surprised by the matrix verb when it had expected the matrix subject to follow the ambiguous noun phrase. This was predicted to occur if the parser constructed the embedded object parse, in which the ambiguous noun phrase was the direct object of the equibiased verb. The fact that these two effects occurred together suggests that the parser builds both the matrix subject and embedded object parses and maintains them in parallel.

5.1.2. Transitive verbs

The goal of Chapter 4 was to determine how strong the influence of dependency formation is on ambiguity resolution. To this end, equibiased verbs used in Chapter 3 were replaced with transitively biased verbs. The transitive bias of the verbs was intended to increase the strength of the influence from the lexical verb frames, as well as boosting the influence from the parser's local attachment preference. However, it was found that the parser still posits a dependency between the pronoun and the proper name within the ambiguous noun phrase, as evidenced by a gender mismatch effect in both the ambiguous and unambiguous conditions. From this, I can conclude that the influence exerted by coreference dependency formation is relatively stronger than the influence from either the parser's local attachment preference or the verb frame bias, or both combined.

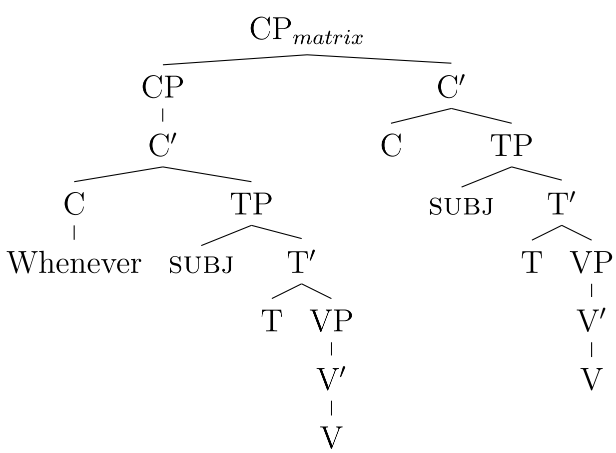
In addition, it was observed that the matrix verb was read faster in the gender-mismatched condition than in either the gender-matched or proper name (no coreference) condition. This replicates a trend observed in Chapter 3, as well. One possible reason for this observation is that, in the gender-mismatched condition, the parser is unsatisfied with the dependency it formed between the cataphoric pronoun and the matrix subject, and thus predicts the next possible location in which a candidate antecedent might appear. Since it has already attempted to form a dependency with the matrix subject position, and the selectional restrictions of English allow the parser to build the framework that supports the presence of the matrix verb, the next candidate antecedent could be located in the object position of the matrix verb. Although without the influence of the dependency search, the parser has the tools to predict the presence of the matrix verb, it is only with the impetus to find another candidate antecedent that the content of the matrix verb is anticipated. Thus, when the parser reaches the matrix verb in the gender-mismatched condition, the matrix verb is already sufficiently represented in the parse to incur a speedup as compared to the other conditions. Such behavior from the parser is consistent with work on parsing in SOV languages like Japanese (Yoshida, 2006) and parasitic gap constructions (Phillips, 2006).

5.2. The parser's behavior

From these observations, I can describe the way in which the parser seems to be operating. In this section, I will step through the parser's behavior, given the interpretations of the results described in the previous chapters. By and large, these results are consistent with a parser that represents rich, detailed, and connected structure (Slattery et al., 2013; Sturt and Lombardo, 2005), and actively searches for positions for the tail of the coreference dependency (Omaki et al., 2015; Traxler and Pickering, 1996; Wagers and Phillips, 2009), while respecting grammatical constraints (Kazanina et al., 2007; Stowe, 1986). Thus, the first stage of building a parse matches the descriptions given in Sections 3.3 and 4.4.2.

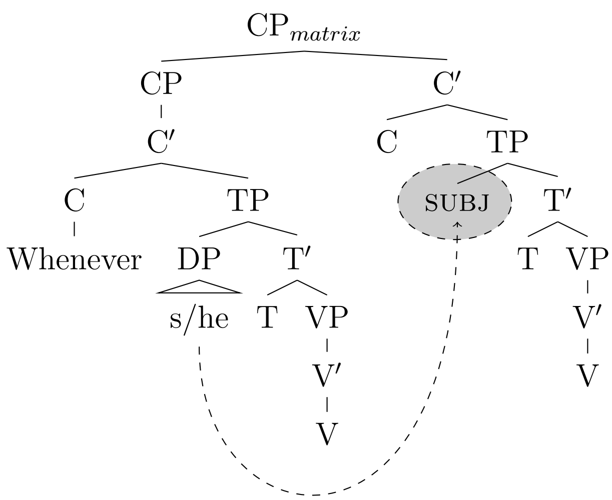
As previously assumed, the parser builds all of the unambiguous structure it is obligated to, upon encountering the first word of the sentence, *Whenever*. Since *Whenever* is located in the head of a CP, but that CP is not the matrix clause of the sentence, it must be dominated by the matrix CP. Both CPs unambiguously select a TP in their complement position. Each TP also selects a VP for its own complement position, thereby projecting down to both an embedded verb head and a matrix verb head. No further structure of either verb is anticipated at this point. Additionally, by the EPP, some type of subject phrase is required in the specifier of each TP. Since the internal content of the subject (besides, potentially, the projection of a single DP) is unknown and ambiguous, the parser does not anticipate the structure in these positions, either. This set of operations results in the tree shown in (1).

(1)



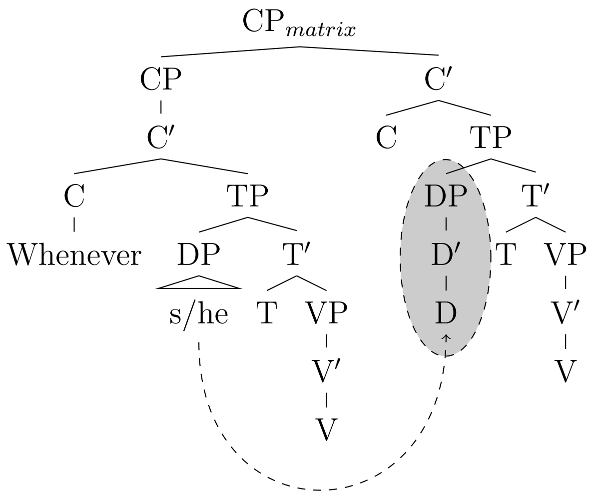
The cataphoric pronoun (either *he* or *she*) that is encountered next triggers a search for its (optional) antecedent. The parser, being active (or hyper-active) (Omaki et al., 2015; Traxler and Pickering, 1996), immediately identifies the closest potential position for a candidate antecedent. Since an object of the embedded verb is neither a grammatically licit position for an antecedent (by Binding Condition C, Chomsky, 1981), nor is it anticipated at this point, there is only one potential site for a candidate antecedent: the matrix subject (shaded in gray in (2)).

(2)



The matrix subject has no definite structure built yet. Even in the case where the parser anticipates the subject containing a projection of DP, the ambiguity associated with predicting the content of the subject will prevent the parser from specifying the precise location of the dependency tail within the subject (3). For instance, if the matrix subject were a proform, the entire content may consist of only this DP, and D could have appropriate properties for an antecedent. However, if the subject were a noun phrase like *the house*, the head D would be inappropriate to consider for coreference, being the article *the*.

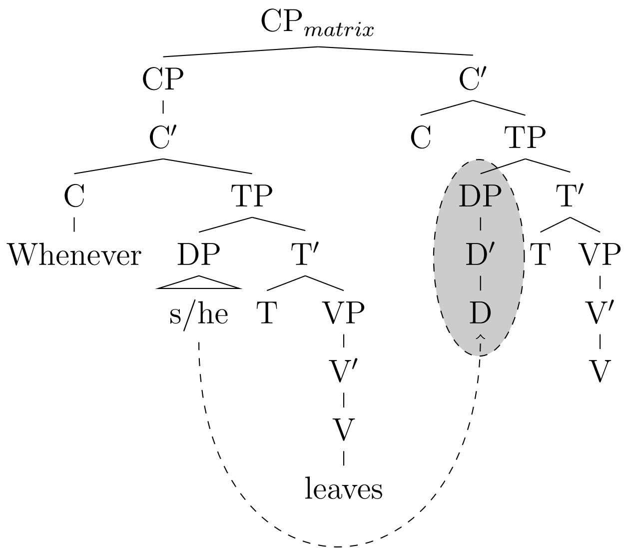
(3)



Moreover, if the matrix subject were a noun phrase like *John's house*, the proper name (*John*) in the specifier position of this top DP would be an excellent candidate antecedent, whereas *'s* and *house* would not (although the parser may consider *house* until its animacy properties are checked). Due to this ambiguity, the parser does not anticipate the structure of the subject DP, and allows the tail of the dependency to remain underspecified.

Next, the embedded verb is encountered. According to the observations made in this dissertation, the frame frequency of the verb does not play a (noticeable) role in the current structure-building process. Therefore, I use the verb *leaves* as an example, without committing to the particular transitive and intransitive frame frequencies associated with it. Any verb that has the potential to be either transitive or intransitive may be an equally adequate example here, in (4).

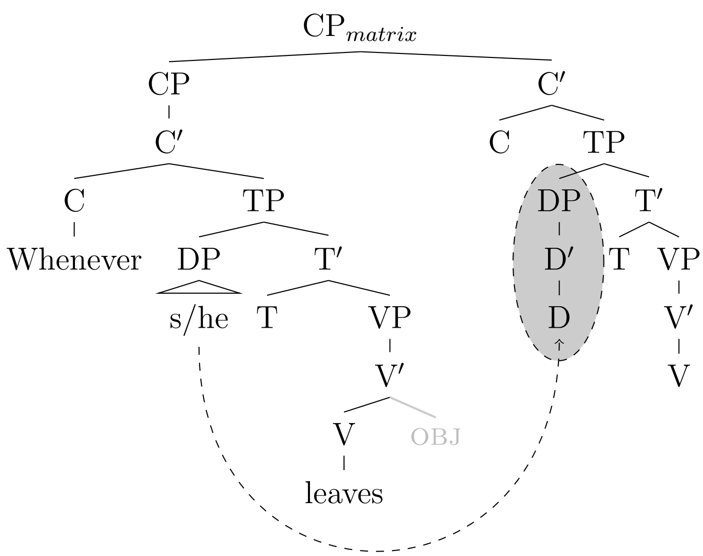
(4)



It is at this point in the representation of the sentence that the ambiguity must be addressed. I suggest that the identification and retrieval of *leaves* is the action that introduces the ambiguity and leads to the parallel representation of the so-called matrix subject and embedded object parses. If one assumes that the broad syntactic category of a word is immediately accessible to the parser, and the argument structure of a verb, once recognized is activated through lexical retrieval, the process may follow this set of operations.

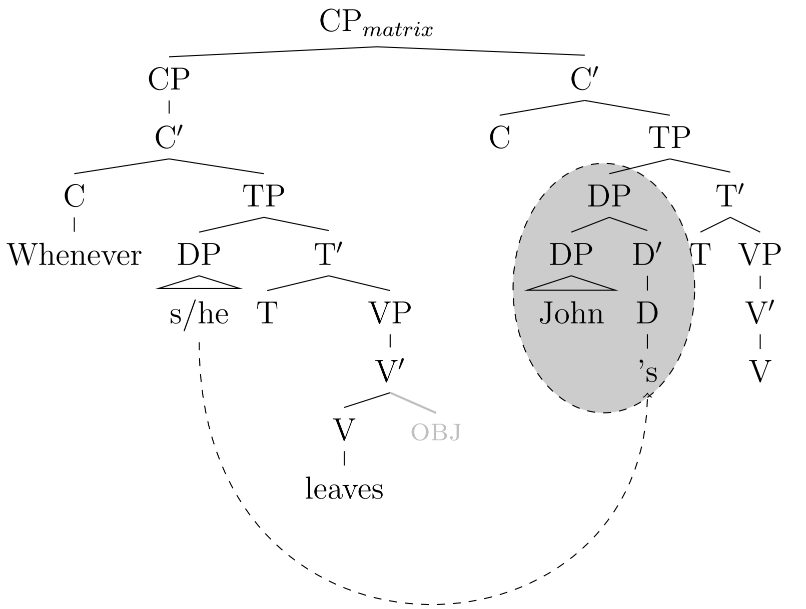
The parser identifies that the incoming word is of the category VERB, and immediately attaches it into the tree (Sturt and Lombardo, 2005). As the finer-grained properties of the verb are retrieved, the parser recognizes that this verb could either have, inclusively, a direct object DP or no object. These two structures are increasingly activated as the word is retrieved, and are incorporated into the tree in parallel, in a similar manner to how lemmas for a lexical form may be activated and retrieved in parallel (Luce and Pisoni, 1998; MacDonald et al., 1994; van Gompel et al., 2012), represented by gray lines and text in (5).

(5)



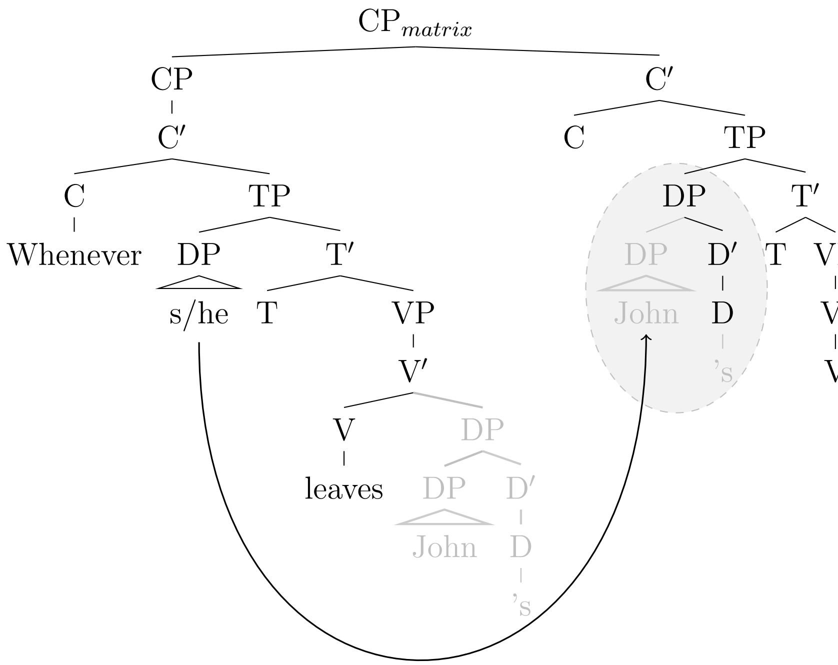
Notably, at this point, the dependency between the cataphor and the matrix subject is unchanged. This sets the stage for the first crucial point in this description. When the parser comes to the beginning of the ambiguous DP, the first word it will encounter is *John's*. This word has a possessive marker, and that marker *'s* is the head of the DP. Since the matrix subject (may) already have the skeleton of a DP built, placing the incoming words here is a low-cost way procedure. This puts *John* in the specifier position of the DP, dominated by its own DP, as shown in (6).

(6)



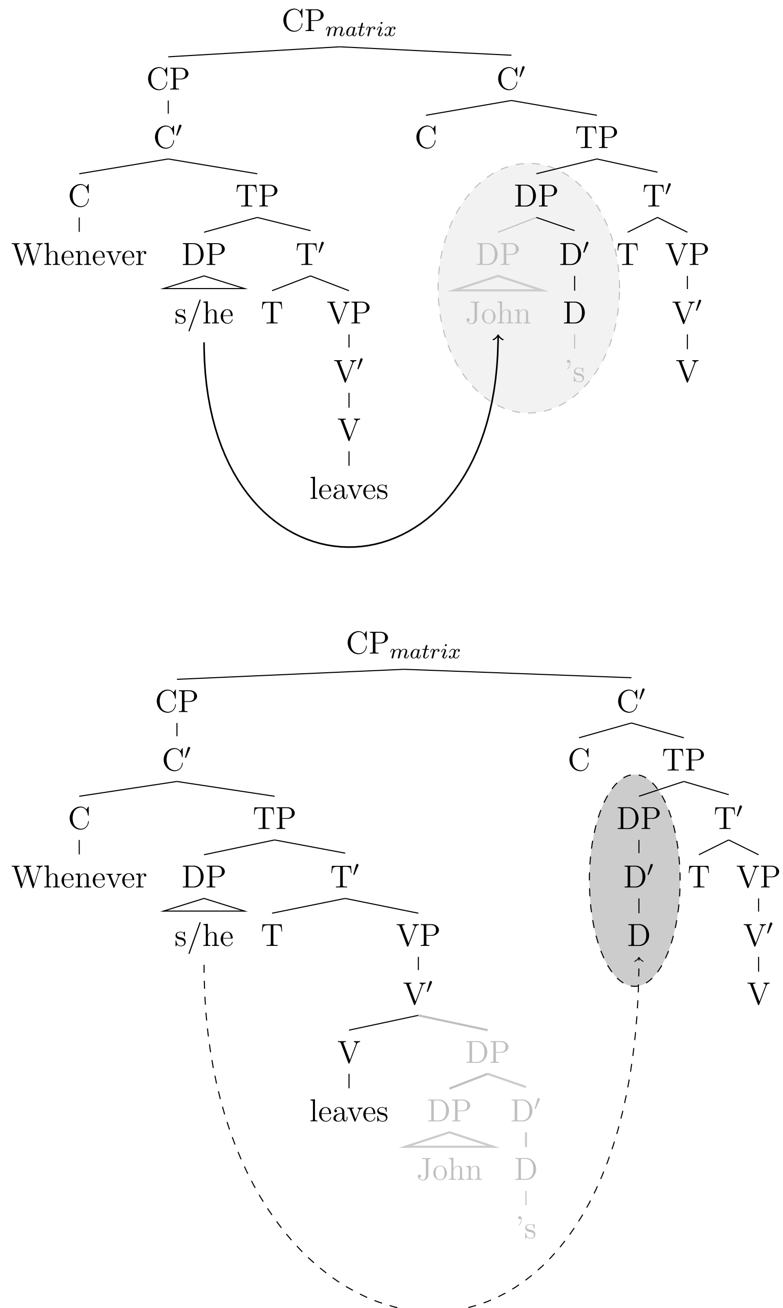
Except, now, there is an item in the tree, within the matrix subject (the shaded area) that is a potential candidate for the tail of the coreference dependency. This word is *John*, which is properly noun-like (i.e., an R-expression), and the parser chooses its position as the location where the tail of the dependency is anchored, shown in (7), with a solid dependency line. At the same time, the frame for *leaves* that selects a direct object DP is also activated, and treats *John's* as the first element in a DP that can serve as this direct object. So, the same structure is built after leaves.

(7)



Any further parallel structures represented in this type of tree diagram will become difficult to follow, so for simplicity's sake, (8) splits (7) into two trees. Note that the structures in black in both trees are shared between the two, and do not need to take up additional space in memory. Note also that the (general) location of the tail of the dependency is the same for both trees: the matrix subject.

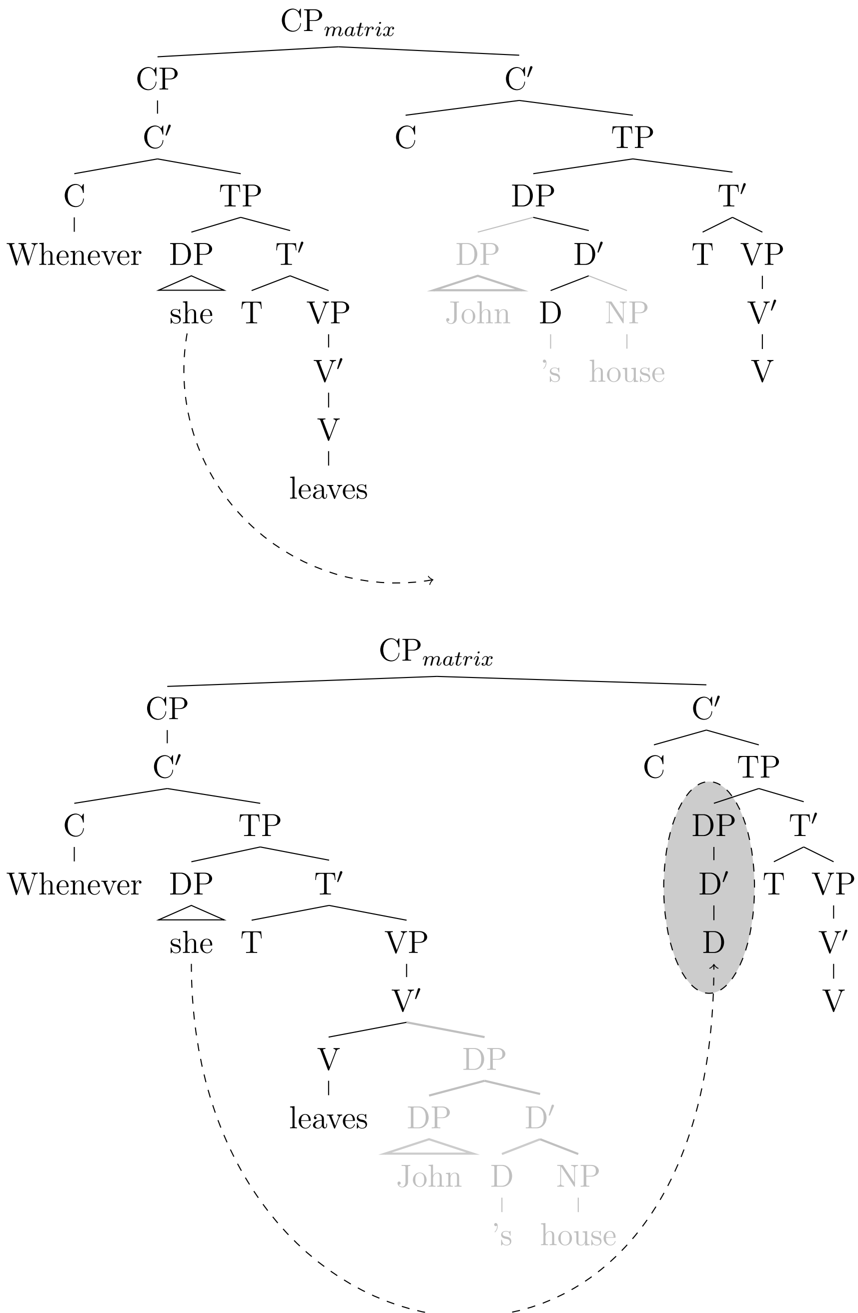
(8)


CP$_{matrix}$
CP
C′
C
TP
Whenever
DP
T′
s/he
T
VP
V′
V
leaves
DP
D′
John
D
’s
CP$_{matrix}$
CP
C′
C
TP
Whenever
DP
T′
s/he
T
VP
V′
V
leaves
DP
DP
D′
John
D
’s

In the matrix subject configuration (the upper tree in (8)), the dependency is linking the cataphor to *John*, so the gender (and number) properties of *John* can now be evaluated for fit (Kazanina et al., 2007; van Gompel and Liversedge, 2003). In the case where the gender matches, processing should proceed without additional cost. In the case where the gender mismatches, there will be an associated reading time slowdown as the dependency is "undone" and reanalyzed. Since gender mismatch is the condition of interest for the current description, the cataphor will be shown as *she* for the remainder of the examples.

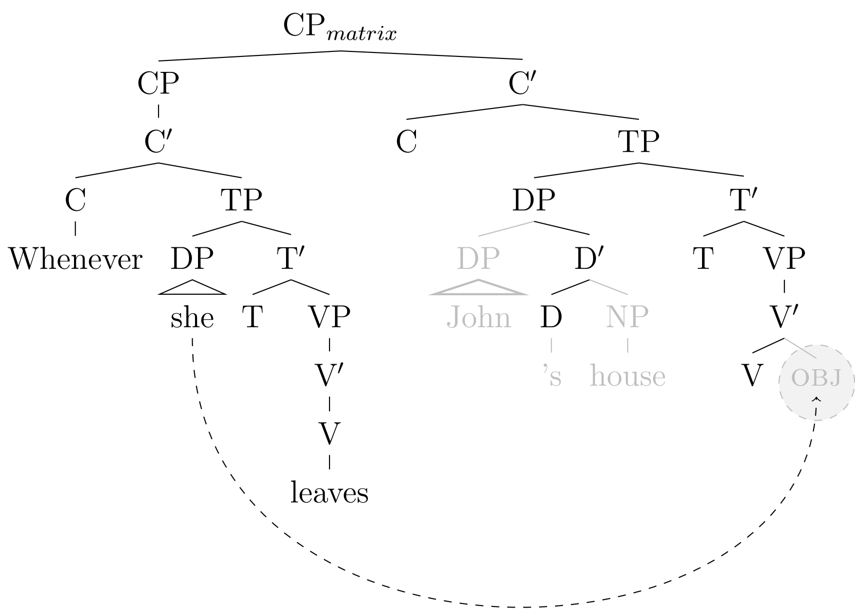
In the lower tree in (9), the embedded object parse, there is no processing cost of integrating *John's* because the tail of the dependency has not been evaluated yet. Furthermore, the position of *house* is not checked because it is also c-commanded by the cataphor. In contrast, in the matrix subject parse in the upper tree, *house* is in a position that makes it a candidate antecedent. The reading times at this position do not show that there is any processing cost associated with reading this position. This could mean that despite the potential for the dependency to form between *she* and *house*, it is not considered. I speculate that this could be due to either of two reasons. If *house* is temporally too close to *John*, the parser may not have time to dissolve the original link and reform it at this new position before it has processed the animacy properties of *house*. Alternatively, the dependency is dissolved and reformed at the position of *house*, but due to this word occurring immediately preceding another critical region, the effects from "undoing" this link yet again are overshadowed by the effect due to the parser encountering the matrix verb (which will be discussed further on). Anyhow, the lack of slowdown observed at this word in the Match conditions as compared to the Mismatch conditions is not inconsistent with this account.

(9)


CP$_{matrix}$
CP
C′
C
TP
C′
C
TP
Whenever
DP
T′
DP
D′
T
VP
she
T
VP
DP
John
D
NP
V′
V′
V
’s
house
V
V
leaves
CP$_{matrix}$
CP
C′
C′
C
TP
C
TP
DP
T′
Whenever
DP
T′
D′
T
VP
she
T
VP
D
V′
V′
V
V
DP
leaves
DP
D′
John
D
NP
’s
house

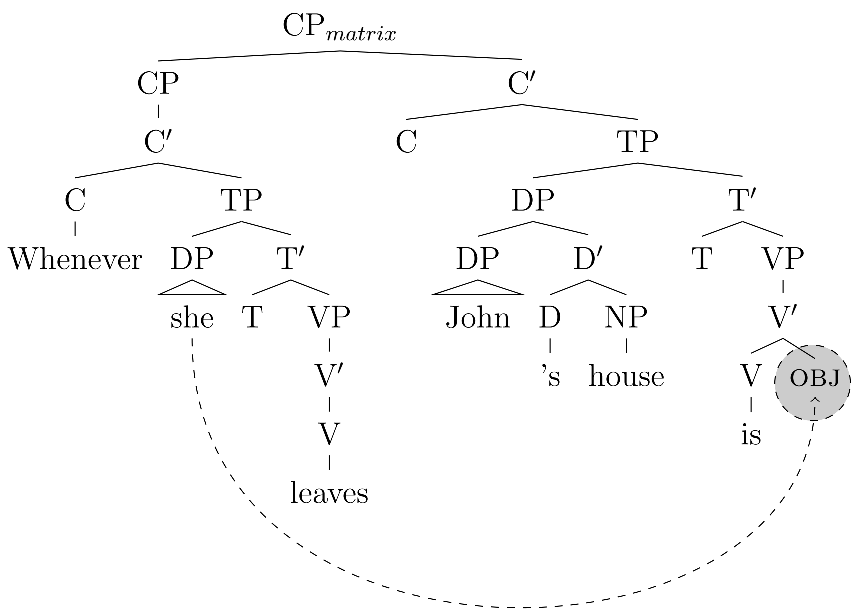
In (10), the coreference dependency has not yet linked the cataphor to an acceptable antecedent. The drive to complete the dependency is strong, so the parser predicts a location for which it has no bottom-up evidence, albeit building an underspecified structure there. This only happens because of the so-called hyper-active nature of dependency formation (Omaki et al., 2015), differentiating this situation from (2), in which there was a potential location for the antecedent in the obligatory structure that the parser (pre-)built.

(10)



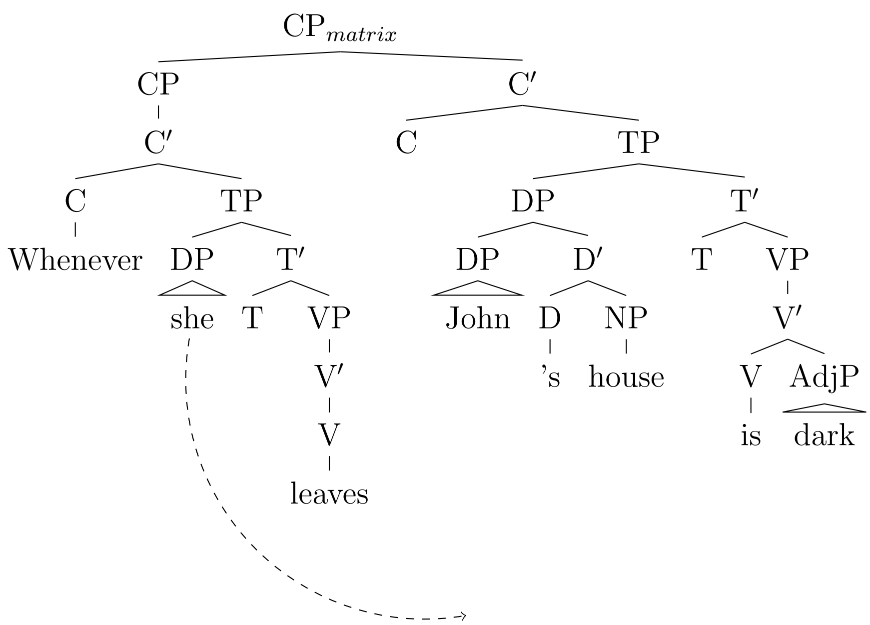
The matrix verb disambiguates the sentence, which may lead to collapsing the two representations back into one (although I will do so primarily for convenience and not make strong claims as to the theoretical motivations for doing so). The parser may realize the embedded object parse is no longer viable when it expects a matrix subject (a likely DP) and encounters a verb, which is the wrong syntactic category. The reading time slowdown observed at the matrix verb and its spillover region could thus be due to a surprise from identifying a verb instead of a subject or the effort needed to prune the embedded object parse (or otherwise reanalyze the representations stored in memory).

(11)



In (12), the parser has identified *is* as the matrix verb, and the matrix subject parse has been adopted. Already, the object position has been constructed and identified as a potential site for the tail of the dependency. Note that this is only the case for the gender mismatched conditions, and not the matched ones. In the gender matched conditions and conditions in which antecedent search was triggered, there is no impetus to predict this location. The prediction of this location means that the processing cost of integrating the verb is reduced since retrieving the lexical item and its associated (transitive) verb frame is boosted by the pre-built structure. This shows up as a reading time speedup in the Mismatch condition.

(12)



At the matrix object position, I found no effects attributable to the failure of the dependency to find an acceptable antecedent. Unlike at *house*, there is a much clearer explanation. In none of the sentences tested in this dissertation was the phrase following the matrix verb a noun phrase. Assuming that the parser has immediate access to the broad syntactic category of an incoming word, it should be able to quickly identify that the matrix object was not a feasibly antecedent and "undo" the dependency before it had identified a specific position with which to link the cataphor. This concludes the positions in which the parser can predict an antecedent without building completely unwarranted structure. I suggest that this pressure to find the tail of an optional antecedent must also decay with time or distance (Gibson and Pearlmutter, 2000; Sturt and Lombardo, 2005). By the time the parser has reached the end of this clause, the pressure may either have been weakened enough to no longer support an active search, or the parser may have exhausted the predictable positions to posit.

(13) Whenever she leaves John's house is dark but Mary doesn't mind.

In either case, when the parser encounters a conjoined clause, as underlined in (13), it may restart its predictions. I found some suggestive but not significant trends at the location of *Mary* in this example, which would be consistent with the parser checking the gender property of *Mary* for coreference in the Mismatch condition. However, the experiments were not designed to test this location for dependency formation, so any statements about the parser's behavior at this point are speculative at best.

5.3. Conclusions

This dissertation set out to distinguish how the processes of dependency formation and local attachment would interact when the parser was dealing with an ambiguous region. Dependency formation, and specifically dependencies formed between a cataphoric pronoun and an antecedent occurring in a (temporarily) structurally ambiguous region appears to have a significant influence on the parser. In fact, the influence exerted by the parser's active search for the tail of the dependency was strong enough to show reading time effects in sentences where none would be traditionally expected. That is, in sentences where the biases of local attachment and verb frame frequency were in collusion toward the embedded object (i.e., transitive) parse of the ambiguous verb, the parser still considered the parse in which the ambiguous DP was a candidate antecedent, despite all the probabilistic information to the contrary. This suggests that the dependency search process is not only a strong influence, but one that the parser relies on to disambiguate or reduce processing cost – even more so than probabilistic information from lexical properties.

Based on previous studies, it was not initially clear whether a serial or parallel model of the parser would be appropriate to assume. Since probabilistic serial models, lexicalist parallel models, and other race-based parallel models make so many similar predictions, it was prudent to start off by laying out how a few models of the parser would function in the type of ambiguity used in this dissertation. To do so, I discussed the ways in which ambiguous verb frames influenced the parser's behavior, and reviewed how the evidence is murky as to whether the parser makes use of verb frame statistics in the initial building of ambiguous structures. I also examined how dependency formation has been used to study the parser's behavior in ambiguity. Wh-dependency formation has consistently been shown to have a strong influence on the parser, whereby the parser will posit a gap to close the dependency well before the position is reached (Traxler and Pickering, 1996). However, coreference dependencies, such as those triggered by cataphora, are not obligatory as wh-dependencies are, thus may not have as strong an influence on initial structure building (Kwon and Sturt, 2014). A parallel model of the parser may, in the case, be able to account for this by building both a parse that is influenced by the dependency search and one that is not, for instance.

The results from this dissertation led to a second surprising conclusion. In addition to observing effects that could only be attributed to the construction of a matrix subject parse during initial structure building, there were also effects that were attributable to the embedded object parse. Moreover, these effects were not mutually exclusive, thus suggesting that both the matrix subject parse and the embedded object parse were being maintained in working memory at the same time. This is strong evidence for a parallel parser because no serial parser of any type would predict that the parser would construct a parse that was successful, but reanalyze it for no reason and construct an ungrammatical parse, and then reanalyze the sentence a third time in order to come to a final coherent interpretation. This is what would be necessary to account for the effects observed due to the matrix subject parse at the ambiguous NP, then later the embedded object parse at the matrix verb and spillover region (and then later again from the matrix subject parse, in late reading times attributable to the parser anticipating an object position of the matrix verb). Given that, I can offer no support of serial parsing models from my analyses. Since this study was not specifically designed to test the serial/parallel distinction, I cannot offer strong support for any subtype of parallel parsing model, either. What I can offer, however, is unequivocal and replicated data showing that the parser can be influenced by two independent and conflicting parses in the same trial which is something that only parallel or multiple output models of the parser can predict.

5.4. Future directions

This dissertation serves as a jumping off point for further study of what sources of information may influence the parser's behavior during ambiguity resolution. I have only examined three influences on the parser so far: the in-built preference for local attachment, a bias introduced by verb frame frequency, and the influence exerted from an active search for the tail of a dependency. I discussed previous literature that made use of plausibility information, and thematic role configurations (separate from lexical frames). Further influences that may bias the parser include myriad unmentioned sources of information.

For instance, I have only investigated the parser's behavior during reading in this study. During listening, additional information is provided by the acoustic signal, including prosodic grouping. In order to test the relative strength of prosody on sentence processing, one possible follow-up study could include a listening component in which prosodic grouping was manipulated. Two possible clause boundaries between the embedded clause and the matrix clause, adapted from the stimuli used in this study may be marked with prosodic breaks, as in (14). The octothorp marks the sentence with infelicitous or incongruous prosody, in which a late closure (embedded object) is signaled by the subsequent late prosodic break, but the sentence is ultimately of the early closure (matrix subject) representation. Since there is a strong intuition that prosodic boundaries disambiguate this type of sentence, examining whether this prevents a coreference dependency from forming would give significant insight into how the parser functions during listening comprehension, as well.

(14) a. Whenever he left || John's house is dark.

b. #Whenever he left John's house || is dark.

This area of research is rich and extremely important to explore. There has been much investigation of how prosodic boundaries influence ambiguity resolution, but by continuing to use gender matching in coreference dependency formation as a probe for rich and detailed structure building will give an added clarity to the (quite often) messy side of prosody. Coreference dependencies are governed by strict grammatical constraints. Thus, if prosodic boundary production and interpretation is highly variable, it may not be a strong influence on the parser. On the other hand, there is much evidence suggesting that the parser (during auditory comprehension) does rely on prosodic boundaries for disambiguation. Future work in this domain can help resolved the question of how the parser prioritizes or uses various sources of information.

5.5. Summary

This dissertation offers three main conclusions. During reading comprehension, the parser is strongly influenced by coreference dependency formation as compared to even the combined influenced from the bias exerted from verb frame frequency information and the local attachment preference seemingly built into the parser. Despite the strength of the influence from this coreference dependency, it is not enough to block the parser from building a representation of the alternate parse that does not allow the coreference dependency to be closed as soon as possible. Rather, both alternatives are built and maintained in memory, at least until the sentence is fully disambiguated, indicating that the parser has the ability to function in a manner described by parallel models. Finally, there is evidence that the parser is very eager to find the antecedent of the coreference dependency. So much so, that when its first choice of candidate antecedent fails due to a mismatch in gender, the parser may predict some detailed content or structure of the matrix clause. In this way, another candidate antecedent may serve as the landing site of the dependency tail when the first is discarded. From these observations, coreference dependency formation may be considered a strong influence on the parser during comprehension.

References

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, 18(1):129–144.

Altmann, G. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Altmann, G. and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3):191–238.

Arppe, A. and Järvikivi, J. (2007). Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3(2):131–159.

Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". *R package version*, 1.4.1.

Badecker, W. and Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4):748–769.

Baldwin, R. S. and Coady, J. M. (1978). Psycholinguistic approaches to a theory of punctuation. *Journal of Literacy Research*, 10(4):363–375.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1):1–48.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In Hayes, R., editor, *Cognition and Language Development*, pages 279–362. Wiley & Sons, New York.

Blodgett, A. R. (2004). *The interaction of prosodic phrasing, verb bias, and plausibility during spoken sentence comprehension*. PhD thesis, The Ohio State University.

Boland, J. E. and Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the Great Divide. *Cognition*, 58(3):309–320.

Boland, J. E., Tanenhaus, M. K., Carlson, G., and Garnsey, S. M. (1989). Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, 18(6):563–576.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., and Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language*, 34(6):774–806.

Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications, Dordrecht.

Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4):368–407.

Connine, C., Ferreira, F., Jones, C., Clifton, C., and Frazier, L. (1984). Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research*, 13(4):307–319.

Cowart, W. and Cairns, H. S. (1987). Evidence for an anaphoric mechanism within syntactic processing: Some reference relations defy semantic and pragmatic constraints. *Memory & Cognition*, 15(4):318–331.

Crain, S. and Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In Dowty, D. R., Karttunen, L., and Zwicky, A. M., editors, *Natural Language Parsing*, pages 320–358. Cambridge University Press.

Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990-present. http://corpus.byu.edu/coca/.

Erlewine, M. Y. and Kotek, H. (2015). A streamlined approach to online linguistic surveys. *Natural Language & Linguistic Theory*, pages 1–15.

Ferreira, F. and Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):555–568.

Ferreira, F. and Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6):725–745.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9(3):427–473.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, France.

Fodor, J. D. and Frazier, L. (1983). Local attachment in a one-stage parser. Unpublished manuscript.

Fodor, J. D. and Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of*

Psycholinguistic Research, 23(5):407–434.

Frazier, L. (1979). *On Comprehending sentences: Syntactic parsing strategies*. PhD thesis, University of Connecticut.

Frazier, L. and Clifton, Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, 4(2):93–126.

Frazier, L. and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4):291–325.

Frazier, L. and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2):178–210.

Frazier, L. and Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26(5):505–526.

Frazier, M., Ackerman, L., Baumann, P., Potter, D., and Yoshida, M. (2015). *Wh*-filler-gap dependency formation guides reflexive antecedent search. *Frontiers in Psychology*, 6.

Fujisaki, H. (1997). Prosody, models, and spontaneous speech. In *Computing Prosody*, pages 27–42. Springer.

Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Gahl, S., Jurafsky, D., and Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons.

Behavior Research Methods, Instruments, & Computers, 36(3):432–443.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., and Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1):58–93.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54(3):363–388.

Gibson, E. and Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, 29(2):231–240.

Gibson, E., Schütze, C. T., and Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, 25(1):59–92.

Gibson, E. A. F. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. PhD thesis, School of Computer Science, Carnegie Mellon University.

Gorrell, P. G. (1987). *Studies of human syntactic processing: Ranked-parallel versus serial models*. PhD thesis, University of Connecticut.

Grillo, N. (2012). Local and universal. In Bianchi, V. and Chesi, C., editors, *Enjoy linguistics!: Papers offered to Luigi Rizzi on the occasion of his 60th birthday*, pages 234–245. CISCL Press.

Juliano, C. and Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research*, 23(6):459–471.

Kazanina, N. (2005). *The acquisition and processing of backwards anaphora*. PhD thesis, University of Maryland.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., and Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56(3):384–409.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2(1):15–47.

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, pages 1178–1183.

Kurtzman, H. S. (1985). *Studies in syntactic ambiguity resolution*, volume 306. Indiana University Linguistics Club.

Kwon, N. and Sturt, P. (2014). The use of control information in dependency formation: An eye-tracking study. *Journal of Memory and Language*, 73:59–80.

Lapata, M., Keller, F., and Im Walde, S. S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, 30(4):419–435.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In *Reanalysis in Sentence Processing*, pages 247–285. Springer.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29(2):241–248.

Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, 19(1):1–36.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703.

McRae, K. and Matsuki, K. (2013). Constraint-based models of sentence processing. In van Gompel, R. P. G., editor, *Sentence Processing*, pages 51–77. Psychology Press, New York, NY.

Miller, G. A. and Chomsky, N. (1963). Finitary models of language users. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology*, pages 419–491. Wiley, New York, NY.

Niikuni, K. and Muramoto, T. (2014). Effects of punctuation on the processing of temporarily ambiguous sentences in Japanese. *Japanese Psychological Research*, 56(3):275–287.

Omaki, A., Lau, E. F., White, I. D., Dakan, M. L., Apple, A., and Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, 6.

Pearlmutter, N. J. and Mendelsohn, A. A. (1998). Serial versus parallel sentence processing. In *The Eleventh Annual CUNY Sentence Processing Conference*, New Brunswick, NJ. Rutgers University.

Pearlmutter, N. J. and Mendelsohn, A. A. (1999). Serial versus parallel sentence comprehension. Unpublished manuscript.

Phillips, C. (1995). Right association in parsing and grammar. *MIT Working Papers in Linguistics*, 26:37–93.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82(4):795–823.

Phillips, C. and Gibson, E. (1997). On the strength of the local attachment preference. *Journal of Psycholinguistic Research*, 26(3):323–346.

Pickering, M. J. and Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4):633–651.

Pritchett, B. L. (1992). Parsing with grammar: Islands, heads, and garden paths. In *Island Constraints*, pages 321–349. Springer.

Qian, Z. and Garnsey, S. (2015). Why do readers answer questions wrongly after reading garden-path sentences? In Noelle, D. C., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., and Maglio, P. P., editors, *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, Austin, TX. Cognitive Science Society.

Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, Massachusetts Institute of Technology.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., and Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2):104–120.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):550–569.

Staub, A., Clifton, C., and Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language*, 54(3):389–406.

Stowe, L. A. (1986). Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3):227–245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3):542–562.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105(2):477–488.

Sturt, P. and Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2):291–305.

Tabor, W., Galantucci, B., and Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4):355–370.

Tabor, W. and Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):431–450.

Tanenhaus, M. K., Boland, J., Garnsey, S. M., and Carlson, G. N. (1989). Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistic Research*, 18(1):37–50.

Traxler, M. J. and Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35(3):454–475.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35(4):566–585.

Trueswell, J. C. and Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language*, 39(1):102–123.

Trueswell, J. C. and Tanenhaus, M. K. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. In Clifton, Jr, C., Frazier, L., and Rayner, K., editors, *Perspectives on Sentence Processing*, pages 155–179. Lawrence Erlbaum Assoc.

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3):528–553.

van Gompel, R. P. G., Arai, M., and Pearson, J. (2012). The representation of mono- and intransitive structures. *Journal of Memory and Language*, 66(2):384–406.

van Gompel, R. P. G. and Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1):128–139.

van Gompel, R. P. G., Pickering, M. J., and Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In Kennedy, A., Radach, R., Heller, D., and Pynte, J., editors, *Reading as a Perceptual Process*, pages 621–648. Oxford: Elsevier.

van Gompel, R. P. G., Pickering, M. J., and Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45(2):225–258.

Wagers, M. W. and Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, 45(2):395–433.

Wagers, M. W. and Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, 67(7):1274–1304.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1):1–191.

Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation.* PhD thesis, University of Maryland, College Park.

Appendix

Equibiased verb stimuli

(1) a. Any time she was about to
quickly worship, Pamela's sacred idol was facing eastward,
but Christine couldn't correct the position of the pedestal alone.

b. Any time she was about to
quickly worship Pamela's sacred idol was facing eastward,
but Christine couldn't correct the position of the pedestal alone.

c. Any time he was about to
quickly worship, Pamela's sacred idol was facing eastward,
but Joshua couldn't correct the position of the pedestal alone.

d. Any time he was about to
quickly worship Pamela's sacred idol was facing eastward,
but Joshua couldn't correct the position of the pedestal alone.

e. Any time Joshua was about to
quickly worship, Pamela's sacred idol was facing eastward,
but Christine couldn't correct the position of the pedestal alone.

f. Any time Joshua was about to
quickly worship Pamela's sacred idol was facing eastward,
but Christine couldn't correct the position of the pedestal alone.

(2) a. Any time she was about to
finally phone, Carolyn's business committee was chatting continuously,
but Kimberly knew how to politely interrupt a long conversation.

b. Any time she was about to
finally phone Carolyn's business committee was chatting continuously,
but Kimberly knew how to politely interrupt a long conversation.

c. Any time he was about to
finally phone, Carolyn's business committee was chatting continuously,
but Jeffrey knew how to politely interrupt a long conversation.

d. Any time he was about to
finally phone Carolyn's business committee was chatting continuously,
but Jeffrey knew how to politely interrupt a long conversation.

e. Any time Jeffrey was about to
finally phone, Carolyn's business committee was chatting continuously,
but Kimberly knew how to politely interrupt a long conversation.

f. Any time Jeffrey was about to
finally phone Carolyn's business committee was chatting continuously,
but Kimberly knew how to politely interrupt a long conversation.

(3) a. Any time she was about to
gently dust, Elise's antique bookshelf was recently cleaned,
but Kathleen didn't know what else to do to help out.

b. Any time she was about to
gently dust Elise's antique bookshelf was recently cleaned,
but Kathleen didn't know what else to do to help out.

c. Any time he was about to
gently dust, Elise's antique bookshelf was recently cleaned,
but Anthony didn't know what else to do to help out.

d. Any time he was about to
gently dust Elise's antique bookshelf was recently cleaned,
but Anthony didn't know what else to do to help out.

e. Any time Anthony was about to
gently dust, Elise's antique bookshelf was recently cleaned,
but Kathleen didn't know what else to do to help out.

f. Any time Anthony was about to
gently dust Elise's antique bookshelf was recently cleaned,
but Kathleen didn't know what else to do to help out.

(4) a. Any time she was about to
carefully play, Martha's elaborate card game was clearly rigged,
but Michelle had a sneaky plan to even out the odds.

b. Any time she was about to
carefully play Martha's elaborate card game was clearly rigged,
but Michelle had a sneaky plan to even out the odds.

c. Any time he was about to
carefully play, Martha's elaborate card game was clearly rigged,
but Arthur had a sneaky plan to even out the odds.

d. Any time he was about to
carefully play Martha's elaborate card game was clearly rigged,
but Arthur had a sneaky plan to even out the odds.

e. Any time Arthur was about to
carefully play, Martha's elaborate card game was clearly rigged,
but Michelle had a sneaky plan to even out the odds.

f. Any time Arthur was about to
carefully play Martha's elaborate card game was clearly rigged,
but Michelle had a sneaky plan to even out the odds.

(5) a. Any time he was about to
finally try, Stephen's jigsaw puzzle was missing pieces,
but Marcus was able to put the rest of it together.

b. Any time he was about to
finally try Stephen's jigsaw puzzle was missing pieces,
but Marcus was able to put the rest of it together.

c. Any time she was about to
finally try, Stephen's jigsaw puzzle was missing pieces,
but Shirley was able to put the rest of it together.

d. Any time she was about to
finally try Stephen's jigsaw puzzle was missing pieces,
but Shirley was able to put the rest of it together.

e. Any time Shirley was about to
finally try, Stephen's jigsaw puzzle was missing pieces,
but Marcus was able to put the rest of it together.

f. Any time Shirley was about to
finally try Stephen's jigsaw puzzle was missing pieces,
but Marcus was able to put the rest of it together.

(6) a. Any time he was about to
deftly kick, Christopher's injured ankle was feeling weak,
but Michael was highly proficient in self-defense.

b. Any time he was about to
deftly kick Christopher's injured ankle was feeling weak,
but Michael was highly proficient in self-defense.

c. Any time she was about to
deftly kick, Christopher's injured ankle was feeling weak,
but Jennifer was highly proficient in self-defense.

d. Any time she was about to
deftly kick Christopher's injured ankle was feeling weak,
but Jennifer was highly proficient in self-defense.

e. Any time Jennifer was about to
deftly kick, Christopher's injured ankle was feeling weak,
but Michael was highly proficient in self-defense.

f. Any time Jennifer was about to
deftly kick Christopher's injured ankle was feeling weak,
but Michael was highly proficient in self-defense.

(7) a. Any time he was trying to
casually drink, Matthew's signature cocktail was intensely sweet,
but Gregory pretended to enjoy it to seem polite.

b. Any time he was trying to
casually drink Matthew's signature cocktail was intensely sweet,
but Gregory pretended to enjoy it to seem polite.

c. Any time she was trying to
casually drink, Matthew's signature cocktail was intensely sweet,
but Deborah pretended to enjoy it to seem polite.

d. Any time she was trying to
casually drink Matthew's signature cocktail was intensely sweet,
but Deborah pretended to enjoy it to seem polite.

e. Any time Deborah was trying to
casually drink, Matthew's signature cocktail was intensely sweet,
but Gregory pretended to enjoy it to seem polite.

f. Any time Deborah was trying to
casually drink Matthew's signature cocktail was intensely sweet,
but Gregory pretended to enjoy it to seem polite.

(8) a. Any time he was trying to
deftly sketch, Thomas's preferred landscape was depressingly boring,
but George knew just how to make it look beautiful.

b. Any time he was trying to
deftly sketch Thomas's preferred landscape was depressingly boring,
but George knew just how to make it look beautiful.

c. Any time she was trying to
deftly sketch, Thomas's preferred landscape was depressingly boring,
but Melissa knew just how to make it look beautiful.

d. Any time she was trying to
deftly sketch Thomas's preferred landscape was depressingly boring,
but Melissa knew just how to make it look beautiful.

e. Any time Melissa was trying to
deftly sketch, Thomas's preferred landscape was depressingly boring,
but George knew just how to make it look beautiful.

f. Any time Melissa was trying to
deftly sketch Thomas's preferred landscape was depressingly boring,
but George knew just how to make it look beautiful.

(9) a. Any time she was trying to
carefully knit, Jessica's winter sweater was somewhat oversized,
but Patricia thought it might shrink in the laundry.

b. Any time she was trying to
carefully knit Jessica's winter sweater was somewhat oversized,
but Patricia thought it might shrink in the laundry.

c. Any time he was trying to
carefully knit, Jessica's winter sweater was somewhat oversized,
but Raymond thought it might shrink in the laundry.

d. Any time he was trying to
carefully knit Jessica's winter sweater was somewhat oversized,
but Raymond thought it might shrink in the laundry.

e. Any time Raymond was trying to
carefully knit, Jessica's winter sweater was somewhat oversized,
but Patricia thought it might shrink in the laundry.

f. Any time Raymond was trying to
carefully knit Jessica's winter sweater was somewhat oversized,
but Patricia thought it might shrink in the laundry.

(10) a. Any time he was trying to
gently wash, Timothy's soccer jersey was particularly filthy,
but David didn't have time to scrub all the dirt out.

b. Any time he was trying to
gently wash Timothy's soccer jersey was particularly filthy,
but David didn't have time to scrub all the dirt out.

c. Any time she was trying to
gently wash, Timothy's soccer jersey was particularly filthy,
but Sarah didn't have time to scrub all the dirt out.

d. Any time she was trying to
gently wash Timothy's soccer jersey was particularly filthy,
but Sarah didn't have time to scrub all the dirt out.

e. Any time Sarah was trying to
gently wash, Timothy's soccer jersey was particularly filthy,
but David didn't have time to scrub all the dirt out.

f. Any time Sarah was trying to
gently wash Timothy's soccer jersey was particularly filthy,
but David didn't have time to scrub all the dirt out.

(11) a. Any time he was trying to
casually watch, Jeremy's energetic puppy was very rambunctious,
but Daniel was great at training young dogs.

b. Any time he was trying to
casually watch Jeremy's energetic puppy was very rambunctious,
but Daniel was great at training young dogs.

c. Any time she was trying to
casually watch, Jeremy's energetic puppy was very rambunctious,
but Carol was great at training young dogs.

d. Any time she was trying to
casually watch Jeremy's energetic puppy was very rambunctious,
but Carol was great at training young dogs.

e. Any time Carol was trying to
casually watch, Jeremy's energetic puppy was very rambunctious,
but Daniel was great at training young dogs.

f. Any time Carol was trying to
casually watch Jeremy's energetic puppy was very rambunctious,
but Daniel was great at training young dogs.

(12) a. Any time she was trying to
quickly cook, Stephanie's requested dinner was extremely complicated,
but Nancy was very good at multitasking in the kitchen.

b. Any time she was trying to
quickly cook Stephanie's requested dinner was extremely complicated,
but Nancy was very good at multitasking in the kitchen.

c. Any time he was trying to
quickly cook, Stephanie's requested dinner was extremely complicated,
but Jason was very good at multitasking in the kitchen.

d. Any time he was trying to
quickly cook Stephanie's requested dinner was extremely complicated,
but Jason was very good at multitasking in the kitchen.

e. Any time Jason was trying to
quickly cook, Stephanie's requested dinner was extremely complicated,
but Nancy was very good at multitasking in the kitchen.

f. Any time Jason was trying to
quickly cook Stephanie's requested dinner was extremely complicated,
but Nancy was very good at multitasking in the kitchen.

(13) a. Whenever she was about to
finally sway, Margaret's original opinion was logically sound,
but Elizabeth was a very persuasive debater.

b. Whenever she was about to
finally sway Margaret's original opinion was logically sound,
but Elizabeth was a very persuasive debater.

c. Whenever he was about to
finally sway, Margaret's original opinion was logically sound,
but Walter was a very persuasive debater.

d. Whenever he was about to
finally sway Margaret's original opinion was logically sound,
but Walter was a very persuasive debater.

e. Whenever Walter was about to
finally sway, Margaret's original opinion was logically sound,
but Elizabeth was a very persuasive debater.

f. Whenever Walter was about to
finally sway Margaret's original opinion was logically sound,
but Elizabeth was a very persuasive debater.

(14) a. Whenever he was about to
carefully escape, Harold's medieval dungeon was impossibly confusing,
but Scott had a map of the only route out.

b. Whenever he was about to
carefully escape Harold's medieval dungeon was impossibly confusing,
but Scott had a map of the only route out.

c. Whenever she was about to
carefully escape, Harold's medieval dungeon was impossibly confusing,
but Marie had a map of the only route out.

d. Whenever she was about to
carefully escape Harold's medieval dungeon was impossibly confusing,
but Marie had a map of the only route out.

e. Whenever Marie was about to
carefully escape, Harold's medieval dungeon was impossibly confusing,
but Scott had a map of the only route out.

f. Whenever Marie was about to
carefully escape Harold's medieval dungeon was impossibly confusing,
but Scott had a map of the only route out.

(15) a. Whenever she was about to
deftly fight, Donna's formidable opponent was heavily armed,
but Sharon was able to avoid receiving injuries.

b. Whenever she was about to
deftly fight Donna's formidable opponent was heavily armed,
but Sharon was able to avoid receiving injuries.

c. Whenever he was about to
deftly fight, Donna's formidable opponent was heavily armed,
but Henry was able to avoid receiving injuries.

d. Whenever he was about to
deftly fight Donna's formidable opponent was heavily armed,
but Henry was able to avoid receiving injuries.

e. Whenever Henry was about to
deftly fight, Donna's formidable opponent was heavily armed,
but Sharon was able to avoid receiving injuries.

f. Whenever Henry was about to
deftly fight Donna's formidable opponent was heavily armed,
but Sharon was able to avoid receiving injuries.

(16) a. Whenever he was about to
gently roll, Andrew's lucky bowling ball was poorly balanced,
but Steven could compensate for the awkward weighting.

b. Whenever he was about to
gently roll Andrew's lucky bowling ball was poorly balanced,
but Steven could compensate for the awkward weighting.

c. Whenever she was about to
gently roll, Andrew's lucky bowling ball was poorly balanced,
but Barbara could compensate for the awkward weighting.

d. Whenever she was about to
gently roll Andrew's lucky bowling ball was poorly balanced,
but Barbara could compensate for the awkward weighting.

e. Whenever Barbara was about to
gently roll, Andrew's lucky bowling ball was poorly balanced,
but Steven could compensate for the awkward weighting.

f. Whenever Barbara was about to
gently roll Andrew's lucky bowling ball was poorly balanced,
but Steven could compensate for the awkward weighting.

(17) a. Whenever he was about to
quickly swing, William's tennis racquet was perfectly aimed,
but Kevin nevertheless expected the opponent to win.

b. Whenever he was about to
quickly swing William's tennis racquet was perfectly aimed,
but Kevin nevertheless expected the opponent to win.

c. Whenever she was about to
quickly swing, William's tennis racquet was perfectly aimed,
but Catherine nevertheless expected the opponent to win.

d. Whenever she was about to
quickly swing William's tennis racquet was perfectly aimed,
but Catherine nevertheless expected the opponent to win.

e. Whenever Catherine was about to
quickly swing, William's tennis racquet was perfectly aimed,
but Kevin nevertheless expected the opponent to win.

f. Whenever Catherine was about to
quickly swing William's tennis racquet was perfectly aimed,
but Kevin nevertheless expected the opponent to win.

(18) a. Whenever he was about to
casually sweep, Brian's brickwork patio was already immaculate,
but Nathan felt compelled to go through the motions.

b. Whenever he was about to
casually sweep Brian's brickwork patio was already immaculate,
but Nathan felt compelled to go through the motions.

c. Whenever she was about to
casually sweep, Brian's brickwork patio was already immaculate,
but Maria felt compelled to go through the motions.

d. Whenever she was about to
casually sweep Brian's brickwork patio was already immaculate,
but Maria felt compelled to go through the motions.

e. Whenever Maria was about to
casually sweep, Brian's brickwork patio was already immaculate,
but Nathan felt compelled to go through the motions.

f. Whenever Maria was about to
casually sweep Brian's brickwork patio was already immaculate,
but Nathan felt compelled to go through the motions.

(19) a. Whenever he was trying to
finally sail, Douglas's seaworthy dinghy was fully prepared,
but Edward was too scared of drowning to set out to sea.

b. Whenever he was trying to
finally sail Douglas's seaworthy dinghy was fully prepared,
but Edward was too scared of drowning to set out to sea.

c. Whenever she was trying to
finally sail, Douglas's seaworthy dinghy was fully prepared,
but Sandra was too scared of drowning to set out to sea.

d. Whenever she was trying to
finally sail Douglas's seaworthy dinghy was fully prepared,
but Sandra was too scared of drowning to set out to sea.

e. Whenever Sandra was trying to
finally sail, Douglas's seaworthy dinghy was fully prepared,
but Edward was too scared of drowning to set out to sea.

f. Whenever Sandra was trying to
finally sail Douglas's seaworthy dinghy was fully prepared,
but Edward was too scared of drowning to set out to sea.

(20) a. Whenever she was trying to
deftly race, Cynthia's prize-winning horse was looking sickly,
but Rebecca thought it would be okay to enter it in the race.

b. Whenever she was trying to
deftly race Cynthia's prize-winning horse was looking sickly,
but Rebecca thought it would be okay to enter it in the race.

c. Whenever he was trying to
deftly race, Cynthia's prize-winning horse was looking sickly,
but Peter thought it would be okay to enter it in the race.

d. Whenever he was trying to
deftly race Cynthia's prize-winning horse was looking sickly,
but Peter thought it would be okay to enter it in the race.

e. Whenever Peter was trying to
deftly race, Cynthia's prize-winning horse was looking sickly,
but Rebecca thought it would be okay to enter it in the race.

f. Whenever Peter was trying to
deftly race Cynthia's prize-winning horse was looking sickly,
but Rebecca thought it would be okay to enter it in the race.

(21) a. Whenever he was trying to
gently start, Joseph's chiropractic massage was excessively painful,
but Robert was a skilled acupuncturist.

b. Whenever he was trying to
gently start Joseph's chiropractic massage was excessively painful,
but Robert was a skilled acupuncturist.

c. Whenever she was trying to
gently start, Joseph's chiropractic massage was excessively painful,
but Janet was a skilled acupuncturist.

d. Whenever she was trying to
gently start Joseph's chiropractic massage was excessively painful,
but Janet was a skilled acupuncturist.

e. Whenever Janet was trying to
gently start, Joseph's chiropractic massage was excessively painful,
but Robert was a skilled acupuncturist.

f. Whenever Janet was trying to
gently start Joseph's chiropractic massage was excessively painful,
but Robert was a skilled acupuncturist.

(22) a. Whenever she was trying to
quickly pass, Dorothy's final exam was incredibly tricky,
but Amanda had studied hard for the last week.

b. Whenever she was trying to
quickly pass Dorothy's final exam was incredibly tricky,
but Amanda had studied hard for the last week.

c. Whenever he was trying to
quickly pass, Dorothy's final exam was incredibly tricky,
but Larry had studied hard for the last week.

d. Whenever he was trying to
quickly pass Dorothy's final exam was incredibly tricky,
but Larry had studied hard for the last week.

e. Whenever Larry was trying to
quickly pass, Dorothy's final exam was incredibly tricky,
but Amanda had studied hard for the last week.

f. Whenever Larry was trying to
quickly pass Dorothy's final exam was incredibly tricky,
but Amanda had studied hard for the last week.

(23) a. Whenever she was trying to
carefully drive, Brenda's dilapidated minivan was rattling loudly,
but Maria was confident that it could make the trip.

b. Whenever she was trying to
carefully drive Brenda's dilapidated minivan was rattling loudly,
but Maria was confident that it could make the trip.

c. Whenever he was trying to
carefully drive, Brenda's dilapidated minivan was rattling loudly,
but Kenneth was confident that it could make the trip.

d. Whenever he was trying to
carefully drive Brenda's dilapidated minivan was rattling loudly,
but Kenneth was confident that it could make the trip.

e. Whenever Kenneth was trying to
carefully drive, Brenda's dilapidated minivan was rattling loudly,
but Maria was confident that it could make the trip.

f. Whenever Kenneth was trying to
carefully drive Brenda's dilapidated minivan was rattling loudly,
but Maria was confident that it could make the trip.

(24) a. Whenever she was trying to
casually hum, Annie's favorite melody was hauntingly beautiful,
but Karen couldn't do it justice all alone.

b. Whenever she was trying to
casually hum Annie's favorite melody was hauntingly beautiful,
but Karen couldn't do it justice all alone.

c. Whenever he was trying to
casually hum, Annie's favorite melody was hauntingly beautiful,
but Jimmy couldn't do it justice all alone.

d. Whenever he was trying to
casually hum Annie's favorite melody was hauntingly beautiful,
but Jimmy couldn't do it justice all alone.

e. Whenever Jimmy was trying to
casually hum, Annie's favorite melody was hauntingly beautiful,
but Karen couldn't do it justice all alone.

f. Whenever Jimmy was trying to
casually hum Annie's favorite melody was hauntingly beautiful,
but Karen couldn't do it justice all alone.

Transitive biased verb stimuli

(1) a. Any time she was about to
finally visit Pamela's sacred idol was facing eastward
but Christine couldn't correct the position of the pedestal alone.

b. Any time he was about to
finally visit Pamela's sacred idol was facing eastward
but Joshua couldn't correct the position of the pedestal alone.

c. Any time Amy was about to
finally visit Pamela's sacred idol was facing eastward
but Christine couldn't correct the position of the pedestal alone.

(2) a. Any time she was about to
quickly confirm Carolyn's business committee was chatting continuously
but Kimberly knew how to politely interrupt a long conversation.

b. Any time he was about to
quickly confirm Carolyn's business committee was chatting continuously
but Jeffrey knew how to politely interrupt a long conversation.

c. Any time Sue was about to
quickly confirm Carolyn's business committee was chatting continuously
but Kimberly knew how to politely interrupt a long conversation.

d. Any time he was about to
quickly confirm Carolyn's business committee was chatting continuously
but Jeffrey knew how to politely interrupt a long conversation.

(3) a. Any time she was about to
casually save Elise's term paper was still unchanged
but Kathleen didn't know what else to do to help out.

b. Any time he was about to
casually save Elise's term paper was still unchanged
but Anthony didn't know what else to do to help out.

c. Any time Meg was about to
casually save Elise's term paper was still unchanged
but Kathleen didn't know what else to do to help out.

d. Any time Meg was about to
casually save Elise's term paper was still unchanged
but Kathleen didn't know what else to do to help out.

(4) a. Any time she was about to
finally lose Martha's card game was clearly rigged
but Michelle had a sneaky plan to even out the odds.

b. Any time he was about to
finally lose Martha's card game was clearly rigged
but Arthur had a sneaky plan to even out the odds.

c. Any time Ana was about to
finally lose Martha's card game was clearly rigged
but Michelle had a sneaky plan to even out the odds.

d. Any time Ana was about to
finally lose Martha's card game was clearly rigged
but Michelle had a sneaky plan to even out the odds.

(5) a. Any time he was about to
finally buy Stephen's jigsaw puzzle was missing pieces
but Marcus was able to put the rest of it together.

b. Any time she was about to
finally buy Stephen's jigsaw puzzle was missing pieces
but Deborah was able to put the rest of it together.

c. Any time Tom was about to
finally buy Stephen's jigsaw puzzle was missing pieces
but Marcus was able to put the rest of it together.

(6) a. Any time he was about to
deftly attack Christopher's left ankle was feeling weak
but Michael was highly proficient in self-defense.

b. Any time she was about to
deftly attack Christopher's left ankle was feeling weak
but Jennifer was highly proficient in self-defense.

c. Any time Jon was about to
deftly attack Christopher's left ankle was feeling weak
but Michael was highly proficient in self-defense.

(7) a. Any time he was trying to
casually accept Matthew's tropical cocktail was intensely sweet
but Gregory pretended to enjoy it to seem polite.

b. Any time she was trying to
casually accept Matthew's tropical cocktail was intensely sweet
but Shirley pretended to enjoy it to seem polite.

c. Any time Tim was trying to
casually accept Matthew's tropical cocktail was intensely sweet
but Gregory pretended to enjoy it to seem polite.

(8) a. Any time he was trying to
deftly study Thomas's panoramic landscape was depressingly boring
but George knew just how to make it look beautiful.

b. Any time she was trying to
deftly study Thomas's panoramic landscape was depressingly boring
but Melissa knew just how to make it look beautiful.

c. Any time Ben was trying to
deftly study Thomas's panoramic landscape was depressingly boring
but George knew just how to make it look beautiful.

(9) a. Any time she was trying to
carefully mend Jessica's winter sweater was somewhat oversized
but Patricia thought it might shrink in the laundry.

b. Any time he was trying to
carefully mend Jessica's winter sweater was somewhat oversized
but Raymond thought it might shrink in the laundry.

c. Any time Liz was trying to
carefully mend Jessica's winter sweater was somewhat oversized
but Patricia thought it might shrink in the laundry.

(10) a. Any time he was trying to
carefully find Timothy's soccer jersey was particularly filthy
but David didn't have time to scrub all the dirt out.

b. Any time she was trying to
carefully find Timothy's soccer jersey was particularly filthy
but Sarah didn't have time to scrub all the dirt out.

c. Any time Don was trying to
carefully find Timothy's soccer jersey was particularly filthy
but David didn't have time to scrub all the dirt out.

(11) a. Any time he was trying to
casually observe Jeremy's energetic puppy was very rambunctious
but Daniel was great at training young dogs.

b. Any time she was trying to
casually observe Jeremy's energetic puppy was very rambunctious
but Carol was great at training young dogs.

c. Any time Bill was trying to
casually observe Jeremy's energetic puppy was very rambunctious
but Daniel was great at training young dogs.

(12) a. Any time she was trying to
quickly eat Stephanie's requested dinner was extremely complicated
but Nancy was very good at multitasking in the kitchen.

b. Any time he was trying to
quickly eat Stephanie's requested dinner was extremely complicated
but Jason was very good at multitasking in the kitchen.

c. Any time Ann was trying to
quickly eat Stephanie's requested dinner was extremely complicated
but Nancy was very good at multitasking in the kitchen.

(13) a. Whenever she was about to
gently criticize Margaret's political opinion was logically sound
but Elizabeth was a very persuasive debater.

b. Whenever he was about to
gently criticize Margaret's political opinion was logically sound
but Walter was a very persuasive debater.

c. Whenever Jen was about to
gently criticize Margaret's political opinion was logically sound
but Elizabeth was a very persuasive debater.

(14) a. Whenever he was about to
carefully guard Harold's medieval dungeon was impossibly confusing
but Scott had a map of the only route out.

b. Whenever she was about to
carefully guard Harold's medieval dungeon was impossibly confusing
but Marie had a map of the only route out.

c. Whenever Dan was about to
carefully guard Harold's medieval dungeon was impossibly confusing
but Scott had a map of the only route out.

(15) a. Whenever she was about to
deftly kill Donna's savage opponent was heavily armed
but Sharon was able to avoid receiving injuries.

b. Whenever he was about to
deftly kill Donna's savage opponent was heavily armed
but Henry was able to avoid receiving injuries.

c. Whenever Jane was about to
deftly kill Donna's savage opponent was heavily armed
but Sharon was able to avoid receiving injuries.

(16) a. Whenever he was about to
gently grasp Andrew's bowling ball was poorly balanced
but Steven could compensate for the awkward weighting.

b. Whenever she was about to
gently grasp Andrew's bowling ball was poorly balanced
but Barbara could compensate for the awkward weighting.

c. Whenever Beth was about to
gently grasp Andrew's bowling ball was poorly balanced
but Steven could compensate for the awkward weighting.

(17) a. Whenever he was about to
quickly strike William's fencing foil was perfectly aimed
but Kevin nevertheless expected the opponent to win.

b. Whenever she was about to
quickly strike William's fencing foil was perfectly aimed
but Catherine nevertheless expected the opponent to win.

c. Whenever Dave was about to
quickly strike William's fencing foil was perfectly aimed
but Kevin nevertheless expected the opponent to win.

(18) a. Whenever he was about to
finally reveal Brian's brickwork patio was already visible
but Nathan felt compelled to go through the motions.

b. Whenever she was about to
finally reveal Brian's brickwork patio was already visible
but Maria felt compelled to go through the motions.

c. Whenever Al was about to
finally reveal Brian's brickwork patio was already visible
but Nathan felt compelled to go through the motions.

(19) a. Whenever he was trying to
casually block Douglas's seaworthy dinghy was fully prepared
but Edward was too scared of drowning to set out to sea.

b. Whenever she was trying to
casually block Douglas's seaworthy dinghy was fully prepared
but Sandra was too scared of drowning to set out to sea.

c. Whenever Mark was trying to
casually block Douglas's seaworthy dinghy was fully prepared
but Edward was too scared of drowning to set out to sea.

(20) a. Whenever she was trying to
deftly chase Cynthia's prized horse was looking sickly
but Rebecca thought it would be okay to enter it in the race.

b. Whenever he was trying to
deftly chase Cynthia's prized horse was looking sickly
but Peter thought it would be okay to enter it in the race.

c. Whenever Deb was trying to
deftly chase Cynthia's prized horse was looking sickly
but Rebecca thought it would be okay to enter it in the race.

(21) a. Whenever he was trying to
carefully remember Joseph's therapeutic massage was excessively painful
but Robert was a skilled acupuncturist.

b. Whenever she was trying to
carefully remember Joseph's therapeutic massage was excessively painful
but Janet was a skilled acupuncturist.

c. Whenever Paul was trying to
carefully remember Joseph's therapeutic massage was excessively painful
but Robert was a skilled acupuncturist.

(22) a. Whenever she was trying to
quickly understand Dorothy's final exam was incredibly tricky
but Amanda had studied hard for the last week.

b. Whenever he was trying to
quickly understand Dorothy's final exam was incredibly tricky
but Larry had studied hard for the last week.

c. Whenever Lucy was trying to
quickly understand Dorothy's final exam was incredibly tricky
but Amanda had studied hard for the last week.

(23) a. Whenever she was trying to
gently investigate Brenda's second-hand minivan was rattling loudly
but Maria was confident that it could make the trip.

b. Whenever he was trying to
gently investigate Brenda's second-hand minivan was rattling loudly
but Kenneth was confident that it could make the trip.

c. Whenever Kate was trying to
gently investigate Brenda's second-hand minivan was rattling loudly
but Maria was confident that it could make the trip.

(24) a. Whenever she was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Karen couldn't do it justice all alone.

b. Whenever he was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Jimmy couldn't do it justice all alone.

c. Whenever Mia was trying to
gently imitate Annie's dissonant melody was hauntingly beautiful
but Karen couldn't do it justice all alone.