PROCESSING OF LF STRUCTURE

PROCESSING OF LF STRUCTURE: EVIDENCE FROM RECONSTRUCTION FOR BINDING

BY
CASSANDRA CHAPMAN, M.A., B.A. Linguistics (University of Ottawa)

A THESIS
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS & LANGUAGES
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2018) (Linguistics & Languages)	McMaster University Hamilton, Ontario, Canada

TITLE:	Processing of LF structure: Evidence from reconstruction for binding
AUTHOR:	Cassandra Chapman B.A., M.A. Linguistics (University of Ottawa)
SUPERVISOR:	Dr. Ivona Kučerová
NUMBER OF PAGES:	xxviii, 387

Abstract

This thesis investigates whether the human sentence processing mechanism (the parser) is sensitive to effects of interpretation in real-time processing. We aim to better understand how the parser assigns meaning to a given string of words as the words are presented one at a time in left-to-right parsing. We assumed Chomsky (1995)'s Y-model of the grammar following which syntax plays a central role. In this model, the phonological component (Phonological Form, PF) and the interpretative component (Logical Form, LF) are considered separate in the architecture of the grammar. In processing experiments, the parser is presented with the string of words in its pronounced form (roughly, the PF form). To interpret the string of words, it must derive an interpretable structure for that string (LF structure). This thesis examines whether the parser is sensitive to the LF component of the grammar and if so, how it derives interpretable structures in processing. We investigated constructions that are argued to have distinct PF and LF representations (PF-LF mismatch). In the theoretical literature, these cases are known under the label *reconstruction*. To systematically investigate potential effects of reconstruction, we used constraints on referential dependencies (Binding Theory, Chomsky, 1981), which are argued to be an LF constraint (Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004). More precisely, we examined constructions in which binding constraints seem to be violated based on the PF level of representation but are satisfied at LF. In a series

of processing experiments, we manipulated whether the phrase containing a determiner phrase could be interpreted upon first encounter or whether it needed to reconstruct to a syntactically lower position to receive an interpretation. Our results suggest that the parser is sensitive to reconstruction effects in the domain of referential processing, providing evidence that the parser is sensitive to PF-LF mismatches. To our knowledge, the experiments reported in this thesis are the first to directly investigate whether the parser is sensitive to reconstruction operations required for semantic interpretation in processing.

Acknowledgements

I began writing these acknowledgements before I finished writing my thesis because I knew just how many people I owed extreme gratitude to for many aspects of this work. While theses must be written by individuals to fulfill their PhD requirements, I would never have been able to do so without the professional and personal support I had around me over the past 5+ years.

First and foremost, I would like to thank my thesis supervisor, Ivona Kučerová, who took me under her wing when I was a very young and very naïve student fresh out of a one-year Master's program. I feel very privileged to be able to call myself Ivona's first PhD student (even though I did not manage to graduate first). I think we have learned so much together and I know how lucky I am to have had such a supportive mentor during the most mentally challenging time in my life. It took me a really long time to find this thesis topic, partly because I was interested in so many different things, many of which would not have been good thesis topics. I spent a long time working on topics that did not make it into this thesis. Ivona was patient with me while always gently reminding me that I had a thesis to write. She (very quickly!) read everything I wrote over the past five years, including thesis chapters, papers, abstracts, cover letters, and research statements. Many of these documents were sloppily written but Ivona took the time to try and understand what I was trying to say. At the beginning, I would often feel overwhelmed by the number of comments on my electronic documents. I soon learned how to (mostly) achieve clarity in my writing and the number of comments decreased. Ivona has also been tremendously supportive on both a personal and professional level, especially at times when I really needed it. I would like to thank her for giving me so much of her time and energy. I have become a better linguist under her guidance (even though I still have much to learn).

Secondly, I would like to thank the members of my (internal) dissertation committee: Catherine Anderson, Susana Béjar and Martin Hackl (in alphabetical order). Catherine joined my committee at the end of my fourth year but she had been an inspiration to me since my first day at McMaster. She is an inspiring teacher and mentor and I am very grateful for the long chats we have had over the years. Even though Catherine says she does not do psycholinguistic research anymore, her work on quantifier scope ambiguity is still highly cited and very influential. Her thesis has definitely influenced the current one. Susana has been on my committee since year one. She has seen the thesis take on many different forms and go in several different directions. I would like to thank Susana for being patient with me and supporting my work year after year. Susana has a particular way of asking those big questions and encouraging me to find the big picture in lots of messy data. Her comments have helped to shape this thesis into its current form and it would not be the same without her feedback. Martin came to McMaster to give a Lecture Series talk in March 2014. His work was inspiring to me, especially the way in which he combined high level theoretical questions with cutting edge experimental methodology. Martin agreed to supervise my visiting term at MIT in Winter 2015. It was during my term away that I narrowed down my thesis topic. Martin introduced me to the world of Amazon Mechanical Turk and self-paced reading. He has a way of asking very detailed questions about language processing and theoretical syntax and semantics that changed my outlook on things in a matter of minutes. He always gave me the opportunity to speak my mind and when I was lost and confused, he found a way to say exactly what I wanted to say in a more clear and concise manner. I would like to thank Martin for his patience with me and his attention to detail. Working with him has made me a better psycholinguist and a better scientist.

Thirdly, I would like to thank my external examiner, Brian Dillon, for his willingness to read a very long thesis and provide constructive comments and feedback. Brian's work on language processing is extensive and uses a wide range of methodologies. His work on anaphoric (i.e., reflexive) processing helped me to clarify some of my own ideas. I would like to extend my gratitude to Brian for his time and for his interesting and thought provoking questions about my work.

Fourthly, I owe thanks to other faculty members at McMaster, many of whom I have had the opportunity to collaborate with: John Connolly, Victor Kuperman, and Lili Service. In particular, I would like to thank Victor for teaching me everything I know about statistics in R and for answering all of my stats questions over the years. The members of the *Syntax Lab*, both current and alumna, have greatly influenced this work: Malaree Baraniuk, Jitka Bartošová, Diane Doran, Rachael Hardy, Olena Kit, Heather Stephens, Sara Sturino, and Chelsea Whitwell. Thanks for listening to messy practice talks and poorly worked out ideas. Your feedback has improved this thesis. I would also like to thank these ladies (also known as the *Female Linguistic Army*) for providing countless grammaticality judgements and for reading over many stimulus lists. Thanks for giving me so much of your time over the years. Most importantly, thanks for your friendship.

I joined McMaster as part of a large cohort in September 2012. I had the privilege to be surrounded by people with diverse backgrounds who have helped me think about the consequences of my research in many different ways. Each year, new students joined the program and contributed to its diversity. Thanks to Laura Beaudin, Rober Boshra, Kaitlin Faulkaukas, Mike Greencorn, Angela Harrison, Gina Henry, Amanda Ho, Connie Imbault, Sam Kramer, Richard Mah, Ana Nunes, Kyle Ruiter, Sheryl Sawyer, Daniel Schmidtke, Edalat Shekari, Bryor Snefjella, Narcisse Torshizi, and Zoë Wälchli (in alphabetical order). Thanks especially to Richard, Rober, and Dan for helping me with some of the technical aspects of this thesis.

While at MIT, I had the privilege of joining a lively and engaging department. I learned so much in my one semester there. I also had the opportunity to join the Experimental Syntax and Semantics Laboratory. Thanks to the members of the Department of Linguistics and Philosophy for welcoming me into your community and for providing me with your feedback on some of the early aspects of this work. Thanks especially to Leo Rosenstein for showing me the ropes and for continuing to provide long distance technical support. These experiments would not have been put online without Leo's help.

My interest for linguistics began when I was a very young undergraduate student at the University of Ottawa. I had always been interested in graduate school but I did not know if I had what it took to actually be a graduate student. I would like to thank Eric Mathieu for seeing something in me and approaching me so many years ago. Eric encouraged me to complete a guided research project with him in the summer of my third year as an undergraduate. This was my first opportunity to do my own original research. If I had not had this opportunity, I am not convinced I would have pursued graduate school. Eric instilled confidence in me and I will be forever grateful for his guidance. I went on to complete a Master's degree under Eric and Laura Sabourin's supervision. Thanks to Laura for giving me my first real exposure to psycholinguistic work.

I would like to thank the Social Sciences and Humanities Research Council of Canada for supporting me financially. In particular, this work was conducted under SSHRC Canada Graduate Scholarship #767-2014-1827 and the Michael Smith Foreign Study Supplement #771-2014-1100. This work was also partially supported by SSHRC Insight Grant #435-2012-1567 (PI: I. Kučerová; The relation between definite and indefinite articles and free word order: a typological study), SSHRC Insight Grant #435-2013-1756 (PI: S. Béjar; co-investigators: Kučerová & Kahnemuyipour; Copular agreement systems) and SSHRC Insight Grant #435-2016-1034 (PI: I. Kučerová; Grammatical vs semantic features: the semantics-morphology mapping, and its consequences for syntax). I also owe gratitude to the Department of Linguistics and Languages at McMaster and the Graduate Student Association at McMaster for providing me with travel funding to attend both national and international conferences.

On a personal level, thanks to my friends near and far for always reminding me that this PhD is something I have always wanted to do, even when times were stressful and difficult. Thanks for the long chats and for distracting me when it was required. Thanks in particular to Lauren, Kristin, Briana, and Brendan. My family has always been very supportive of me and throughout this PhD, they were no different. Thanks to Mom, Dad, Marcel, Christiane, Memere, Andrew, Marie-Eve, Hollie, and Sebastien for always having my back. (Dad: Finally, I will not be a student anymore!)

My life has changed significantly since I began writing this thesis. Thanks to my newest little love, Simon, for bringing a new light to my life and for making me slow down and enjoy life's smallest moments. Last but most definitely not least, thanks to my amazing husband, Bryan, without whom I would have never finished this degree. Words cannot describe how much you mean to me.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

ACD Antecedent Contained Deletion

CI Confidence Interval

CI interface Conceptual-Intentional Interface

CP Complementizer Phrase

CQ Comprehension Question

CU Cost Unit

DLT Dependency Locality Theory

DP Determiner Phrase

LF Logical Form

M Mean

N Number

NP Noun Phrase

PF Phonological Form

PP Prepositional Phrase

QR Quantifier Raising

R-expression Referential expression

RT Reading Time

SD Standard Deviation

SE Standard Error

USD United States Dollars

VP Verb Phrase

Chapter 1

Introduction

1.1 What this thesis is about

This thesis investigates whether the human sentence processing mechanism (the parser) is sensitive to effects of interpretation in real-time processing. More precisely, this thesis seeks to better understand how the parser assigns meaning to a given string of words when the words are presented one at a time from left-to-right in real-time. As a starting point, we follow Chomsky (1995)'s Y-model of the architecture of the grammar. Under the Y-model, narrow syntax plays a central role. Syntactic operations (such as Merge and Move) occur before the derivation reaches Spell-Out domains (see e.g., Chomsky, 2000, 2001b). We define Spell-Out domains as phases (namely, *v*P and CP). Upon reaching a Spell-Out domain, the derivation is sent to two interfaces: i) Phonological Form (PF) where the derivation is linearized and readied for pronunciation by the Articulatory-Perceptual system and ii) Logical Form (LF) where the derivation is readied for semantic interpretation by the Conceptual-Intentional system. Crucially, structures that are accessible at the edge of the phase remain available to the syntactic module for further syntactic operations. Following Chomsky (1995)'s original formulation of the Y-model, there is no direct communication between the two interfaces (PF and LF). Thus, the two interfaces can be viewed as separate components of the grammar.

In this dissertation, we argue that real-time sentence processing experiments can be used as a tool to investigate whether the parser is sensitive to the LF component of the grammar. In a real-time processing experiment, the parser is presented with a string of words in its PF form, or the way in which it is pronounced. To interpret that string, the parser must derive an interpretable LF representation. Using a series of psycholinguistic experiments, this thesis seeks to understand how the parser takes the linearized string as input and derives an interpretable structure that is readable by the Conceptual-Intentional system. If a string fails to be assigned a semantic interpretation, it will be filtered out by the grammar. The results of the experiments reported in this thesis will enable us to better understand whether the parser is sensitive to LF structure. If it turns out the the processing mechanism is sensitive to the interpretative component of the grammar, it will need to build LF structures that will be readable by the Conceptual-Intentional system. Consequently, such results would provide evidence that the parser builds structures that are interpretable at LF and thus, that it makes a distinction between the string it receives as input and the structure it must derive to interpret that string.

To directly investigate how LF structure is built in real-time processing, we use constraints on referential dependencies, namely Binding Theory (Chomsky, 1981), as a tool. We assume that Binding Theory is a constraint on LF structure (see e.g., Chomsky, 1995; Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004, among others). More precisely, we investigate constructions in which binding constraints are seemingly violated based on the surface linear string that the parser is given as input (the PF representation). In the theoretical linguistics literature, such cases are known under the label *reconstruction* (see e.g., Lebeaux, 1990; Heycock, 1995; Fox, 1998, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004). Theoretical linguists argue that in order to satisfy binding principles at LF structure, the phrase must *reconstruct* to a lower structural position at LF. Reconstruction will ensure that the structure is readable by the semantic component. Crucially, in order for the lower structural position to be available at LF, we must assume that such a position was built in the syntax before the derivation reached Spell-Out and was sent to the interfaces. LF will only have access to structural positions that were previously built in the course of the derivation. In general, we propose that the parser has access to spell-out domains. In the domain of real-time processing, we follow previous work in assuming that the parser prefers the simplest possible parse of a given string of words (see e.g., Frazier and Fodor, 1978). This work has generally found that there is a preference to resolve open dependencies as soon as possible. If we extend this preference to the domain of reconstruction, it follows that it should be simpler to derive a structure that can be directly readable by the semantic component, without the need for additional operations, such as reconstruction. If a string of words can be directly readable by the semantic component, the parser is able to interpret the words in the string as soon as they are read. In contrast, if the words cannot be interpreted as soon as they are encountered and the parser must wait until it has received further input, the parser must keep these words in memory in order to interpret them further to the right. This latter option should be more difficult to derive since the parser is unable to satisfy its preferences. Thus, a LF structure that does not require reconstruction should be simpler to derive compared to a structure that requires reconstruction in order to be interpreted.

1.1.1 Previous work on interpretation in processing: PF-LF match

Investigating how the parser operates in real-time is not novel to the current work by any means. However, the work presented in this thesis differs from previous work in the domain of interpretation in processing both in its focus and experimental design. Previous work investigating effects of interpretation in real-time processing has been generally limited to investigating how the parser chooses the correct parse for a given string of words when it is presented with two alternatives. An abundance of work in psycholinguistics has focused on so-called *garden path* sentences (Bever, 1970; Frazier and Fodor, 1978, among many others). A famous example from Bever (1970) is shown in (1).

(1) The horse raced past the barn fell.

In (1), the verb *raced* can either be interpreted as the main verb of the sentence or as the head of a reduced relative clause. Without receiving further evidence, either of these parses is available to the parser. Upon reaching the word *fell*, the structure is disambiguated and the parser discovers that *raced* must be interpreted as the head of a reduced relative clause in order to derive an interpretable structure. Interpreting *raced* as the main verb of the sentence will lead to an illicit parse. Previous work on the resolution of garden path sentences has found that processing difficulty is incurred on *fell*, suggesting that the parser originally pursues an analysis in which *raced* is interpreted as the main verb of the sentence (Bever, 1970; Frazier and Fodor, 1978). If the parser had pursued an analysis in which *raced* was interpreted as the head of the relative clause, no processing difficulty on *fell* would be expected. This result has been taken as evidence that the parser pursues the simplest possible parse of a sentence.

Thus, research on structural ambiguities (see e.g., Bever, 1970; Frazier and Fodor, 1978), such as (1), has found that the parser prefers to adopt the *simplest* possible parse of a given string of words. In the current thesis, we follow this previous work and share this assumption. Since previous work has focused on cases in which the syntactic structure is ambiguous, the simplest possible parse is one in which the words in the linear string can be structurally integrated as soon as possible. In (1), upon reaching the verb *raced*, the parser has already built the determiner phrase (DP) corresponding to the subject and it is looking for the main verb of the sentence. It is easier to attach *raced* as the main verb of the sentence than to begin building a new phrase to build the reduced relative clause. If *raced* is interpreted as the head of a reduced relative clause, the parser needs to continue looking for the main verb of the sentence. That parse is therefore more complex than the one in which *raced* is the main verb because the parser must build more structure and must keep dependencies open. Crucially, the structure will be interpreted differently depending on which structural analysis the parser assigns to the words in the sentence. This means that the parser must be able to build two separate structures which correspond to two separate interpretations, i.e., *raced* occupies a distinct structural position when it is interpreted as the main verb compared to if it is interpreted as the head of a reduced relative.

Processing work on structural ambiguities can be understood as demonstrating that a given linearized string of words (PF representation) can be interpreted in two distinct ways and that each of these interpretations corresponds to a distinct syntactic structure. Thus, for a given PF representation, there can be two distinct LF structures. In structurally ambiguous constructions, processing difficulty arises when the parser pursues the incorrect LF structure for a given string. Since the LF structure can be directly built from the PF representation, we can view these constructions as consisting of a *PF-LF match*. In other words, it is possible to build the LF structure based on the order in which the words appear in the linear string, even if there is more than one possible LF structure.

1.1.2 Current work: PF-LF mismatch

In the current thesis, we focus on investigating constructions in which it is not always possible to construct an appropriate LF structure based on the order in which the words appear in the linear string. Constructions involving *reconstruction* represent one such case. In a structure involving reconstruction, words are pronounced in distinct positions from where they can be interpreted. As a result, we can view such constructions as consisting of a *PF-LF mismatch*. In other words, the position in which the words appear in the linear string has no direct correspondence to where they must be interpreted at LF. Previous work on LF processing (see e.g., Tunstall, 1998; Anderson, 2004; Conroy, 2008) has generally focused on cases involving semantic ambiguity, i.e., a phrase could either be interpreted in its surface position or it could be interpreted in a distinct structural position at LF. The cases addressed in this thesis differ from this work in that the phrase cannot be interpreted in its PF position (see also Hackl et al., 2012).

While the theoretical linguistics literature has significantly advanced our understanding of reconstruction and binding constraints on LF structure (Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004), we still know very little about how the left-to-right parser processes such constructions. This topic is interesting to address from a psycholinguistic perspective because it allows us to investigate the potential real-time effects that may reflect the principles outlined in the theoretical literature. In this thesis, we use binding theory and the constraints on different types of DPs in English combined with filler-gap constructions as a tool to examine how the parser resolves a PF-LF mismatch in real-time processing.

1.2 Empirical focus

In a series of self-paced reading experiments, we show that the parser is indeed sensitive to a mismatch between the surface string it receives as input and the LF structure it must derive for the string to be interpreted. It is important to note that this thesis is not the first attempt to investigate PF-LF mismatches in real-time processing. Previous work on filler-gap dependencies can be characterized as reflecting a type of mismatch between the surface linear string (PF) and the LF structure. The current work therefore adds to a body of previous literature on filler-gap dependencies (see e.g., Crain and Fodor, 1985; Stowe, 1986, among many others) but investigates the question of whether the parser is sensitive to PF-LF mismatches using a distinct empirical domain, namely, that of referential dependencies. This work also explores whether the two interfaces might interact via narrow syntax (i.e., based on spell-out domains). More precisely, if the parser is able to reconstruct a phrase to a lower syntactic position at LF, this position must have been created through syntactic structure building operations in the narrow syntax.

1.2.1 Component 1: *Wh*-constructions

The constructions of interest in this thesis are embedded *wh*-questions. As a result, we can use the received wisdom from the filler-gap literature to make predictions for the current experiments. These constructions involve syntactic movement. Upon finding the moved phrase (also called the *filler*) at the beginning of the sentence, the parser's task is to identify the integration position (i.e., the gap site) of the moved phrase in order to determine its thematic role within the sentence. For example, in (2), upon finding the *wh*-phrase at the beginning of the sentence, the parser must identify the syntactic position in which the phrase is assigned its thematic role further to the right in the linear string. In this case, the filler phrase must be integrated in object position after the verb *see*.[1]

(2) Who$_i$ did you see ~~who~~$_i$?

Understanding precisely how the task of identifying the gap site takes place requires an understanding of how movement structures are derived, which will be discussed in Section 1.3.1.

The relevant distinction between canonical filler-gap constructions and the constructions we are investigating in this thesis is whether the moved phrase found further to the left in the string can be interpreted in its surface position. To be interpreted, we propose that the phrase must have a semantic denotation. If the phrase can be assigned a denotation in its surface position, it can be semantically interpreted in that position. This would mean that the linear positions of the words in the PF representation have a direct correspondence to the positions of the words in the LF structure. In other words, the LF structure can be built based on the order in which the words appear in their PF form.[2] In contrast, if the moved phrase cannot be assigned a semantic denotation in its surface position, it cannot be interpreted in that position. If the phrase cannot be assigned a semantic denotation in its surface position, it will need to be interpreted further downstream (further to the right in the linear string). If the phrase must be interpreted further to the right, it must reconstruct to that position for interpretation reasons. In such cases, the parser must interpret the phrase in a distinct position from where it appears on the surface. Thus, there is a *mismatch* between the PF and LF positions of the phrase.

[1] We assume the copy theory of movement, which we will discuss in Section 1.3.1.1.

[2] We put aside for the moment the integration operation that must take place at the gap site. Since all moved phrases must be integrated at the gap site, no interesting differences between different types of moved phrases would be predicted at this syntactic position.

1.2.2 Component 2: Type of DP embedded in *wh*-phrase

To systematically investigate whether a moved phrase (always *wh*-phrases in our experiments) could be interpreted in its surface position or whether it needed to be interpreted in a distinct position at LF, we manipulated the type of DP embedded in the *wh*-phrase. We used three different types of DPs: anaphors (also called reflexive pronouns, *himself, herself*), pronouns (*him, her*), and R-expressions (*John, Mary*), as shown in (3) for the anaphor condition.

(3) [Which picture of [himself]$_j$]$_i$ did John$_j$ see ~~[which picture of [himself]$_j$]$_i$~~?

Using different types of DPs embedded in *wh*-phrases enabled us to control for whether the phrase containing the DP (the complement of *wh*) could be interpreted in its surface position (high) or whether it needed to be interpreted lower in the structure via reconstruction. In order to understand whether the phrase containing the DP can be interpreted when it is encountered, we need to first understand how DPs are interpreted. As a starting point, we will first consider whether the DP can be interpreted on its own, irrespective of its syntactic position in the sentence. After establishing how DPs are interpreted on their own, we will then consider how DPs are interpreted within the larger structural configuration. The relevant question is: how do we know if a DP is interpretable? As briefly mentioned, we propose that in order for a phrase to be interpretable, it must have a semantic denotation. Thus, in order for a DP to be interpretable on its own, it must have a semantic denotation. We will discuss this in more detail in Section 1.3.2.

To summarize, to better understand the different components involved in the structures investigated in this thesis, it is useful to think about two separate questions: i) how is movement derived structurally? and ii) how and when is a DP assigned a semantic interpretation? These questions are necessarily interconnected but it is important to discuss them separately before we delve into how they interact.

1.3 Theoretical Background

1.3.1 How is movement derived?

In theoretical linguistics, *wh*-constructions are argued to arise via syntactic *movement*. More precisely, the *wh*-phrase that appears at the beginning of the sentence in the PF representation is base-generated in the position in which it is assigned its thematic role. The phrase then moves to the beginning of the sentence, where it is pronounced. For example, in (4), the *wh*-phrase is base-generated in the object position of the verb and moves to the beginning of the sentence. As a result, there is a mismatch between the position in which the phrase is generated, i.e., where it is first merged, and where it is merged second (or internally merged, Chomsky, 2001a). The position in which the phrase is internally merged also corresponds to the position in which it will be pronounced.

(4) Who$_i$ did you see ~~who~~$_i$?

Following the Minimalist Program (Chomsky, 1995), syntactic operations, such as Merge and Move, occur cyclically. Once the derivation reaches a phase boundary (namely, *v*P or CP), that part of the derivation is spelled out (cyclic Spell-Out). Upon reaching Spell-Out, the derivation is sent to two interfaces: i) Phonological Form where the structure is linearized and then readied for pronunciation and ii) Logical Form where the structure is readied for interpretation by the Conceptual-Intentional system. As previously mentioned, there is currently little understanding of how (or whether) the two interfaces can directly interact with one another and they are generally viewed as separate components of the grammar. We adopt this view in the current thesis and assume that the pronounced form of a string of words (PF) does not necessarily directly correspond to the form that is interpreted (LF). In other words, there may not be a one-to-one correspondence between the order that the words appear in the linear string and the hierarchical structure that the parser must derive to interpret that string. As previously mentioned, we assume that the parser must have access to the Spell-Out domains derived in the syntax. In order for a phrase to be interpreted in a distinct position from where it appears on the surface, the phrase must have occupied that position at some point in the course of the derivation. As a result, the parser must be able to access the form of the derivation before it was spelled out.

1.3.1.1 Copy theory of movement

In the current work, the constructions of interest are always embedded *wh*-constructions. Thus, they all involve movement. To have a clearer understanding of movement structures at LF, we adopt Chomsky (1993)'s *copy theory of movement*. This theory allows us to have an explicit model of how movement structures are computed at LF and it enables us to make clear predictions about processing effects at specific regions in the sentences. Following copy theory, when a phrase moves in the syntax, it leaves behind a *copy* of that phrase in its original position. For example, consider again (4) in which the *wh*-phrase is base-generated in the object position of *see* and moves to the beginning of the sentence. As it moves, a copy of the phrase is left behind in its original position and in any intermediate landing sites in which the phrase stops on its way to its final position, following successive cyclic movement (Chomsky, 1995). As a result, there is a copy of the phrase in the position it occupies on the surface, in its base-generated position, and in any intermediate landing sites. Together, the copies form one complex syntactic object (or a *chain*). In English, the structurally higher copy (*head of the chain*) gets pronounced by the phonological component and the lower copies get phonologically deleted at PF (indicated by striking out the text in the example). The lowest copy corresponds to the position in which the phrase is assigned its thematic role and is called the *tail of the chain*.

Crucially for the current work, since there are copies of the *wh*-phrase in distinct positions at LF, it should be in principle possible for the phrase (or parts of the phrase) to be interpreted in any of those positions. We argue that the position in which a copy can be interpreted depends on the structure in which it appears.

However, it is not the case the entire *wh*-phrase needs to be interpreted in a single copy position. Since all of the stimuli used in the current thesis are *wh*-questions, in order to interpret the sentence as a question, the *wh*-operator must necessarily be interpreted high, i.e., in its surface position. If *wh* is not interpreted high, the sentence will not be interpreted as an interrogative. However, it is not the case that the whole higher copy of the *wh*-phrase needs to be interpreted in this position. Consider (5) (from Sauerland and Elbourne, 2002) in which the pronoun *hers* is bound[3] by the quantified phrase *every student*. In order for the pronoun to be bound by the quantified phrase, it must be interpreted in a position that is structurally lower than *every student*.

(5) Which relative of hers$_i$ did every student$_i$ invite?

As argued in Sauerland and Elbourne (2002), the copy theory of movement (Chomsky, 1993; Fox, 1999; Sauerland, 1999) can account for this reading. If *wh* must be interpreted high but the pronoun must be interpreted below the quantified phrase, this means that the complement of *wh* can be interpreted in a lower copy position (see Fox, 2002; Sauerland

[3]We will discuss Binding Theory further in Section 1.3.2.1.

and Elbourne, 2002). In other words, the complement of *wh* can reconstruct to the lower copy position at LF, leaving the *wh*-operator in the higher position. The LF for (5) is given in (6).

(6) Which$_i$ did every student$_j$ invite [relative of hers$_j$]$_i$

Whether the complement of *wh* can be interpreted in the higher copy position or whether it must reconstruct to a lower copy position depends on the structure in which the phrase appears. One way to control for the interpretation position of the complement of *wh*is to manipulate the linguistic elements that the complement phrase contains.[4] As previously mentioned, in the constructions investigated in this thesis, we manipulated the type of DP that was embedded in the complement of *wh*. Different types of DPs in English are subject to different interpretative and structural constraints, as we will discuss in the next two sub-sections. Thus, to understand if the complement of *wh* can be interpreted when it is encountered, we need to understand how DPs are assigned semantic interpretations.[5] Specifically, it is important to first establish whether the DP can be assigned a semantic interpretation on its own, irrespective of the structural configuration in which it appears. After establishing whether the DP can be interpreted on its own, we must consider whether the DP can be interpreted (or integrated) within the structure.

[4]Another way is to manipulate the larger structural configuration. For example, in scope environments, the complement of *wh* could either take low or high scope with respect to a scopal operator. This possibility will be discussed in Section 1.3.4.2.

[5]For reasons that will be outlined in Chapter 2, we remain agnostic about parsing at the base-generated position. Since all of our constructions were embedded questions, the parser must locate the tail of the chain for the phrase to be assigned its thematic role. Since this operation is obligatory across all conditions, irrespective of the type of DP embedded in the *wh*-phrase, we should not observe any effects at the lower copy position.

1.3.2 How are DPs interpreted?

We propose that in order for a DP to be interpretable, it needs to have a semantic denotation. Assuming that *wh* is always interpreted high, we must consider whether the complement of *wh* can also be interpreted high or whether it must be interpreted in a lower copy position. In the current work, the complement of *wh* was always a DP embedded with either an anaphor, pronoun, or R-expression. To determine where the complement of *wh* can (or must) be interpreted, we need to understand the properties of the DP complement. We will consider each type of DP separately.

If the DP is embedded with an anaphor, as in (3), the anaphor is a bound variable (Heim and Kratzer, 1998). The value of its assignment function depends on its antecedent. This means that there must be an antecedent for the anaphor in the structure. In other words, a DP containing an anaphor is not interpretable on its own; its interpretation is dependent on another DP in the structure. As a result, it cannot be assigned a semantic denotation without its antecedent. Consider now a DP containing a pronoun, as in (4). In this case, the pronoun can be interpreted in two ways: i) as a bound variable, just like the anaphor, or ii) as a free pronoun. If it is interpreted as a bound variable, its assignment function will depend on its antecedent, similarly to the case discussed for anaphors. However, if it is interpreted as a free pronoun, it will receive its interpretation from the context and it could be assigned a semantic denotation on its own, i.e., not dependent on an antecedent. Thus, for a DP embedded with a pronoun, there are in principle two possible interpretations. Finally, if the DP contains an R-expression, as in (7-c), the R-expression can be assigned a semantic denotation. Its meaning is not dependent on an antecedent.

(7) a. pictures of himself

b. pictures of him

c. pictures of John

To summarize, if we consider only whether a DP can be assigned a semantic denotation on its own, the three types of DPs differ. An R-expression has a denotation as soon as it is encountered but an anaphor does not; its denotation depends on its antecedent. If the DP contains a pronoun, there are two possibilities. Either the pronoun is interpreted as referential and it refers to a salient entity in the discourse or it is interpreted as a bound variable and its interpretation depends on an antecedent in the structure. We assume in the referential case that the pronoun can be assigned its semantic denotation as soon as it is encountered. In contrast, if the DP is interpreted as bound variable, it can only be assigned a semantic denotation when it is encountered if there is a structurally appropriate antecedent in the linguistic context. Thus, in the case of a DP embedded with a pronoun, there is potentially a choice in interpretation, which we will argue depends on its structural position within the sentence. In order to make precise predictions for word-by-word parsing, it is important to understand the syntactic distributions of the three different types of DPs in English. We follow Chomsky (1981)'s Binding Theory.

1.3.2.1 Binding Theory

English anaphors, pronouns, and R-expressions have different syntactic distributions, which have been accounted for by the Binding Theory, proposed by Chomsky (1981). We assume that Binding Theory is an LF condition (Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004), meaning the binding principles must be satisfied at LF in order for the structure to be interpretable. At a very general level, in mono-clausal environments, a DP can only co-refer with an anaphor (8-a) but never with a pronoun (8-b) or an R-expression to its right (8-c). If two DPs co-refer, they refer to the same real world referent. We will mark co-reference between two DPs through matching indices in the examples.

(8) a. John$_i$ likes himself$_i$.
 b. *John$_i$ likes him$_i$.
 c. *He/John$_i$ likes John$_i$.

If the DPs are in separate clauses, as in (9), a different pattern emerges.[6] The main clause subject, *John*, can only co-refer with the pronoun in a separate clause, (9-b), but can no longer co-refer with the anaphor if it is found in a separate clause, (9-a). The anaphor in (9-a) can only co-refer with the embedded subject, *Harry*, which is found in the same clause as the anaphor. The R-expression cannot co-refer with another R-expression, whether they are found within the same clause, (8-c), or not, (9-c).

(9) a. John$_i$ thinks that Harry$_j$ likes himself$_{*i/j}$.
 b. John$_i$ thinks that Harry$_j$ likes him$_{i/*j}$.
 c. He/John$_i$ thinks that Harry$_j$ likes likes John$_{*i}$.

The distribution of these three types of DPs is therefore constrained depending on their relation to another DP in the sentence.[7] *Binding* is one way that two DPs can be in a referential dependency, such that the denotation of one DP depends on the denotation of the other DP. Binding has two components. If two DPs are in a binding relation, they must

[6]For the purposes of this thesis, we assume that the relevant binding domain is the smallest clause within which the DP is contained. This is a simplistic view of binding theoretic notions but it is sufficient for the current work.

[7]In addition to the syntactic binding proposals, it has also been proposed that binding is a semantic requirement (see e.g., Heim and Kratzer, 1998). For simplicity, we put such proposals aside for the current thesis and assume that binding is a syntactic requirement (see e.g., Charnavel and Sportiche, 2016).

i) bear the same index (they are *co-indexed*) and ii) one DP (called the *antecedent DP*) must c-command the other DP (see e.g., Reinhart, 1976). A basic definition of c-command is provided in (10), adapted and simplified from Reinhart (1983) (p. 50).

(10) *C(onstituent)-command*:

A node A c-commands a node B iff the first branching node α that dominates A also dominates B.

If DP_1 c-commands DP_2 and if the two DPs bear the same index, then DP_1 binds DP_2. However, these restrictions on the relation between two DPs are not sufficient to explain the whole pattern of distribution. For example, in (9-a), both *John* and *Harry* c-command the anaphor, *himself*, yet the anaphor can only co-refer with *Harry*. Thus, there is also a locality restriction on the distribution of the DPs with respect to their antecedents. To account for their distribution, the following three binding principles were proposed by Chomsky (1981).[8]

(11) a. *Principle A*: An anaphor must be bound in its binding domain (roughly, the clause).

b. *Principle B*: A pronoun cannot be bound in its binding domain.

c. *Principle C*: An R-expression cannot be bound at all.

Binding Principle A is able to explain why the anaphor, *himself*, can only co-refer with *Harry* and not with *John* in (9-a). *Harry* and *himself* are in a local domain (the clause) and thus, *Harry* locally binds *himself*, satisfying Principle A. In contrast, Binding Principle B states that pronouns cannot be bound locally (i.e., within their binding domain). This principle explains why *him* can only co-refer with *John* in (9-a) and not with *Harry*. Since

[8]These are simplified versions of the binding principles but they are sufficient for the current purposes.

Harry and *him* are within the same binding domain, the pronoun cannot be bound by this DP. Finally, an R-expression can never be bound by another DP in the sentence, irrespective of its binding domain, explaining the ungrammaticality of (8-c) and (9-c).

1.3.2.2 Reconstruction for binding

At first glance, interrogative constructions, such as (12), seem to violate binding principles. This construction seemingly violates Binding Principle A: the moved *wh*-phrase contains an anaphor but the anaphor does not seem to be bound by its antecedent. In (12), the antecedent for the anaphor is *Mary* but this DP does not c-command the anaphor, which would violate Binding Principle A.

(12) Which picture of herself$_i$ did Mary$_i$ like?

This analysis of (12) only considers the surface string of the words and assumes that the LF structure is derived directly from the linear ordering of the words in their pronounced form. Following the copy theory of movement, in *wh*-constructions, there are two copies of the *wh*-phrase at LF, as shown in (13).

(13) Which picture of herself$_i$ did Mary$_i$ like ~~which picture of herself$_i$~~?

Even though the anaphor is not bound in its surface position (because it is not c-commanded by *Mary*), the anaphor is bound in its base-generated position, i.e., at the tail of the chain. In this case, the minimal phrase containing the anaphor (complement of *wh*) can only be interpreted in its base-generated position.[9] In order for the anaphor to be interpreted and assigned a semantic denotation, the phrase containing the anaphor must *reconstruct* to a

[9]It might also be possible for the phrase containing the anaphor to be interpreted in an intermediate position structurally below the antecedent for the anaphor, provided this is a position the phrase had occupied at some point in the derivation, following successive cyclic movement (Chomsky, 1995).

structurally lower syntactic copy position at LF (see e.g., Chomsky, 1995; Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004, among others). In this position, the anaphor is structurally bound by its antecedent and the structure is interpretable at LF. Note that if the phrase does not reconstruct, the anaphor is uninterpretable in its surface position because it does not have a structurally appropriate antecedent, which is a violation of Binding Principle A.

1.3.3 Putting all the pieces together: binding and reconstruction

We follow Binding Theory and assume that whether an anaphor or a pronoun can be associated with an antecedent in the linguistic context depends on the structural configuration in which the two DPs appear. In all of our stimulus items, the DP containing that anaphor, pronoun or R-expression was embedded in a *wh*-phrase. There was no DP in the preceding string of words that could serve as the antecedent for the anaphor/pronoun. Thus, the *wh*-phrase always linearly preceded a potential antecedent for the anaphor/pronoun. The *wh*-operator must always be interpreted in the position in which it appears, i.e., in the higher copy position, in order for the structure to be interpreted as an interrogative. The relevant question is whether the complement of *wh* can also be interpreted in this position or whether it can (or must) be interpreted in a lower copy position. The position in which the phrase can or must be interpreted depends on where the DP can be assigned a semantic denotation. In other words, we assume that a phrase can only be integrated within the structure if it has been assigned a semantic denotation.

We will consider the simplest case first. If the complement of *wh* contains an R-expression, it can be interpreted where it is encountered because the R-expression can be assigned a semantic denotation as soon as it is encountered. The interpretation of an

R-expression is not dependent on another DP in the sentence. This means that when the complement of *wh* contains an R-expression, it can be interpreted in the high position without the need for reconstruction.

We are also able to make clear predictions for the complement of *wh* containing an anaphor. If the complement of *wh* contains an anaphor, it cannot be interpreted in the higher copy position with the *wh*-operator. The anaphor cannot be assigned a semantic denotation when it is encountered because it is a bound variable and its denotation depends on its antecedent.[10] Since the *wh*-phrase linearly precedes a potential antecedent for the anaphor, the complement of *wh* must be interpreted in a lower copy position (structurally below the antecedent). In principle, the complement of *wh* could either reconstruct to the lowest copy position (base-generated position) or to an intermediate copy position, depending on the position of the antecedent and the structural configuration in which it appears. Crucially, the phrase containing the anaphor must reconstruct to a lower copy position from which it would be structurally bound by its antecedent (Binding Principle A).

The case with pronouns is more complicated. Upon encountering a pronoun embedded in a *wh*-phrase, the parser can either interpret the DP containing the pronoun in the higher copy position or in a lower copy position, depending on the assignment function of the pronoun. If the pronoun is to be interpreted as a free pronoun, then it can be interpreted as soon as it encountered, similarly to R-expressions. Free pronouns can be assigned semantic denotations on their own since their interpretation does not depend on another DP. However, if the pronoun is to be interpreted as a bound variable, its denotation depends on its antecedent, similarly to the case with anaphors. Since the only potential antecedent is

[10]The phrase containing the anaphor would need to be kept in memory so that it can be integrated further to the right in the linear string. Thus, it might be possible for the parser to derive a partial interpretation of the phrase in order to integrate it in the lower copy position. Crucially, however, the phrase containing the anaphor cannot be fully interpreted on its own since its meaning is dependent on its antecedent.

found in a position linearly to the right of the *wh*-phrase in our stimuli, the phrase containing the pronoun will need to be interpreted structurally below the antecedent. As a result, it will need to reconstruct to a lower copy position in order to receive its denotation and be interpreted. Crucially, in order to abide by Binding Principle B, the phrase containing the pronoun would need to reconstruct to a copy position from which it would not be locally bound by its antecedent. In principle, either of these options should be possible for the parser.

1.3.3.1 Binding versus co-reference

We have been discussing co-reference between two DPs as a strictly structural relation, following the binding principles proposed by Chomsky (1981). As a result, we have been assuming that if two DPs are interpreted as co-referential, they must be in a particular structural configuration. In an influential proposal, Reinhart (1983) argues that co-reference between two DPs can take on two different forms, only one of which depends on structural properties, i.e., binding. If two DPs are in a *binding* relation, the antecedent must c-command the co-indexed pronominal (typically a pronoun or anaphor).[11] Thus, in order for binding to occur, the DPs must be in a specific structural configuration. Following Reinhart (1983)'s proposal, if two DPs are in a binding relation, the pronominal is interpreted as a bound variable. In contrast, if two DPs are in a *co-reference* relation, the pronominal receives its interpretation from the discourse, through reference assignment. Co-reference relations are therefore not restricted structurally. When the pronominal is an anaphor, only a binding relation is possible between the anaphor and its antecedent following Principle A.[12] The situation is different with pronouns because they can either be in a co-reference

[11] We put aside quantificational DPs for the purposes of this thesis.

[12] We put aside cases of logophoricity. See Pollard and Sag (1992); Reinhart and Reuland (1993) for further details.

or binding relation. For the pronoun to be in a binding relation with another DP, the antecedent must be co-indexed with and c-command the pronoun. To abide by Principle B, the antecedent and the pronoun cannot be in a local relation, i.e., they cannot be in the same clause. It is important to understand the difference between binding and co-reference in the domain of pronoun resolution since only one of these relations depends on structure. If we were to find that a pronoun is interpreted as co-referential with another DP in the sentence but the structure seems to violate Binding Principle B, we would need to consider whether a co-reference analysis between the two DPs is possible.

A co-reference analysis is typically used to explain the distribution of pronouns that are co-referential with another DP within the sentence but do not seem to abide by Binding Principle B. Under a co-reference analysis, co-indexation of a pronoun with another DP can seemingly violate Binding Principle B, provided that the structural configuration does not violate Binding Principle C, i.e., the pronoun cannot c-command an R-expression if they bear the same index. However, co-reference relations are not unrestricted. Theoretical linguistic research has shown that the grammar prefers to establish a binding relation over a co-reference relation (see Reinhart, 1983; Reinhart and Reuland, 1993; Reinhart, 2006; Heim and Kratzer, 1998; Fox, 1998; Roelofsen, 2010 for details). This means that the grammar generally prefers to establish a structural anaphoric dependency relation between two DPs over a non-structural one. It follows that if a pronoun and an R-expression bear the same index, a co-reference relation can only be established if binding is not possible. If a binding relation between two DPs is possible but this parse is not pursued, then the only possible interpretation of the utterance will be one in which there is no co-reference between those two DPs. Reinhart (1983) argues that because a different option was chosen than the one provided by the grammar, i.e., since co-reference is restricted by the discourse, the two DPs cannot be interpreted as co-referential and this parse is ruled out.

To capture the distributional differences between anaphors and pronouns, Reinhart (1983) argues that anaphors (in her terms, *R-pronouns*) must always be interpreted as anaphoric but that pronouns (or *non-R-pronouns*) are *referentially ambiguous*. Consider (14).

(14) Mary$_i$ likes her$_{*i/j}$.

In principle it should be possible to adopt a co-reference analysis in (14): the pronoun could be construed as being co-referential with the DP subject. However, this interpretation is not possible and the pronoun must refer to a DP outside the sentence. Reinhart (1983)'s proposal is able to account for this example and the fact that the pronoun cannot be interpreted as co-referential with *Mary*, even under a co-reference analysis. In this case, the grammar provides the option to use an unambiguous form, namely *herself*, to convey a meaning in which the two DPs are co-indexed. If a speaker uses *her* instead of *herself*, the speaker must intend a different meaning, ruling out the co-reference reading of the pronoun in (14). Reinhart (1983) argues that speakers prefer to avoid ambiguity. In the domain of referential dependencies, the clearest way to avoid ambiguity is to use constructions in which two DPs are in a binding relation, whether the bound DP is an anaphor or a pronoun.

However, it is not always possible to avoid ambiguity in natural language. Consider (15).

(15) Mary$_i$ likes her$_{i/j}$ friends.

(15) can have both a binding or co-reference interpretation. In constructions like (15), Reinhart (1983) argues that speakers should choose the best option, depending on what is

permitted by the grammar. If a bound anaphora reading is possible, this option should be chosen to avoid ambiguity. However, if it is not possible to establish a bound anaphora reading, the interpretation of the pronominal must be determined by discourse properties. More precisely, a co-reference relation can only be established if the pronoun and its referent are in a local domain and the R-expression c-commands the pronoun, e.g., (14). If they are in non-local configuration, i.e., if they are in separate clauses, binding should always be preferred, e.g., (16).

(16) Mary$_i$ said that Jane$_k$ liked her$_{i/j/*k}$.

One major advantage to Reinhart (1983)'s proposal is that there are no principles restricting non-coreference. Instead, such cases are restricted by the discourse or *pragmatically*. As we have already alluded to, there is one exception to the general rule that binding is preferred over co-reference. This exception stems from the assumption that the grammar prefers simple structures and that more than one LF for a particular string of words cannot be postulated unless the distinct LFs would lead to distinct interpretations (Fox, 1995 1998, 2000). In other words, the grammar prefers the most economical representations (Fox, 1998). Since binding relations are preferred over co-reference relations (Reinhart, 1983), it follows that two distinct LFs for a given string, one containing a binding relation and the other containing a co-reference relation, will only be postulated if they would lead to two separate interpretations for that string (Rule I, Grodzinsky and Reinhart, 1993). See Sekerina et al. (2004) for psycholinguistic evidence from both adults and children showing that binding relations in which reference is established in the syntax are less costly than co-reference relations in which reference is established outside the syntax (i.e., based on discourse properties).

To summarize, the interpretation of a pronominal can occur in two ways: i) through binding which is a relation that is structurally restricted, and ii) through co-reference which is restricted pragmatically. Crucially, for our purposes, binding is a constraint on LF structure. To satisfy binding constraints in interrogative constructions, a phrase can reconstruct to a lower copy position at LF. Following Reinhart (1983), even though both binding and co-reference analyses are in principle possible, the grammar has a preference to adopt a structural analysis (binding) over a non-structural analysis (co-reference). This means that a co-reference analysis will only be possible if i) a binding relation is not possible, or ii) the co-reference analysis provides a different interpretation than the binding analysis.

1.3.4 Another reason for reconstruction at LF: *De dicto* interpretations

Reconstruction at LF has also been used to explain how we are able to get two interpretations of the sentence in (17), originally discovered by Obenauer (1984).

(17) How many books does Chris want to buy?

This construction can either mean that there are particular books that Chris wants to buy and the question is asking how many such books there are (*de re* reading) or it can mean that Chris wants to buy a certain number of books but does not know the identity of those books and the question is asking how many such books there are (*de dicto*). Crucially, as we will describe in the next subsection, in order to derive the *de dicto* reading of this sentence, the complement of *wh* must reconstruct below the scope-bearing operator, *want*, at LF. Thus, reconstruction for the *de dicto* reading provides another way to test for whether the parser is sensitive to PF-LF mismatches in processing.[13]

1.3.4.1 *De re* versus *de dicto* readings in *how many* questions

Obenauer (1984) observed that *how-many* questions such as (18) are ambiguous between two interpretations: *de re*, as in (18-a), and *de dicto*, as in (18-b). More precisely, *many-x* can either take high or low scope with respect to a scope-bearing operator, such as *want*. If *many-x* takes high scope with respect to *want*, it is interpreted in a position that is structurally higher than *want* at LF. This is also called the *surface scope* interpretation because the relevant phrases are interpreted at LF in the positions in which they appear in the PF representation. In contrast, if *many-x* takes low scope with respect to *want*, it is interpreted in a position that is structurally below *want* at LF. This reading is also known as the *narrow scope* reading because *many-x* and the scope bearing operator (here, *want*) are interpreted in distinct syntactic positions at LF from where they appear in the PF representation of the string.

(18) How many books does Chris want to buy?

a. *De re, surface scope (many-x > want)*:

Paraphrase: What is the number *n* such that there are *n* books that Chris wants to buy?

b. *De dicto, narrow scope (want > many-x)*:

Paraphrase: What is the number *n* such that Chris wants to buy *n* books?

(Rullmann, 1993)

[13] Unfortunately, this manipulation did not show significant effects in our studies, as we will detail in Chapters 2 and 3. We discuss the results of this manipulation in Chapters 2 and 3. Since this manipulation did not show the predicted effects, we removed it from subsequent experiments, reported in Chapters 3 and 4. The consistent manipulation used across studies was the type of DP embedded in the *wh*-phrase.

If (18) is interpreted as *de re*, *many-x* takes scope over *want*, e.g., (18-a). This corresponds to the surface scope interpretation of the sentence. If (18) is interpreted as *de dicto*, *want* takes scope over *many-x*. This corresponds to the narrow scope interpretation, e.g., (18-b). In order to get the narrow scope interpretation, the verb needs to appear in a structural position that is higher than *many-x* at LF. This means that the positions of the relevant phrases in the PF representation mismatches their positions in the LF representation. The mechanism that has been used to explain how we are able to derive the *de dicto* reading is syntactic reconstruction (see e.g., Lebeaux, 1990; Heycock, 1995; Fox, 1999, 2000; Fox and Nissenbaum, 2004, among others).[14] Following a reconstruction analysis, *many-x* reconstructs to a lower position (i.e., to a position structurally below the scope-bearing operator) in order to be interpreted at LF (see e.g., Heycock, 1995; Fox and Nissenbaum, 2004). Following the copy theory of movement, in order to derive the *de dicto* reading, the lower copy of the phrase must be interpreted at LF. LF structures for the two possible interpretations of (18) are given in (8).[15]

(19) a. Surface: [wh many books]$_1$ Chris wants [~~many books~~]$_1$ to buy [~~many books~~]$_1$ (*many-x* > *want*)

b. Narrow: [wh]$_1$ Chris wants [n_1 many books]$_2$ to buy [~~n_1 many books~~]$_2$ (*want* > *many-x*)

[14]There are a few different alternative hypotheses about the form that reconstruction takes. One possibility is movement following which the phrase covertly lowers to its base position at LF (Quantifier Lowering). Another alternative is that the parser interprets a lower copy at LF, which does not imply any sort of movement operation. A third possibility is semantic reconstruction, which we leave aside for the purposes of this thesis since all of our cases necessarily involve syntax. For the purposes of this thesis, we remain agnostic about which mechanism is at play. What is crucial for this thesis is that the *de dicto* reading depends on the syntactic position in which the *many-x* DP appears at LF.

[15]In (19-b), it is in principle possible to interpret *many-x* in either of the lower copy positions since in both cases, the phrase would be interpreted below *want*. For simplicity, we have indicated that the phrase is interpreted in the lowest copy position but either interpretation position would yield the appropriate analysis.

In order to derive the *de dicto* interpretation, an additional operation, i.e., syntactic reconstruction, must take place at LF. In order to derive the *de dicto* interpretation, parts of the string (phrases once the parser has built the structure) must be interpreted further to the right (lower in the structure) than where they appear on the surface. This means that there is a mismatch between the PF and LF representations of the sentence. Reconstruction is not needed to derive the surface scope interpretation because the relevant phrases are interpreted in the structural positions in which they appear in the linear string (there is a PF-LF match). From a processing perspective, we predict that the surface scope interpretation should be simpler to derive at LF compared to the narrow scope interpretation. This means that the *de dicto* reading should incur a processing cost compared to the *de re* reading. Note that this reasoning follows the same logic outlined in Chapter 1 when we discussed the predicted processing cost associated with reconstruction for anaphoric binding (Principle A of the Binding Theory).

In many cases, either the *de re* or *de dicto* reading of a *how many* question is possible. As a result, it can be difficult to use these constructions to test for processing effects of reconstruction. If both readings are possible, this means that either the *de re* or *de dicto* LF would derive an appropriate interpretation. Since we assume that the parser prefers to adopt the simplest possible structure of a given string, we might predict that it should be simpler to derive the *de re* reading and that a processing cost should be incurred if the parser adopts the *de dicto* reading instead. However, if either parse would yield an appropriate interpretation, it might be difficult to determine which interpretation has been adopted. Crucially, in some instances, the *de re* reading is not a possible interpretation and reconstruction of *many-x* at LF is obligatory. Such cases enable us to directly investigate whether a cost is incurred when the parser is forced to adopt the *de dicto* reading. If we were to find such a cost, this would suggest that the parser is sensitive to PF-LF mismatches and thus, that it has a preference to interpret phrases in their surface positions (i.e., as soon as they are encountered). *How-many* questions with creation verbs provide us with a way to test for this preference, as we will discuss in the next sub-section.

1.3.4.2 Creation verbs

Heycock (1995) observed that *how-many* questions embedded with verbs of creation force the reconstructed reading, i.e., the *de dicto* reading. In other words, when a verb of creation is embedded in a *how many* question, as in (20), the *de re* reading is unavailable. The question in (20) is only compatible with the *de dicto* reading; compare (20-a) to (20-b). The LF structures for both readings are provided in (21).

(20) How many books does Chris want to write?

a. *De re, surface scope*:

#What is the number *n* such that there are *n* books that Chris wants to write?

b. *De dicto, narrow scope*:

What the the number *n* such that Chris wants to write *n* books?

(21) a. #[wh many books]$_1$ Chris wants [~~*many books*~~$_1$] to write [~~*many books*~~$_1$]

b. [wh]$_1$ Chris wants [~~n_1 many books~~]$_2$ to write [n_1 *many books*$_2$]

To explain these facts, Heycock (1995) argues that the *de re* reading in (20-a) is incompatible with the semantics of the verb of creation. The object of a creation verb, i.e., *books* in the case of (20-a), only exists after the creation event, i.e., the writing event, has been completed (see von Stechow, 2001, among others). In (20-a), the books that John wants to write only exist after the writing event has been completed. In contrast, the surface scope interpretation depends on this object already existing before the event takes place. More precisely, the *de dicto* reading presupposes that the books that Chris wants to write at a time later than t already exist at t, i.e., the utterance time. Therefore, *how-many* questions with non-creation verbs, such as *buy*, are ambiguous between a *de re* and *de dicto* interpretation, but *how-many* questions with creation verbs are unambiguous and allow only the reconstructed, or *de dicto*, interpretation. These facts are summarized in Table 1.1.

Table 1.1: Interpretation differences between *how-many* questions with and without creation verbs

Sentence	*de re*	*de dicto*
How many books does Chris want to buy?	✓	✓
How many books does Chris want to write?	✗	✓

Creation verbs can therefore be used to force the reconstructed (*de dicto*) interpretation of *many-x*. This means that even though the parser prefers to pursue the simplest possible interpretation of a given string of words (in this case, the surface scope or *de re* reading), upon reaching a creation verb further to the right in the string of words, it should determine that the surface scope reading is not possible and that it must adopt the *de dicto* reading. We predict that adopting the *de dicto* reading should incur a processing cost since the reconstructed reading is not the preferred reading of the string.

Note that this is the same logic used to explain the preference to interpret the complement of *wh* in its surface position, as described in section 1.3.3 and as we will discuss further in section 1.4.1.3. We propose that the parser will attempt to interpret the complement of *wh* as soon as it is encountered, unless it is given evidence that such an analysis is not possible. When the *wh*-phrase is embedded with different types of DPs, the parser will know right away if the DP can be assigned a semantic denotation in its surface position. In contrast, if the structure can only be interpreted as *de dicto*, the parser will not know that the *de re* reading is not possible until it finds the creation verb further to the right in the string.

1.3.5 Two reasons for reconstruction

We now have two different motivations for reconstruction: i) reconstruction for binding, and ii) reconstruction for scope, specifically for the *de dicto* interpretation in *how many*. Crucially, reconstruction for the *de dicto* interpretation involves the same structural process as reconstruction for binding. As a result, we might predict that these two different triggers for reconstruction might interact with each other. We specifically investigated this question in Experiments 1 and 2, reported in Chapters 2 and 3.

Before providing further details about the psycholinguistic studies that were conducted to address the question of how the parser processes a PF-LF mismatch in the domain of referential processing, we will first go over the relevant psycholinguistic background, from which we borrow several assumptions.

1.4 Psycholinguistic Background

1.4.1 Parsing assumptions

1.4.1.1 Preference for simpler parses

In this thesis, we are concerned with parsing of written texts. We do not make any claims about parsing of speech. At a very general level, we follow much previous work in assuming that the left-to-right parser always prefers the simplest possible parse of a given sentence (Frazier and Fodor, 1978). Evidence for the parser preferring simpler structures over more complex ones has been found in previous psycholinguistic work (Ferreira and Clifton, 1986; Ferreira and Henderson, 1991; Frazier and Fodor, 1978; Frazier, 1987; Frazier and Rayner, 1982) demonstrating that processing difficulty is incurred if a more complex parse needs to be adopted as the sentence is read. For example, work on structural attachment ambiguity (Frazier and Fodor, 1978) demonstrates that the parser prefers to adopt the simplest possible parse of a sentence. More specifically, this work shows that the parser will incorporate subsequent words into the current hierarchical structure that it is building as soon as they are read. If a word is introduced into the structure that does not conform to the currently adopted structural representation, the parser must reanalyze its initial parse and processing difficulty is incurred (as reflected for example by longer reading times compared to a baseline condition).

1.4.1.2 Incremental processing

We also follow previous work (see e.g., Altmann and Steedman, 1988; Tanenhaus et al., 1989, among others) in assuming that sentences are processed in an incremental, left-to-right fashion in English. Moreover, we assume that processing occurs in a word-by-word fashion as each new word in an utterance is read. Previous work has shown that in English, the parser anticipates (or predicts) which grammatical elements it will encounter further to the right in the string,[16] depending on the input it has received so far in the course of the derivation (see e.g., Marslen-Wilson, 1973, 1975; Frazier and Fodor, 1978; Altmann and

[16]To our knowledge, whether or not the parser predicts structures that it will encounter further to the right is still a topic of debate and thus, we remain agnostic about whether structures are anticipated or whether the parser simply predicts the types of grammatical elements that it will find further to the right. In work on filler-gap dependencies, processing studies (Crain and Fodor, 1985; Stowe, 1986) have shown that the parser predicts that it will find a gap for a *wh*-filler in the linearly closest syntactic position. However, it remains unclear whether the parser is predicting a particular structure or if it is simply predicting that it will find a syntactic element that can host a gap site, i.e., a verb site.

Steedman, 1988, among others). As each new word is read, the parser revises its predictions about what type of elements it will encounter further to the right in the derivation (see e.g., Hale, 2001; Levy, 2008).

1.4.1.3 PF as the parser's input

It is important to make a distinction between the input that the parser receives and the linguistic form that can be interpreted by the semantic component. The parser receives as input the linearized string of words that can be roughly understood as the PF representation of an utterance. Since meanings are derived compositionally, based on structure (see e.g., Heim and Kratzer, 1998), the parser must derive an interpretable structure from the string of words it receives as input. Thus, the parser's task in real-time processing is to identify (and build) the LF structure that can be read by the semantic component. In many cases, the linear order in which the words are presented in the pronounced form corresponds directly to the order of words in LF hierarchy.[17] However, in some cases, the linear order in which the words are presented does not correspond directly to the hierarchical organization of the words in the LF. As a result, there can be a *mismatch* between the position of a word in the linear string (sometimes called the *surface string*) and the position the word is assigned in the hierarchical structure that must be built to derive an appropriate interpretation, what

[17]This is a simplistic way of presenting the issue at hand. Crucially, this statement is only true depending on constituency. Within a given constituent, it could be true that the order of the words in the PF representation has a direct correspondence to the order of words at the LF level of representation. However, this statement would not be true across constituents, for example, because there is a particular structure assigned to each constituent that would not be apparent in the pronounced form of the sentence. This statement could be applied to English, under certain circumstances, but it cannot be applied to all languages, i.e., those in which the syntactic structure and the word order are more distinct. It would also not be true for cases of adjunction because the word order of the constituents does not directly correspond to the hierarchical structure, i.e., the word order does not clearly indicate the attachment position of the adjunct, whether it is a phrasal or head adjunct. While this statement is simplistic, we will adopt it for the purposes of the current thesis and propose that when the parser can (roughly) build an LF structure based on the order of the words in the pronounced form of the sentence, that will be simpler than cases where words must be interpreted in completely distinct syntactic positions.

we call a PF-LF mismatch. We propose that a hierarchical structure that can be assigned based on the linear order in which words appear in the linearized input should be simpler to derive than a structure that cannot be built based on the linear string of words, i.e., when words need to be assigned to syntactic positions further to the right/left than where they appear in the PF representation. Under such an assumption, one would predict that a PF-LF mismatch should incur a processing cost.

As each new word in a string is read, the parser makes and revises previous predictions about the types of elements it expects to find further to the right in the string. In the domain of interpretation, the parser should try to derive an interpretable structure as each word is processed. To do so, it will need to build an LF structure that can be read by the semantic component. we assume that there is a general parsing preference for each word to be integrated into the structure that the parser is building as soon as possible, based on the evidence it has received so far. However, it is not always possible for the parser to adhere to this parsing preference. If the parser encounters grammatical elements that cannot be interpreted in their surface positions (i.e., PF-LF mismatch), it should make predictions about the types of elements it will find further to the right in the string in order to be able to interpret that string. For example, if the parser finds an anaphor that does not yet have an antecedent in the surface representation of an utterance, it should predict that it will find an antecedent for that anaphor further to the right in the string. Failing to find the antecedent would yield an uninterpretable structure. If this hypothesis about how the parser makes predictions in real-time are on the right track, it should be possible to find real-time effects of a PF-LF mismatch in the domain of interpretation.

One important question that remains is: how does the parser know what parse is simpler? We assume that this knowledge comes from the preferences of the grammar.

The critical assumption is that the parser is sensitive to the grammar's preferences.[18] For example, work on structural ambiguity resolution has shown that it is simpler to build a structure in which less phrases need to be built. In a garden path sentence, such as *The horse raced past the barn fell*, it is simpler to interpret *raced* as the main verb in the sentence because this parse would not require as many phrases to be built. In contrast, if *raced* is interpreted as the head of a reduced relative clause (the correct parse in this case), the parser must build more phrases to be able to interpret the sentence. As a result, the less operations (e.g., merge, move, copy, etc.) that the parser needs to use to derive a parse, the simpler the parse.

In the next two sub-sections, we will summarize the main findings from previous psycholinguistic work in the domains of filler-gap dependencies and referential dependencies. The experiments that we report in the upcoming chapters of this thesis build on previous work in both of these domains.

1.4.2 Previous psycholinguistic studies

1.4.2.1 Filler-gap dependencies

Psycholinguistic work over the past several decades has shown that *wh*-dependency formation in real-time processing is anticipatory. Upon finding a *wh*-filler (or the head of the chain), the left-to-right parser postulates a gap site (the tail of the chain) further to

[18] A relevant question is whether the parser *is* narrow syntax. What is crucial for the current work is that the form of the sentence at Spell-Out must be accessible to the parser. In other words, the parser must be able to take the string of words at PF and determine what that string looked like when it reached Spell-Out. Once the parser has determine the Spell-Out form of the sentence, it can start building the LF representation. Thus, the parser must be sensitive to the same form of the derivation as narrow syntax. However, the parser is also subject to general processing constraints that arise as a string of words is being processed in real-time. Narrow syntax would not be subject to such constraints. Consequently, we assume that the parser must be able to access the derivation in the same way as narrow syntax but it can be influenced by other extra-linguistic factors. For more on the parser versus grammar debate, see Phillips (1996).

the right in the structure. More precisely, upon finding the *wh*-filler, the parser predicts that it will find the gap site in the linearly closest syntactic position in which a gap site could be found, i.e., at a potential verb site. If the linearly closest potential gap site[19] is instead filled with an overt element (and therefore, not containing a gap), processing difficulty is incurred. The evidence comes from studies showing that when the linearly closest potential gap site is filled with an overt element (*filled-gap* effect), longer reading times are found compared to minimally different constructions in which there is no syntactic dependency, and thus, no gap site.

In a foundational self-paced reading study, Stowe (1986) (see also Crain and Fodor, 1985, among others) compared constructions with a *wh*-dependency to constructions without one, as in (22-a) and (22-b). The critical items differed only in the type of complementizer but otherwise contained the same words. Longer reading times were found on *us* in (22-a) compared to the same word in the control sentence, which did not have a *wh*-dependency, as in (22-b).

(22) a. My brother wanted to know who$_i$ Ruth will bring **us** home to t_i at Christmas.
b. My brother wanted to know if Ruth will bring **us** home to Mom at Christmas.

Assuming that longer reading times reflect greater processing difficulty, these results suggest that the parser had initially predicted that the gap site for *who* would be found following the first verb, *bring*. When the predicted gap site is instead filled with an overt DP, *us*, processing difficulty arises. Since these results suggest that the parser does not wait until the gap site is found to establish a syntactic dependency between the filler and its gap site, such findings have been termed an *active search* effect.

[19]This work concentrates on linearity because it provides evidence that the parser has a preference to resolve open dependencies as soon as possible. The closer the gap site, the sooner it can resolve the dependency.

Another important finding from previous work on filler-gap dependencies is that the parser is sensitive to the fact that such dependencies are not completely free and must abide by grammatical constraints, most notably syntactic islands (Ross, 1967). Experimental work shows that the parser is sensitive to such constraints and will not postulate a gap site in positions in which grammatical constraints, such as islands, would be violated (Stowe, 1986; Traxler and Pickering, 1996).[20]

To sum up, there are two main findings from previous work in the filler-gap literature: i) as soon as a filler is found, the parser postulates a gap in the linearly closest syntactic position from the filler in which a gap would be licit and does not wait until later in the structure to determine whether or not a gap is actually found in that position (*active* search), explaining why we find processing difficulty when the linearly closest position contains an overt element, instead of a gap, and ii) the parser will not postulate gaps in positions where they would violate grammatical constraints (i.e., islands), suggesting that the parser is sensitive to the grammatical constraints imposed on syntactic dependencies in real-time processing.

1.4.2.2 Cataphoric dependencies

Another type of dependency that has been investigated in the psycholinguistic literature is the referential dependency that holds between a pronominal and its antecedent. Upon encountering a pronoun, a referential dependency must be made between the pronoun and

[20] Similar effects have also been found in plausibility studies, demonstrating that processing difficulty arises if the parser attempts to link a filler with an implausible gap site (Traxler and Pickering, 1996; Tanenhaus et al., 1989; Boland et al., 1990; Stowe et al., 1991; Garnsey et al., 1989). Related research has demonstrated the parser is highly sensitive to the licensing conditions imposed on gap sites and will not posit gaps in constructions where they cannot be licensed (Phillips, 2006), and moreover, that it will only posit a gap site when it is obligatory and not when it is optional (Wagers and Phillips, 2009).

its referent, also called the antecedent DP.[21,22] This dependency can be formed in two ways: i) the pronoun can linearly follow its antecedent DP, called forwards anaphora, as in (23-a), or ii) the pronoun can linearly precede its antecedent DP, called backwards anaphora, as in (23-b). For the purposes of the current paper, we will call the former case *anaphoric processing* and the latter case *cataphoric processing*.

(23) a. *Anaphoric processing*:

Mary$_i$ stayed up late to finish the paper even though she$_i$ was tired.

b. *Cataphoric processing*:

Even though she$_i$ was tired, Mary$_i$ stayed up late to finish the paper.

In (23-a), when the pronoun, *she*, is found in the adverbial clause, a dependency must be created between the pronoun and the antecedent DP, *Mary*. Note that, in both cases, it is also possible for the pronoun to receive its referent from the discourse context but for current purposes, we assume that pronouns prefer to find their referent within the sentence (see Van Gompel and Liversedge, 2003 for experimental evidence). For the dependency between the anaphor and its antecedent DP to occur, *Mary* must be retrieved from memory and linked to the pronoun. Such a dependency cannot be made until the pronoun is found in the adverbial clause. When the antecedent is found earlier in the sentence, there is no indication that a referential dependency will be required later in the sentence. In other words, upon reading the antecedent DP in (23-a), the parser cannot predict that a referential dependency will need to be made between this DP and a pronoun later in the structure. The

[21] By *antecedent*, we mean that this DP is co-indexed with the pronoun and they are interpreted as referring to the same referent. Whether or not there is also a structural dependency between the pronoun and the antecedent is a non-trivial question (see Section 1.3.2) and we put aside this discussion for the time being. Note that anaphoric dependencies holding between quantifiers, which do not refer, are beyond the scope of the current work.

[22] For simplicity, we focus on cases where the pronoun refers to a linguistic antecedent DP and we put aside so-called "free pronoun" cases.

situation is different in (23-b). When the pronoun, *she*, is found early in the sentence, it cannot be linked to a co-referent until later in the sentence. Upon finding the pronoun in (23-b), the parser has not yet encountered the antecedent DP. Therefore, when the cataphor is found at the beginning of the sentence, the parser has not yet found a referent for that pronoun. In processing terms, the pronoun linearly precedes its antecedent and cannot be fully interpreted until the antecedent is found further to the right in the linear string. This construction therefore resembles the filler-gap cases discussed earlier in which the filler linearly precedes its gap site.

Active search strategy In terms of linearity, filler-gap and cataphoric dependency constructions resemble each other: in both cases, an element that appears earlier (further to the left) in the structure, i.e., the filler or cataphor, must be associated with another element found later (further to the right) in the structure, i.e., gap or antecedent, in order to satisfy the *wh* or referential dependency. If the parser adopts a similar processing strategy for both types of dependencies, it may predict that upon finding a cataphoric pronoun, its antecedent will be found later in the sentence, i.e., further to the right. Such an effect could be analyzed as an *active search* effect, similar to the effect found in filler-gap studies. Indeed previous work on cataphoric processing provides evidence in favour of the parser adopting an active search strategy for antecedents when it finds a cataphor further to the left in the string.

Preference to find antecedent within the sentence In an eye-tracking while reading study, Van Gompel and Liversedge (2003) investigated sentences containing cataphoric pronouns which either matched (24-a) or did not match (24-b) the first-mentioned DP in gender, i.e, *the boy/the girl*.

(24) a. *Gender match*:

When he$_i$ was fed up, **the boy**$_i$ visited the girl very often.

b. *Gender mismatch*:

When he$_i$ was fed up, **the girl**$_{*i}$ visited the boy very often.

The researchers found that the word following the first-mentioned DP, i.e., *visited*, was read longer in first-pass reading times when it mismatched the cataphoric pronoun in gender, i.e., longer reading times were found on *visited* in (24-b) compared to this same word in (24-a). Kazanina et al. (2007) argue that these results are similar to the filled-gap effect found in earlier studies and suggest that these results may also point to the parser pursuing an active search strategy when processing cataphoric dependencies.[23] More precisely, upon encountering the cataphoric pronoun in a sentence-initial adverbial phrase, the parser is able to reliably predict that a main clause will follow the adverbial. If the parser also knows that the main clause must always contain a subject, it is able to reliably predict that it will find an antecedent for the cataphor in the main clause.[24,25] Therefore, the parser may engage in an active search for an antecedent as soon as it finds a cataphoric pronoun.[26]

[23]Note that the goal of Van Gompel and Liversedge (2003) was to investigate whether the parser considers morphological information during the initial stage of processing during pronoun resolution or whether it only considers structural information at this stage.

[24]Upon encountering the pronoun, the parser does not know that it is cataphoric and therefore, there is no guarantee that an antecedent will actually be found later in the sentence. Thus, the parser may adapt a different strategy and not postulate a later dependency until the antecedent has actually be found. Generally, previous work on referential dependencies has shown that the parser prefers to find a referent within the sentence (Van Gompel and Liversedge, 2003), at least within an experimental setting. For simplicity, we will make this assumption for the studies reported in this thesis.

[25]Note that Kazanina et al. (2007) do not discuss cases of quantificational DPs which bring up added complications. Such cases are beyond the scope of the current thesis and therefore, we leave them aside.

[26]For the current purposes, we are concerned with cases of backwards pronominal dependencies, i.e., where the pronoun precedes its antecedent in the linear order, because they enable the investigation of how the parser resolves constructions search for an appropriate antecedent (see Kazanina et al. (2007) for a similar argument). In forwards pronominal dependencies, the antecedent always precedes the pronominal referent and thus, once the pronoun has been found later in the sentence, it already has an antecedent. Thus, with forwards anaphora, no active search for an antecedent is expected.

Sensitivity to grammatical restrictions in cataphoric dependencies Similarly to the effects found in filler-gap studies, psycholinguistic studies investigating the processing of cataphoric dependencies have also shown that the parser is sensitive to the grammatical restrictions placed on co-referential DPs. This work demonstrates that the parser will not create a referential dependency between two DPs if doing so would incur a binding violation (Cowart and Cairns, 1987; Nicol and Swinney, 1989; Clifton et al., 1997; Kazanina et al., 2007 but see Badecker and Straub, 2002; Sturt, 2003 for results showing that inaccessible antecedents can still affect processing). In a series of self-paced reading experiments, Kazanina et al. (2007) investigated whether the parser is sensitive to Principle C effects[27] upon encountering cataphoric pronouns, as in (25-a)-(25-d). The researchers used the gender-mismatch paradigm proposed by Van Gompel and Liversedge (2003) to investigate whether the parser considers potential antecedents for cataphoric pronouns, even if doing so would incur a Principle C violation.

(25) a. *Principle C, match*:

Because last semester she$_{i/*j}$ was taking classes full-time while Kathryn$_j$ was working two jobs to pay the bills, Erica$_i$ felt guilty.

b. *Principle C, mismatch*:

Because last semester she$_i$ was taking classes full-time while Russell was working two jobs to pay the bills, Erica$_i$ felt guilty.

c. *No constraint, match*:

Because last semester while she$_i$ was taking classes full-time Kathryn$_i$ was

[27] Note that an alternative analysis is that the while-clause could attach above the first subject, i.e., *she*, even in the Principle C conditions. Kazanina et al. (2007) do not consider such cases and manipulate only the surface positions of the relevant DPs. The fact that the parser could interpret the while-clause in a structurally higher position might explain why their results only showed marginal effects for the Principle C cases in their Experiment 1.

working two jobs to pay the bills, Russell never got to see her.

d. *No constraint, mismatch*:

Because last semester while she$_i$ was taking classes full-time Russell was working two jobs to pay the bills, Erica$_i$ promised to work part-time in the future.

In (25-a), co-indexing the cataphoric pronoun, *she*, with the first-mentioned DP, *Kathryn*, would incur a Principle C violation because the pronoun would c-command the DP. The pronoun can only be co-indexed with the embedded subject, *Erica*, because the pronoun does not c-command the embedded subject. In (25-b), the first-mentioned DP mismatches the pronoun in gender, thus co-indexation is not possible both for Principle C and due to mismatching gender. In the control conditions, (25-c)-(25-d), the first-mentioned DP can be co-indexed with the cataphoric pronoun, provided they match in gender, because there would be no Principle C violation. In (25-c), co-indexation between the pronoun and *Kathryn* would not incur a Principle C violation because the pronoun is not c-commanded by the DP. The critical assumption in the gender mismatch paradigm is that gender mismatch effects reflect the parser establishing a referential dependency between the pronoun and the antecedent. Such effects appear when the first mentioned DP mismatches the cataphoric pronoun in gender, suggesting that the parser had expected to create the referential dependency in this position. In the control conditions, (25-c)-(25-d), gender mismatch effects should be found because co-indexation between the cataphor and the first mentioned DP would be interpretable. More precisely, *Russell* in (25-c) should be read longer than *Kathryn* in (25-d). However, if the parser is sensitive to Principle C violations, no gender mismatch effect is expected in (25-a)-(25-b) since co-indexation between the pronoun and the first-mentioned DP would incur a Principle C violation. In other words, (25-a) and (25-b) should be read similarly.

Kazanina et al. (2007) found a significant interaction between constraint type (Principle C/no constraint) and gender (match/mismatch with the cataphor) at the second subject DP (*Kathryn/Russell*), demonstrating that there was no effect of gender mismatch in the Principle C conditions but there was a marginal effect in the no-constraint conditions. More precisely, longer reading times were found when the gender of the second DP subject mismatched the gender of the pronoun (i.e., the gender mismatch effect) but only in the no-constraint condition. In other words, no gender mismatch effect was found if co-indexation of the pronoun and the subject DP would incur a Principle C violation.[28] The researchers interpret these results as demonstrating that the parser did not consider the second subject DP as an antecedent for the cataphoric pronoun when this would incur a Principle C violation.[29] In addition, Kazanina et al. (2007) argue that their results provide evidence in favour of the parser adopting an active search mechanism in cataphoric processing: as soon as the parser finds a cataphoric pronoun, it actively searches for an antecedent, except in conditions where a grammatical principle would be violated (Principle C). As a result, the parser seems to be adopting a similar strategy when resolving cataphoric dependencies as it did when resolving *wh*-dependencies.

1.4.2.3 Summary of previous psycholinguistic studies

To summarize, previous work on the processing of filler-gap and cataphoric dependencies has led to three main discoveries: i) upon finding a phrase that needs to satisfy a syntactic

[28]A similar effect was also found in offline judgements: participants were significantly less likely to accept co-reference between the cataphoric pronoun and the second subject DP in the Principle C conditions compared to the no-constraint condition.

[29]Such effects were also replicated in two follow-up self-paced reading studies which controlled for different structural properties of the stimuli. These studies were reported as Experiments 2 and 3 in Kazanina et al. (2007).

dependency that cannot be satisfied based on its surface position (i.e., a *wh*-filler or cataphor), the parser engages in an active search for an element (i.e., gap/antecedent) that can satisfy that dependency; ii) the parser assumes that the syntactic dependency will be satisfied in the linearly closest position (first potential gap/antecedent site); and iii) the parser will not postulate gap/antecedent sites if doing so would lead to a grammatical violation (island or binding principle violation).

1.5 Reconstruction in real-time processing

Whether or not the parser is sensitive to reconstruction effects in real-time has remained relatively unexplored thus far in the experimental literature. We only know of one study that investigated potential effects of reconstruction in real-time processing using *wh*-constructions in which the anaphor linearly preceded its antecedent.[30] Using self-paced reading, Frazier et al. (1996) investigated interrogative constructions in which an anaphor was embedded in a *wh*-phrase that linearly preceded its antecedent. In their Experiment 1, they manipulated the structural positions of the antecedent for the anaphor, as shown in (26). The antecedent either appeared as the subject of the matrix clause, as in (26-a), as the embedded subject, as in (26-b), or both subjects were possible antecedents for the

[30] Another relevant study on reconstruction was conducted by Omaki (2010). The goal of Omaki (2010)'s study was to compare the processing of anaphors appearing in either *wh*-arguments or *wh*-predicates. In this experiment, the matrix subject either matched or did not match the anaphor in gender. For *wh*-argument conditions, if the matrix subject matched the anaphor in gender, it was able to structurally bind the anaphor. However, if it did not match the anaphor in gender, the phrase containing the anaphor would need to be bound by the embedded subject, appearing further to the right in the linear string. Since we did not examine *wh*-predicates in this thesis and since we did not manipulate whether the matrix subject could be interpreted as the antecedent for the anaphor, we cannot compare our results to this work. What is relevant to the current work is that in Omaki (2010)'s studies, upon finding the anaphor in the *wh*-phrase, the parser attempted to bind the anaphor to the matrix subject and that a cost was incurred (longer reading times) when the matrix subject was not an appropriate antecedent (did not match in gender). This result suggests that the parser attempts to find antecedents for anaphors in the linearly closest position.

anaphor, as in (26-c). Frazier et al. (1996) were interested in whether participants preferred to find the antecedent for the anaphor within the clause in which the anaphor appears in its surface position (i.e., *the actress* in (26-a)) or whether the antecedent could be found in the reconstructed position (i.e., *the actress* in (26-b)).[31] In the terms discussed in this thesis, if participants interpret the matrix subject as the antecedent for the anaphor, there must be an intermediate reconstruction position available that is found structurally below the matrix subject. If the phrase containing the anaphor reconstructs to this intermediate copy position, the matrix subject would be able to locally bind the anaphor. However, if participants interpret the embedded subject as the antecedent for the anaphor, they must have a preference for reconstructing the phrase containing the anaphor to a lower copy position in which the embedded subject could locally bind the anaphor.

(26) a. Which rumor about herself did the actress claim the newspaper made up?
b. Which rumor about herself did the newspaper claim the actress made up?
c. Which rumor about herself did the matron claim the actress made up?

Overall, Frazier et al. (1996) found that participants preferred to interpret the anaphor as referring to the matrix subject, e.g., *the actress* in (26-a) or *the matron* in (26-c). This was reflected in both reading time data and question accuracy (see Plunkett, 1991, as cited in Frazier et al., 1996, for offline data supporting this preference). Thus, the results reported in Frazier et al. (1996) could be taken as evidence that the parser prefers to interpret the linearly closest feature-matched DP as the antecedent for the anaphor. We will return to this discussion in Chapter 4.

[31] Frazier et al. (1996) also investigated how binding would interact with scopal effects. They included conditions in which a quantified determiner phrase, e.g., *every actress* appeared either as the matrix or embedded clause. Since quantified phrases are beyond the scope of the current thesis, we leave these conditions aside.

1.5.1 Summary

To summarize the background relevant for this thesis, we have argued that previous psycholinguistic work on filler-gap dependencies can be understood as demonstrating the parser's sensitivity to one type of PF-LF mismatch. *Wh*-dependencies represent one type of PF-LF mismatch because the parser's task of integrating words into the structure when they are encountered cannot be completed until it has built further structure. Since a *wh*-filler forms a complex syntactic object with its gap site (head and tail of the chain), the parser must locate that gap site to be able to fully integrate the phrase. In the case of *wh*-dependencies, the parser needs to *syntactically* integrate the filler at the tail of the chain. In other words, the parser must locate the tail of the chain to determine the grammatical role of the *wh*-filler. We argue that cases involving filler-gap dependencies represent only one type of PF-LF mismatch.

Following copy theory of movement, as a phrase moves, a copy is left behind in each position it occupies through the movement chain. The *wh*-operator must be interpreted in its surface position (high) but whether the complement of *wh* can also be interpreted high depends on the properties of the linguistic elements it contains. In canonical *wh*-constructions, it is difficult to determine whether the complement of *wh* was interpreted in the higher or lower copy position because either position would yield an appropriate semantic interpretation at LF.

To systematically investigate whether the complement of *wh* could be interpreted in the higher copy position or whether it needed to reconstruct to a lower copy position, we manipulated the type of DP embedded in the *wh*-phrase. Even though the *wh*-operator must be interpreted in its surface position, the complement of *wh* must reconstruct to a lower copy position in order to be assigned a semantic interpretation. Thus, *wh*-dependencies involving reconstruction potentially present the parser with two separate types of integration tasks: i) it must syntactically integrate the *wh*-phrase in the lowest copy position (base-generated position) so that the phrase is assigned a grammatical role, and ii) it must reconstruct the complement of *wh* to a lower copy position (which could be the base-generated position or could be an intermediate position) so that the phrase can be assigned a semantic interpretation. Crucially, the task of locating the tail of the chain to be able to syntactically integrate the phrase is distinct from the task of reconstructing the complement of *wh* to a lower copy position for interpretation reasons.

In our experiments, we manipulated the PF positions of different types of DPs by embedding them in embedded *wh*-filler phrases. Anaphors and pronouns were placed in positions in which they linearly preceded potential antecedents in the PF representation. Our baseline condition consisted of filler phrases embedded with R-expressions since they do not need antecedents. While anaphors must find local antecedents, pronouns cannot be bound by local antecedents (Principles A and B). Nevertheless, we followed previous experimental work (Van Gompel and Liversedge, 2003; Kazanina et al., 2007) in assuming that the parser prefers to find the antecedent for a pronoun within the sentence. Crucially, even when embedded in a filler phrase, pronouns and R-expressions can be interpreted in their surface positions. In contrast, in order for anaphors embedded in *wh*-fillers to be interpreted, the phrase containing the anaphor (the complement of *wh*) must reconstruct to a lower copy position at LF.

We also investigated whether the parser shows effects of reconstruction for the *de dicto* interpretation in *how many* questions. Since we assume that the parser adopts the simplest possible parse of a given string of words, the parser should always adopt a *de* *re* interpretation in *how many* questions, unless it is given evidence to pursue a different parse. If the parser adopts the *de re* reading, it is able to interpret *many-x* as soon as it is encountered and this parse is simpler than one in which the parser must interpret *many-x* further to the right in the string. To force the *de dicto* interpretation, we used creation verbs as embedded verbs in *how many* questions. If the parser is sensitive to reconstruction for the *de dicto* interpretation, it should show a processing cost on the creation verb compared to a non-creation verb. Upon reaching the creation verb, the parser should determine that it must reconstruct *many-x* to derive the *de dicto* interpretation. Failure to reconstruct would result in an uninterpretable string. Since we had two different motivations for reconstruction: i) binding, and ii) scope (*de dicto* reading), we predicted that they might interact in a processing experiment, building on a paradigm developed by Hackl et al. (2012). More precisely, if the complement of *wh* contains an anaphor, the parser is given an early signal that it must reconstruct for binding. Finding a creation verb further to the right in the string will signal to the parser that it must also reconstruct for scope. However, since it had already reconstructed the phrase for binding, it should be less costly to engage in a second reconstruction operation, compared to structures in which the parser only finds out that it must reconstruct for scope upon reaching the creation verb. We go over the precise predictions of this experimental paradigm in Chapter 2.

In the current thesis, we investigate potential effects of reconstruction on LF structure by combining two well studied domains in the psycholinguistic literature, namely filler-gap and referential dependencies. Constructions involving reconstruction are particularly interesting to investigate in the domain of sentence processing because they involve a mismatch between the surface string that the parser is given as input and the LF structure it needs to build to derive an appropriate semantic interpretation. Following the assumption that the parser prefers the simplest possible parse of a given string of words, the parser should attempt to build an LF structure that corresponds most directly to the surface string it is given as input (the PF representation). However, cases involving reconstruction require that the parser builds an LF that is distinct from the order of words in the surface string and thus, they should be pursued as a last resort. If the parser is sensitive to effects of reconstruction, we predict that we should be able to observe real-time processing effects when reconstruction is required compared to when it is not. Since reconstruction is an additional operation that is not always required, we predicted that a structure involving reconstruction should incur a processing cost compared to a structure that does not involve reconstruction. We used reading studies to investigate this question. In such studies, longer reading times are assumed to reflect greater processing costs. Consequently, assuming that reconstruction is costly, we predicted longer reading times in constructions involving reconstruction compared to constructions that did not require reconstruction.

1.5.2 How this thesis is structured

This thesis is structured as follows: In Chapter 2, we report the results of a self-paced reading experiment in which we compared the processing of anaphors embedded in *wh*-phrases compared to R-expressions found in the same syntactic positions. The antecedent for the anaphor was always found in a syntactic position further to the right in the linear string. The logic behind this experiment was to investigate how the parser arrives at an interpretation for an anaphor that cannot be semantically interpreted, i.e., assigned a denotation, in its surface position. Since the anaphor cannot be interpreted in the higher copy position, the phrase containing the anaphor must reconstruct to a lower copy position that is structurally below the antecedent for the anaphor. This operation will ensure that the anaphor is assigned a semantic denotation and that it can be integrated within the structure. It also ensures that the structure abides by Binding Principle A. We also manipulated whether the embedded verb was a creation or non-creation verb. We assumed that the parser must reconstruct the complement of *wh* (*many-x*) upon reaching the creation verb in order to derive an interpretable structure (*de dicto*). In contrast, the complement of *wh* can be interpreted in its surface position when the verb is a non-creation verb, since these verbs are compatible with *de re* readings. We predicted that if the parser is sensitive to both motivations for reconstruction, we should find a facilitation effect when the complement of *wh* contains an anaphor and the embedded verb is a verb of creation. The anaphor further to the right in the string should signal to the parser that reconstruction must occur further to the right. Reconstruction for the *de dicto* interpretation of creation verbs should be less costly in anaphor conditions since the parser has already reconstructed the phrase for binding. In contrast, if the parser has not been given an early signal that it must reconstruct further to the right (i.e., when the complement of *wh* contains an R-expression), reconstruction should be more costly at the embedded verb of creation. The results show a processing cost in anaphor conditions compared to R-expression conditions. We argue that these results are compatible with a reconstruction analysis of binding following which the parser reconstructs the phrase containing the anaphor to a lower copy position in order to satisfy Binding Principle A at LF. However, the results are also compatible with a non-structural explanation following which the parser simply looks for a feature-matched antecedent upon encountering an anaphor that does not have an antecedent in the preceding linguistic context, i.e., it must look to the right to locate an antecedent for the anaphor but it is not concerned with the structural position of the antecedent. Under this analysis, it is sufficient that there is a feature-matched antecedent within the sentence. We found no reliable effects of verbs of creation and as a result, we cannot make any claims about parsing at this position in the sentence.

In Chapter 3, we report the results of a self-paced reading experiment designed to directly address the structural versus non-structural explanations of the results reported in Chapter 2. In Chapter 3, we investigate whether we can find stronger evidence in favour of a structural explanation. The relevant manipulation in this experiment was whether a pronoun or an R-expression occurred in the filler phrase. Since pronouns can also serve as embedded subjects, we also manipulated whether the embedded subject was a pronoun or R-expression. We found a processing cost for R-expressions in both syntactic positions. Crucially, pronouns were not costly, in contrast to what we found with anaphors. We interpret these results as supporting a reconstruction analysis of the results reported in Chapter 2. Since anaphors were costly but pronouns were not and phrases containing anaphors must reconstruct to find an appropriate antecedent, our results suggest that the parser is sensitive to reconstruction for anaphoric binding. We again investigated whether the parser is sensitive to reconstruction for the *de dicto* interpretation of creation verbs and we failed to find an effect of reconstruction. As a result, we dropped this manipulation in subsequent experiments. In a follow-up study reported in Chapter 3, we manipulated the result demonstrating that R-expressions are more costly than pronouns. However, in this follow-up study, we also manipulated whether the R-expression was "familiar" (name of a famous person) or "unfamiliar" (a made-up name). We found that unfamiliar R-expressions were more costly than familiar ones but they were both still more costly than pronouns. We argue that this result argues against a purely lexical explanation of our results, i.e., R-expressions are more costly because they carry more discourse properties, and that they provide support for a structural explanation, i.e., binding.

In Chapter 4, we report a series of self-paced reading experiments and one eye-tracking while reading experiment designed to investigate how the parser processes potential antecedents that appear in positions from which they could not structurally bind anaphors at LF. More precisely, potential antecedents either appeared in adjunct positions from which they could not structurally bind the anaphor or they appeared in subject position from which they could structurally bind the anaphor. In these experiments, the filler phrase either contained an anaphor or a pronoun and we manipulated whether the potential antecedents matched or did not match the anaphor/pronoun in gender. Overall, we find different gender effects based on the structural position of the DP, suggesting that the parser is sensitive to the structural positions of potential antecedents.

In Chapter 5, we propose a series of computational metrics to explain the collected data based on general processing constraints and grammatical constraints proposed in the theoretical linguistics literature. We discuss where these metrics correctly predict the results reported in this thesis and where they fall short. We also report the results of a follow-up study on the processing of DPs in positions where anaphors and pronouns have antecedents in the previous linguistic context. This study was conducted to further refine the proposed metrics accounting for the complexity of different types of DPs.

In Chapter 6, we conclude the thesis and discuss some avenues for future work as well as some open questions that this thesis was not able to address.

Chapter 2

Reconstruction for anaphoric binding

2.1 Introduction

This chapter reports the results of a naturalness judgement task and a self-paced reading experiment investigating whether the parser is sensitive to potential effects of syntactic reconstruction. As discussed in Chapter 1, we used constructions argued to involve syntactic reconstruction in an attempt to investigate whether the left-to-right parser is sensitive to a PF-LF mismatch in real-time processing. In Chapter 1, we discussed the possibility that the parser might reconstruct a phrase in order to satisfy binding principles at LF. We also investigate whether an early signal for reconstruction for binding facilitates later reconstruction for scope.

2.1.1 Experimental setup

Previous work in both theoretical linguistics and psycholinguistics has argued that structural economy plays a role in the way that grammatical representations are derived. The assumption has been that additional operations should not be pursued unless doing so would give rise to a distinct interpretation compared to the more economical parse (for theoretical work, see e.g., Fox, 1995, 1998, 2000; for experimental evidence, see e.g., Anderson, 2004). Moreover, if a more complex (less economical) parse is pursued, it will incur a cost (see e.g., Anderson, 2004 for a cost associated with quantifier scope ambiguity). Assuming that the parser is sensitive to the grammar's preferences, we should be able to observe evidence for such a cost in the processing domain. *How-many* questions with creation verbs provide one case where the *de re* (or simpler) parse is not available and the sentence must be interpreted as *de dicto*. From a processing perspective, the reading that is forced with creation verbs is unusual because reconstruction, or the more complex structure, is obligatory. Thus, the parser cannot adopt the simplest possible structure in these cases. If we assume, though, that the parser will adopt the simplest structure unless it is given other information to pursue a different parse, we expect that the parser will always assume the *de re* interpretation unless it has been given an earlier indication that it needs to pursue a different interpretation. In the *how-many* questions with creation verbs, the parser will not know that it has pursued the incorrect parse of the sentence until it reaches the creation verb further to the right in the linear string. Thus, when the creation verb is found, we expect to observe a processing cost (a reanalysis effect) because the parse that was being entertained turned out to be infelicitous with the semantics of the creation verb.

In the current experiment, we use *how-many* questions with and without creation verbs to be able to test for potential real-time effects of a PF-LF mismatch in real-time processing. If the parser assumes the simplest possible structure, therefore interpreting *many-x* in its surface position, we should see a processing cost once it encounters the verb of creation because at this point, the parser will need to reanalyze its original parse of the sentence.

However, it will not know that its parse is incorrect until the actual verb site is reached (creation, non-creation) because there would be no earlier indication that a creation verb is going to be found further to the right in the string.

Crucially, there is a potential confound brought in by the semantics of the creation verbs themselves. Semanticists have argued that creation verbs are semantically more complex than non-creation verbs (see Dowty, 1979; Krifka, 1989; Kratzer, 1994; Zepter, 2000; von Stechow, 2001, among others). This is because the semantics of creation verbs must ensure that the object exists only after the event has been completed. No such requirement is needed for non-creation verbs. If it is true that creation verbs are semantically more complex than non-creation verbs, this presents a potential confound for using them as a tool to test for reconstruction. If we find a cost incurred at the verb site in *how-many* questions with creation verbs, it is possible that this cost may be attributable to a PF-LF mismatch (i.e., deriving the *de dicto* interpretation) but it may also simply reflect the semantic complexity of the verb itself. Therefore, we need a strategy to investigate the *de re* versus *de dicto* interpretations without the added confound brought in by the semantics of the creation verbs. The strategy used to systematically investigate this question built on a paradigm developed by Hackl et al. (2012), which we will explain briefly in the next sub-section.

2.1.1.1 Hackl et al. (2012)

Hackl et al. (2012) investigated whether finding an object quantifier further to the left in the linear string would facilitate the resolution of an antecedent contained deletion (ACD) site further to the right in the string. It has been argued that quantifiers in object position cannot be interpreted in their surface position, unlike other DPs in this same position (Heim and

Kratzer, 1998; May, 1977). For example, the definite object DP in (1-a) can be interpreted in its surface position but the quantified object DP in (1-b) cannot. To be semantically interpreted, object quantifiers must undergo quantifier raising (QR) at LF, as the shown in (1-c).

(1) a. The child loves the park.
    b. The child loves every toy.
    c. [every toy] The child loves ~~[every toy]~~

A similar operation has also been argued to occur in ACD constructions. More precisely, the DP that hosts the ACD must QR at LF in order to be interpreted (Sag, 1976). For example, compared the elided construction in (2-a) to the ACD construction in (2-b).

(2) a. VP ellipsis: John read the book and Mary did ~~read the book~~ too.
    b. ACD: John read the book [that Mary did ~~read the book~~].

In the ACD construction, the elided material is contained within the relative clause structure. However, the antecedent for the elided material is the matrix VP, similarly to non-ACD VP ellipsis. Therefore, on the surface, the elided structure is found within its antecedent. This presents a problem for how to interpret the elided structure. As argued by Sag (1976), at LF, there is a licensing condition imposed on ellipsis: the antecedent and the elided phrase must be able to establish an identity relation. However, if the elided phrase is contained within its antecedent, then these elements cannot be identical, violating the licensing requirement on ellipsis. Since ACD constructions are grammatical, there must be a way to satisfy the licensing requirement. The solution proposed is that the DP that hosts the relative structure must move covertly at LF to a position above the VP. If this happens, then the elided phrase is no longer contained within its antecedent, i.e., it is found in a higher position. Once the DP moves, it leaves behind a trace, which can be bound by the moved element and this satisfies the licensing requirement. The movement operation that must occur to resolve the ACD site is similar to the QR operation that must occur to interpret quantifiers in object position.

There are now two motivations for QR: i) interpretation of an object quantifier; ii) resolution of an ACD site. Crucially, QR for ACD resolution must occur regardless of what type of DP heads the relative clause, i.e., whether it is a quantifier or not. In their experiment, Hackl et al. (2012) investigated whether an object quantifier hosting an ACD site would facilitate resolution of the ACD site. If both an object quantifier and an ACD site are found in the same sentence, QR must occur for two independent reasons.

Hackl et al. (2012) predicted that finding an object quantifier hosting the ACD site will provide an early indication that QR needs to occur, compared to a construction where a definite determiner hosts the ACD site. They predicted that when the ACD is found later in the sentence, it will be easier to resolve when QR had to occur for another reason (i.e., interpreting the object quantifier), compared to cases where there was no early indication (e.g., with a definite determiner). More precisely, when the ACD site is found, the parser needs to find an antecedent in order to resolve the site. Hackl et al. (2012) predicted that it would be easier to find this antecedent if QR has already occurred. Sample stimuli is shown in (3):

(3) The understaffed general hospital was negotiating with ...

a. No QR or ACD:

the doctor that the nonprofit medical organization funded

b. ACD but no QR:

the doctor that the nonprofit medical organization was

c. QR but no ACD:

every doctor that the nonprofit medical organization funded

d. QR and ACD:

every doctor that the nonprofit medical organization was

... in order to arrange for free vaccination clinics.

The stimuli was created using a 2x2 design crossing determiner (*every* vs. *the*) and verb type (main verb, elided verb). The relative clause always occurred in object position. When the relative clause was hosted by a quantifier (e.g., (3-c), (3-d)), the quantifier needed to undergo QR in order to be interpreted. QR was not required when the relative clause was hosted by a definite determiner since definite DPs can be interpreted in their surface positions (e.g., (3-a), (3-b)). Similarly, when the verb phrase was elided within the relative clause, thus introducing an ACD site, QR was required to resolve the ACD site (e.g., (3-b), (3-d)). In contrast, if a main verb was used in the relative clause, no QR operation was required (e.g., (3-a), (3-c)). Hackl et al. (2012) predicted that quantifiers in object position facilitate ACD resolution, then we should find faster reading times on the *every-was* condition, where QR occurs for two independent reasons, compared to *the-was*, where QR only occurs for ACD resolution. This prediction was borne out: faster reading times were found for *every-was*, e.g., (3-d), compared to *the-was*, e.g., , on the third and fourth words following the verb (e.g., *to* and *arrange* in), suggesting that finding an object quantifier earlier in the sentence, which indicated that QR needed to occur, did facilitate ACD resolution. Moreover, longer reading times were observed for *the-was* compared to the conditions where a main verb was found. Additionally, no significant differences were found between *every-was* and the main verb conditions. The results from Hackl et al. (2012)'s study demonstrate that effects of covert movement can be found in real-time processing and that the parser is sensitive to these grammatical restrictions. When the parser finds an object quantifier earlier in the sentence, this indicates that QR needs to occur which makes resolving the ACD site easier. However, when the parser only finds a definite determiner, it is not given any early information about QR and it only finds out that QR needs to occur when it reaches the ACD site, incurring a processing cost.

The question of interest in this thesis shares some properties with the paradigm used in Hackl et al. (2012)'s study. Similarly to ACD resolution, the parser will not know that the *how-many* phrase cannot be interpreted in its surface position until the creation verb has been reached further to the right in the string. Building on Hackl et al. (2012)'s paradigm, we need a tool to facilitate the reconstructed reading of the *many-x* phrase. One tool that we can use to facilitate reconstruction is Binding Theory, specifically, anaphoric binding,[32] as discussed in Chapter 1.

2.1.1.2 Reconstruction for binding

As described in Chapter 1, different types of DPs in English (namely, anaphors, pronouns, and R-expressions) are restricted by distinct structural constraints with respect to their antecedents. Presenting the DPs in positions in which they linearly precede (potential) antecedents allows us to investigate whether the left-to-right parser is sensitive to these

[32]See Dillon et al. (2013) for experimental evidence suggesting that the parser uses syntactic information to determine antecedents for anaphors. The experiments reported in Dillon et al. (2013) are different in both design and motivation from the experiments reported in this thesis. More precisely, the researchers investigated whether intrusive DPs appearing in inaccessible structural positions affected how participants resolved subject-verb agreement and antecedent resolution for reflexive pronouns. While they observed intrusion effects for their subject-verb agreement condition, no such effects were found in the reflexive pronoun condition. They argue that their results suggest that participants use syntactic information for antecedent resolution but they use both morphological and syntactic information for subject-verb agreement.

structural constraints (binding constraints). Of particular interest is the way in which the parser processes anaphors when they appear in positions in which they linearly precede their antecedents. On the surface, such constructions appear to violate Binding Principle A, as shown in (4). As argued in Chapter 1, in such cases, the complement of *wh* must be interpreted in a lower copy position, i.e., one in which the anaphor appears in a structurally lower position than its antecedent at LF. This enables the anaphor to be assigned a semantic denotation, a requirement to be able to integrate the phrase in the structure, and also satisfy Binding Principle A. If the phrase containing the anaphor is not interpreted in a lower copy position, it will not be interpretable because the anaphor does not have a structural antecedent in its surface position and consequently, cannot be assigned a semantic denotation since its assignment function is dependent on its antecedent.

(4) How many pictures of himself$_i$ did John$_i$ take ~~pictures of himself~~?

Crucially, reconstruction of *many-x* to satisfy Binding Principle A in (4) is obligatory. Manipulating the type of DP in the *wh*-phrase informs the parser early on (further to the left) if the complement of *wh* is interpretable. If the parser cannot assign the DP a semantic denotation (i.e., as in the case of anaphors and perhaps pronouns), it will be provided with an early indication that it will need to interpret the lower copy. As a result, manipulating the type of DP in the *wh*-phrase enables us to control for the issue brought up earlier with respect to the *de re* versus *de dicto* readings of creation verbs. We showed that the parser will not know that *many-x* cannot be interpreted as *de re* when the embedded verb is a creation verb until the verb is found further to the right in the sentence. As a result, the parser should pursue the *de re* interpretation until it finds the creation verb further to the right. Upon finding the creation verb, it should determine that its initial parse is incompatible with the verb and it must interpret the phrase as *de dicto*. By manipulating the type of DP found in the *many-x* phrase, we can manipulate whether the parser has been provided with an early signal about reconstruction or whether it needs to wait until it has found the creation verb further to the right to determine that reconstruction is required. More precisely, upon finding the anaphor in the *many-x* phrase, the parser will determine that the anaphor cannot be interpreted in its surface position and thus, that the phrase containing the anaphor must reconstruct and be interpreted in the lower copy position. Since the phrase containing the anaphor cannot be interpreted in its surface position, albeit for a different reason, the parser should not try to interpret the phrase as *de re* either. Since the parser is already interpreting the phrase in a lower copy position, it should be easier for it to also interpret the phrase as *de dicto* upon reaching the creation verb further to the right in the string.

2.2 Experiment 1

Broadly, the goal of Experiment 1 was to investigate whether the parser is sensitive to LF structures that cannot be built directly based on the order in which the words appear in the linear string that the parser receives as input (PF-LF mismatch). The simplest possible representation is one in which the LF structure can be built based on the order of the words in the linear string. We manipulated whether such a structure could be derived in this experiment by using constructions that have been argued to require syntactic reconstruction in order to be interpretable by the semantic component. Assuming that the parser is sensitive to PF-LF mismatches, our more specific goal was to investigate whether an early signal for reconstruction for anaphoric binding would facilitate reconstruction for semantic scope, specifically for the *de dicto* interpretation of creation verbs.

In Experiment 1, constructions had the possibility of having two independent reasons for reconstruction at LF: i) anaphoric binding, and ii) scope. As anaphors can be embedded within *wh*-phrases, they provide us with a way to test for potential facilitation effects when a creation verb is found further to the right in the sentence. If the parser finds an anaphor early in the sentence (i.e., further to the left), and if the parser is sensitive to binding requirements, the anaphor embedded in the *wh*-phrase will provide an early indication that the *many-x* phrase needs to reconstruct to a syntactically lower position at LF. If the parser finds a creation verb as the embedded verb and if the parser is sensitive to the *de re/de dicto* distinction, the parser will determine that *many-x* must reconstruct below the creation verb in order to derive the *de dicto* interpretation. Thus, either finding an anaphor in the *many-x* phrase or finding a creation verb as the embedded verb will signal to the parser that it must reconstruct *many-x* in order to derive an appropriate interpretation. The difference between the two types of reconstruction is that the anaphor appears early in the sentence (i.e., further to the left) while the creation verb appears relatively late in the sentence (i.e., further to the right). Following the logic from Hackl et al. (2012), if the parser has already encountered an anaphor in the *many-x* phrase, it has already been given an early signal that it must reconstruct the phrase containing the anaphor to a structurally lower position, i.e., below its antecedent. If the parser also finds a creation verb further to the right in the structure, it should be less costly to continue to reconstruct *many-x* to a structurally lower position in order to adopt the *de dicto* interpretation. In other words, since the parser had reconstructed *many-x* for binding purposes, it should be less costly to continue reconstructing *many-x* to a (potentially) second reconstruction position. In contrast, if there is no early indication that reconstruction needs to occur, i.e., if there is an R-expression in the *wh*-phrase, we expect processing difficulty to arise when a creation verb is found later in the sentence. In this latter case, the parser had been given no prior indication that reconstruction needed to occur and thus, it must reanalyze its parse and adopt the less preferred structure (i.e., reconstruction) when it finds the creation verb later in the sentence.

2.2.1 Design

Two variables were manipulated in our stimuli: DP type and Verb type. The DP was either a third person, singular anaphor, e.g., *himself/herself*, or an R-expression, which was always a first name, e.g., *John, Mary*. The verb was either a creation verb, e.g., *write, invent*, or a non-creation verb, e.g., *edit, deny*. Crucially, all of the non-creation verbs carried a presupposition of the existence of the object, i.e., they could not have been construed as creation verbs. For example, the object of the non-creation verb *edit* must already exist before the editing event takes place. The different types of DPs were always embedded inside a *picture DP*, e.g., picture of herself, lie about himself, etc. We used picture DPs because they can be embedded in *wh*-fillers and are compatible objects with both creation and non-creation verbs.[33]

In order to set up the sentential context in a natural way, we used embedded questions. The subject of the main clause was always a plural DP, ensuring that the main clause subject could never be construed as a possible antecedent for the third person singular anaphor. The scope taking element in our stimuli was always the verb *ask*, which occurred in the present progressive to facilitate a future-oriented reading of the embedded verb, i.e., to ensure that

[33]We are aware of the literature on picture DPs following which anaphors embedded in picture DPs have been analyzed as logophors which are exempt from Principle A of the Binding Theory (Pollard and Sag, 1992; Reinhart and Reuland, 1993). Thus, there is a potential confound in our stimuli because our anaphors could be analyzed as logophors since they are all embedded in picture DPs. However, work in this domain has also shown that for an anaphor to be exempt from the binding principles, its antecedent must be the perspective holder of the event (Charnavel and Sportiche, 2016). In all of our stimulus items, the perspective holder was the pronoun *you*. Thus, we argue that the anaphors used in Experiment 1 are true anaphors and they are not exempt from Principle A.

the creation verb was being interpreted as intended such that its object could not have existed before the creation event. The subject of the asking event was always the pronoun *you* to ensure that it could not be misinterpreted as the binder of the anaphor. Finally, the antecedent of the anaphor was always a proper name, which matched the anaphor in gender. We ensured that all names used in the experiment were not gender neutral and thus that the embedded subject was necessarily the antecedent for the anaphor. The experimental design was therefore 2x2. Sample stimuli are shown in (5).

(5) The reporters wondered ...

a. *Anaphor, creation*

how many lies about herself$_i$ you are asking Alexa$_i$ to invent

b. *Anaphor, non-creation*

how many lies about herself$_i$ you are asking Alexa$_i$ to deny

c. *R-expression, creation*

how many lies about Sean you are asking Alexa to invent

d. *R-expression, non-creation*

how many lies about Sean you are asking Alexa to deny

... for tomorrow's live television interview.

2.2.1.1 Predictions

Region 1: DP region The first region of interest in our stimuli was the DP region in the *many-x* phrase. The DP embedded in the *wh*-phrase was either an anaphor or an R-expression, as shown in (5). When the parser finds an anaphor in the complement of *wh*, it should determine that the DP is not interpretable in its surface position because there is no feature-matched (gender, number) antecedent in the previous linguistic context. As a result, the DP cannot be assigned a semantic denotation. In all cases, the anaphor linearly preceded its antecedent. If the parser is sensitive to Binding Principle A and thus knows that anaphors must be structurally bound by feature-matched antecedents, processing difficulty should be incurred when the anaphor cannot be interpreted in its surface position and must find an antecedent further to the right in the string of words. If the parser is sensitive to this constraint, we expect a slowdown (longer reading times) on the words following the anaphor in (5-a)-(5-b) but not in conditions in which an R-expression is found in this position, e.g., (5-c)-(5-d). Unlike the anaphor, the R-expression can be interpreted (i.e., assigned a semantic denotation) in its surface position and thus, we do not predict any processing difficulty on the words following the R-expression. The R-expression conditions can thus be viewed as the baseline condition. Since the DP containing the anaphor cannot be interpreted in its surface position, the parser must interpret the phrase in a lower copy position. Consequently, upon processing the complement of *wh* further to the left in the string of words, the parser will be left with an uninterpretable phrase. Since the phrase does not have a semantic denotation, the parser is unable to integrate it within the structure. Therefore, there is an open dependency that the parser must resolve before the sentence can be interpretable. One possibility is that the parser creates a partial denotation for the phrase containing the anaphor and that it keeps this partial denotation in memory until it finds an antecedent for the anaphor further to the right. Upon finding an anaphor, the parser reconstructs the complement of *wh* to a copy position that is structurally lower than the antecedent. Since the parser must keep the phrase containing the anaphor in memory until it finds the antecedent, we expect a slowdown on words following the anaphors compared to those same words following the R-expression until the antecedent (embedded subject) is found further to the right in the linear string. Once the parser has processed the antecedent, a binder has been found for the anaphor and it can be interpreted, i.e., it now has a semantic denotation and it can be structurally integrated within the sentence. As a result, we expect that any effect found in the anaphor conditions should disappear at the embedded subject.[34]

The stimuli used in this experiment were all embedded questions. Assuming successive cyclic movement of *wh*-phrases (see Chomsky, 1986, among many others), the *wh*-phrase must stop at each embedded CP before reaching its final position (i.e., its surface position). Since our stimuli consisted of embedded questions, this means that the *wh*-phrase must stop at an intermediate landing site before moving to its final position. Recall that as a *wh*-phrase moves in the derivation, it leaves behind a copy in each landing position. If the phrase containing the anaphor needs to reconstruct to be bound by its antecedent (Principle A), it should be possible for the phrase containing the anaphor to reconstruct to any of the lower copy positions, provided that this position would enable the anaphor to be structurally bound by its antecedent. For example, in (6), the *wh*-phrase containing the anaphor can either reconstruct to its base-generated position, i.e., following the verb *invent*, or it can reconstruct to the intermediate copy position, as shown in the example. In this position, the *wh*-phrase would be c-commanded by its antecedent, obeying Binding Principle A. However, if the embedded verb is a verb of creation, the phrase containing the anaphor must be interpreted below the creation verb in order to derive the *de dicto* reading. Since reconstruction for binding and reconstruction for the *de dicto* reading are two distinct motivations for interpreting a lower copy, it might be possible that the phrase containing the anaphor is interpreted in the intermediate copy position for binding reasons and then it

[34]Another possibility is that structurally integrating the complement of *wh* containing an anaphor at the embedded subject incurs a processing cost. If the parser reconstructs the phrase containing the anaphor to a position structurally below the antecedent, this operation might be costly and thus we would expect processing difficulty at this position.

is interpreted in the lowest copy position for the *de dicto* reading.

(6) ... how many lies about herself$_i$ you are asking Alexa$_i$ ~~lies about herself$_i$~~ to invent ...

The lowest copy position (base-generated position) is the region directly following the verb. Since we manipulated the type of verb that appeared in the embedded clause (i.e., creation vs. non-creation), potential effects found after the verb has been presented might be attributable to the verb itself and might not reflect any potential effects of the type of DP found in the *wh*-phrase. More precisely, any effects found at the verb site might not reflect a PF-LF mismatch but might simply reflect lexical properties of the different types of verbs used in the experiment. However, the intermediate copy position is found several words before the verb site, before the parser has even processed the verb. This means that this copy position is found before the parser is able to determine the type of verb found in the embedded clause. Thus, any effects found at the intermediate copy position could only be attributable to the type of DP found in the *many-x* phrase and not to the type of verb found further to the right in the string. More precisely, if we were to find that anaphor conditions are more costly than R-expression conditions at the intermediate copy position, this effect could be interpreted as reflecting a PF-LF mismatch.

Region 2: Embedded verb The second region of interest in this experiment is the embedded verb region, which either contained a creation or non-creation verb, as shown in (5-c)-(5-d). While the non-creation verb allows both the *de re* and *de dicto* interpretations of *many-x*, the creation verb forces the parser to adopt a *de dicto* interpretation, i.e., the creation verb is incompatible with a *de re* interpretation. As previously discussed, we assume a general parsing preference to interpret phrases in their surface positions, whenever this is possible. Extending this idea to the current stimuli, the parser should attempt to interpret *many-x* as soon as it is encountered, unless it has been given evidence to suggest that a different parse should be pursued. Thus, there should be a general preference to interpret *how many* questions as *de re*. However, this parse will only be possible if the embedded verb is a non-creation verb. Since our stimuli consisted of embedded questions with *many-x* linearly preceding the verb site, the parser will not know that the *de re* reading is unavailable until the creation verb is reached further to the right in the sentence. If the parser prefers to interpret phrases in their surface positions whenever this is possible, it will attempt to interpret *many-x* as *de re*. Upon finding the creation verb further to the right in the string, it should determine that the *de re* reading is unavailable and it will need to reanalyze its parse and adopt a *de dicto* interpretation via reconstruction. Since a *de dicto* interpretation is not the preferred interpretation, this reanalysis effect should incur a processing cost.

Facilitation for later reconstruction Following Hackl et al. (2012)'s paradigm, we also investigated whether the two motivations for reconstruction interact with one another. Since we manipulated the type of DP found in the *many-x* phrase, we were also able to investigate whether an early signal for reconstruction for anaphoric binding (to interpret the phrase containing the anaphor) facilitated reconstruction for the *de dicto* interpretation of creation verbs. If *many-x* must be interpreted as *de dicto* due to the sentence containing an embedded creation verb, the parser will not know that the phrase will need to reconstruct until it finds the creation verb further to the right. Thus, processing difficulty is expected further downstream when the parser is forced to reanalyze its original parse of the sentence. By manipulating the type of DP found in the *many-x* phrase, the parser is being given an early signal that it must reconstruct the phrase containing the anaphor in order for it to be given a semantic interpretation at LF. Since the parser has already been given a signal further to the left in the string that it must reconstruct *many-x* for binding purposes, upon finding a creation verb further to the right in the string of words, it should be easier for the parser to reanalyze its parse compared to a case where it had not been given an early signal, i.e., when the complement of *wh* contains an R-expression. In other words, even though the *de dicto* interpretation of creation verbs is the less preferred parse of the sentence, it should be easier for the parser to reanalyze its parse in cases where the parser already needed to reconstruct for another reason (in this case, for anaphoric binding).

If finding the anaphor early in the sentence facilitates reconstruction for the *de dicto* interpretation, we expect faster reading times at the verb site for (5-a) compared to (5-c). In (5-c), there is no early signal for reconstruction because the *many-x* phrase contains an R-expression which does not need to reconstruct. While we expect to see effects of anaphoric binding at a copy position in (5-b), we do not expect to see any differences at the verb site between conditions (5-b) and (5-d) because in both of these cases, the verb is a non-creation verb, which does not require reconstruction. In other words, when the embedded verb is a non-creation verb, the parser can adopt the *de re* interpretation and interpret *many-x* in its surface position. Since no reconstruction is required, we do not expect these conditions to incur processing difficulty. Importantly, since the predicted reconstruction effects for anaphoric binding and for the *de dicto* interpretation of creation verbs differ in their syntactic positions (i.e., at a copy position, which could be an intermediate copy position, or at the verb site), we expect any potential effects of these distinct reconstruction operations to also appear in distinct syntactic positions.

2.3 Naturalness judgement task

Since the sentences are very complex, one might expect them to be difficult to process. To ensure that these sentences were not too difficult to process, we ran a naturalness judgement task before running the self-paced reading experiment. We assumed that lower naturalness ratings would indicate a greater processing load.

2.3.1 Specific Predictions: Judgement task

Our stimuli consist of constructions in which anaphors linearly precede their antecedents (cataphoric anaphors). Such constructions are less commonly found compared to sentences in which the anaphor follows its antecedent. Moreover, in these cases, binding needs to occur between the anaphor and the antecedent, i.e., when the anaphor is found early in the sentence, the participant would not yet have a salient referent. Since this additional binding operation is required only for the anaphor conditions (and not the R-expression ones), one might expect that the conditions containing anaphors are more complex than those containing R-expressions. If this complexity measure is reflected in naturalness ratings, we predict that the anaphor conditions should be given lower naturalness ratings than the R-expression conditions.

An alternative explanation is that anaphors as lexical items are less natural and more abstract than R-expressions. We would still predict that the anaphor conditions would be more complex to process and thus, we expect lower naturalness ratings for the anaphor conditions. If we find that the anaphor conditions are rated lower, either of these explanations (structural versus lexical) could explain the data. The online reading experiment would enable us to distinguish between these two explanations.

We also manipulated the type of verb found as the embedded verb in our stimuli. If the parser prefers to interpret phrases as soon as possible, it should prefer the constructions with non-creation verbs and those sentences should be judged as more natural than their creation verb counterparts. However, since participants are being presented with the whole sentence as a unit in the naturalness study, it is also possible that the parser will not be sensitive to the verb effects. Participants only need to make a judgement about the sentence once they have read and processed the whole sentence. Effects of the creation verb might only be detectable in the reading study when participants are presented with the string in smaller chunks (in this case, word-by-word).

2.3.2 Materials

Thirty-two sentence templates were created following the paradigm shown in (5). Target sentences were counterbalanced across four lists of stimuli combined with 40 sentences from an unrelated experiment and 36 fillers (16 of the fillers were created to resemble the target sentences, and 20 fillers resembled the unrelated experiment's items). In total, participants read and judged 108 stimulus sentences. Yes-no comprehension questions were asked after each sentence to ensure that participants were reading and processing the sentences. For the critical test items, 50% of the answers to the questions were "no".

2.3.2.1 Creation versus non-creation verb descriptive statistics

The creation verbs used in Experiment 1 (as well as Experiment 2) had a mean length of 5.86 characters while the non-creation verbs had a mean length of 6.59 characters. Descriptively, the creation verbs contained less characters than the non-creation verbs. Some of this variation is accounted for by our residual reading times (explained below). Frequencies for the creation and non-creation verbs were collected from the NOW corpus

(News on the Web, Davies, 2013). This corpus contains 5.7 billion words from newspapers and magazines published on the web from 2010 until now. The corpus is updated daily. The creation verbs had a mean log frequency of 11.18 (579 015 untransformed) while the non-creation verbs had a mean log frequency of 9.83 (133 484). As result, the creation verbs were more frequent than the non-creation verbs.

2.3.3 Participants

Fifty nine participants completed the acceptability judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). Thirteen participants were removed from the analysis because they scored less than 75% on the comprehension questions throughout the whole experiment. Three additional participants were removed from the analysis because they scored less than 75% on the target sentences. Thus, in the results section, the data for the remaining 43 participants (23 males, 20 females) will be presented. All participants were self-reported native speakers of English, aged between 19 and 60 years of age (M = 33.7 years, SD = 9.74). Participants were paid $3.00 USD for their participation and were naïve as to the purposes of the experiment.

2.3.4 Procedure

Participants were asked to rate the naturalness of a series of sentences presented to them on a screen on a scale of 1 (extremely unnatural) to 5 (perfectly natural). Judgement choices were made by selecting the appropriate button below each sentence on their computer screen. They were instructed to work quickly and not spend too much time thinking about their judgements. Once participants had selected their judgement, they were presented with a new screen with a yes-no comprehension question. Participants were asked to indicate their answer by selecting the appropriate button on the screen. They were given feedback if they answered the question incorrectly: "Oops! That was the wrong answer!". Moreover, as the sentences being presented to participants were particularly difficult, we told them only consider the information suggested in the sentences to answer the question. To ensure that they read the sentences for comprehension, we also told participants that they needed to score at least 75% in order for us to be able to use the data. Item presentation was pseudo-randomized by participant such that each test item was preceded by a filler item. In addition, as the participants' screen sizes could not be controlled for the online experiment, the test sentences were always presented with a line break after the head noun of the picture DP, e.g., *pictures, lies, photos*, as shown in (7). Similar line breaks appeared in the test sentences. Participants provided informed consent prior to beginning the experiment.

(7) The reporters wondered how many lies
about herself/Sean you are asking Alexa to deny for tomorrow's live television interview.

2.3.5 Results

2.3.5.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis.[35] In total, 8 fillers and 1 target sentence was removed from the analysis. After eliminating the low scoring item, accuracy on the comprehension questions for the target sentences was high at 87.3%. However, in order to accurately investigate naturalness judgements on

[35]We used this low cut off to fully eliminate trials that were not interpreted as intended by participants. If the average score for an item was as low as 60%, we deemed this item to be unreliable. Items that individual participants might not have understood as intended were removed from the analysis through other elimination methods, as explained in this sub-sub-section.

trials where participants were correctly interpreting the sentence as intended, trials in which the participant answered the comprehension question incorrectly were removed from the analysis.[36]

2.3.5.2 Statistical analyses

We used generalized linear mixed effects regression models with participants and items as random effects (Baayen et al., 2008; Baayen, 2008; Pinheiro and Bates, 2000) as implemented in the lme4 package (version 1.1-12, Bates and Sarkar, 2007; Bates et al., 2015b) for R (version 3.2.3, R Core Development Team, 2015), whenever they significantly improved the model, as determined by the likelihood ratio test. This statistical method allows multiple factors to be explored at once and it also allows for any variance between participants or items to be accounted for. Models were initially fitted with a maximal random effects structure (Barr et al., 2013), specified to include all fixed effects as random slopes for participants and items, whenever the models converged and the added random effect structure significantly improved the model's goodness of fit, as determined by likelihood ratio tests (see Bates et al., 2015a). In order to reduce skewness, we also removed outliers from the datasets: any data points that were +/- 2.5 SDs from the residual error of the model were removed.[37] Once we removed these data points, the models were refitted. We report here the fixed and random effects that remained in the final models after trimming.

The dependent variable for this experiment was the naturalness judgement from 1-5.

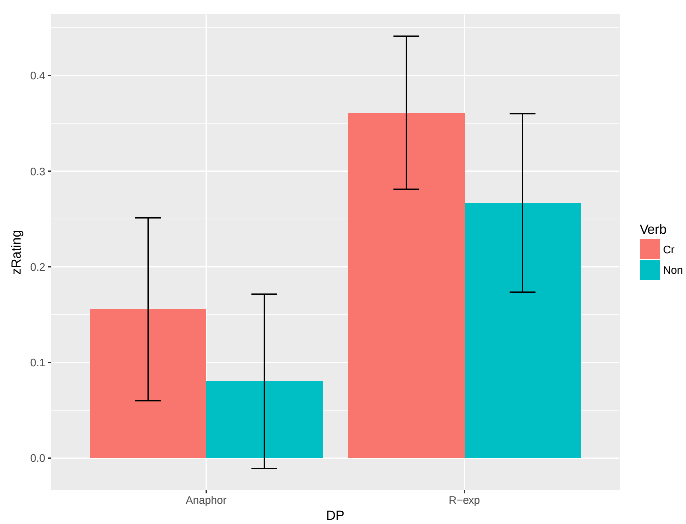
[36]Note that the results do not change if the incorrectly answered items are included in the analyses. We removed incorrectly answered trials to investigate the most conservative data set and we report those results here.

[37]Trimming the models in this way removes any extreme values that might be influencing the results (either towards significance or making a significant effect not come out that way). In our results tables, we report the number of data points that were included before and after model criticism. For more information on this type of model criticism, see Baayen and Milin, 2010, p. 26.

To account for the variation in how participants interpret the scale, the naturalness data was *z*-transformed by participant (Schütze and Sprouse, 2014). All statistical analyses were calculated on the *z*-scored data.

Mean *z*-scored naturalness judgements are plotted in Figure 2.1. The main question of interest was whether we would find a difference in naturalness judgements based on the type of DP found in the *wh*-phrase (anaphor vs. R-expression) or based on the type of verb found as the embedded verb (creation vs. non-creation). Descriptively, the figure demonstrates that R-expression conditions (right panel) were judge as more natural than anaphor conditions (left panel). Creation verbs (pink bars) seemed to also be judged as more natural than non-creation verbs (blue bars), both when *many-x* contained an anaphor or an R-expression.

Figure 2.1: Mean naturalness judgements (in *z*-scores) for DP type (anaphor, R-expression) and verb type (creation, non-creation) in Experiment 1. Error bars represent 95% confidence intervals.



We fitted a linear mixed effects regression model to the *z*-scored data using DP type and Verb type as fixed effects and the random effects described earlier (see Table 2.1).

We found a main effect of DP type such that conditions with R-expressions found in the *many-x* phrase were more likely to be rated higher than conditions with anaphors in the *many-x* phrase ($p < 0.001$). However, there was no statistically reliable effect of verb type. While non-creation verbs were rated numerically lower than creation verbs, this effect was not significant at the 5% threshold ($p = 0.067$). Preliminarily, these results suggest that participants found the conditions with anaphors linearly preceding their antecedents to be less natural than the conditions with R-expressions. This result could be attributed to the fact that the anaphors always linearly preceded their antecedents which are less commonly found than sentences where the anaphor follows its antecedent. Following the assumption that lower ratings reflect greater processing difficulty, the results from the naturalness judgement task suggest that the anaphor conditions were more difficult to process than the R-expression conditions. Based solely on the naturalness judgement results, either a structural explanation, i.e., the anaphor must be locally bound by its antecedent and this occurs via reconstruction at LF, or a lexical explanation, i.e., anaphors are lexically more abstract than R-expressions making them harder to process, would be compatible with our data. Online reading times from a self-paced reading experiment might allow us to differentiate between these two possible explanations of the judgement results.

Table 2.1: Final mixed-effects model for naturalness judgement task (N = 1164 before trimming, 1133 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	0.212	0.062	3.416	0.001
R-expression	0.160	0.037	4.352	0.000
Non-creation	-0.067	0.037	-1.835	0.067

Random effects: by-participant intercept (SD = 0.343), and by-item intercept (SD = 0.053).

2.4 Self-paced reading experiment

To be able to compare our materials across the judgement and reading studies, no items from the naturalness judgement task were removed for the self-paced reading version of Experiment 1. We used the moving window word-by-word self-paced reading methodology (Just et al., 1982). In this methodology, sentences are presented one word at a time and participants are asked to press a button (or key) or reveal each subsequent word in the sentence. Before the words are presented, sentences appear as a series of dashes revealing the number of words in each sentence. As each new word is presented, the previous word in the sentence disappears. This methodology has been previously shown to be sensitive to effects of syntactic movement (see e.g., Stowe, 1986, among many others) and it has also been used to test for interpretation effects in real-time (see e.g., Anderson, 2004; Hackl et al., 2012). The critical dependent variable in this methodology is reading times on each word. The assumption is that slowdowns on particular regions of the sentence, measured by longer reading times compared to a baseline condition, reflect greater processing difficulty at that region.

2.4.1 Participants

One hundred and nine participants were tested for the self-paced reading experiment. They were all recruited from Amazon Mechanical Turk and completed the experiment using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm). Seventeen participants were removed from the analysis because they scored less than 75% on the comprehension questions throughout the whole experiment. Six additional participants were removed from the analysis because they scored less than 70% on the target sentences. Two additional participants were removed from the analysis because they lost more than 20% of data when we trimmed the reading times (explained below). Thus, in the results section, the data for the remaining 84 participants (33 males, 51 females) will be presented. All participants were self-reported native speakers of English, aged between 20 and 70 years of age (*M* = 36.48 years, *SD* = 11.87). Participants were paid $6.00 USD for their participation and were naïve as to the purposes of the experiment. Participants for the self-paced reading version of the experiment did not complete the naturalness judgement task.

2.4.2 Stimuli

The same stimuli and filler items were used for the self-paced reading version of the experiment on Amazon Mechanical Turk, except that 40% of the comprehension questions were removed. These questions were removed to decrease the length of the experiment as we anticipated this version taking longer due to the fact that participants needed to read words one at a time, and we did not want the experiment to take more than one hour to complete.

2.4.3 Procedure

Participants were instructed that they would be reading sentences presented one word at a time. In order to see the next word of the sentence, they were instructed to press the spacebar. The sentences appeared on the screen as a series of dashes, revealing the number of words in the sentence. Once the participant selected the spacebar to move onto the next word in the sentence, the word that they were reading disappeared. Participants were told that after some questions, they would be asked a yes-no comprehension question. As in the judgement task, they were told that they needed to score 75% on the questions and were given feedback when they answered questions incorrectly. Participants who completed the self-paced reading version of the experiment did not complete the judgement task. Item presentation followed the same procedure as in the naturalness judgement study.

2.4.4 Results

2.4.4.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 4 fillers were removed from the analysis. We also removed from the analysis any reading times that were less than 90 ms and over 3000 ms. Accuracy on the target sentences was high at 83.6%. When possible, in order to accurately investigate the reading times on trials where participants were interpreting the sentence as intended, trials where the participant answered the question incorrectly were also removed from the analysis.[38]

The dependent variable in this experiment was continuous (reading times). As reading times allow for a great deal of variation, we transformed the data in four ways: raw reading times, residual reading times, log reading times, and residuals of the log reading times. These measures allow us to reduce skewness and control for the number of characters in each word across trials. For all transformations, data points that were +/- 2 SDs away from the mean by participant, condition, and word number in the sentence were removed (approximately 5% of possible data were trimmed across all transformations). The statistical models were run on all of these data transformations, and revealed similar effects, unless otherwise indicated.

[38]The statistical analyses were also run on a data set that included all trials which revealed the same effects. Consequently, we report here the more conservative data set.

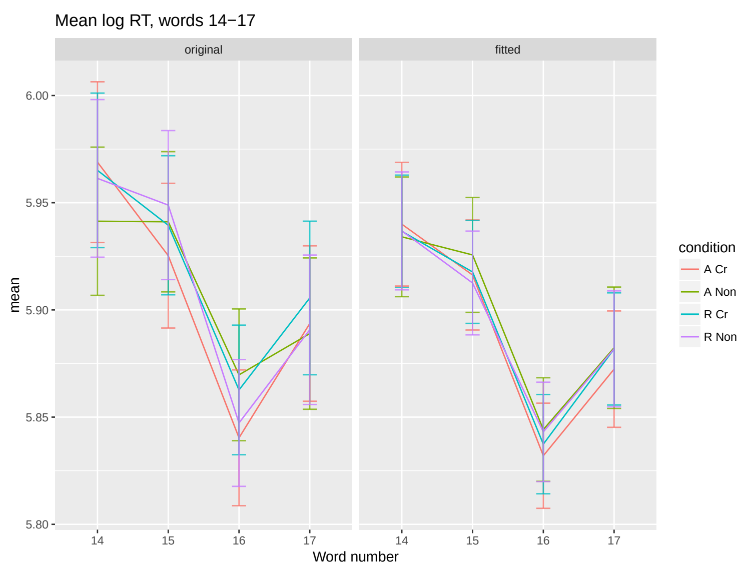
We fitted a linear mixed effects regression model to the data with DP type (anaphor, R-expression) and verb type (creation, non-creation) as fixed effects and the random effects described earlier. We also added DP type and Verb type as a random slope for participant whenever the model converged.

2.4.4.2 Verb site

In all of the test sentences, the critical verb appeared at word 14. We fitted linear mixed effects models for the reading time data on the verb site as well as for three words following the verb site, due to potential spillover effects. At word 14, when the verb site is introduced, the non-creation verb was read at a numerically faster rate than the creation verb. On words 15-16, one and two words following the verb site, the non-creation verb was read numerically longer than the creation verbs across all transformations. At word 17, the non-creation verb is again read at a numerically faster rate than its creation verb counterpart. Since none of these effects were significant at the 5% threshold, we do not deem them to be reliable. In the same region, there were no significant effects of DP type and there were also no significant interactions. Mean log reading times for this region are plotted in Figure 2.2. The left panel shows the original mean reading times, before they have been fitted to the linear mixed effects model. The right panel shows the mean reading times after the linear model was fitted. The fitted values have been adjusted for variance between participants and items.

Overall, the reading time results suggest that our hypothesis that finding an anaphor earlier in the sentence (i.e., in the *many-x* phrase) might facilitate reconstruction for the *de dicto* interpretation of creation verbs was not borne out. Notice, however, that there was no main effect of verb type at this region, suggesting that the creation verb did not incur a greater processing load compared to the non-creation verbs. We will discuss possible explanations for these results in the Discussion section.

Figure 2.2: Mean log reading times at the Verb site by DP (anaphor, R-expression) and Verb (creation, non-creation) type for words 14-17 in Experiment 1. Original log mean values appear in the left panel and fitted log mean values appear in the right panel. Error bars represent 95% confidence intervals.

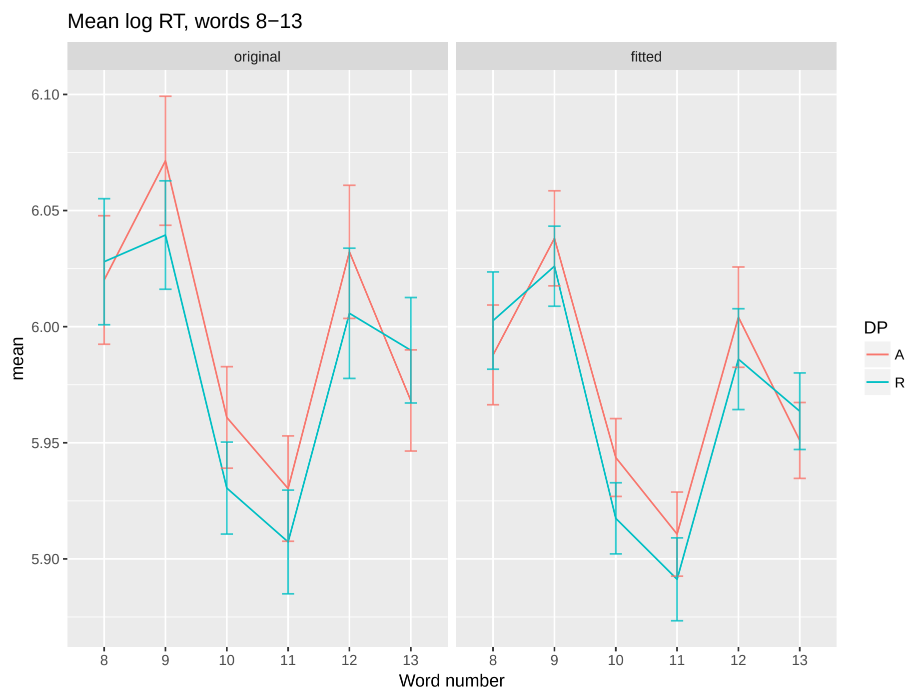


2.4.4.3 DP site

We also fitted models to investigate the potential effect of the type of DP found in the *many-x* phrase, adding DP as a random slope by participant whenever the model converged. The anaphor or R-expression was introduced at word 8 and the antecedent occurred at word 12. We fitted models on words 8-13. The mean log reading times are plotted in Figure 2.3.

We found no reliable effects on word 8, i.e., when the anaphor/R-expression was introduced. At word 9, one word following the anaphor/R-expression, there was a significant effect of DP type in the raw (β = -16.9, SE = 7.42, t = -2.279, p = 0.03, model

Figure 2.3: Mean log reading times at DP site from words 8-13 in Experiment 1. Original log mean values in the left panel and fitted log mean values in the right panel. Error bars represent 95% confidence intervals.



not shown) and residual reading times (β = -14.17, SE = 6.81, t = -2.081, p = 0.04, model not shown), such that R-expression conditions were read faster than anaphor conditions. At word 10, two words following the anaphor/R-expression, we found significant effects across all data transformations, showing that the R-expression conditions were read faster than the anaphor conditions ($p \leq 0.05$), as shown in Table 2.2 for the log reading times. At word 11, three words following the anaphor/R-expression, we find the same effect in the raw (β = -7.54, SE = 3.33, t = -2.264, p = 0.02, model not shown) and residual reading times (β = -7.78, SE = 3.61, t = -2.157, p = 0.03, model not shown). This effect reached marginal significance in the log reading times, as shown in Table 2.3.[39] This effect was also

[39]Note that this effect reaches significance at the 5% threshold prior to model trimming.

found on word 12, when the embedded subject is introduced, in the raw reading times (β = -17.62, SE = 7.16, t = -2.46, p = 0.02, model not shown) and residual reading times (β = -17.71, SE = 7.19, t = -2.46, p = 0.02, model not shown). Crucially, the effect disappeared across all transformations at word 13, one word following the embedded subject.

Table 2.2: Final mixed-effects model for log reading times at word 10 (N = 2336 before trimming, 2281 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.943	0.033	182.94	0.000
R-expression	-0.025	0.009	-2.85	0.006

Random effects: by-participant intercept (SD = 0.293), by-item intercept (SD = 0.005), by-participant random slope for DP type (SD = 0.0025), and the correlation between by-participant and slope (r = -0.68).

Table 2.3: Final mixed-effects model for log reading times at word 11 (N = 2347 before trimming, 2286 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.908	0.035	168.87	0.000
R-expression	-0.016	0.009	-1.886	0.06

Random effects: by-participant intercept (SD = 0.316) and by-item intercept (SD = 0.01).

2.5 Discussion and chapter conclusions

2.5.1 Summary of this chapter

In Experiment 1, we investigated whether we could find real-time evidence for the parser's sensitivity to a PF-LF mismatch using syntactic reconstruction. We examined two different motivations for reconstruction: i) reconstruction for anaphoric binding and ii) reconstruction for the *de dicto* interpretation of creation verbs. We predicted that potential effects of these two reasons for reconstruction would appear in distinct syntactic positions in the stimulus items. More precisely, if the parser reconstructs *many-x* containing an anaphor for Binding Principle A, we predicted that this effect would appear on the embedded subject because this syntactic position corresponds to an intermediate copy site. If the parser reconstructs the phrase containing the anaphor to this position, the anaphor would be structurally bound by its antecedent at LF, enabling the phrase to be assigned a semantic denotation and be integrated into the structure. Interpreting the complement of *wh* at this position would also satisfy Binding Principle A. Assuming that the parser creates a partial denotation of the DP containing the anaphor when it encounters the phrase at the beginning of the sentence and that it must keep this partial denotation in memory until it finds the antecedent, our DP effects are compatible with a reconstruction analysis of the results. We found that conditions containing anaphors were read longer on the words following the anaphor, compared to those same words following R-expressions. This result suggests that the parser is unable to interpret the phrase containing the anaphor and that processing difficulty is incurred until the antecedent is found. Upon reaching the antecedent, the DP is interpretable (assigned a denotation and structurally integrated), and processing difficulty is ameliorated. These effects provide preliminary evidence that the parser is sensitive to a mismatch between the surface string and the LF structure that it needs to build to interpret the sentence. These results are compatible with a reconstruction analysis of anaphoric binding at LF.

2.5.2 Alternative explanation of the results

We have been pursuing a strictly structural explanation of the DP results, which is compatible with reconstruction to an intermediate copy position. An alternative explanation of these results is that the parser looks for an antecedent for the anaphor but that it is not concerned with the structural position of the antecedent, i.e., it is sufficient that the antecedent match the anaphor in relevant phi-features (gender, number). More precisely, under the alternative hypothesis, it is still the case that anaphors come with a licensing condition that requires that the anaphor have an antecedent that matches the anaphor in phi-features and is an accessible subject. By "accessible subject," we follow standard Binding Principle A. However, this alternative view does not require reconstruction to satisfy Binding Principle A. Instead it is sufficient that such an antecedent (phi-feature matching and in an "accessible" position) but it does not depend on reconstruction to satisfy binding. This alternative explanation is indeed compatible with our results and could be viewed as a shortcut that the parser takes in processing to satisfy Binding Principle A or it could be viewed as a new definition of Binding Theory. Either way, it would be distinct from traditional Binding Theory which requires c-command.[40] In Experiment 1, there was only ever one possible antecedent for the anaphor. This antecedent also occurred in an appropriate structural configuration for binding to occur, i.e., the antecedent always occurred in a structural position in which it would c-command the anaphor at LF under a reconstruction analysis. We call such an explanation of the results *non-structural*, compared to the strictly structural explanation discussed so far. The results from Experiment 1 alone are not enough to rule out an alternative non-structural explanation of the results following which the parser is simply looking for a feature-matched antecedent.

[40]Thanks to Martin Hackl for his help in clarifying this alternative explanation of the results.

We will address this concern in Chapter 3, when we compared the processing of cataphoric pronouns to R-expressions.

Note that the effects found in Experiment 1 cannot simply be explained by a processing account of *wh*-dependencies. Since all the stimulus sentences were embedded questions, any effect of dependency formation, i.e., associating the filler phrase with its gap site or associating the head of the chain with the tail of the chain and completing the complex syntactic object, would be consistent across all stimulus items. This effect would not depend on the type of DP found in the filler phrase because in both cases, the DP is embedded in a *wh*-filler phrase which would need to find its gap site. Furthermore, as previously mentioned, the antecedent site is found in a distinct position from the integration site of the filler phrase. We argue that our results reflect interpretation effects and specifically, the cost associated with determining a semantic interpretation for an anaphor that is uninterpretable in its surface position. At the very least, the results suggest that the parser is sensitive to the fact that anaphors need antecedents in order to be interpreted.

2.5.3 More on creation verbs

We had also predicted a reconstruction effect to the lowest copy position (tail of the chain) if the parser reconstructs for the *de dicto* interpretation of creation verbs. If the parser is sensitive to the fact that *many-x* cannot be interpreted as *de re* with creation verbs, we expected to find a reanalysis effect at the creation verb. The parser does not know that the embedded verb is a creation verb until reaching the verb site further to the right in the sentence. Thus, we predicted that conditions with creation verbs would be more difficult to process, as reflected by longer reading times, compared to conditions with non-creation verbs. This prediction was not borne out: we found no reliable differences between the verb types in this experiment. Consequently, we remain agnostic about parsing effects at the lowest copy position.

One possible explanation for the lack of an effect of the verb type is that participants were not sensitive to the creation versus non-creation verb distinction, i.e., they were not interpreting them differently. However, this possibility was ruled out in a follow-up experiment (reported in Chapter 3) in which we specifically asked participants about the creation event in the comprehension questions. More precisely, we asked participants about whether the object of the creation verb already existed before the event had taken place. If participants were interpreting these verbs as intended, i.e., verbs of creation, they should answer *no* to this type of question since the object of the verb of creation cannot exist until after the creation event has taken place. In contrast, with non-creation verbs, it is perfectly acceptable (and sometimes necessary) that the object of the verb already exists before the event has taken place. In Experiment 2, we found that participants only presupposed the existence of the object of the verb of creation 20% of the time, compared to 90% of the time with non-creation verbs. These results suggest that participants were interpreting the verbs as intended and they were sensitive to the distinction between a creation and non-creation verb. Moreover, we also observed a marginal effect of verb type (p = 0.07) in the naturalness judgement task reported in this chapter. This effect would be rather surprising if participants were not making a distinction between the two verb types. Another possibility for the absence of an effect of verb type is that the self-paced reading methodology is not sensitive enough to capture any potential distinctions at the verb site. A final possibility is that participants were also interpreting the non-creation verbs as *de dicto*. Recall that this interpretation is not ruled out with the non-creation verbs because they are ambiguous between *de re* and *de dicto*. In this experiment, we assumed that the *de* *re* interpretation is preferred because under this interpretation, *many-x* can be interpreted in its surface position, provided it does not contain an anaphor. Other than this general parsing preference, there is nothing ruling out a *de dicto* interpretation of non-creation verbs. If *de dicto* interpretations were being pursued with both verb types, we would not expect to observe processing differences based on the type of verb, irrespective of whether or not the parser needed to reconstruct *many-x*. Unfortunately, the results of the present experiment are not able to address these questions directly.

It is important to note that under standard copy theory of movement (Chomsky, 1993), both the higher (surface) and lower (base-generated) copies are interpreted. We have assumed that there is a preference for *de re* interpretations because interpreting the higher copy enables the parser to interpret the copy as soon as it is encountered. If we were to translate this assumption into copy theory, we would predict that there should be a difference in structural complexity in the two copy positions. More precisely, the higher copy should somehow be more complex and there should be a preference to interpret as much as possible in this higher copy position. This would mean that the two copies are not completely identical and the lower copy would be simpler (for example, the lower copy might not contain the DP restrictor).[41]

In contrast, *de re* and *de dicto* readings are not canonically predicted to differ based on structure (see e.g., Fox and Nissenbaum, 2004). Under both interpretations, the lower copy contains the complement of *wh* (in our cases) or a definite description. When the string is to be interpreted as *de dicto*, the world variable on the DP is bound by the scope bearing operator. When the string is to be interpreted as *de re*, the world variable is not bound, see the LF structures in (8).

[41]Thanks to Martin Hackl for bringing up this issue and for his suggestion that the lower copy would need to be less complex in order to derive a preference for *de re* over *de dicto* in parsing.

(8) a. *De re*: [wh many books]$_1$ [Chris wants [t_1] to buy [t_1]]]

b. *De dicto*: [wh]$_1$ [Chris wants $_{w3}$ [[n_1 many books $_{w3}$]$_2$ to buy [t_2]]]

Therefore, under standard assumptions, there is no structural distinction between the two readings. Since we did not observe an effect of verb type in Experiment 1, we do not have evidence for a structural difference between the two readings. Consequently, we also do not have evidence for a preference for *de re* over *de dicto*, or vice versa. This could mean that the two readings do not differ based on structure but we cannot commit to such an interpretation based on our results. We leave this issue aside for the time being.

2.5.4 Concluding remarks

To conclude, we argue that the results from this experiment provide evidence in support of a reconstruction analysis for anaphoric binding at LF. When the anaphor is not bound in its surface position, and therefore cannot be interpreted, the parser must find an appropriate antecedent to bind the anaphor further to the right in the linear string. One mechanism the parser can utilize is reconstruction following which it interprets the phrase containing the anaphor in a lower copy position (in this case, at the embedded subject). Reconstruction of the phrase containing the anaphor incurs a processing cost, as reflected by longer reading times on the words following anaphors compared to those same words following R-expressions, which do not need to find within-sentence antecedents.

Our results, however, do not shed any light on the LF structure associated with creation verbs. We have suggested several possibilities for why this could be the case. One might wonder if the results from this experiment could be interpreted as demonstrating that reconstruction for binding is different from reconstruction for scope. Since we were not able to find any effects of creation versus creation verbs, it is unclear if these results indicate that the two types of reconstruction reflect different operations altogether which may or may not be detectable using the self-paced reading methodology or whether the manipulations used in the current experiment were not sensitive enough to target the effect directly. We discuss this issue further when we discuss some open questions in Chapter 6.

From a methodological standpoint, the paradigm used in this experiment, inspired by a paradigm developed by Hackl et al. (2012), is the first of its kind. In this study, the linear order of the sentence played a crucial role and this paradigm allowed us to investigate how the parser copes with a surface structure that is uninterpretable. The results suggest that the parser is able to accurately compute grammatical principles that hold of the interpretation level of the structure, providing new evidence for the parser's sensitivity to a PF-LF mismatch in real-time processing. In the rest of the thesis, we attempt to further refine this paradigm and determine the precise types of LF effects are active in online parsing. As the anaphoric effects were the most reliable in this study, we focus on investigating these effects in more detail in the following chapters. Our hope is that this work will help us to better understand the effect of LF structure on real-time processing and specifically, on the processing of structure-dependent DPs.

Chapter 3

Does the parser also reconstruct for Binding Principle B?

3.1 Introduction

3.1.1 Summary of Experiment 1 results

In Experiment 1, we asked whether the parser was sensitive to PF-LF mismatches in sentence processing. To directly investigate this question, we used embedded *wh*-questions and manipulated whether the *wh*-phrase contained an anaphor or an R-expression. The complement of *wh* containing an R-expression could be assigned a semantic denotation, and thus interpreted, as soon as it was encountered, i.e., in the higher copy position. However, if the complement of *wh* contained an anaphor, the phrase could not be interpreted in the higher copy position since the assignment function of the anaphor is dependent on its antecedent. Crucially, there was no antecedent for the anaphor in the preceding linear string. As a result, the parser could only interpret the phrase containing the anaphor in a lower copy position that was structurally below the antecedent for the anaphor. In other words, the parser needed to reconstruct the phrase containing the anaphor to the lower position in order to derive an interpretation for that phrase. If the complement of *wh* containing the anaphor was interpreted in a lower copy position, it could be assigned a semantic denotation and be structurally integrated into the structure. Interpretation of the copy in this lower position also satisfied Binding Principle A.

3.1.1.1 DP type results

In Experiment 1, we found significant effects based on the type of DP (anaphor vs. R-expression) found in the complement of *wh* phrase such that words following anaphors were read longer than words following R-expressions. Crucially, there was no reliable difference between conditions containing anaphors and R-expressions once the embedded subject, i.e., the antecedent for the anaphor, was found further to the right in the linear string. We interpreted this result as demonstrating that the parser is sensitive to the fact that the phrase containing the anaphor cannot be interpreted in its surface position and that it must be interpreted in a lower copy position at LF. Since the parser must locate an antecedent for the anaphor in order to interpret the phrase, we observed a processing cost (longer reading times) until the antecedent was found further to the right in the linear string. Assuming Binding Principle A, we suggested that these results are compatible with a reconstruction analysis for anaphoric binding at LF. The parser must locate an antecedent to the right but this antecedent must also be found in an appropriate structural position for the parser to be able to structurally integrate the phrase.

3.1.1.2 Verb type results

Experiment 1 also sought to investigate whether we could find real-time evidence of reconstruction for *de dicto* interpretations of creation verbs at LF (Heycock, 1995; Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004). We followed previous work in assuming that in *how-many* questions with creation verbs, the *many-x* DP (complement of *wh*) can only be interpreted as *de dicto* because a *de re* interpretation would be incompatible with the semantics of verbs of creation. For *many-x* to be interpreted as *de dicto*, *many-x* must reconstruct to a syntactic position that is structurally lower than the creation verb at LF. In other words, it would need to be interpreted in the lowest copy position. However, our results did not shed light on this question as we did not find a significant effect of verb type (creation vs. non-creation) in our reading time measures. Possible explanations for why this might be the case were discussed at the end of Chapter 2.

It is important to note that the region following the verb site is also the lowest copy position (base-generated position) for the complement of *wh*. Crucially, the significant effect of DP type was found in the intermediate copy position (i.e., before the embedded subject or antecedent for the anaphor) and not in the lowest copy position, where we would have predicted to observe an effect of verb type. If our Experiment 1 results can be interpreted as evidence for reconstruction for binding, the fact that the effect was found before the base-generated position provides preliminary evidence in favour of an intermediate copy position and that the parser is able to reconstruct to this position at LF. More precisely, these results are consistent with a view following which the parser syntactically integrates the phrase containing the anaphor in the intermediate copy position, i.e., right after the subject. The parser is able to integrate the phrase since the anaphor can now be assigned a semantic denotation.[42] Furthermore, since the base-generated position also corresponds to the syntactic integration position for the *wh*-filler, i.e., where it is assigned its thematic role, and since we did not observe any effects in the embedded object position, our results found in the intermediate copy position cannot be attributed to effects of syntactic integration of the filler for thematic reasons at the base-generated position. Since we did not observe processing differences at the base-generated position, we remain agnostic about parsing at this region of the sentence.

In general, we interpret our Experiment 1 results as demonstrating that the parser is sensitive to PF-LF mismatches in processing. More precisely, the parser is sensitive to constructions in which a phrase cannot be semantically interpreted in its surface position and must be interpreted in a distinct position at LF. Thus far, we have assumed that the parser is sensitive to structural constraints on antecedent resolution for pronouns, i.e. that it will only co-index an anaphor with an antecedent that is in an appropriate structural position to bind the anaphor at LF (Principle A).

3.1.2 Goals of Experiment 2

In Experiment 2, we aim to further investigate the questions asked in Experiment 1 by using similar conditions but with pronouns embedded in the complement of *wh* instead of anaphors. Similarly to Experiment 1, pronouns linearly preceded potential antecedents, which were found in the embedded subject position. Unlike anaphors, pronouns cannot be locally bound by their antecedents (Binding Principle B). Nevertheless, previous

[42]One might wonder why we do not observe a processing cost after the embedded subject if the parser is only able to syntactically integrate the phrase containing the anaphor upon finding the embedded subject. We do not have an explanation for this. One possibility is that the parser creates a partial denotation for the DP in its surface position and then keeps this denotation in memory until it finds the antecedent and it is able to complete the denotation. Keeping the phrase in memory might be costly and this cost is diminished when the antecedent is found further to the right in the string.

experimental work investigating how the parser establishes a co-referent for a pronoun has found that the parser prefers to find a co-referent for the pronoun within the sentence (see e.g., Van Gompel and Liversedge, 2003). As we briefly discussed in Chapter 1, there are two possibilities for how the parser might process pronouns in positions in which they linearly precede potential antecedents, which we elaborate on below. Upon finding a pronoun in the complement of *wh*, the parser can either interpret the phrase in its surface position or it can wait until it has received further input and interpret the phrase containing the pronoun further to the right.

3.1.2.1 Option 1: Pronoun is referential

Upon finding a pronoun in the complement of *wh*, the parser has the option of interpreting the phrase in its surface position. If the pronoun is interpreted as referential, meaning that it gets its interpretation from the context, the complement of *wh* can be assigned a semantic denotation as soon as it is encountered. Thus, if the phrase is interpreted in its surface position (higher copy position), the pronoun must refer to an antecedent outside of the sentence. This analysis would abide by the parser's preference to interpret phrases as soon as they are encountered but it would violate the parser's preference to adopt structural relations between two DPs (binding).

3.1.2.2 Option 2: Pronoun is a bound variable

In contrast, following previous work on referential processing, if the parser prefers to find antecedents for pronouns within the sentence, it might be possible for the pronoun to be interpreted as a bound variable. This would mean that the parser would need to find an antecedent for the pronoun further to the right (lower) in the structure, similarly to cases involving anaphors. However, even if there is a preference for finding antecedents for pronominals within the experimental sentence, there are two possibilities for how the parser might go about finding an antecedent for a pronoun.

The first possibility is that it treats pronouns similarly to anaphors since they would both be interpreted as bound variables. If this is the case, we would expect to find the same effects in Experiment 2 as we did in Experiment 1. Note that this result would suggest that the parser is not sensitive to the structural requirements on different types of DPs because it treats anaphors and pronouns similarly.

The second possibility is that the parser is sensitive to the fact that anaphors and pronouns are restricted by different structural constraints (binding principles). If so, it should treat pronouns and anaphors differently in processing. Consequently, we should find different effects in Experiment 2 compared to those we found in Experiment 1. The syntactic environment from Experiment 1 should incur a Binding Principle B violation if the complement of *wh* contains a pronoun because the pronoun would be locally bound by its antecedent. It is also possible that the parser waits until it has actually found an antecedent for the pronoun further to the right before it establishes a co-referential relation between the two DPs. In this case, we would predict to observe a processing cost once the antecedent is found. This result would be different than what we found with anaphors because in Experiment 1, we found evidence of the parser looking for an antecedent and not waiting until it has found an antecedent to establish a co-referential relation. Note that a result showing that the parser has waited until it has found an antecedent to establish a co-referential relation between that antecedent and the pronoun would provide evidence for a structural account of referential processing since the parser does not treat anaphors and pronouns in a parallel way, even though both can find antecedents within the sentence. We will discuss more precise predictions after we explain the relevant background literature and the experimental details.

3.2 Background

Previous psycholinguistic studies have shown that the parser is sensitive to the restrictions placed on co-referential DPs. This work demonstrates that the parser will not create a dependency between two co-referential DPs if doing so would incur a binding violation (Cowart and Cairns, 1987; Nicol and Swinney, 1989; Clifton et al., 1997; Kazanina et al., 2007, but see Badecker and Straub, 2002; Sturt, 2003 for results showing that inaccessible antecedents can still affect how the co-referent is processed). Previous work investigating the processing of co-referential DPs has consistently investigated constructions in which binding principles would be violated on the surface, i.e., the structure that is pronounced (roughly, the PF representation of the sentence). Binding Theory has been argued to be an LF constraint (Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004). In other words, binding conditions do not need to be satisfied in the surface representation but they must be satisfied at LF, the structure that is interpreted. To our knowledge, no previous studies have investigated how the parser processes constructions in which the parser must derive an appropriate LF structure for a given string of words and how constraints on binding may or may not restrict the LF structure that it can build.[43]

[43] Conroy (2008) investigated how the parser derives an LF structure for a given string of words but examined this question using quantifier scope ambiguity. Thus, it was possible to build an LF structure from the string of words that the parser receives as input (PF) but this LF only corresponded to one of the possible meanings of that string of the words. More precisely, one of the meanings could be derived through a PF-LF match but the other could only be derived by assuming a PF-LF mismatch. In the current thesis, we investigate cases in which it is not possible to derive an interpretable LF structure from the linear string of words. Thus, our cases can be viewed as PF-LF mismatches. See also Hackl et al. (2012), as discussed in Chapter 2.

3.2.1 How are pronouns interpreted?

In Experiment 1, reconstruction of the complement of *wh* to a lower copy position was obligatory in order to assign the phrase containing an anaphor a semantic denotation and satisfy Binding Principle A. In Experiment 2, we investigate cases where reconstruction of the complement of *wh* could lead to a binding violation. Consider the examples in (1). In (1-a), if the complement of *wh* reconstructs to a structural position below the antecedent, this should incur a Principle B violation since the pronoun would be locally bound by the embedded subject at LF. Similarly, in (1-b), if the complement of *wh* reconstructs below the embedded subject, Binding Principle C will be violated because the R-expression would be bound by the embedded subject.[44]

(1) a. *How many pictures of him$_i$ did you ask John$_i$ to take?

b. *How many pictures of John$_i$ did you ask him$_i$ to take?

Importantly, since the complement of *wh* does not contain an anaphor in these examples, it should be possible for the phrase to be interpreted in its surface position. If the phrases are interpreted in their surface positions, we do not predict any binding violations. This is especially true for the R-expression condition. In this condition, upon reaching the R-expression in the complement of *wh*, the parser should have no reason to pursue an analysis in which the phrase reconstructs to a structurally lower position. When it is encountered in the linear string, the phrase containing the R-expression can be assigned a semantic denotation, and thus, it can be interpreted in this position. In fact, if the parser was to pursue a reconstruction analysis when the complement of *wh* contains an R-expression,

[44] The embedded subject in these constructions is actually PRO, which is controlled by the matrix object, i.e., *John* in (1). For simplicity, the indices in the examples abstract away from the issue of PRO since it is non-overt and thus does not make clear predictions for processing.

this analysis would not be economical since it would create an open dependency that is not required for interpretation.

In principle, it is also possible for the phrase containing the pronoun to be interpreted in its surface position. If such an analysis is adopted, the pronoun is interpreted as referential. In contrast, if the parser prefers to find within-sentence antecedents for the pronouns (following Van Gompel and Liversedge, 2003), it might pursue a reconstruction analysis when the complement of *wh* contains a pronoun. This could happen in two ways. The parser could treat the pronoun similarly to anaphors and immediately start looking for an antecedent further to the right. Alternatively, the parser could wait until it has actually found an antecedent further to the right before it pursues a reconstruction analysis. If the parser was to not find an antecedent further to the right, it would still be possible for the pronoun to receive a semantic denotation in its surface position (the higher copy position).

3.2.1.1 Avoiding logophoricity

Crucially for the current experiment, it is important to note that there is an alternative analysis in (1-a). In this example, it might also be possible for the parser to interpret the pronoun as an anaphor when it appears in this syntactic position, i.e., a logophoric position (Reinhart and Reuland, 1993; Pollard and Sag, 1992). If the parser is able to interpret the pronoun as an anaphor, any results found in this position would not be reflective of how the parser processes different types of pronominals, i.e., because it would be treating the two types of pronominals similarly. To avoid this potential confound, in Experiment 2, the pronoun/R-expression appeared inside a verbal adjunct on the picture noun, as in (2).

(2) a. ??How many pictures showing him$_i$ with a paintbrush did you ask John$_i$ to take?

b. *How many pictures showing John$_i$ with a paintbrush did you ask him$_i$ to take?

In (2-a), there is a potential for ungrammaticality if the parser adopts a binding relation between the two DPs. If the parser reconstructs the complement of *wh* containing a pronoun below the embedded subject, this would incur a Principle B violation since the embedded subject would c-command the pronoun at LF. However, as previously discussed, pronouns can also be interpreted in their surface positions. In this case, the pronoun would not be co-referential with the within-sentence DP.

3.2.1.2 Preference for binding > co-reference

Even though it is possible for the pronoun to be interpreted as referential in (2-a), it is still possible for some speakers to adopt a co-referential analysis between the two DPs in this example. At first glance, this parse should be ruled out since it violates Principle B of the Binding Theory. To explain how it is still possible to get a co-referential reading, despite the apparent binding violation, we appeal to Reinhart (1983)'s proposal that two co-referential DPs can enter a co-reference relation if binding is not possible.

We assume that the parser is sensitive to the grammar's preferences and that, following Reinhart (1983), there is a grammatical preference to adopt binding relations over co-reference relations.[45] This means that whenever a binding relation might be possible, the parser should try to adopt that relation. Thus, upon finding a pronoun in the complement

[45] A recent psycholinguistic study by Cunnings et al. (2014) investigated whether (variable) binding is preferred over co-reference when the potential binder is a quantified DP. Since quantified DPs are quite different from other types of DPs, most notably because they do not refer, we cannot directly compare the results reported in Cunnings et al. (2014) to the current results. It is worth mentioning though that the researchers did not find that co-reference is preferred over binding or vice versa. Instead they found that participants interpreted a pronoun as co-referential with the linearly closest DP, whether it was a quantified DP or an R-expression. In other words, Cunnings et al. (2014) found that participants interpreted the linearly closest DP as the antecedent for the pronoun, irrespective of whether it occurred in a position from which it could bind the pronoun or not.

of *wh*, the parser should attempt to reconstruct the phrase containing the pronoun. Since there is no antecedent in the previous linguistic context, the parser will need to reconstruct the complement of *wh* in order to adopt a binding relation between the pronoun and its antecedent. In (2-a), reconstruction of the complement of *wh* to a lower copy position would incur a Binding Principle B violation at LF because *John* would c-command the phrase containing the pronoun.[46]

In (2-b), if the complement of *wh* reconstructs below the verb site, this would incur a Principle C violation because pronoun in the main clause will locally bind the R-expression at LF.[47] If the embedded verb was not a verb of creation in (2-b), we would see no reason for the parser to pursue an analysis in which it would need to reconstruct the complement of *wh* when it contains an R-expression. R-expressions can be interpreted when they are encountered so reconstruction of the complement of *wh* containing an R-expression would be unmotivated. Since the embedded verb is a verb of creation, reconstruction is obligatory for the *de dicto* interpretation. Reconstruction of the complement of *wh* containing an R-expressions incurs a Principle C violation and this sentence should be ruled out.

If reconstruction of the complement of *wh* incurs a binding violation, we expect to observe reanalysis effects further to the right in the linear string, i.e., structurally below the potential antecedent. For example, if the parser reconstructs *many-x* in (2-a) below the subject DP, *John*, this would incur a Principle B violation. If the parser is sensitive to this grammatical restriction, then we expect to find increased processing difficulty at this position, as reflected by increased reading times. This effect can be

[46]An alternative analysis is that the verbal adjunct creates its own binding domain and thus even when *many-x* reconstructs, Principle B would not be violated because the pronoun would be in its own binding domain. See Reuland (1983) for an analysis of DP-ing gerunds, e.g., Michael counted on *them finishing the book soon*, as their own clauses and thus, their own binding domains.

[47]Note that this prediction holds even if we analyze the verbal adjunct as its own binding domain. If we view ungrammaticality on a scale, this means that (2-b) would be more ungrammatical than (2-a), under a reconstruction analysis.

interpreted as a reanalysis effect, i.e., the parser initially interpreted the pronoun and R-expression as co-indexed and adopted a binding interpretation and it must now reevaluate this interpretation.[48]

3.2.1.3 Creation verbs in the embedded verb position

Note that we are making a distinction between reconstruction for binding and reconstruction for scope (i.e., the *de dicto* reading of creation verbs). When the complement of *wh* contains an R-expression, the phrase does not reconstruct for binding since the R-expression can be interpreted in its surface position and reconstruction would lead to a Principle C violation. Nevertheless, it is still possible that the phrase might need to reconstruct for scope, i.e., in the case of creation verbs. In (2-b), the creation verb should signal to the parser that reconstruction is obligatory to the lowest copy position since this sentence must be interpreted as *de dicto*. If the parser reconstructs the complement of *wh* to the base-generated position, this should incur a Principle C violation. As a result, (2-b) should be ruled out if the embedded verb is a verb of creation. A similar issue arises in (2-a). The creation verb should signal to the parser that the complement of *wh* containing a pronoun must reconstruct below the creation verb for the *de dicto* interpretation. Under the co-indexed reading, we judge (2-a) to be questionable but it is not as ungrammatical as (2-b). It is not clear to me why this is the case. One possibility is that we are able to get the co-reference reading in (2-a) but this reading is not possible in (2-b). As we will see in the results

[48] These predictions rely on the parser interpreting the pronoun and the R-expression as co-indexed, which we have already discussed. However, co-indexation between the R-expression and the pronoun is not the only possible interpretation of the sentence. It is also possible for the pronoun and the R-expression to be interpreted as referring to separate entities. If this is the case, then no binding violations are predicted once *many-x* has reconstructed at LF. Our Experiment 2 results (reported in this chapter) suggest that participants interpret the two DPs as co-referential the majority of the time, providing further evidence for the hypothesis that the parser prefers to finds antecedents for pronominals within the sentence (Van Gompel and Liversedge, 2003).

section, participants judged the two DPs as co-referential in our Experiment 2 which either suggests that this sentence does not violate Binding Principle B (i.e., this is not a Principle B environment, despite appearances) or that they are able to get a co-reference relation.[49]

3.2.2 Cataphoric vs anaphoric processing

As previously mentioned, in Experiment 2, we manipulated the type of DP found in the complement of *wh*, which was either a pronoun, as in (3-a) or an R-expression, as in (3-b). Since pronouns can serve as embedded subjects, we also manipulated whether the embedded subject was an R-expression or a pronoun, as in (3-a)-(3-b), respectively. If the DP found in the *many-x* phrase was a pronoun, the embedded subject was an R-expression and vice versa. The two DPs always matched in gender. Following previous work, we assumed that the parser prefers to find an antecedent for a pronoun within the sentence. Thus, (3-a) can be characterized as a cataphoric dependency because the pronoun linearly precedes its antecedent while (3-b) can be characterized as an anaphoric dependency because the pronoun linearly follows its antecedent.[50]

[49] If we consider the unmoved version of the sentence, as shown in (i-a), the sentence seems acceptable to some speakers under a co-indexed reading. In contrast, the unmoved version of (2-b), as shown in (i-b), is ungrammatical due to a Principle C violation.

(i) a. ?? I asked John$_i$ to take 5 pictures showing him$_i$ with a paintbrush.
    b. *I asked him$_i$ to take 5 pictures showing John$_i$ with a paintbrush.

Thanks to Martin Hackl for bringing up this issue. These examples seem to suggest that, at least for some speakers, the construction containing a pronoun does not create a Binding Principle B violation. Since this sentence does not seem to be completely acceptable to all speakers, we propose that this manipulation still enables us to investigate the processing of pronouns appearing in positions in which they linearly precede their antecedents. However, we acknowledge that there might be some speaker variation in terms of acceptability of Principle B violations. We discuss this issue in more detail in Chapter 4.

[50] Under a reconstruction analysis, it might be possible to view these relations in the opposite way, i.e., (3-a) might be seen as anaphoric and (3-b) might be seen as cataphoric. To be consistent with previous literature, we will characterize anaphora and cataphora based on the positions in which the pronouns and their antecedents appear in the surface representation of the sentence.

(3) The reporters wondered ...

a. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to deny

b. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to deny

... for tomorrow's live television interview.

3.2.2.1 Experimental work on cataphoric processing

The results from previous psycholinguistic experiments investigating cataphoric processing (Van Gompel and Liversedge, 2003; Kazanina et al., 2007), suggest that cataphora are processed differently than anaphora. However, most of these experiments did not compare the processing of both types of dependencies in a single experiment. One exception is a study by Kennison et al. (2009) that directly investigated how the parser processes anaphoric and cataphoric dependencies in the same experiment, as in (4). In the anaphoric condition, (4-a), the pronoun linearly follows its R-expression antecedent whereas in the cataphoric condition, (4-b), the pronoun linearly precedes its R-expression antecedent.

(4) a. *Anaphoric*:

After Ted$_i$ arrived, he$_i$ asked for a cup of coffee.

b. *Cataphoric*:

After he$_i$ arrived, Ted$_i$ asked for a cup of coffee.

In Experiment 1, Kennison et al. (2009) investigated the processing of anaphoric and cataphoric dependencies using the self-paced reading methodology. All stimulus items consisted of two clauses, as in (5). In the anaphoric conditions, a proper name occurred in the first clause and a pronoun occurred in the second clause, as in (5-a)-(5-b). In the cataphoric conditions, a pronoun occurred in the first clause and a proper name occurred in the second clause, as in (5-c)-(5-d). Kennison et al. (2009) used the gender-mismatch paradigm (Van Gompel and Liversedge, 2003), as discussed in Chapter 1 of this thesis. Recall that the assumption in this paradigm is that effects of gender mismatch reflect the parser's attempt to establish a referential dependency between the two DPs. Thus, in Kennison et al. (2009)'s experiment, the pronoun and the proper name either matched, e.g., (5-a)-(5-c), or did not match, e.g., (5-b)-(5-d), in gender. The researchers predicted that the parser would associate the pronoun with the proper name in all conditions (Kennison, 2003; Van Gompel and Liversedge, 2003). Processing difficulty was expected when the pronoun and R-expression mismatched in gender because in such cases, the parser cannot find an antecedent for the pronoun within the sentence.

(5) a. *Gender congruent, anaphoric*:
   After Ted arrived at the party, he decided very quickly to leave.
 b. *Gender incongruent, anaphoric*:
   After Sue arrived at the party, he decided very quickly to leave.
 c. *Gender congruent, cataphoric*:
   After he arrived at the party, Ted decided very quickly to leave.
 d. *Gender incongruent, cataphoric*:
   After he arrived at the party, Sue decided very quickly to leave.

To make predictions about the differences between the anaphoric and cataphoric conditions, Kennison et al. (2009) appealed to discourse processing theories. Following Gordon and Hendrick (1998), cataphoric conditions should incur greater processing difficulty compared to anaphoric conditions because the cataphoric conditions violate discourse processing norms. While pronouns typically refer to previously introduced discourse entities, proper names normally introduce new discourse entities and can act as antecedents for later pronominal referents (see also Gordon et al., 1993; Kennison and Gordon, 1997; Swaab et al., 2004). In an anaphoric dependency, both the proper name and the pronoun are found in positions that adhere to typical discourse properties, i.e., the proper name introduces a new discourse entity and the pronoun refers back to this entity later in the sentence. In contrast, in a cataphoric dependency, both the proper name and the pronoun occur in positions which violate the expected discourse properties, i.e., the introduction of the new discourse entity follows the pronominal referent. Gordon and Hendrick (1998)'s theory of discourse processing thus predicts that associating the anaphoric pronoun with its antecedent should occur quickly and be easier to process but that associating the cataphoric pronoun with its antecedent cannot occur until the antecedent has been found later in the sentence and thus, it should incur greater processing difficulty.

In contrast, following Gernsbacher (1990)'s model, cataphoric conditions should be processed more quickly than anaphoric conditions because the cataphoric pronoun in the first clause creates a strong expectation for a later referent whereas the early proper name does not create an expectation that a referent will be found further to the right in the linear string (see also Gernsbacher and Jescheniak, 1995; Gernsbacher and Shroyer, 1989; Jescheniak, 2000). More precisely, the researchers argue that cataphors increase their activation in the discourse and thus allow them to be retrieved more quickly than non-cataphoric elements. Gernsbacher and colleagues' model also predicts that resolving cataphoric dependencies may occur more quickly than resolving anaphoric dependencies because the DP found further to the right in the string is only expected with cataphoric pronouns but not with anaphoric pronouns. In other words, there is no indication earlier (further to the left) in the linear string that a pronoun will be found further to the right in an anaphoric string. In contrast, upon encountering a cataphoric pronoun, the parser must necessarily find an antecedent further to the right if it is to find the referent for the pronoun within the sentence.[51]

Kennison et al. (2009) found that reading times in the second clause were longer when the second DP mismatched the first DP in gender, e.g., (5-b)-(5-d) compared to conditions in which the two DPs matched in gender, e.g., (5-a)-(5-c). This effect replicates the results of Van Gompel and Liversedge (2003). Kennison et al. (2009) found that the difference between the gender matched and mismatched conditions was larger for the anaphoric conditions, e.g., (5-a)-(5-b), compared to the cataphoric conditions, e.g., (5-c)-(5-d). More precisely, when the pronoun and proper name matched in gender, the second clause was read faster in anaphoric conditions compared to cataphoric conditions. This result suggests that it was easier to resolve the anaphoric dependency compared to the cataphoric dependency, i.e., (5-a) compared to (5-c). However, when the pronoun and proper name mismatched in gender, cataphoric conditions were read faster than anaphoric conditions, suggesting that mismatched gender incurred greater processing difficulty for anaphoric conditions than for cataphoric conditions.

Kennison et al. (2009) also examined reading times on the region when the pronoun/proper name was first introduced and found that reading times on the cataphoric pronoun were faster than those on the proper name, supporting Gordon and Hendrick (1998)'s hypothesis that DPs with more discourse properties are more difficult to process. This result is also contra Gernsbacher (1990)'s view that a cataphoric pronoun will create a stronger expectation of a later referent and incur processing difficulty. Thus, Kennison et al. (2009)'s results

[51] Note that the studies conducted by Gernsbacher and her colleagues did not specifically address cataphoric pronouns but they found effects with cataphoric determiners, e.g., *this* vs. *a*. It is possible that the cataphoric use of *this* may not be directly comparable to cataphoric pronouns because *this* is topic-marked in these constructions. Thanks to Ivona Kučerová for bringing up this fact.

demonstrate that when the cataphoric pronoun is first introduced, it is read faster than when a proper name is found in this position. However, processing difficulty in cataphoric sentences is incurred when the antecedent for the cataphoric pronoun is found later in the sentence, compared to this same region in anaphoric sentences. These results suggest that the resolution of cataphoric and anaphoric dependencies in real-time processing is fundamentally different and thus, that the parser may pursue different strategies when confronted with different types of anaphora in real-time.

3.2.2.2 Summary of previous work on cataphoric processing

Based on the results of Kennison et al. (2009)'s study and the cataphoric processing experiments reported in Chapter 1, previous psycholinguistic work on cataphoric processing has consistently found that the parser prefers to find a co-referent for a cataphoric pronoun within the sentence, at least within an experimental setting (Cowart and Cairns, 1987; Van Gompel and Liversedge, 2003; Kennison, 2003; Kennison et al., 2009). Therefore, the parser will not look outside of the sentence for a co-referent, unless co-reference would incur a grammatical violation such as Principle C (Kazanina et al., 2007). Work specifically investigating how cataphoric and anaphoric dependencies are processed (Kennison et al., 2009) suggests that the parser uses different strategies for cataphors and anaphors. More precisely, processing difficulty for cataphoric processing is incurred when the antecedent is found later in the sentence, i.e., when a dependency must be created between the previously introduced anaphor and its DP antecedent. No such processing difficulty is observed in anaphoric processing, suggesting that it is easier to retrieve a previously mentioned proper name from memory and link it to a pronoun, compared to retrieving a pronoun from memory when a DP antecedent is found. Thus, cataphoric processing has been shown to incur greater processing difficulty compared to anaphoric processing.

3.3 Experiment 2

3.3.1 Anaphors vs pronouns in complements of *wh*

In Experiment 2, we investigated how pronouns are interpreted in positions in which they linearly precede potential antecedents. Previous work on cataphoric pronouns has found that the parser prefers to associate pronouns with the first DP it finds in the sentence (Cowart and Cairns, 1987; Van Gompel and Liversedge, 2003; Kennison, 2003; Kazanina et al., 2007; Kennison et al., 2009). Thus, we assume that the parser will always prefer a parse in which it can find an antecedent for the pronoun within the sentence. As mentioned in the Discussion section of Chapter 2, the results from Experiment 1 are compatible with both a structural analysis and non-structural analysis. Under the structural analysis, the parser reconstructs the phrase containing the anaphor (complement of *wh*) to a lower copy position at LF. In this position, the anaphor can be structurally bound by its antecedent. Under the non-structural analysis, the parser is simply looking for a feature-matched antecedent for the anaphor but it does not care about its structural position. This means that it would be sufficient for the parser to find a feature matched antecedent in the sentence, even if this DP cannot structurally bind the anaphor. In Experiment 2, we aim to further investigate which of these analyses is better able to explain our results.

If the results from Experiment 1 can be explained by the non-structural analysis, we predict that we should be able to find similar effects with pronouns as we did with anaphors. In both cases, the parser is looking for a feature-matched antecedent with which it can associate the pronominal. If the parser does not care about structural constraints (such as binding constraints), the type of DP found in the complement of *wh* should not make a difference.

In contrast, if our previous results can be explained by a structural analysis, we might expect to find different processing effects with pronouns compared to anaphors. Recall that when the complement of *wh* contains a pronoun, it can either be interpreted high or low, depending on the assignment function associated with the pronoun. Crucially, unlike anaphors, pronouns can be assigned semantic denotations, and thus interpreted, in their surface positions. This means that the complement of *wh* can be interpreted in the higher copy position. If the phrase containing the pronoun is interpreted as soon as it is encountered, the pronoun would refer to a DP outside of the sentence since there is no possible antecedent in the preceding linguistic context. Under this parse, the parser would satisfy its preference to interpret phrases as soon as possible.

However, if the parser prefers to find antecedents for pronouns within the sentence, as suggested by Van Gompel and Liversedge (2003)'s work, it may still prefer to associate the pronoun with a DP found further to the right in the string. Finding a DP antecedent for the pronoun further to the right would also enable the parser is satisfy the preference for binding over co-reference relations (Reinhart, 1983). In the pronoun cases, the parser will not know if a binding relation is a possible analysis until it finds the DP subject further to the right in the string. Upon finding the DP subject, the parser may revise its parse to adopt an analysis in which the pronoun and the embedded DP are co-indexed. Thus, we might not expect to find any effects in the pronoun condition until the embedded subject is found. At this point, the parser is able to consider the possibility of co-indexing the pronoun with a within-sentence DP. Prior to reaching the embedded subject, the parser would not know if such a DP would be found further to the right in the string. If the parser prefers to associate the pronoun with a within-sentence DP and if binding is preferred over co-reference, the parser should reconstruct the complement of *wh* to a lower copy position at LF. Thus, we would expect to find an effect of reconstruction at the embedded subject position in this condition.

To summarize, we expect anaphors and pronouns to show different processing profiles when they linearly precede their (potential) antecedents if the structural analysis of our previous results is on the right track. Anaphors must find local antecedents within the sentence while pronouns cannot have local antecedents. Furthermore, pronouns can be interpreted when they are encountered since their interpretation does not necessarily depend on another DP. It is possible that the parser will not pursue a binding analysis in the pronoun conditions until it finds an antecedent further to the right. If this was the case, we would expect to find effects of the pronoun at the embedded subject position. Overall, if our previous results reflect a mismatch between the PF and LF representations, we expect anaphors and pronouns to be processed differently. In contrast, if our previous results can be explained by a non-structural analysis, we expect that anaphors and pronouns should be processed similarly since they are both looking for within-sentence antecedents, irrespective of their structural position.

3.3.2 Reconstruction for *de dicto* interpretations

In Experiment 2, we also investigated whether the parser is sensitive to reconstruction for the *de dicto* interpretation of creation verbs. As discussed in Chapter 2, if the embedded verb is a creation verb, reconstruction of the complement of *wh* is obligatory because the string must receive a *de dicto* interpretation. However, the parser will only discover that reconstruction needs to occur for *de dicto* interpretations when the creation verb is found further to the right in the sentence. Upon reaching the creation verb, if the parser is sensitive to LF structure, then it will realize that *many-x* needs to reconstruct below the verb in order to be interpreted.

In Experiment 1, we predicted that finding an anaphor in the complement of *wh* would signal to the parser that it must reconstruct the phrase containing the anaphor below its antecedent. In these conditions, we predicted that upon finding a creation verb, the parser would be able to recover more easily than in conditions where there was no early signal for reconstruction because it had already reconstructed the phrase containing the anaphor. This prediction was not borne out and we found no reliable effect of verb type in Experiment 1.

In Experiment 2, we make two critical assumptions: i) the parser prefers to find within-sentence antecedents for pronominals (Van Gompel and Liversedge, 2003), and ii) binding relations are preferred over co-referential relations (Reinhart, 1983). If these are the parser's preferences, we might expect that as soon as the parser finds a pronoun, it will try to bind this pronoun to a DP within the sentence, unless doing so would incur a binding violation. This will be the case whether the pronoun precedes or follows its antecedent. Thus, finding a pronoun in the complement of *wh* would provide an early indication that reconstruction may need to occur further to the right in the sentence. However, as previously mentioned, there is also the possibility that the parser will wait until it has located an antecedent further to the right before it attempts to reconstruct and establish a binding relation. In this case, finding a pronoun early in the sentence would not necessarily signal to the parser that it must reconstruct that phrase. Experiment 2 was designed to help us better understand the results from Experiment 1 and in particular, help us clarify whether the Experiment 1 results truly reflect reconstruction operations or whether they can simply be explained as an effect of the parser looking for a feature-matched DP associate.

3.3.3 Design

The experimental design was adapted from Experiment 1 and manipulated the type of DP in the *wh*-filler phrase and the type of verb (creation, non-creation). Sample items are shown in (6).

(6) The reporters wondered ...

a. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to invent

b. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to deny

c. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to invent

d. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to deny

... for tomorrow's live television interview.

The DP was either a third person, singular pronoun, e.g., *him/her*, or an R-expression, which were always first names, e.g., *John, Alexa*. The verb was either a creation verb, e.g., *write, invent*, or a non-creation verb, e.g., *edit, deny*. In contrast to Experiment 1, the first target DP was embedded in a verbal adjunct that followed the picture-DP, e.g., *discrediting her/Alexa as a witness*. The verbal adjunct was always 5 words in length and of a similar syntactic structure across experimental items. The pronoun/R-expression was embedded in a verbal adjunct to eliminate the possibility that the pronoun could be interpreted as a logophoric anaphor (Reinhart and Reuland, 1993; Pollard and Sag, 1992)[52] and to ensure

[52]Note that the perspective holder of the asking event was always the pronoun, *you*, which also reduced the possibility of interpreting the pronominal logophorically. In the theoretical literature on logophors (Reinhart and Reuland, 1993), one test that has been used to demonstrate whether an anaphor can be interpreted logophorically is if it can be replaced by a pronoun and the meaning does not change, as in (i).

(i) a. That picture of her$_i$ on the wall bothers Mary$_i$.

b. That picture of herself$_i$ on the wall bothers Mary$_i$.

To avoid the possibility that participants might have preferred a logophor as the PP object in the picture DP, that it was being correctly interpreted as a pronoun. Since pronouns, unlike anaphors, can themselves serve as subjects, we also manipulated whether a pronoun or R-expression served as the embedded subject, e.g., *her* in (6-c)-(6-d). In (6-a)-(6-b), the pronoun is therefore cataphoric (its referent linearly follows it) whereas in (6-c)-(6-d), the pronoun is anaphoric (its referent linearly precedes it).[53] We again used embedded questions with a plural DP as the matrix subject to ensure that the pronoun could never be co-indexed by the matrix subject. Within each sentence, the pronoun and R-expression always matched in gender. We also ensured that an equal number of male and female trials were found across the experiment. The proper names used in the experiment were not gender neutral.

As mentioned in section 3.2.1.2, (6-c) is predicted to be ungrammatical when the embedded verb is a verb of creation. The creation verb forces reconstruction of the complement of *wh* at LF. If the R-expression is co-indexed with the embedded subject, reconstruction would yield a Principle C violation. As a result, it should not be possible for speakers to interpret the R-expression and the pronoun as co-referential in this example. We also predict a binding violation in (6-b). The creation verb forces reconstruction for the *de dicto* interpretation but reconstruction yields a Principle B violation. If we were to find that participants are able to get a co-referential reading between the pronoun and the embedded subject, this would either suggest that this is not a Principle B environment or that they are able to adopt a co-reference analysis in this example.

we embedded the pronouns in verbal adjuncts. In doing so, the pronoun could not be misconstrued as a logophor:

(ii) a. The reporters wondered how many lies discrediting her$_i$ as a witness you are asking Mary$_i$ to deny...
b. *The reporters wondered how many lies discrediting herself$_i$ as a witness you are asking Mary$_i$ to deny...

[53] Note that it is of course possible for these pronouns to get their referent from outside the sentence but following much previous work (e.g., Van Gompel and Liversedge, 2003; Kazanina et al., 2007), we presume that the parser prefers to find its referent within the sentence.

3.3.4 Naturalness judgement task

As in Experiment 1, we first ran a naturalness judgement task before running the self-paced reading experiment. As in Experiment 1, the sentences were very complex and thus, we expected them to be difficult to process. Assuming that lower naturalness ratings reflect greater processing complexity, the naturalness judgement task enabled us to get a general measure of the processing difficulty of these sentences. In addition, in this version of the task, we directly asked comprehension questions targeting participants' interpretation of the DPs in the sentence. This allowed us to measure whether naïve speakers do or do not have a preference is to interpret the pronoun (cataphoric or anaphoric) as referring to the same real world referent as another DP within the same sentence, as predicted by previous processing studies.

3.3.4.1 Specific predictions

Generally, cataphoric pronouns are less common than anaphoric pronouns (as suggested by Kennison et al., 2009) and therefore, they may be more difficult to process. If cataphors incur greater processing difficulty, we expect that the cataphoric constructions will receive lower ratings compared to constructions containing anaphoric pronouns. This follows the assumption that lower ratings reflect greater processing difficulty. One possible explanation for why cataphoric pronouns might be more difficult to process is that the referent for the cataphor has not yet been encountered when the cataphor is read. However, since ratings are provided once the whole sentence has been read and processed, the parser has already located the antecedent for the cataphor before a rating is provided. As a result, this measure might not be sensitive enough to detect any differences between the two constructions and they might be rated similarly.

3.3.4.2 Materials

Thirty-two sentence templates were created following the paradigm shown in (6). Target sentences were counterbalanced across four lists of stimuli combined with 30 sentences from an unrelated experiment and 40 fillers (16 of the fillers were created to resemble the target sentences, and 24 fillers resembled the items from the unrelated experiment). In total, participants read and judged 102 stimulus sentences. Comprehension questions were asked after each sentence to ensure that participants were paying attention.

As co-indexation is not obligatory between the pronouns and the R-expressions in the test items, we asked CQs directly targeting how participants were interpreting the DPs. For example, for (6), a CQ of this type would be *Was Alexa being asked to invent lies about herself for the interview?* If participants were interpreting the pronoun and R-expression as co-indexed, then we expect them to answer *yes*. Based on our results from Experiment 1, in which we did not find significant differences based on the type of verb used in the sentence, we also asked CQs directly targeting participants' interpretation of the event. For example, for the sentence in (6), a CQ asking about the creation event would be *Did the lies already exist?* If participants were interpreting the creation verb as intended, we expected a *no* answer because the object of a creation verb does not exist until the event has been completed. For the non-creation verbs, we expected a *yes* answer because the objects of non-creation events carry a presupposition of existence. There were a total of 8 CQs asking about the creation event, 12 CQs asking about co-indexation and 12 CQs with clear yes-no answers (50% with a "no" answer). When the answer to the CQ depended on how participants were interpreting the sentence, no feedback was provided and either answer was accepted. If there was a clear yes-no answer, we provided feedback and informed participants when they were answering questions incorrectly.

3.3.4.3 Participants

Seventy-three participants completed the naturalness judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). Five participants were removed from the analysis because they scored less than 75% on the comprehension questions throughout the whole experiment. We did not eliminate participants based on their comprehension scores on the test trials because there were so few trials with correct answers in this version of the experiment. In the results section, the data for the remaining 68 participants (29 males, 39 females) will be presented. All participants were self-reported native speakers of English, aged between 19 and 61 years of age (M = 34.5 years, SD = 11.19). Participants were paid $3.50 USD for their participation.

3.3.4.4 Procedure

The experimental procedure was the same as in Experiment 1 with the exception that we accepted both answers when there was not a clear yes-no answer for a CQ. As in Experiment 1, the test sentences were always presented on two lines with the line break appearing after the head noun of the picture DP, e.g., *pictures, lies, photos*, as shown in (7).

(7) The reporters wondered how many lies
discrediting her as a witness you are asking Alexa to deny for tomorrow's live television interview.

3.3.5 Results

3.3.5.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 6 fillers were removed from the analysis. After trimming, mean accuracy on the comprehension questions for the test trials was high at 83.5 % and 90.4% for all trials. In order to only investigate judgements on trials where participants were interpreting the sentence as intended, trials in which CQs were answered incorrectly were removed from the analysis, whenever possible.[54] This resulted in a loss of 7.6% of the data.

3.3.5.2 Statistical analyses

We used the same statistical analyses as described in Experiment 1 with the same trimming procedures. We report here the final trimmed models. As in Experiment 1, all statistical analyses were calculated on the z-scored data across all trials.

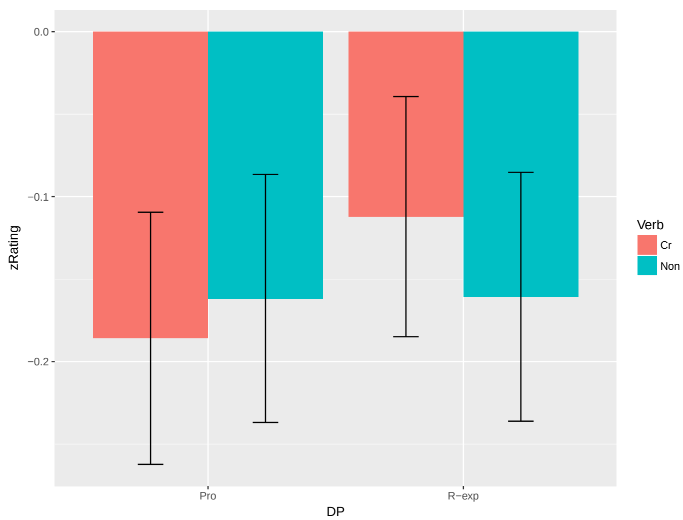
3.3.5.3 Naturalness judgements results

Mean z-scored naturalness judgements are plotted in Figure 3.1. The mean judgements for pronoun conditions are plotted in the left panel and for R-expressions in the right panel. There were no reliable effects of either verb type (creation, non-creation) or DP type (pronoun, R-expression), suggesting that participants did not show a preference for the linear order of the DPs in the critical sentences, i.e., pronoun preceding R-expression or vice versa, nor did they show a preference for either the creation verb (pink bars) or non-creation verb (blue bars) stimuli. The DP results suggest that the cataphoric conditions were

[54] For trials where there was no correct answer, all trials were kept in the analysis.

not more difficult to process than the anaphoric conditions, at least not in the naturalness ratings.

Figure 3.1: Mean naturalness judgements (in z-scores) by DP type (pronoun, R-expression) and verb type (creation, non-creation) in Experiment 2. Note that DP type indicates the first DP in the sentence, i.e., the DP found in the filler phrase. Error bars represent 95% confidence intervals.



Note that the absence of a DP effect in the judgements in Experiment 2 is distinct from the effect found in the naturalness ratings in Experiment 1, where we found that R-expression conditions were rated as more natural than the anaphor conditions. In Experiment 1, we argued that the differences in judgements between the two conditions, which differ only based on the type of DP found in the filler phrase, suggested that processing difficulty is greater when the anaphor linearly precedes its antecedent compared to a condition in which no referential dependency is required. In Experiment 2, both of the conditions contained referential dependencies but they differed based on the linear position of the pronoun with respect to its antecedent. In the cataphoric condition, the pronoun linearly preceded a potential antecedent. In contrast, in the anaphoric condition, the pronoun linearly followed a potential antecedent. The judgement results from Experiment 2 suggest that neither condition is overall more difficult to process than the other. Thus, these results suggest that cataphoric dependencies do not cause greater processing difficulty than anaphoric dependencies.

Numerically, creation verb conditions were given higher naturalness ratings when there was an R-expression in the picture DP, compared to a pronoun. However, this effect did not reach significance at the 5% threshold and thus, we deem it uninterpretable. The results from the real-time reading study might enable us to interpret these results. No significant differences in judgements were observed in the non-creation verb conditions, i.e., the non-creation verb items were given similar naturalness ratings for both the anaphoric and cataphoric conditions.

3.3.5.4 Comprehension question results

CQs about co-indexation between the pronoun and R-expression We were also interested in how participants were interpreting the sentences. Since co-indexation between the pronoun and R-expression was not obligatory, we asked CQs directly targeting whether participants were interpreting the two DPs as co-indexed. For example, for the stimulus sentences provided in (6), repeated below as (8), a co-indexation question would take the form of *Was Alexa being asked to invent/deny lies about herself for the interview?* If participants responded *yes*, this meant that they were interpreting the two DPs as co-referential. However, if they responded *no*, they were not interpreting the two DPs as co-referential.

(8) The reporters wondered ...

how many lies discrediting her_i/Alexa_i as a witness you are asking Alexa_i/her_i to invent/deny

... for tomorrow's live television interview.

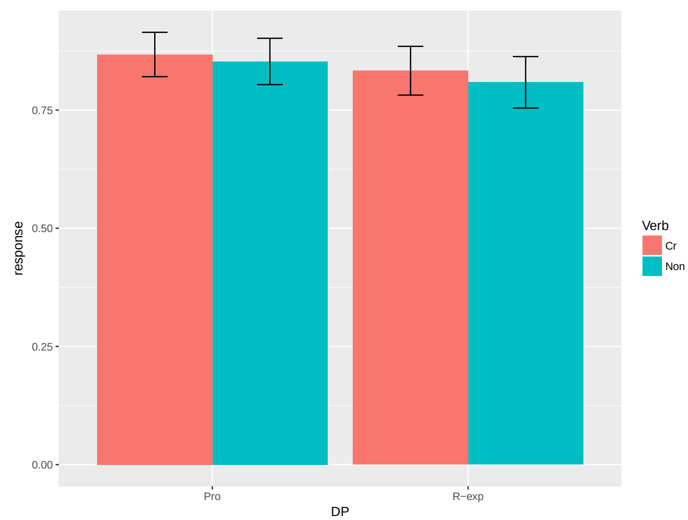
The percentage of *yes* responses to CQs asking about whether or not the pronoun and the R-expression were co-indexed is plotted in Figure 3.2.[55] Across conditions, participants responded *yes* 84.1% of the time. When the pronoun preceded the R-expression (left panel), participants answered *yes* 86% of the time compared to 82.1% of the time when the R-expression preceded the pronoun (right panel). This effect was significant but only after trimming the residual error of the model[56] (β = -0.978, SE = 0.28, *t* = -3.490, $p < 0.001$, model not shown). This effect suggests that participants were more likely to interpret the two DPs as co-indexed when the sentence contained a cataphoric pronoun compared to an anaphoric pronoun. The mean percentage of *yes* responses by DP type only are plotted in Figure 3.3.

We can only speculate about possible reasons for why participants seemed more likely to co-index the two DPs in cataphoric conditions compared to anaphoric conditions. The results suggest that participants are more likely to find a referent outside the sentence

[55] As pointed out to me by Martin Hackl, we might expect to observe an interaction between the type of answers provided to the creation verb questions and the type of answers participants provide for the binding questions. For example, if participants respond *no* to a question asking if the object of the verb existed before the event, meaning that they interpret the verb as a creation verb, we might expect that they will not interpret the pronoun and R-expression as co-referential. This is because the creation verb would force reconstruction at LF and reconstruction of the phrase containing the pronoun or R-expression would yield binding violations. To investigate whether we get such an interaction would require that the same stimulus items are associated with more than one question across participants. Unfortunately, our experiment was not designed in such a way so we can only speculate about this possibility.

[56] We trimmed the residual error of the model to remove outliers. This trimming ensured that the fit of the model was not influenced by particular data points (i.e., outliers). See section 2.3.5.2 in Chapter 2 for more information.

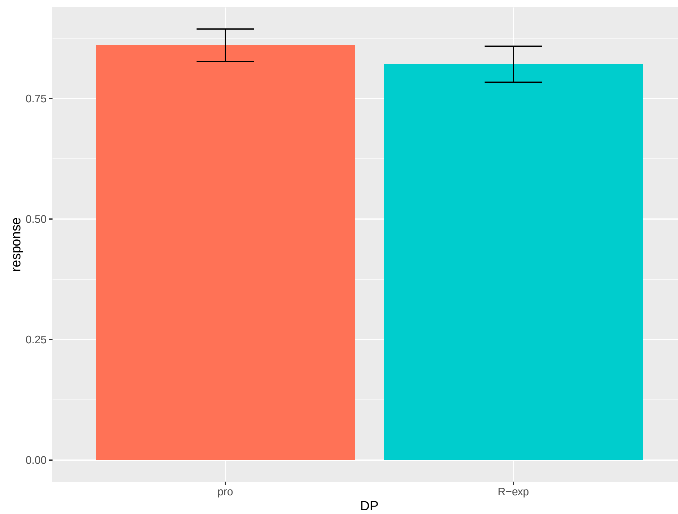
Figure 3.2: Percentage of *yes* responses to the co-indexation CQs in Experiment 2. The DP variable refers to the first DP found in the sentence, i.e., the DP appearing in the filler phrase. Error bars represent 95% confidence intervals.



when the pronoun linearly follows a potential antecedent compared to cases where the pronoun linearly precedes the potential antecedent. In both constructions, a binding relation between the pronoun and the R-expression is not possible, thus violating the parser's preference for binding over co-reference (Reinhart, 1983; Heim and Kratzer, 1998; Fox, 1998; Reinhart, 2006; Roelofsen, 2010).[57] To illustrate, consider the critical sentences used in this experiment, repeated here as (9).

[57] See the discussion in Chapter 1.

Figure 3.3: Percentage of *yes* responses to the co-indexation CQs in Experiment 2. The DP variable refers to the first DP found in the sentence, i.e., the DP appearing in the filler phrase. Error bars represent 95% confidence intervals.



(9) The reporters wondered ...

a. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to deny

b. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to deny

... for tomorrow's live television interview.

In the cataphoric construction, (9-a), the only way for a binding relation to be established between the cataphor and the embedded subject R-expression is for the parser to reconstruct the complement of *wh* to a structural position that is below the embedded subject (its antecedent) at LF. In doing so, the parser would incur a Principle B violation at LF because the pronoun would be locally bound by its antecedent. Consequently, this parse of the string should be ruled out by binding constraints. In the anaphoric construction, (9-b), it is also not possible for a licit binding relation to be established. Since the *wh*-phrase contains an R-expression, it can be interpreted in its surface position and there is no need for reconstruction. If it is interpreted in its surface position, the R-expression does not c-command the pronoun at LF and thus no binding relation is established. The only way for the two DPs to be in a binding relation would be if the phrase containing the R-expression (complement of *wh*) reconstructed below the pronoun at LF. Not only does this operation seem unmotivated, but it would also incur a Principle C violation at LF (the pronoun would c-command the R-expression). As a result, in both cases, the parser needs to adopt a co-reference interpretation to be able to interpret the two DPs as co-indexed.[58,59]

One possible explanation for the result that co-indexation was more likely in cataphoric conditions is that the co-indexation analysis is easier for the parser to establish with cataphoric pronouns compared to anaphoric pronouns. Why? Upon encountering the cataphor, the parser is already able to set up an expectation that will find an antecedent further to the right in the linear string. As a result, in cataphoric conditions, the parser prefers to find an antecedent within the sentence (Van Gompel and Liversedge, 2003). Upon finding the embedded subject, the parser determines that this DP is an appropriate match for the cataphor and establishes co-indexation. Due to binding principles, it cannot

[58]We are treating reconstruction for binding as a separate operation from syntactic integration of the filler phrase in its base-generated position. Recall from Chapter 1 that these we are treating these two operations as separate identification tasks. It is possible that the position to which the parser reconstructs the complement of *wh* is the same as the base-generated position but it can also reconstruct the phrase to an intermediate copy position, provided this position is structurally lower than the antecedent.

[59]It is important to note that even though it seems possible to co-index the pronoun with the embedded subject in (9-a), repeated below as (i-a), the pronoun is unable to be bound by a quantified phrase in the same syntactic environment, as in (i-b).

(i) a. The reporters wondered how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to deny.
  b. *The reporters wondered how many lies discrediting her$_i$ as a witness you are asking every girl$_i$ to deny.

In order for the pronoun to be bound by the quantified phrase, it would need to be found in the c-command domain of the quantifier at LF. One way that this could occur in (i-b) is via reconstruction. Since binding seems impossible in (i-b) but we are still able to get a co-referential interpretation in (i-a), this suggests that the two DPs in (i-a) must be in a co-referential relation and not in a binding relation, where Principle B should be violated.

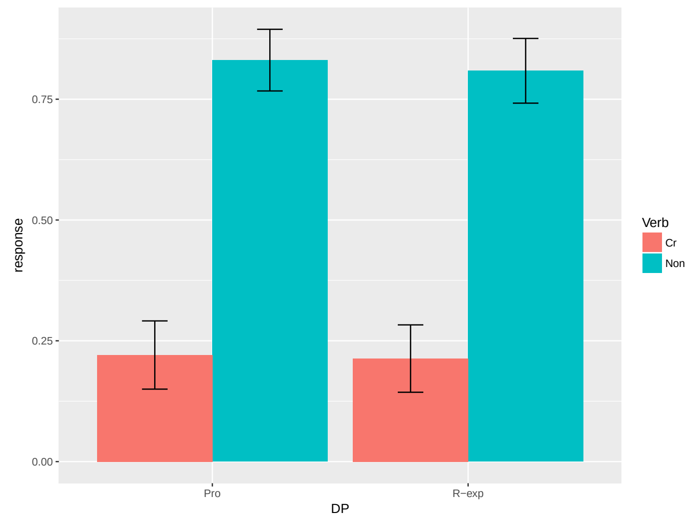
establish a binding relation between the two DPs and it must instead establish a co-reference interpretation. In the anaphoric condition, the parser does not have any prior indication that this sentence might contain a referential dependency until it encounters the pronoun in the embedded subject position. The parser can either look back on the structure it has built so far and establish a co-reference relation between the R-expression and the pronoun (binding is again not possible) or it can find a referent for the pronoun outside the sentence. However, since *yes* responses to the co-indexation questions were high across conditions, we interpret this result as demonstrating that participants prefer to find an antecedent for a pronoun (whether cataphoric or anaphoric) within the sentence and they prefer to not look outside the sentence to find an antecedent. The results of our self-paced reading study might allow us to better understand why we observed differences based on the linear order of the DPs.

We did not observe significant differences in the answers to the co-indexation CQs based on the type of verb found in the sentence (creation or non-creation). If the sentence contained a creation verb, participants responded *yes* 85% of the time, compared to 83.1% of the time if the sentence contained a non-creation verb (see Figure 3.2). Based on their answers to the co-indexation CQs, participants did not seem to interpret the sentences differently based on whether or not there was a creation verb in the sentence. Recall that in the creation verb conditions, the prediction is that *many-x* must reconstruct upon reaching the creation verb because the verb forces a *de dicto* interpretation of *many-x*. Based solely on these CQ results, we cannot conclude whether or not we have evidence for this reconstruction operation with creation verbs. Since the CQs are presented after the whole sentence has already been presented, participants have already needed to process and interpret the sentence before answering the CQ. Thus, these results may not provide us with the most accurate measure of how they are interpreting these sentences as they are being presented to them.

CQs about the creation event Recall that in Experiment 1, we did not observe any significant differences based on the type of verb found in the sentence. In the Discussion section of Chapter 2, we asked whether these results might suggest that participants were not interpreting the verbs as intended. In other words, perhaps participants were not interpreting the creation verbs as events that bring about the existence of an object, i.e., the object of the creation event. In an effort to determine whether our previous results could be explained as an effect of participants not interpreting the verbs as intended or whether they might suggest that participants are not sensitive to the reconstruction operation required for the *de dicto* reading of creation verbs, in Experiment 2, we asked CQs that directly targeted participants' interpretation of the sentential event.

For example, for the stimulus item in (9), we asked, *Did the lies already exist?* We expected participants to respond *yes* when the verb was non-creation but *no* when the verb was a creation verb. Recall that with creation verbs, the object does not exist before the creation event has been completed. The percentage of *yes* responses to the CQs targeting participants' interpretation of the sentential event is plotted in Figure 3.4. Our prediction was borne out: when the verb was a non-creation verb, participants answered *yes* 82% of the time but only 21.7% of the time when the verb was a creation verb. This effect was significant at the 5% threshold (β = 6.65, SE = 0.82, t = 8.08, $p < 0.001$, model not shown) and did not differ based on the type of DP found in the *many-x* phrase ($p > 0.05$). These results suggest that participants were able to detect the difference between creation and non-creation events in the context of the experiment. The lack of an effect of verb type in Experiment 1 cannot be attributed to these verbs not being interpreted as intended.

Figure 3.4: Percentage of *yes* responses to the creation verb CQs in Experiment 2. Error bars represent 95% confidence intervals.



3.3.6 Self-paced reading experiment

No items were removed for the self-paced reading experiment as we wanted to be able to compare the results across the two different versions of Experiment 2 (naturalness judgement task and self-paced reading).

3.3.6.1 Specific predictions

Sample stimuli items are repeated in (10).

(10) The reporters wondered ...

a. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to invent

b. how many lies discrediting her$_i$ as a witness you are asking Alexa$_i$ to deny

c. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to invent

d. how many lies discrediting Alexa$_i$ as a witness you are asking her$_i$ to deny

... for tomorrow's live television interview.

DP in the complement of *wh* The first region of interest in our stimuli for Experiment 2 was the DP occurring in the verbal adjunct, which was either a pronoun, as in (10-a)-(10-b), or an R-expression, as in (10-c)-(10-d). Assuming that the parser prefers to find a within-sentence antecedent for cataphors, the parser should look for a within-sentence antecedent for the pronoun, in a similar way to what we saw in Experiment 1 with anaphors. If the parser initially adopts the same strategy with pronouns as it did for anaphors, we expect greater processing difficulty (longer reading times) on words following pronouns compared to those same words following R-expressions, i.e., we should see the same pattern of results as what we found in Experiment 1. This effect would be compatible with a non-structural explanation of the results from Experiment 1 because it would suggest that the parser is not sensitive to the structural differences between anaphors and pronouns and therefore, treats them the same way. In contrast, if the parser is sensitive to the different structural constraints imposed on anaphors and pronouns, we should find different effects when a pronoun appears in the complement of *wh*.

There is an important distinction to make between the conditions containing anaphors that linearly precede their antecedents and those containing cataphoric pronouns. When anaphors precede their antecedents, they are uninterpretable in their surface positions, i.e., they do not have semantic denotations. As a result, the parser must find an antecedent for the anaphor further to the right to derive an interpretable structure. In contrast, the phrase containing the pronoun can either be interpreted in its surface position (head of the chain) or in a lower copy position, such as an intermediate copy position or the base-generated position (tail of the chain). While there is an overall preference for the pronoun to find its antecedent within the sentence, the parser may not attempt to establish a referential dependency until it finds a matching antecedent. In other words, it may not attempt to associate the pronoun with an antecedent until it reaches the embedded subject in our stimuli. If this is the case, we would not expect to observe any processing difficulty when a cataphoric pronoun is encountered. If we were to find processing difficulty when the pronoun is encountered in the complement of *wh*, this might be taken as evidence that it is being interpreted in that position.

From a purely lexical standpoint, we might expect that R-expressions may incur longer reading times compared to pronouns because they carry more discourse properties (see e.g., Gordon and Hendrick, 1998). If greater discourse properties contribute to greater processing difficulty, we expect R-expressions to incur longer reading times compared to pronouns when they are initially found in the filler phrase.

Embedded subject DP The second region of interest in Experiment 2 was the subject of the embedded clause, which again was either a pronoun, as in (10-c)-(10-d), or an R-expression, as in (10-a)-(10-b). The second DP always matched the previously introduced DP in gender. If the parser prefers to find a within-sentence antecedent for the cataphoric pronoun and this is reflected by longer reading times after the pronoun is first introduced, we expect this effect to disappear once the antecedent is found, i.e., the R-expression subject, similarly to the effect found in Experiment 1 in the anaphor conditions. However, a different possibility is that the parser does not establish a referential dependency between a pronoun and another DP until it has located the second DP in the sentence.

Recall that we are following previous work in assuming that the grammar prefers binding relations between two DPs over co-reference relations (see e.g., Reinhart, 1983, 2006; Heim and Kratzer, 1998; Fox, 1998; Roelofsen, 2010, among others). If the parser is sensitive to this preference, it should first try to bind (non-locally) the cataphoric pronoun with the embedded R-expression. Co-reference relations in which the pronoun is not in a structural relation with another DP in the sentence should only be postulated as a last resort. The only way for the parser to establish a binding relation between the cataphor and the embedded R-expression is if it reconstructs the phrase containing the cataphor below the R-expression. However, doing so would incur a Principle B violation and this parse should be ruled out. Provided that the parser still interprets the two DPs as co-indexed, the parser will need to reanalyze its parse and adopt a co-reference interpretation. Reanalysis should be costly and this cost should be reflected by longer reading times at the embedded subject region. Thus, we predict that if the parser is sensitive to the binding over co-reference preference, longer reading times should be incurred in the cataphoric condition on the embedded R-expression subject. What about in the anaphoric condition? In this case, we do not predict reconstruction and thus we do not predict a reanalysis effect. Since the complement of *wh* contains an R-expression, it can be interpreted in its surface position and does not need to reconstruct. In fact, reconstruction of the complement of *wh* containing an R-expression would be unmotivated because there is no grammatical preference or constraint motivating this analysis.

Verb region The third and final region of interest is the verb region, which was either a creation verb, as in (10-a)-(10-c), or non-creation verb, as in (10-b)-(10-d). The non-creation verb is compatible with both *de re* and *de dicto* interpretations of *many-x* but the creation verb forces the *de dicto* interpretation, i.e., *many-x* cannot be interpreted as *de re*. In order for a *de dicto* interpretation to be derived, the parser must reconstruct *many-x* when the creation verb is found further to the right in the linear string. In this experiment, unlike in Experiment 1, the parser has not been given any prior indication that it will need to reconstruct *many-x* later in the sentence. In Experiment 1, finding an anaphor in the *many-x* phrase provided the parser with an early indication that this phrase must reconstruct in order to be interpreted. When the filler phrase contains a pronoun or R-expression, the parser has not been given this early indication because both of these types of DPs can be interpreted in their surface positions. Thus, we might expect a cost of reconstruction at the creation verb, compared to the non-creation verb. Increased reading times at the verb site are expected for (10-a) and (10-c) but not for (10-b) and (10-d).[60]

3.3.6.2 Participants

Ninety-nine participants were tested for the self-paced reading version of Experiment 2. Eleven participants were removed from the analysis because they scored less than 75% on the comprehension questions throughout the whole experiment. As in the naturalness study, we did not trim based on the comprehension accuracy on the critical trials because there were so few trials with clear yes-no answers. In the results section, the data for the remaining 88 participants (40 males, 48 females) will be presented. All participants were self-reported native speakers of English, aged between 19 and 64 years of age (M = 33.67 years, SD = 10.59). Participants were paid $6.00 USD for their participation and were naïve as to the purposes of the experiment. Participants who completed the self-paced reading study did not complete the naturalness judgement study.

[60] We assume an analysis in which reconstruction for scope and reconstruction for binding can target distinct structural positions. It is also possible that these reconstruction operations interact. If so, it might be possible to observe reanalysis effects in the base-generated position. If the parser reconstructs *many-x* to this position when the verb is a creation verb and if it has not yet interpreted the complement of *wh*, we predict that this parse will incur binding violations both when the phrase contains an R-expression (Principle C) and when it contains a pronoun (Principle B). However, if the complement of *wh* can be interpreted in a distinct position, either in its surface position or in an intermediate copy position, we only predict to observe a cost of verb type in the base-generated position.

3.3.6.3 Stimuli

The same stimuli and filler items used in the natural judgement task were used for the self-paced reading version of the experiment on Amazon Mechanical Turk. Unlike in Experiment 1, we kept all comprehension questions in this version of the experiment since there were already so few critical trials with clear yes-no answers.

3.3.6.4 Procedure

The procedure was the same as in the self-paced reading study of Experiment 1.

3.3.6.5 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 6 fillers and 1 critical trial were removed from the analysis.[61] We also removed from the analysis any reading times that were less than 90 ms and over 3000 ms in length, as these reading times would be considered outliers. When possible, in order to accurately investigate the reading times on trials where participants were interpreting the sentence as intended, trials where the participant answered the question incorrectly were also removed from the analysis.[62] Accuracy on the comprehension questions was high at 90.5%, after trimming.

As reading times allow for a great deal of variation, we transformed the data in four ways: raw reading times, residual reading times, log reading times, and residuals of the log reading times. For all transformations, data points that were +/- 2 SDs away from the

[61] In addition, due to a typographical error, one condition from one of the lists also had to be removed. To ensure that this did not affect the results, statistical analyses were run excluding the whole item but reveal the same effects. As a result, only the condition with the typographical error was removed from the reported analyses.

[62] As in the previous experiments, all statistical analyses were also run on a data set that included all items. These analyses revealed the same effects. We report here the more conservative data set.

mean by participants were removed (less than 3% of data across all transformations). The statistical models were run on all of these data transformations, and revealed similar effects, unless otherwise indicated.

3.3.6.6 DP site 1

We fitted models to investigate the effect of the first-introduced DP, which occurred within the complement of *wh*. The type of DP (pronoun, R-expression) was added as a random slope by participant whenever the model converged.[63] The first-mentioned pronoun or R-expression was introduced at word 8. We fitted models on words 8-12 to investigate how the first-introduced DP is initially processed. The mean log reading times are plotted in Figure 3.5.

Table 3.1: Final mixed-effects model for log reading times at word 9, one word following the pronoun or R-expression (N = 2510 before trimming, 2438 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

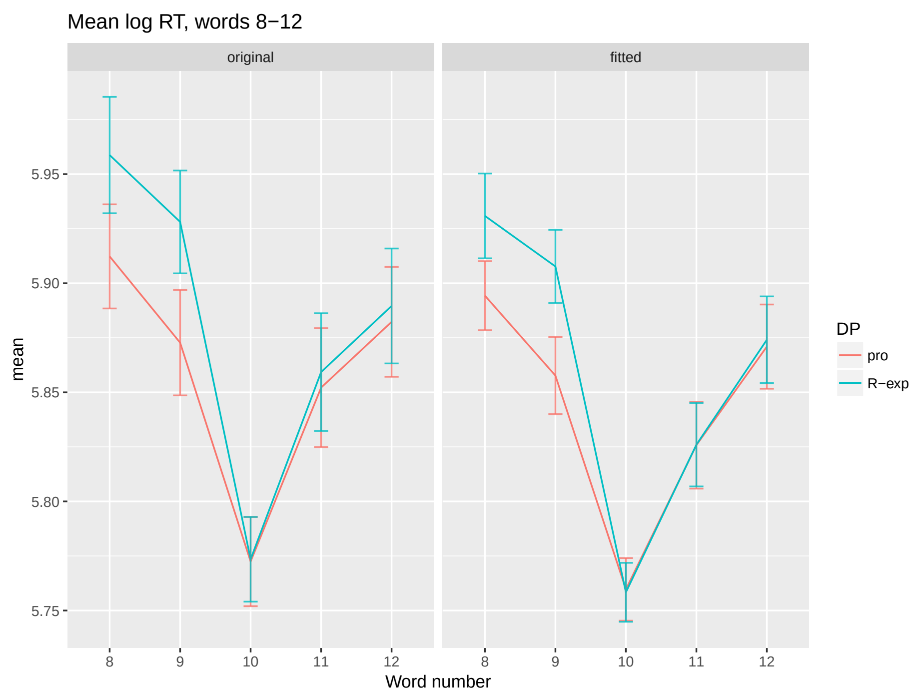
Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.858	0.355	165.13	0.000
R-expression	0.051	0.012	4.412	0.000

Random effects: by-participant intercept (SD = 0.325), by-item intercept (SD = 0.007), by-participant random slope for DP type (SD = 0.046), and the correlation between by-participant and slope (r = -0.56).

At word 8, when the DP was first introduced, we did not find any reliable differences between conditions. Conditions containing R-expressions (pink line in the figure) were read numerically longer than conditions containing pronouns (blue line) but this effect reached only significance at the 5% threshold in the raw reading times (β = 21.62, SE

[63] DP was also added as a random slope by item but the model either did not converge or the effect did not improve model fit.

Figure 3.5: Mean log reading times at DP site from words 8-12 in Experiment 2. Original log mean values in the left panel and fitted log mean values in the right panel. The DP variable refers to the DP found in the filler phrase. Error bars represent 95% confidence intervals.



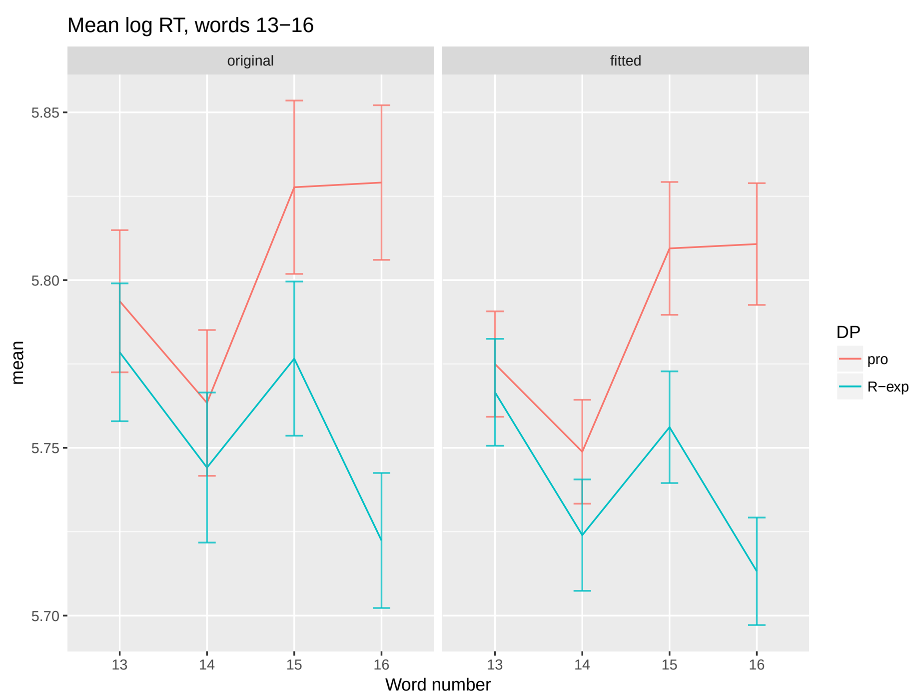
= 5.92, t = 3.652, $p < 0.001$, model not shown) and in the log reading times (β = 0.04, SE = 0.01, t = 2.875, p = 0.004, model not shown). At word 9, one word following the pronoun/R-expression, we found significant effects of DP type across transformations such that the R-expressions conditions were read significantly longer than the pronoun conditions ($p < 0.01$, across transformations). The model for the log reading times on word 9 is shown in Table 3.1.[64] We found no reliable effects on words 10 to 12, suggesting that any initial difficulty with the R-expression is easily resolved.

[64] The data was pooled across verb types at this region since the filler phrase linearly appeared before the verb site and sentences were presented one word at a time. As a result, participants would not have known the type of verb used in the experiment at this region and thus verb type would not have had an effect.

3.3.6.7 DP site 2

A second manipulation in this experiment was the type of DP presented as the subject of the embedded clause, which occurred at word 15 in all critical items. The subject DP was either a pronoun or an R-expression. Recall that if the first-mentioned DP was a pronoun, then the DP subject of the embedded clause was an R-expression, which always matched the pronoun in gender (cataphoric condition). Likewise if the first-mentioned DP was an R-expression, then the DP subject of the embedded clause was a pronoun (anaphoric condition). The mean log reading times are plotted in Figure 3.6.

Figure 3.6: Mean log reading times at DP site 2 from words 13-16 in Experiment 2. The DP variable refers to the first DP found in the sentence, i.e., the DP in the filler phrase. Original log mean values in the left panel and fitted log mean values in the right panel. Error bars represent 95% confidence intervals.



We found significant effects of DP type at word 14, one word preceding the embedded subject, in the raw reading times (β = -11.89, SE = 3.56, t = -3.34, p = 0.001, model not shown), residual reading times (β = -12.04, SE = 3.52, t = -3.419, p = 0.001, model not shown), and log reading times (β = -0.03, SE = 0.01, t = -2.703, p = 0.008, model not shown) such that conditions containing R-expressions as the embedded subject were read longer than conditions containing pronouns in this position. We interpret this effect as an anticipatory effect: as all of our test sentences followed the same structure, participants could reliably predict that the R-expression subject would be found on the following word in these conditions.[65] This DP type effect is also found on word 15, i.e., on the embedded subject itself, in the raw reading times (β = -27.91, SE = 6.06, t = -4.607, $p < 0.001$, model not shown) and log reading times (β = -0.06, SE = 0.01, t = -3.989, $p < 0.001$, model not shown). This effect is reliable across transformations at word 16, i.e., one word following the embedded subject, indicating that R-expressions were read longer than pronouns when they were found as the embedded subject. We report the model for the log reading times for word 16 in Table 3.2. The effect disappears at word 17 across transformations. Similarly to the effect found on the first DP site, this suggests that any difficulty found on the R-expression embedded subject is easily resolved.

[65] A visual inspection of the plot suggests that the effect on word 14 is rather surprising. The effect is significant but it has a relatively small effect size (314 ms vs. 305 ms, a difference of 9 ms). It is thus unclear to us if the effect is reliable at word 14 and we concentrate our discussion on the effects found at later regions. It is also important to note that the plot shows the confidence intervals for each condition separately whereas the statistical model examines the contrast between the variables. As a result, the contrast will have its own confidence intervals which are not reflected in the plot.

Table 3.2: Final mixed-effects model for log reading times at word 16 (N = 2489 before trimming, 2416 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.811	0.035	164.3	0.000
R-expression	-0.097	0.012	-7.859	0.000

Random effects: by-participant intercept (SD = 0.327), by-item intercept (SD = 0.008), by-participant random slope for DP type (SD = 0.087), and the correlation between by-participant and slope (r = -0.55).

3.3.6.8 Verb site

In all of the critical sentences, the creation/non-creation verb occurred at word 17. The mean log reading times for the verbal region are plotted in Figure 3.7. As in Experiment 1, we fitted linear mixed effects models for the reading time data on the verb site as well as three words after, due to potential spillover effects. We found reliable effects at word 18, across all transformations. At word 18, the word following the verb, conditions with non-creation verbs were read significantly longer than conditions with creation verbs (raw reading times: $\beta = 9.865$, SE = 3.88, $t = 2.542$, $p = 0.01$, residual reading times: $\beta = 12.16$, SE = 3.92, $t = 3.091$, $p = 0.003$, models not shown). The model for the log reading times at word 18 is shown in Table 3.3. No other effects were significant at the 5% threshold. There were also no reliable effects of DP type nor were there any significant interactions between DP type and Verb type on these words. The results of this experiment differ from those observed in Experiment 1, where we did not find any significant effects based on the type of verb in the sentences, i.e., creation and non-creation verbs were read similarly. Our original hypothesis was that creation verbs should be more difficult to process because they force the parser to adopt a *de dicto* interpretation of *many-x*. We hypothesized that this reconstruction operation should incur a processing cost because it does not allow the parser to interpret phrases in the position in which they are encountered: if *many-x* is interpreted in its surface position, we will end up with a *de re* reading but this reading is not possible with creation verbs. The effect observed in Experiment 2 is in fact the opposite of the effect predicted, i.e., non-creation verbs were read longer than creation verbs. We will discuss possible reasons for this result in the General Discussion section.

Figure 3.7: Mean log reading times at the Verb site by Verb (creation, non-creation) and DP (R-expression, pronoun) type from words 17-20 in Experiment 2. Original log mean values in the left panel and fitted log mean values in the right panel. Error bars represent 95% confidence intervals.

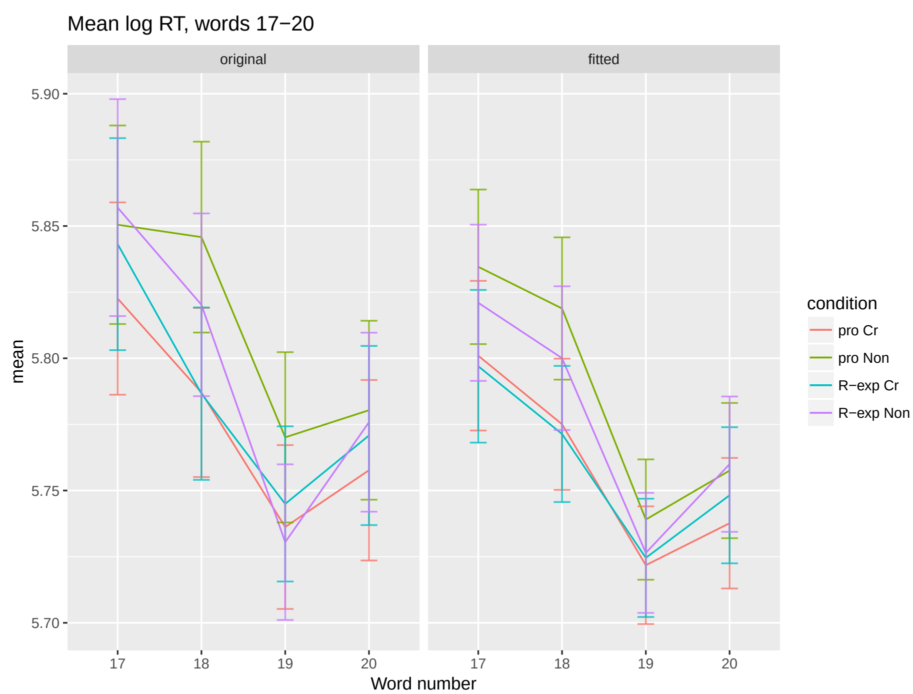


Table 3.3: Final mixed-effects model for log reading times at word 18 (N = 2499 before trimming, 2432 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.783	0.036	159.46	0.000
R-expression	-0.013	0.01	-1.319	0.187
Non-creation	0.03	0.011	2.725	0.008

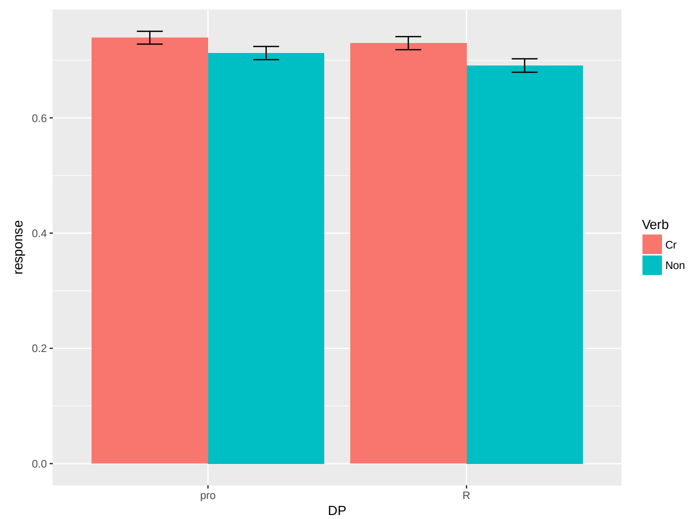
Random effects: by-participant intercept (SD = 0.325), by-item intercept (SD = 0.038), by-participant random slope for Verb type (SD = 0.051), and the correlation between by-participant and slope (r = 0.18).

3.3.6.9 Comprehension questions

As in the naturalness judgement task, we also asked comprehension questions targeting participants' interpretation of the sentences. We asked two types of questions which directly probed for participants' interpretation of the stimulus sentences: i) questions asking about co-indexation between the two DPs, and ii) questions about whether or not the object of the verb existed before the completion of the event. For examples see section 3.3.5.4.

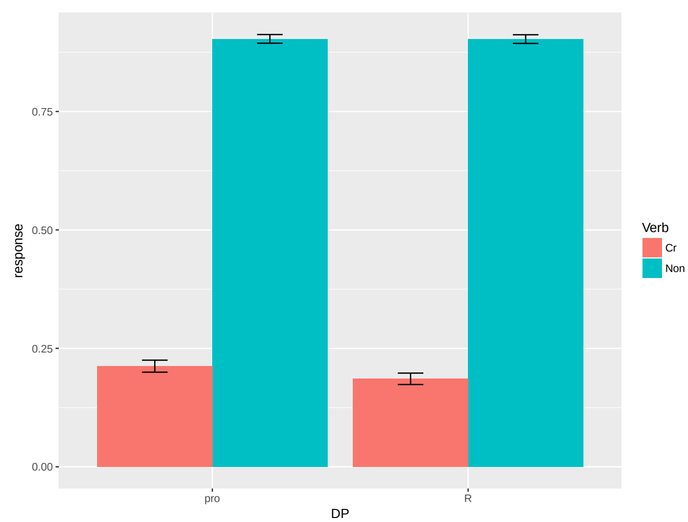
Co-indexation questions The results for the co-indexation questions were similar to those of the judgement task: participants judged the two DPs as referring to the same entity 71.8% of the time, across conditions. The mean percentage of *yes* responses to these types of CQs across conditions is plotted in Figure 3.8. These results suggests that participants were most often interpreting the pronoun and the R-expression as co-indexed, irrespective of their syntactic positions, i.e., the interpretation did not differ for anaphoric versus cataphoric pronouns. The differences between conditions did not reach significance at the 5% threshold (Verb type variable: β = -0.262, SE = 0.161, t = -1.63, p = 0.103, DP type variable: β = -0.204, SE = 0.205, t = -0.1, p = 0.318). Overall, these results suggest that the parser prefers to find an antecedent within the sentence for both cataphoric and anaphoric pronouns, as has been assumed in previous work (see e.g., Van Gompel and Liversedge, 2003).

Figure 3.8: Percentage of *yes* responses to the co-indexation CQs in Experiment 2 self-paced reading. The DP variable refers to the first DP mentioned in the sentence, i.e., the DP in the filler phrase. Error bars represent 95% CIs.



Questions about the creation/non-creation event The other relevant CQs asked about whether the object of the verb existed before the event, e.g., *Did the lies already exist?* The mean percentage of *yes* responses by condition is plotted in Figure 3.9. If the verb was a creation verb, participants answered *yes* to this question 19.9% of the time. In contrast, if the verb was a non-creation verb, they answered *yes* 90.3% of the time. The difference between creation and non-creation verbs was significant at the 5% threshold (β = 22.82, SE = 1.17, t = 19.55, $p < 0.001$). These results provide evidence that participants were interpreting the creation verbs as intended, i.e., they were sensitive to the fact that the object of the creation verb does not exist before the completion of the creation event and that the opposite is true for non-creation verbs. In light of these results, we cannot explain the lack of a processing cost with verbs of creation as resulting from participants not interpreting the creation verbs as intended. Based on the results of the comprehension questions, participants had no trouble differentiating between a creation and non-creation verb, at least with respect to whether or not the verb carries a presupposition of existence. We will discuss possible explanations for why we did not observe a processing effect with creation verbs in the General Discussion section of this chapter.

Figure 3.9: Percentage of *yes* responses to the creation verb CQs in Experiment 2 self-paced reading. The DP variable refers to the first DP mentioned in the sentence, i.e., the DP in the filler phrase. Error bars represent 95% CIs.



3.3.7 Experiment 2 General Discussion

In Experiment 2, we aimed to investigate how pronouns are processed in positions in which they linearly precede potential antecedents. In this position, pronouns do not (yet) have antecedents. We used similar stimuli as in Experiment 1, except that we replaced the anaphors with pronouns in the picture DP and we also manipulated whether the embedded subject was a pronoun or an R-expression. As in Experiment 1, we used two different verb types, creation and non-creation verbs, with the assumption that creation verbs force *many-x* phrases to reconstruct because they need to be interpreted as *de dicto*. The relevant contrast in this experiment was that unlike anaphors, pronouns do not necessarily need to reconstruct in order to be interpreted, i.e., they can be interpreted in positions in which they linearly precede potential antecedents. As a result, the parser may not have been given an early indication that reconstruction is required in these stimulus items. The parser can only determine that reconstruction of *many-x* is needed upon reaching the creation verb further to the right in the linear string. Moreover, reconstruction of the complement of *wh* embedded with a pronoun would yield binding violations at LF because the pronoun would be c-commanded by its antecedent at LF (Principle B).

3.3.7.1 Effect of verb type

We did find significant effects of verb type in Experiment 2: non-creation verbs were read longer than creation verbs. This effect is unexpected following the hypothesis that reconstruction is costly and that it is obligatory in creation verbs but not in non-creation verbs because *many-x* cannot be interpreted as *de re* when it is the object of a creation verb. Thus far, we have been assuming that obligatory reconstruction should incur a cost, as reflected by longer reading times. The reading time results suggest that the parser was either not sensitive to this reconstruction operation or that our methodology was not sensitive enough to detect this effect. This lack of effect cannot be explained by hypothesizing that participants were simply not sensitive to the difference between the two types of verbs. The results from the CQs directly probing for participants' interpretation of the event provide evidence that participants do make a distinction between the two different verb types. In this experiment, we specifically asked participants about the creation event. These questions were coded as correct, regardless of how participants answered the questions. Crucially, participants only presupposed the existence of the object of the verb of creation 20% of the time, compared to 90% of the time with non-creation verbs. There were no significant differences based on DP type, suggesting that participants were interpreting the sentences as *de re* for non-creation verbs but *de dicto* with creation verbs, irrespective of the type of DP found in the sentence.

Is reconstruction costly? Another possibility is that reconstruction itself is not costly but that needing to wait to adopt an interpretation is costly (see e.g., Frazier and Rayner, 1990). The interpretation of *many-x* in a *how many* question containing a non-creation verb is ambiguous. *Many-x* can either be interpreted as *de re* or *de dicto*. We have been assuming that the parser initially pursues the *de re* interpretation upon encountering *many-x* because it can assign the phrase an interpretation as soon as it is encountered and it is simpler to interpret phrases as soon as they are encountered. However, it is also possible that when the parser is presented with an ambiguous string, it does not commit to an interpretation of the sentence until it receives further evidence about which interpretation should be adopted. In the current case, the parser may not interpret the sentence until it reaches the verb site. Upon reaching the verb site, the parser will need to decide if *many-x* should be interpreted as *de re* or *de dicto*. With the creation verbs, no decision needs to be made because creation verbs are unambiguous and only the *de dicto* reading is available. Under such an analysis, creation verbs would be less costly than creation verbs since there is only one interpretation of the sentence. In contrast, if the verb is non-creation, the parser must make a decision about which interpretation it should adopt and needing to make this decision is costly. Unfortunately, due to the constraints of our experiment, we are unable to determine if this hypothesis is on the right track.

Different types of CQs Recall that in Experiment 1, we did not find any effects at the verb site, i.e., creation and non-creation verbs were read similarly. Why did we find effects of verb type in Experiment 2, despite using the same verbs in both experiments? Previous work on pronominal ambiguity resolution (Stewart et al., 2007) and syntactic attachment ambiguity (Swets et al., 2008) suggests that the types of comprehension questions that participants are asked in an experiment can actually have an effect on the level of processing in which they engage. The experiments conducted by Stewart et al. (2007) are particularly relevant to the current experiments because the researchers investigated how participants interpret ambiguous pronouns that could be associated with two previously mentioned antecedents. The researchers were interested in whether the level of processing (deep versus shallow) affected reading times on these ambiguous pronouns. The level of processing was manipulated using comprehension questions that either forced participants to adopt a deep reading of the sentence or allowed them to only pursue a shallow understanding of the sentence and still answer the questions correctly. The assumption was that the comprehension questions could guide participants towards a particular level of processing. Participants could engage in shallow processing if the comprehension questions did not require participants to interpret the ambiguous pronoun whereas participants needed to engage in deep processing if the comprehension questions required that participants determine the referent of the ambiguous pronoun. More precisely, shallow processing was measured with "superficial" comprehension questions that did not require participants to associate an ambiguous pronoun with a referent. Deep processing was measured using comprehension questions that always asked about the referent of the pronoun and therefore forced participants to interpret the pronoun.

Stewart et al. (2007) found that participants in the deep processing condition, i.e., those who were only asked comprehension questions about the association between the pronoun and its antecedent, read the ambiguous pronoun significantly longer than participants in the shallow processing condition, i.e., those who were only asked superficial questions and were not required to find a referent for the pronoun. Stewart et al. (2007)'s results show that the types of comprehension questions that participants are asked can affect how they process sentences in an experiment. The comprehension questions in our Experiment 1 could be viewed as "superficial" because they did not ask participants about their interpretation of the sentence. For example, we asked questions of the type *Was it the reporters who wondered how many lies about Sean Alexa was being asked to deny?* To accurately answer this question, participants only needed to remember the stimulus sentence but they did not need to actually interpret any relevant parts of the sentence. However, in Experiment 2, we directly asked participants about the creation event, e.g., *Did the books already exist?* These types of questions might have drawn their attention to the semantics of the two verb types, which might explain why we found different effects on the verb site in Experiments 1 and 2. An important question is why the non-creation verbs were more difficult than the creation verbs, which is the opposite of what we predicted. One possibility is that participants found the non-creation verbs to be less natural in our sentence templates compared to the creation verbs. Since all of our sentences involved picture DPs, which are all semantically plausible with verbs of creation, participants may have been expecting a verb of creation in the embedded clause. When they instead find a non-creation verb, this surprises them and processing difficulty is incurred.

3.3.7.2 Effect of DP type

We can make clearer conclusions about the effects observed on the DP sites. We found significant effects at both DP sites in Experiment 2.

DP site 1 When the R-expression/pronoun was first introduced in the complement of *wh* phrase, we observed longer reading times on the R-expression compared to the cataphoric pronoun. We attribute this effect to the fact that the R-expression has more discourse properties compared to the pronoun (Gordon and Hendrick, 1998). More discourse properties lead to a greater processing cost when the R-expression is initially read. One might ask why we did not also find this effect in Experiment 1, where we also investigated the processing of R-expressions. We propose that we did not observe effects of the R-expression in Experiment 1 because anaphors are also difficult to process. A relevant question is whether the two types of DPs are equally costly or whether one of the DPs is more costly than the other. If anaphors and R-expressions are equally costly, we would not expect to observe a difference in reading times on the DPs. However, if one of the DPs is more costly than the other, we expect that the more costly DP should incur longer reading times. In Experiment 1, we did not observe a difference in reading times when the DP was initially introduced, however, we found longer reading times in anaphor conditions two and three words following the DP. This result suggests that anaphors are more costly than R-expressions. In Experiment 1, we attributed this effect to the fact that upon finding an anaphor further to the left in the linear string, the parser must locate its antecedent further to the right. This requirement makes anaphors more costly than R-expressions. Combining the results from Experiments 1 and 2, it seems as though pronouns are the least costly, followed by R-expressions followed by anaphors. We specifically investigate the processing cost of the three types of DPs in Experiment 7, reported in Chapter 6.

Unlike in Experiment 1, in Experiment 2, we did not observe any prolonged effects after the initial DP was introduced. This suggests that any initial difficulty with R-expressions is resolved easily. Since we did not find prolonged processing difficulty with pronouns that linearly precede their antecedents, our results suggest that the parser makes a distinction between pronouns and anaphors appearing further to the left than their antecedents. More precisely, the parser seems to be sensitive to the PF-LF mismatch with respect to interpretation in the case of anaphors but that no such mismatch (necessarily) exists in the pronoun conditions. Recall that pronouns, like R-expressions, can be interpreted in the higher copy (surface) position and do not need to reconstruct. It is also possible that we might only observe an effect of reconstruction for the pronoun condition when the embedded subject is reached. Reconstruction might only become an option for the parser once it determines that there is an appropriate DP antecedent. Since we did not observe the same effects with pronouns as we did with anaphors, we argue that the results from Experiment 2 are compatible with the *structural* analysis proposed in the Discussion section of Chapter 2.

DP site 2 We also found effects at the embedded subject position in Experiment 2, such that conditions containing an R-expression in this position, i.e., conditions which had a pronoun in the complement of *wh*, were read longer than conditions containing a pronoun, i.e., conditions which had an R-expression in the complement of *wh*. Therefore, in both syntactic positions, R-expressions incurred greater processing difficulty compared to their pronominal counterparts.[66,67] Note, however, that if the second DP region contains an R-expression, this was the cataphoric condition.

We argue that this effect can be explained by a structural account of reference resolution which follows from the grammar's binding over co-reference preference. Upon finding a cataphoric pronoun in the complement of *wh*, the parser has two options. It has the option of interpreting the pronoun in its surface position. Since the meaning of a pronoun does not necessarily depend on an antecedent, the pronoun could be assigned a semantic denotation in the higher copy position. This would mean that the antecedent for the pronoun is found outside the sentence. This analysis would abide by the parser's preference to interpret phrases as soon as possible but it would not abide by the preference that binding relations are preferred over co-reference relations. We turn to this preference next.

The second option is that the parser waits until it has received further evidence to determine if a binding relation between the pronoun and a later DP would be possible. Upon encountering the cataphoric pronoun in the *wh*-phrase, the parser will not know if there will be a feature-matched DP further to the right that could serve as an antecedent for the pronoun. It is possible that the parser does not even consider the possibility of a binding

[66] Note that this effect cannot simply be attributed to the fact that R-expressions contained more characters than pronouns. The effect was consistently found across data transformations, including the residual reading times which take into consideration the number of a characters in each word. The effect was also consistently found on the word following the DP which contained the same number of characters across conditions.

[67] Note that the effects found at the DP regions in the current experiment replicate Kennison et al. (2009)'s effects in their gender-matched conditions. Kennison et al. (2009) investigated how cataphoric and anaphoric dependencies are processed and their results suggest that the parser uses different strategies for cataphors and anaphors. More precisely, processing difficulty for cataphoric processing was incurred when the antecedent is found later in the sentence, i.e., when a dependency must be created between the previously introduced anaphor and its DP antecedent. No such processing difficulty was observed in anaphoric processing, suggesting that it is easier to retrieve a previously mentioned full DP from memory and link it to a pronoun, compared to retrieving a pronoun from memory when a DP antecedent is found. Thus, cataphoric processing incurs greater processing difficulty compared to anaphoric processing. We do not think that this explanation is sufficient to explain our results because it is unclear why retrieving a pronoun from memory would incur a greater processing cost once the antecedent was found. One might actually predict the opposite, i.e., that it would be more difficult to retrieve an R-expression when a pronoun is found further to the right.

relation between the pronoun and another DP until it has determined that such a DP exists in the string of words. In our stimulus sentences, the possibility of binding only presents itself to the parser when the embedded subject is found further to the right in the string. Upon reaching the embedded subject, the parser determines that the pronoun can enter a binding relation with that DP. In order for two DPs to enter a binding relation, the antecedent must c-command the pronoun. Thus, the parser reconstructs the phrase containing the pronoun to a lower copy position that is structurally below the embedded subject. However, this parse would incur a Principle B violation since the antecedent would locally c-command the pronoun. As a result, the parser must reanalyze its parse and adopt a co-reference interpretation.[68] Co-reference is the less preferred analysis and thus incurs a processing cost.

Another possibility is that a greater processing cost is incurred on the embedded subject because the parser now needs to consider whether it must reconstruct the phrase containing the pronoun below the embedded subject. In other words, we do not find earlier processing difficulty because the parser assumes that the pronoun is interpretable in its surface position. Upon finding a feature-matched DP further to the right, the parser is presented with a different possibility, i.e., that this DP might serve as an antecedent for the pronoun. Thus, the slowdown at the embedded subject may not reflect a Principle B violation but it might just reflect the fact that the parser must now consider an alternative parse. This explanation might explain why some speakers may not seem sensitive to the predicted Binding Principle B violation, even in the conditions with creation verbs where reconstruction is predicted to be obligatory. Unfortunately, our experiment was not designed in a way that we can

[68]Note that it would also be possible for the pronoun to receive its interpretation from the context but our CQ results indicate that participants prefer to interpret the pronoun and R-expression as referring to the same real world referent.

distinguish between these two possibilities.[69]

Is there a general cost associated with R-expressions? One possible objection to the interpretation that the results at the second DP region can be explained by the binding over co-reference preference is that our results from both regions show that R-expressions incurred a greater processing cost compared to pronouns. Since it seems as though R-expressions are overall more costly, can we use the same explanation of the results for both syntactic positions? For example, we argued that the effect found on the first DP region could be attributed to the fact that R-expressions carry more discourse properties compared to pronouns. Could this interpretation not also explain the effects at the second DP region? If this analysis was on the right track, we might expect the effect sizes to be the same in both DP regions.[70] In other words, if the results observed on both regions reflect the same processing cost, they should show similar effect sizes. In contrast, if the results reflect different processing costs, we might expect the effect sizes to be different. The effect size was much larger in the second DP region compared to the first DP region (0.051 log ms, 18.25 raw ms vs. 0.097 log ms, 30.84 raw ms). The differences in effect sizes suggest that the results found on these two regions reflect different processing costs. We might interpret the larger effect size on the second DP region as demonstrating that this region is more costly than the first DP region. To further substantiate our claim, we ran an analysis crossing DP type (pronoun, R-expression) by syntactic position in the sentence (early, i.e.,

[69] One way to distinguish between these two alternative explanations would be to conduct a follow-up study investigating cases where there is a clear Principle B violation, under a reconstruction analysis. If the binding $>$ co-reference analysis is on the right track, we would predict that cases in which Principle B is violated should be more costly than cases in which it is not. In contrast, if the alternative analysis is better able to explain the effects, we would not necessarily predict a cost of Binding Principle B since the parser would not attempt to reconstruct the phrase if such an analysis would lead to a Principle B violation. This is an interesting question but it is beyond the scope of the current work. We leave it to future work to investigate this question further.

[70] Thanks to Martin Hackl for suggesting this.

word 9, or late, i.e., word 16). The DP type variable was coded based on which lexical item participants saw in that particular syntactic environment, i.e., if they read a pronoun on word 9, it was coded as such. We expected to observe an interaction between these two factors if the DP type effects found on the two regions could be explained by different processes. More precisely, if the syntactic position in which the DP is found matters, we should find that the DP type effect (R-expressions are more costly than pronouns) differs based on the syntactic position of the DP. If, however, our results can be simply explained as a lexical effect, i.e., R-expressions are more costly than pronouns, irrespective of their syntactic position, no such interaction should be found. Our predictions were borne out and we found a significant interaction between DP type and the position of the DP in the sentence, as shown in Table 3.4.

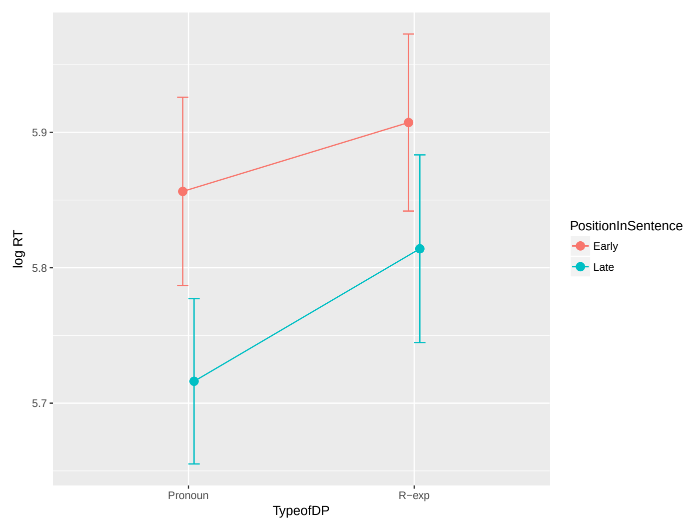
Table 3.4: Final mixed-effects model for log reading times with an interaction between DP type (pronoun, R-expression) and syntactic position in the sentence (early (word 9) or late (word 16)) (N = 4999 before trimming, 4854 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.856	0.036	165.11	0.000
R-expression	0.051	0.012	4.413	0.000
Late syntactic position (word 16)	-0.014	0.014	-10.354	0.000
R-expression * Late position	0.047	0.016	2.988	0.004

Random effects: by-participant intercept (SD = 0.326), by-item intercept (SD = 0.01), by-participant random slope for DP type (SD = 0.063), by-participant random slope for position in sentence (SD = 0.092), by-participant random slope for DP type by position in sentence interaction (SD = 0.08), the correlation between by-participant intercept and slope for DP type (r = -0.41), the correlation between by-participant intercept and slope for position in sentence (r = -0.56), and the correlation between by-participant intercept and slope for the interaction (r = 0.71).

As shown in Figure 3.10, the interaction shows that the effect of DP type, i.e., that R-expressions incurred a greater cost compared to pronouns, was stronger in the late syntactic position (word 16, one word following the embedded subject) compared to the early syntactic position (word 9, one word following the DP itself). This result suggests that the two effects could be explained by two separate mechanisms. On the first DP region, we argued that the effect arises due to the discourse properties associated with R-expressions. This result could therefore be viewed as a lexical effect. In contrast, we argued that the effect observed on the second DP region arises because the parser has to reanalyze its parse. Reanalysis of the parse is predicted to be more costly than a lexical effect brought in by a single lexical item. This prediction is borne out.

Figure 3.10: Interaction between DP type (pronoun, R-expression) and syntactic position of the DP in the sentence (early, i.e., word 9, or late, i.e., word 16). The DP variable refers to the type of DP that participants saw in that particular syntactic environment. Error bars represent 95% CIs.



3.3.7.3 Experiment 2 takeaway

The takeaway from our Experiment 2 results is that we did not replicate the Experiment 1 results when pronouns were embedded in *wh*-fillers, even though the parser still preferred to find an antecedent for the pronoun within the sentence (as measured by our comprehension questions directly targeting participants' interpretation of the DPs). Pronouns and anaphors are subject to different structural constraints (Binding Principles A and B). Since we found processing difficulty on words following anaphors but not on words following pronouns, we argue that our results are compatible with a structural explanation of binding at LF. If the complement of *wh* contains an anaphor, it must reconstruct below the antecedent at LF and this is reflected by a processing cost in anaphor conditions. In other words, the phrase containing the anaphor cannot be interpreted in its PF position and must be interpreted in a distinct position at LF (PF-LF mismatch). Pronouns do not incur this processing cost because they do not need to reconstruct to find an antecedent (PF-LF match). Our results therefore suggest that the parser is sensitive to a PF-LF mismatch in the domain of referential dependencies.

3.4 Experiment 3

Experiment 3 directly investigated whether our results on the DP regions in Experiment 2 could be explained as a lexical effect. It is important to note that in Experiment 2, all of the R-expressions used in the stimulus items were new referents that had not been previously introduced to the participants. Since the R-expressions were unfamiliar to participants, this might explain why they incurred greater processing difficulty compared to the pronouns (for formal semantic work on *familiarity*, see e.g., Heim, 1982, among many others). Even though the pronouns did not have an antecedent when they were introduced in the *wh*-phrase, it might have been easier for the parser to accommodate the pronoun compared to accommodating an unfamiliar R-expression. By "accommodate," we mean that the parser is able to assign the pronoun an unspecified denotation. Under such an analysis, it is sufficient that the a pronoun like *her* be interpreted as *a female referent who this sentence is about.* This would mean that the parser does not need to fully interpret the pronoun and it can receive an underspecified representation (see Ferreira et al., 2002's good enough work for results showing that participants do not always fully specify syntactic representations). One way to test for whether or not R-expressions incur greater processing difficulty because they are unfamiliar is to manipulate the familiarity of the R-expressions used in the experiment. If our Experiment 2 results can be simply attributed to the R-expressions being unfamiliar to participants and thus incurring greater processing difficulty compared to the pronouns, we expect the processing cost to be alleviated when familiar R-expressions are used in the stimulus items. In contrast, if our previous results can be explained as an effect of R-expressions generally carrying more discourse properties in the first DP position and reflecting the parser's preference for binding over co-reference in the second DP position, we should find similar effects when familiar R-expressions are used in the stimuli. The goal of Experiment 3 was to test these hypotheses.

3.4.1 Methods

In Experiment 3, we used the same stimuli as in Experiment 2, except that we modified whether the R-expression was familiar or unfamiliar. Familiarity was manipulated using the names of famous people. Since the primary goal of this experiment was to investigate whether familiar names show a different effect than unfamiliar names and whether familiar names behave like pronouns, we did not include the verb manipulation. Sample stimuli are shown in (11). As in Experiment 2, we manipulated whether a pronoun or R-expression appeared in the *wh*-phrase or as the embedded subject. R-expressions were either familiar, as in (11-a)-(11-c), or unfamiliar, as in (11-b)-(11-d).

(11) The reporters wondered ...

a. how many lies discrediting him$_i$ as a witness you had asked Charlie Sheen$_i$ to deny

b. how many lies discrediting him$_i$ as a witness you had asked Kevin Holmes$_i$ to deny

c. how many lies discrediting Charlie Sheen$_i$ as a witness you had asked him$_i$ to deny

d. how many lies discrediting Kevin Holmes$_i$ as a witness you had asked him$_i$ to deny

... for the live television interview.

Since the goal of this experiment was to investigate whether our previous reading time results could be explained by the fact that the R-expressions were unfamiliar to participants, we did not run a naturalness judgement task in Experiment 3.

3.4.1.1 Specific predictions

DP 1 region The first region of interest in this experiment was the first DP region, which occurred in the complement of *wh*. The DP was either a pronoun, as in (11-a)-(11-b), or an R-expression, as in (11-c)-(11-d). In this experiment, we manipulated whether the R-expression was "familiar," using the names of famous people, as in (11-c), or "unfamiliar," using names that were matched based on the number of syllables to the names of the famous people, as in (11-d). We predicted that if our results at this DP region in Experiment 2 could be simply explained as an effect of R-expressions being unfamiliar to participants, then the effect should be alleviated when the stimulus items contain the names of famous people. In other words, we should only see a processing cost with R-expressions in (11-d) but not in (11-c). If the effect is solely attributable to familiarity, familiar R-expressions should pattern with pronouns and unfamiliar R-expressions should be different from both pronouns and familiar R-expressions. In contrast, if our previous results can be explained as resulting from R-expressions carrying more discourse properties than pronouns when they are initially introduced, we should still see a processing cost when the R-expression is familiar to participants. It is possible that there should be less of an effect when the R-expression is the name of a famous person because in such cases, the R-expression would be much more easily accommodated compared to an R-expression that is completely unfamiliar to participants.

DP 2 region The second region of interest was the DP occurring as the subject of the embedded clause. As in Experiment 2, this DP was either a pronoun, as in (11-c)-(1-d), or a pronoun, as in (11-a)-(11-b). We manipulated whether the R-expression was "familiar," as in (11-a), or "unfamiliar," as in (11-b). As in Experiment 2, if the DP in the *many-x* phrase was a pronoun, the embedded subject DP was an R-expression, and vice versa. The two DPs always matched in gender. We predicted that if our previous results at this region in Experiment 2 could be simply attributed to the fact that we used names that were unfamiliar to participants, the effect should be alleviated when familiar names are used in the stimulus items. More precisely, the processing cost should be alleviated in (11-a) but not in (11-b). In contrast, if our results at this region in Experiment 2 can be explained as resulting from a reanalysis effect because the parser cannot pursue a binding relation between the pronoun and the R-expression, we should replicate this effect when the embedded subject is an R-expression, irrespective of whether it is familiar or not to participants. This effect should not be dependent on whether or not participants can accommodate the R-expression.

3.4.1.2 Participants

Eighty-one participants were tested for the self-paced reading experiment. One participant was removed from the analysis because they scored less than 60% on the comprehension questions throughout the whole experiment[71] and four participants were removed because they scored less than 70% on the comprehension questions asked for the critical trials.[72] In the results section, the data for the remaining 76 participants (42 males, 34 females) will be presented. All participants were self-reported native speakers of English, aged between 18 and 71 years of age (*M* = 35.57 years, *SD* = 11.4). Participants were paid $6.00 USD for their participation and were naïve as to the purposes of the experiment. Participants who completed Experiment 3 did not complete any of our previous experiments.

3.4.1.3 Stimuli

Thirty-two sets of sentences were created for Experiment 3. The sentences followed the same paradigm used in Experiment 2, except that we did not manipulate the type of verb used in the sentence. The relevant manipulation in this experiment was whether the R-expression was a familiar name or a made-up name. To control for the familiarity of the names, we used the first and last names of famous people, e.g., *Taylor Swift, Johnny Depp, Barbara Walters*, etc. We matched the names of famous people to made-up names that had the same number of syllables in both the first and last name. We also added a familiarity task at the end of the experiment to ensure that the names coded as familiar were indeed known to our participants and that they did not identify the made-up names as being familiar to them. The familiarity task will be explained in more detail below but overall,

[71] This lower comprehension score cut off was used because using a higher cut off eliminated a much higher number of participants and there were several trials without a correct answer.

[72] Note that these scores were calculated after the item trimming (explained below).

we found that the familiarity manipulation accurately reflected participants' impressions of the names. In total, we used four conditions, manipulating the familiarity of the names as well as the syntactic positions of the R-expression and the pronoun. As in Experiment 2, the pronoun either preceded or followed the R-expression. Similarly to Experiment 2, the pronoun and R-expression always matched in gender and there was an equal number of male and female trials across the experiment. Thirty-two filler trials were also used in this experiment. The fillers were constructed to have a similar syntactic structure to the critical items. Half of the filler trials also contained familiar names, i.e., the names of famous people.

3.4.1.4 Procedure

The procedure was the same as in the self-paced reading studies of Experiments 1 and 2 except that participants were also asked to complete a familiarity judgement task (explained below) after completing the self-paced reading experiment.

3.4.2 Results

3.4.2.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 7 fillers and 2 test trials were removed. We also removed any reading times that were less than 90 ms and over 3000 ms in length, as these reading times would be considered outliers. When possible, in order to accurately investigate the reading times on trials where participants were interpreting the sentence as intended, trials where the participant answered the question incorrectly were also removed from the analysis.[73] Accuracy on the comprehension questions was high at 83.4% across all trials and 88.8% in the critical trials, after trimming.

As reading times allow for a great deal of variation, we transformed the data in four ways: raw reading times, residual reading times, log reading times, and residuals of the log reading times. For all transformations, data points that were +/- 2 SDs away from the mean by participants were removed (less than 4.5% of data across all transformations). The statistical models were run on all of these data transformations, and revealed similar effects, unless otherwise indicated.

3.4.2.2 DP site 1

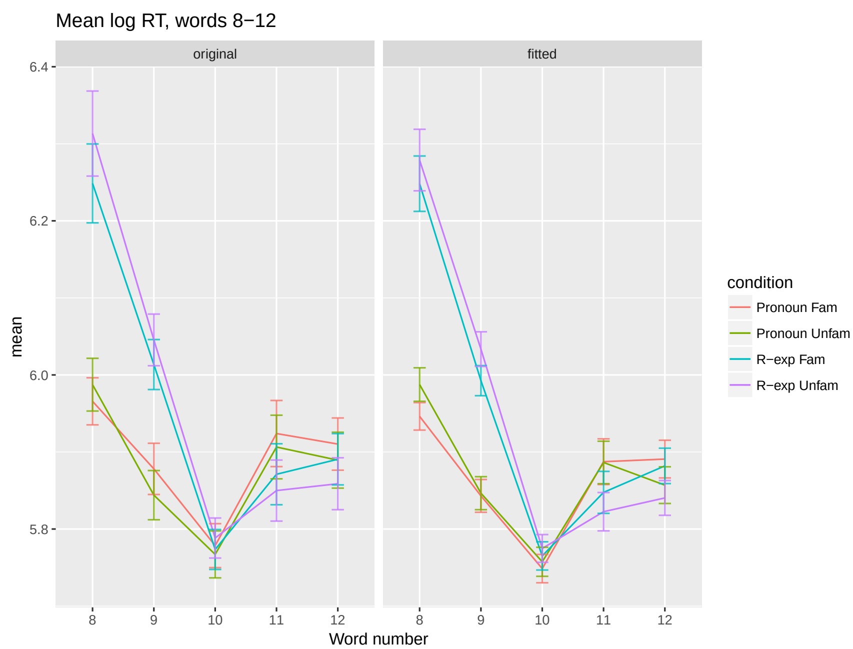
We fitted linear mixed effects regression models to investigate the effect of the first-introduced DP, which occurred within the complement of *wh*. The type of DP (pronoun, R-expression) and the familiarity variable (familiar, unfamiliar) were added as a random slope by participant whenever the model converged.[74] The first-introduced pronoun or R-expression was found at word 8 across items. We fitted models on words 8-12 to investigate how the first-introduced DP is initially processed. The mean log reading times by DP Type and Familiarity are plotted in Figure 3.11.

At word 8, there was a main effect of both DP type and familiarity. The main effect of DP type revealed that items containing R-expressions in the complement of *wh* were read significantly longer than items containing a pronoun in this position (raw reading times: β = 189.79, SE = 26.7, t = 7.11, $p < 0.001$, log reading times: β = 0.316, SE = 0.02, t

[73] As in the previous experiments, all statistical analyses were also run on a data set that included all items. These analyses revealed the same effects. We report here the more conservative data set.

[74] DP was also added as a random slope by item but the model either did not converge or the effect did not improve model fit.

Figure 3.11: Mean log reading times by DP type and Familiarity at the first DP site (in the complement of *wh*) from words 8-12 in Experiment 3. Original log mean values in the left panel and fitted log mean values in the right panel. Error bars represent 95% confidence intervals.



= 8.83, $p < 0.001$, models not shown). However, this effect did not reach significance in the residual reading times, which take into account the number of characters in the word. Consequently, we do not deem this effect to be reliable at this word position as it seems to be driven by word length. The main effect of the familiarity variable revealed that unfamiliar names were read longer than familiar names when they were first introduced (raw reading times: β = 23.85, SE = 8.57, t = 2.78, $p < 0.01$, residual reading times: β = 24.88, SE = 8.78, t = 2.83, $p < 0.01$, log reading times: β = 0.04, SE = 0.02, t = 2.46, p = 0.01, models not shown). This suggests that the parser was sensitive to the familiarity variable, as intended.[75] On word 9, one word following the DP, R-expression conditions were read

[75]Note that the Familiarity variable collapses the results for pronouns and R-expressions. Only R-expressions can be coded as familiar or unfamiliar. As a result, we proceed cautiously with our interpretation

significantly longer than pronoun conditions (raw reading times: β = 61.71, SE = 6.52, t = 9.46, $p < 0.001$, residual reading times: β = 61.53, SE = 4.48, t = 13.75, $p < 0.001$, models not shown). The model for the log reading times on word 9 is shown in Table 3.9. Since this effect reached significance across transformations, we deem it to be a reliable effect. Note that this effect replicates the results found in Experiment 2 at this syntactic position. In Experiment 2, we argued that R-expressions are more costly when they are introduced because they carry a greater number of discourse properties.

Table 3.5: Final mixed-effects model for log reading times at word 9, one word following the pronoun or R-expression (N = 2105 before trimming, 2061 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.839	0.301	194.31	0.000
R-expression	0.168	0.017	9.9	0.000
Unfamiliar	0.011	0.013	0.84	0.402

Random effects: by-participant intercept (SD = 0.249), by-item intercept (SD = 0), by-participant random slope for DP type (SD = 0.112), and the correlation between by-participant intercept and slope (r = -0.27), by-participant random slope for Familiarity (SD = 0.055, the correlation between by-participant intercept and slope (r = 0.16).

We found no reliable effects on word 10, two words following the pronoun/R-expression. On word 11, we found that conditions containing pronouns were read longer than conditions containing R-expressions (raw reading times: β = -24.47, SE = 8.67, t = -2.82, $p < 0.01$, residual reading times: β = -23.01, SE = 8.45, t = -2.72, $p < 0.01$, models not shown). The model for the log reading times on word 11 is shown in Table 3.6. Note that this effect differs from the results found in Experiment 2 where we did not find that conditions containing pronouns were read longer than those containing R-expressions. We will discuss our interpretation of this finding in the General Discussion section. We found

of this result at this word position.

no reliable effects on word 12.

Table 3.6: Final mixed-effects model for log reading times at word 11, three words following the pronoun or R-expression (N = 2103 before trimming, 2052 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.894	0.043	136.85	0.000
R-expression	-0.048	0.018	-2.648	0.01
Unfamiliar	-0.011	0.016	-0.687	0.495

Random effects: by-participant intercept (SD = 0.353), by-item intercept (SD = 0.05), by-participant random slope for DP type (SD = 0.111), and the correlation between by-participant intercept and slope (r = -0.43), by-participant random slope for Familiarity (SD = 0.09, the correlation between by-participant intercept and slope (r = -0.44).

We also found significant interactions between the type of DP presented in the complement of *wh* and the familiarity variable. Note that these interactions are only interpretable if they reveal an effect of familiarity in R-expression conditions. Pronouns cannot be categorized as familiar or unfamiliar. The interaction between DP type and familiarity was significant at word 8 (raw reading times: β = 50.67, SE = 24.40, t = 2.08, p = 0.04, residual reading times: β = 39.1, SE = 15.79, t = 2.48, p = 0.01, log reading times: β = 0.062, SE = 0.031, t = 1.98, p = 0.05, models not shown), at word 9 (raw reading times: β = 36.45, SE = 10.46, t = 3.49, $p < 0.001$, residual reading times: β = 32.67, SE = 10.32, t = 3.17, $p < 0.01$, models not shown), and at word 10 (raw reading times: β = 16.7, SE = 6.35, t = 2.16, p = 0.001, residual reading times: β = 16.06, SE = 6.12, t = 2.6, $p < 0.01$, models not shown). The models for the log reading times on words 9-10 are shown in Tables 3.7 and 3.8. The interaction plots for words 9-10 are shown in Figures 3.12 and 3.13. The interaction suggests that while R-expressions incurred a greater processing cost compared to pronouns, the effect was stronger when the R-expression was unfamiliar, compared to when it was familiar.

Table 3.7: Final mixed-effects model for log reading times at word 9, one word following the pronoun or R-expression (N = 2105 before trimming, 2059 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.859	0.032	182.49	0.000
R-expression	0.133	0.023	5.864	0.000
Unfamiliar	-0.033	0.017	-1.976	0.052
R-expression*Unfamiliar	0.08	0.026	3.032	0.003

Random effects: by-participant intercept (SD = 0.263), by-item intercept (SD = 0.15), by-participant random slope for DP type (SD = 0.05), and the correlation between by-participant intercept and slope for DP type (r = -0.43), by-participant random slope for Familiarity (SD = 0.05), the correlation between by-participant intercept and slope for Familiarity (r = -0.36), by-participant random slope for the interaction between DP type and Familiarity (SD = 0.127), the correlation between by-participant intercept and slope for DP type * Familiarity (r = 0.47).

To investigate this effect further, we specifically looked at the data comparing the pronoun conditions, unfamiliar R-expressions, and familiar R-expressions. If our effects from Experiment 2 can be explained as a lexical effect alone, familiar R-expressions should pattern with pronouns. This was not borne out. Familiar R-expressions were significantly different from both pronouns and unfamiliar R-expressions at word 9. More precisely, pronouns were read significantly faster than familiar R-expressions but unfamiliar R-expressions were read significantly longer than familiar R-expressions. The model for the log reading times on word 9 is shown in 3.9. This effect was not significant on word 10, suggesting that any initial processing difficult with unfamiliar R-expressions is easily resolved by the parser.

Overall, the effects of DP type and Familiarity on the first DP region (complement of *wh*) suggest that the processing cost associated with R-expressions when they are first introduced is reduced when the R-expressions are familiar to participants. This result is in

Table 3.8: Final mixed-effects model for log reading times at word 10, one word following the pronoun or R-expression (N = 2107 before trimming, 2057 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.768	0.028	208.82	0.000
R-expression	-0.005	0.016	-0.293	0.770
Unfamiliar	-0.026	0.014	-1.799	0.076
R-expression*Unfamiliar	0.043	0.019	2.239	0.027

Random effects: by-participant intercept (SD = 0.227), by-item intercept (SD = 0.008), by-participant random slope for DP type (SD = 0.08), and the correlation between by-participant intercept and slope for DP type (r = -0.26), by-participant random slope for Familiarity (SD = 0.06), the correlation between by-participant intercept and slope for Familiarity (r = 0.07), by-participant random slope for the interaction between DP type and Familiarity (SD = 0.06), the correlation between by-participant intercept and slope for DP type * Familiarity (r = -0.33).

line with the hypothesis that more familiar names are easier to accommodate and thus that the cost associated with R-expressions carrying more discourse properties than pronouns is ameliorated when the R-expression is a familiar name. However, we still observed a difference between familiar R-expressions and pronouns, suggesting that familiarity does not eliminate the effect altogether. This result is not compatible with a solely lexical analysis of our previous effects.

Figure 3.12: Interaction plot showing the effect of DP Type on Familiarity on word 9 for log reading times in Experiment 3. Error bars represent 95% CIs.

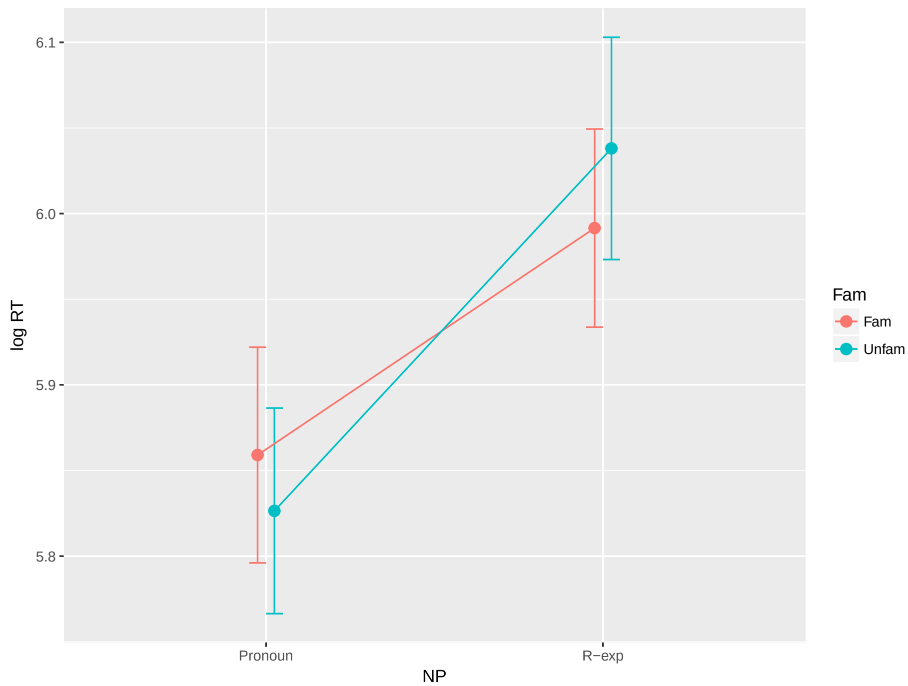
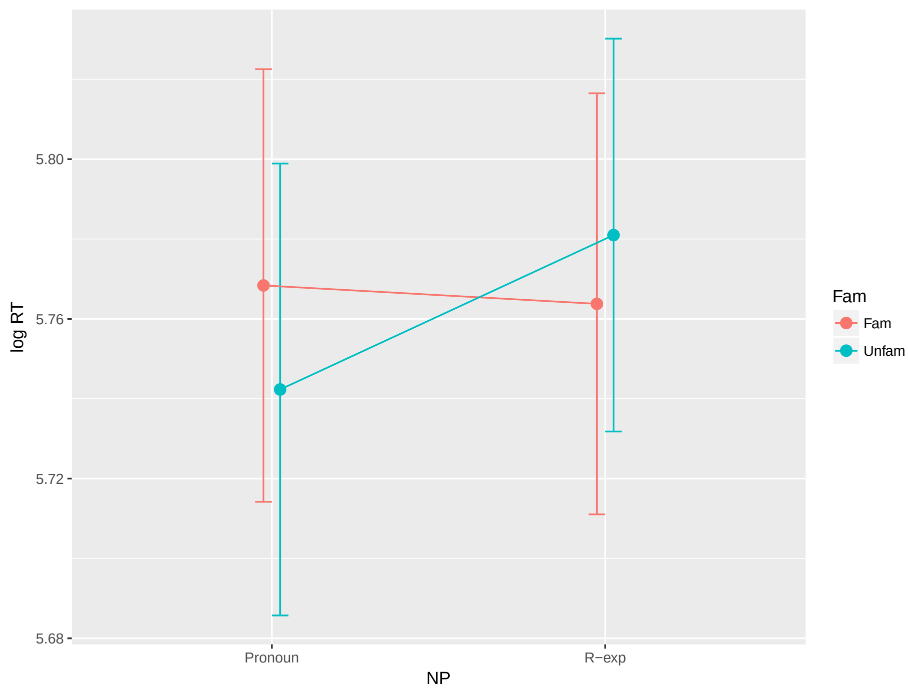


Table 3.9: Final mixed-effects model for log reading times at word 9, one word following the pronoun or R-expression (N = 2105 before trimming, 2062 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.991	0.029	203.17	0.000
Pronoun	-0.147	0.020	-7.339	0.000
Unfamiliar	0.046	0.019	2.401	0.002

Random effects: by-participant intercept (SD = 0.239), by-item intercept (SD = 0), by-participant random slope for Pronoun (SD = 0.13), and the correlation between by-participant intercept and slope for Pronoun (r = -0.15), by-participant random slope for Unfamiliar R-expressions (SD = 0.1), the correlation between by-participant intercept and slope for Unfamiliar R-expressions (r = 0.14).

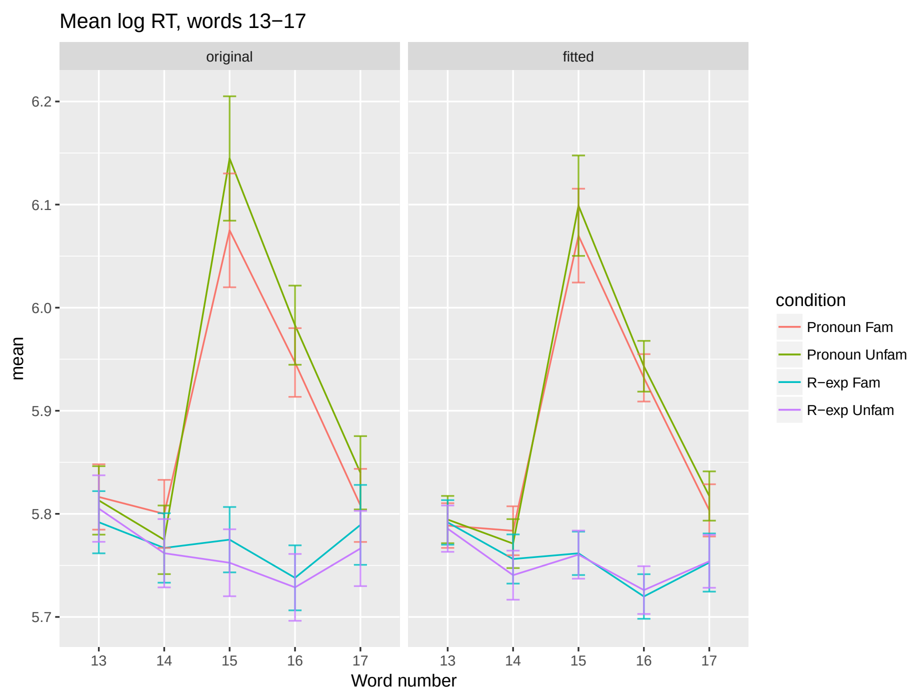
Figure 3.13: Interaction plot showing the effect of DP Type on Familiarity on word 10 for log reading times in Experiment 3. Error bars represent 95% CIs.



3.4.2.3 DP site 2

As in Experiment 2, the subject of the embedded clause was either a pronoun or an R-expression and always occurred at word 15 in all critical items. In Experiment 3, we also manipulated whether the R-expression was a familiar or unfamiliar name in this word position. The mean log reading times by DP type and Familiarity are plotted in Figure 3.14.

Figure 3.14: Mean log reading times at DP site 2 by DP type and Familiarity from words 13-17 in Experiment 3. The DP variable refers to the first DP found in the sentence, i.e., the DP in the filler phrase. Original log mean values in the left panel and fitted log mean values in the right panel. Error bars represent 95% confidence intervals.



We found significant effects of DP type on word 15, i.e., on the embedded subject itself, in the raw reading times (β = -202.45, SE = 32.87, t = -6.159, $p < 0.001$, model not shown) and log reading times (β = -0.33, SE = 0.04, t = -8.033, $p < 0.001$, model not shown). This effect demonstrates that conditions with R-expressions as the embedded subject were read longer than conditions with pronouns as the embedded subject. This effect replicates the effect found in Experiment 2 in this same syntactic position. However, since we did not find the effect in the residual reading times, which control for the number of characters in the word, we do not deem this effect reliable at this position. This effect is reliable across transformations at word 16, i.e., one word following the embedded subject, replicating the results of Experiment 2. We report the model for the log reading times for word 16 in Table 3.10. This effect is also found on word 17, two words following the embedded subject in the log reading times only (β = -0.06, SE = 0.02, t = -3.282, p = 0.002, model not shown). Crucially, the effect disappears at word 18 across all transformations. Note that there was no significant main effect of Familiarity in this region.

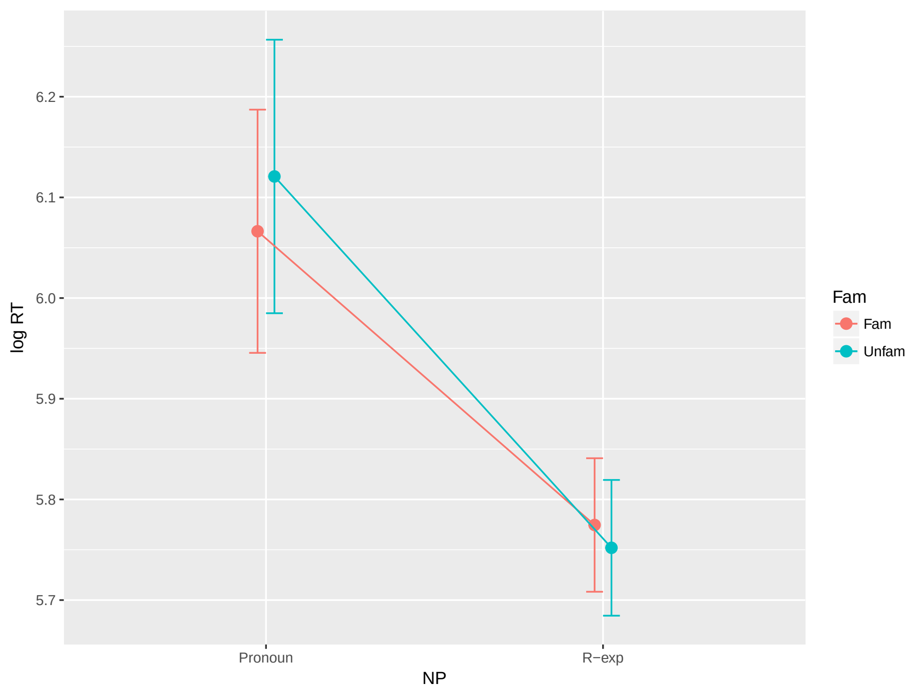
Table 3.10: Final mixed-effects model for residual log times at word 16 (N = 2101 before trimming, 2048 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.934	0.033	182.4	0.000
R-expression	-0.212	0.019	-11.174	0.000
Unfamiliar	0.01	0.011	0.892	0.376

Random effects: by-participant intercept (SD = 0.273), by-item intercept (SD = 0), by-participant random slope for DP type (SD = 0.137), and the correlation between by-participant intercept and slope for DP type (r = -0.32), by=participant random slope for Familiarity (SD = 0.04) and the correlation between by-participant intercept and slope for Familiarity (r = 0.25).

We also found a significant interaction between DP Type and Familiarity at word 15, i.e., when the embedded subject is introduced. This interaction was found across transformations (raw reading times: β = -55.67, SE = 22.27, t = -2.5, p = 0.015, residual reading times: β = -35.90, SE = 17.74, t = -2.023, p = 0.046, and log reading times: β = -0.08, SE = 0.04, t = -2.045, p = 0.044, models not shown). The interaction plot for the log reading times is shown in Figure 3.15. The interaction demonstrates that while conditions containing R-expressions as the embedded subject were read longer than those containing pronouns, this effect was stronger when the R-expression was an unfamiliar name compared to when it was a familiar name. This interaction only reached significance at word 15, i.e. where the embedded subject is found, suggesting that the additional processing cost associated with unfamiliar R-expressions is easily resolved.

Figure 3.15: Interaction plot showing the effect of DP Type on Familiarity on word 15 for log reading times in Experiment 3. The DP variable refers to the first DP found in the sentence, i.e., the DP in the filler phrase. Error bars represent 95% CIs.



Based on the interaction, we again compared log reading times on pronouns, unfamiliar R-expressions and familiar R-expressions on word 15. We used familiar R-expression embedded subjects as our baseline. If the embedded subject was a pronoun, it was read significantly faster than if it was a familiar R-expression (β = -0.304, SE = 0.04, t = -7.554, $p < 0.001$). If the embedded subject was an unfamiliar R-expression, it was read marginally longer than if it was a familiar R-expression (β = 0.054, SE = 0.03, t = 1.826, p = 0.07). This result again seems to suggest that both types of R-expressions are more costly than pronouns but that familiar R-expressions are (marginally) less costly than familiar R-expressions. We will briefly review the results from the familiarity judgement task and then discuss our interpretation of the main findings from Experiment 3 in more detail.

3.4.2.4 Familiarity judgement task

To ensure that our familiarity manipulation was interpreted as intended by participants, i.e., that "familiar" names were indeed familiar to our participants and that "unfamiliar" names were not familiar to them, the self-paced reading experiment was followed by a familiarity judgement task. In this task, we asked participants to rate how familiar the names were to them. They were presented with a series of names on a screen and asked to rate the familiarity of the name on a scale of 1 to 5 where 1 corresponded to "not familiar at all" and 5 corresponded to "extremely familiar." The same participants who completed the reading task also completed the familiarity judgement task. The average familiarity rating for names coded as "familiar" was 4.55 and the average familiarity rating for names coded as "unfamiliar" was 1.8. These results therefore suggest that our familiarity manipulation was in line with participants' interpretation of the names, i.e., they recognized our familiar names but did not recognize our unfamiliar names.

3.4.3 Experiment 3 General Discussion

Experiment 3 was designed as a follow-up study to Experiment 2 and investigated whether our previous results from Experiment 2 could be explained by the fact that all of the R-expressions used in Experiment 2 were unfamiliar to participants. Since participants were unable to assign a referent to the R-expression when it was first introduced (since it was unfamiliar), this could have led to greater processing difficulty compared to conditions in which a pronoun was found in the same syntactic position. In contrast to the R-expressions, the pronouns could be accommodated within the sentence (see e.g., Heim and Kratzer, 1998, among others). We predicted that if this lexical explanation of the results was on the right track, the cost of R-expressions should be alleviated when familiar R-expressions are used in our stimulus items. More precisely, we predicted that reading times on familiar R-expressions should be comparable to those observed on pronouns. We manipulated whether the R-expression could be analyzed as familiar using the names of famous people.

The results from Experiment 3 on the first-introduced DP replicate the results found in Experiment 2. We again found that R-expressions incurred a greater processing cost compared to pronouns in this position, suggesting that even familiar R-expressions incurred a greater processing cost compared to pronouns. We explain this effect as resulting from R-expressions carrying more discourse properties compared to pronouns, as argued in Experiment 2. The results at this region also revealed that the parser was sensitive to the familiarity manipulation used in this experiment. We found a main effect of familiarity such that unfamiliar names were read longer than familiar ones. Crucially, the familiarity manipulation did not eliminate the processing cost associated with R-expressions, suggesting that explaining our Experiment 2 results as an effect of the R-expressions being unfamiliar to participants is not sufficient.[76]

The significant interaction between DP type and familiarity revealed that while R-expressions incurred an overall greater processing cost compared to pronouns, the effect

[76] At the first DP region, we also found that conditions containing pronouns were read longer than those containing R-expressions at word 11, three words following the pronoun/R-expression. We currently have no explanation for this effect. However, since this effect is only found on this one word region, which occurs several words after the initial position of the pronoun/R-expression, we do not think this is an interpretable result so we leave it aside for the time being.

was stronger when the R-expression was unfamiliar (a made-up name) compared to when it was familiar (the name of a famous person). If our previous results could be explained as an effect of the R-expressions being unfamiliar to participants, we predicted that familiar R-expressions should be processed similarly to pronouns. This was not borne out. Pronouns were less costly than familiar R-expressions but unfamiliar R-expressions were more costly than familiar ones. Overall, this effect demonstrates that it was easier for the parser to process the R-expression when it was the name of a famous person (familiar) compared to when it was a made-up name (unfamiliar).

At the second DP region, i.e., the subject of the embedded clause, we again replicated our results from Experiment 2, demonstrating that conditions in which an R-expression appeared as the embedded subject incurred a greater processing cost (longer reading times) compared to those in which a pronoun appeared as the embedded subject. Crucially, the effect of familiarity did not reach significance at this region in the sentence. This suggests that our previous results at this region in Experiment 2 cannot simply be attributed to a lexical effect of R-expressions carrying more discourse properties than pronouns. If our previous results could be explained as a lexical effect, we would expect familiar R-expressions to show the same processing profile as pronouns. Since we observed similar effects in Experiment 3 where we manipulated the familiarity of the R-expression, these results provide further support for our previous interpretation of the results in this region. If the embedded subject is an R-expression, a pronoun appears in the *wh*-filler phrase. Upon reaching the pronoun in the filler phrase, the parser must find an antecedent for that pronoun. Since the parser prefers to find referents within the sentence whenever possible (Van Gompel and Liversedge, 2003; Kazanina et al., 2007), it must locate an antecedent for the pronoun further to the right in the string. However, since the parser also prefers binding relations over co-reference relations, it should determine that it must reconstruct the phrase containing the pronoun in order for it to enter a binding relation with another DP. If the parser reconstructs the phrase containing the pronoun (complement of *wh*) to a position that is structurally below its antecedent (the R-expression embedded subject), it should determine that this analysis incurs a Principle B violation at LF. As a result, the parser must reanalyze its parse to adopt a co-reference relation between the two DPs. Since co-reference is the less preferred interpretation, a processing cost is incurred. Our results from Experiment 3 are compatible with this explanation of the results.

It is important to note that we did find a significant interaction between DP type and familiarity at the embedded subject region, showing that unfamiliar R-expressions were read longer than familiar ones. This effect only reached significance on the embedded subject itself, suggesting that any initial processing difficulty associated with unfamiliar R-expressions is quickly resolved and does not incur a prolonged processing cost. Upon further inspection of this effect, we found that the difference between familiar and unfamiliar R-expressions reached marginal significance at this region.

In Experiment 2, we argued that the effects on the two DP regions could be explained differently because they also showed different effect sizes. In Experiment 3, we again observed differences in effect sizes on the two DP regions. In Experiment 3, the effect size was larger in the second DP region compared to the first DP region on the word following the pronoun/R-expression (0.168 log ms, 62.8 raw ms vs. 0.212 log ms vs. 72.1 raw ms). This pattern of results is similar to what we found in Experiment 2. As argued in the General Discussion section of Experiment 2, if the DP type results at both regions were indicative of a lexical cost associated with R-expressions, we might expect that the effect size would be the same in both DP regions in the sentence.

3.5 Chapter conclusions

We began this chapter asking whether we could find stronger evidence for the structural explanation of our results from Experiment 1. At the end of Chapter 2, we provided two possible explanations for the results observed in Experiment 1. If the results observed with anaphors reflect a structural analysis of reconstruction at LF, we predicted that we should see different effects when pronouns are used in our stimulus items. We investigated this question in Experiment 2 by presenting pronouns in the *many-x* phrase, instead of anaphors. The distribution of pronouns is not restricted in the same way as anaphors (Binding Principles B vs. A). Unlike anaphors, pronouns cannot be structurally bound by a local antecedent. In contrast, if our Experiment 1 results can be explained by a non-structural analysis, in which both types of DP are looking for an antecedent further to the right in the linear string, irrespective of its syntactic position, pronouns should behave similarly to anaphors.

The results from Experiment 2 provide evidence for the structural explanation of our results. In Experiment 2, we found that pronouns did not incur a processing cost compared to R-expressions. Our results show that R-expressions incurred a greater processing cost in both syntactic positions: whether they were found in the filler phrase or as the embedded subject. These results suggest that the parser is sensitive to the structural restrictions on different types of English DPs and they are compatible with a reconstruction analysis of anaphoric binding at LF. If our anaphor results (Experiment 1) could be explained as an effect of the parser looking for an antecedent for the anaphor irrespective of the syntactic position of the antecedent, we would expect to find the same effects with pronouns, which also prefer to find antecedents within the sentence. Our CQ results from Experiment 2 confirmed that the parser does indeed interpret the pronoun as co-indexed with the

R-expression found within the same sentential context. Thus, these results cannot be explained as resulting from the pronoun finding its antecedent outside the sentence.

The results from Experiment 2 revealed that R-expressions incurred a greater processing cost compared to pronouns. We attributed this result to two different explanations, depending on the syntactic position of the DP. When the DP was first introduced in the *wh*-filler, we argued that R-expressions incurred a greater processing cost because they carry more discourse properties than pronouns. At the embedded subject region, however, R-expressions incurred a greater processing cost due to the pronoun found in the complement of *wh*. Since the parser prefers binding relations over co-reference relations, it tries to reconstruct the phrase containing the pronoun to a position below a feature-matched antecedent. However, doing so would incur a Principle B violation at LF. Once the parser reaches the embedded subject and binding becomes an option, it pursues a binding analysis. However, reconstruction leads to a Principle B violation and the parser must reanalyze its parse and adopt a co-reference relation between the two DPs.

In Experiment 2, we again failed to find an effect of reconstruction for the *de dicto* interpretation of creation verbs, even though our CQ results revealed that participants were interpreting these verbs as intended. In fact, in Experiment 2, we found that creation verbs were overall less costly than non-creation verbs.[77] In Section 2.5.3 in Chapter 2, we discussed some possible explanations for the lack of a creation verb effect. In the subsequent experiments reported in thesis, we concentrate on potential reconstruction effects associated with DPs and we leave aside the creation verb manipulation.

We replicated our Experiment 2 findings in Experiment 3, where we additionally manipulated whether the R-expression was familiar or not to participants. Overall we found

[77] As reported in Chapter 2, the creation verbs used in Experiments 1-2 were more frequent and contained less characters than the non-creation verbs. Both of these factors could have contributed to the effect observed in our Experiment 2 reading times.

that unfamiliar R-expressions incurred a greater processing cost compared to familiar ones but that the processing cost of R-expressions was still greater than the cost of pronouns. The effects were again different on the two DP regions, supporting our interpretation of the previous results.

Overall, we argue that the results from Experiments 1-3 provide evidence for the parser's sensitivity to a PF-LF mismatch in real-time processing. When an anaphor linearly precedes its antecedent in the PF representation, the parser must reconstruct the phrase containing the anaphor (complement of *wh*) to a copy position that is structurally lower than the antecedent at LF. In doing so, the anaphor will be structurally bound by its antecedent at LF, abiding by Binding Principle A. In contrast, if a pronoun linearly precedes its antecedent, it can be interpreted in the position in which it appears in the PF representation. As a result, the parser does not need to reconstruct the phrase containing the pronoun at LF. In fact, we have argued that a binding analysis only becomes an option in pronoun conditions once the parser finds a feature-matched embedded subject. Since we observed processing difficulty on the words following anaphors but not on the words following pronouns, we interpret these results as demonstrating the parser's sensitivity to PF-LF mismatches in the domain of referential processing.

It is important to note that in the experiments reported so far, the stimulus items all contained only one feature-matched antecedent for the anaphor/pronoun. This means that in the experimental sentences, there was only one possible antecedent for the anaphor/pronoun. This antecedent also always appeared in a structural position from which it could c-command (and therefore bind) the anaphor, under a reconstruction analysis. A related question is how the parser determines the antecedent for an anaphor/pronoun when there are multiple potential antecedents in the sentence which either appear or do not appear in appropriate structural positions to bind the anaphor at LF. Stronger support for the structural explanation of our results would be results demonstrating that the parser only considers antecedents for anaphors/pronouns if they are found in appropriate structural positions, i.e., ones in which they could c-command the anaphor but not the pronoun at LF. The goal of Experiments 4-6 was to directly investigate these questions by manipulating the number of potential antecedents as well as their structural positions in the sentences. These results are reported in Chapter 4.

Chapter 4

Do gender-matched intervening DPs play a role?

4.1 Introduction

4.1.1 Summary of Experiments 2 and 3 results

In Experiments 2 and 3, we investigated whether pronouns show similar processing effects as anaphors when they linearly precede their antecedents (typically known as cataphoric pronouns). Previous work on cataphoric processing has shown that the parser prefers to find the antecedent for the cataphor within the sentence (Van Gompel and Liversedge, 2003; Kazanina et al., 2007) and building on work on filler-gap dependencies, this work has also shown that the parser is sensitive to grammatical constraints on referential processing (see e.g., Kazanina et al., 2007 for results showing that the parser is sensitive to Principle C effects in cataphoric processing). However, previous work has not investigated whether the parser is sensitive to grammatical constraints on referential processing when the string of words that is presented to the parser does not directly correspond to the structure it must build to be able to interpret the sentence (PF-LF mismatch).

4.1.1.1 Effects of DP type

Building on the results from Experiment 1 (Chapter 2), in Experiments 2 and 3 (Chapter 3), we manipulated whether a pronoun or R-expression appeared in the *many-x* phrase. Since pronouns (unlike anaphors) can also serve as subjects, we also manipulated whether a pronoun or R-expression appeared as the embedded subject, i.e., in the antecedent position for a pronominal. If a pronoun appeared in the *many-x* phrase, an R-expression appeared as the embedded subject, and vice versa. In all cases, the pronoun and R-expression matched in gender, meaning that there were no ungrammatical constructions in our stimuli. Building on our Experiment 1 results, we hypothesized that the parser prefers to adopt structural relations between two DPs in the same sentence instead of non-structural relations. Consequently, we predicted that the parser would prefer to adopt an analysis in which the pronoun is structurally bound by its antecedent, instead of adopting a co-reference relation between the two DPs (see e.g., Reinhart, 1983). Since we always used embedded questions and there was no feature-matched antecedent for the pronoun in the preceding linguistic context, the only way for the parser to be able to adopt a binding relation between the two DPs would be if it reconstructed the phrase containing the pronoun (complement of *wh*) to a position structurally below its antecedent at LF. However, in our stimuli, reconstruction of the phrase containing the pronoun below the embedded subject would yield a Principle B violation at LF. In other words, the parser's preferred analysis of the sentence is to adopt a binding relation between the two DPs. However, since adopting a binding relation between the two DPs would yield a Principle B violation, the parser must reanalyze its parse upon reaching the embedded subject. We predicted that if the parser adopted the reconstruction analysis, we would observe processing difficulty (longer reading times) after the embedded subject. This effect would suggest that the parser has determined that it did not pursued the correct parse of the sentence and that it must revise its parse to adopt a co-reference interpretation between the two DPs.

We found significant main effects of the type of DP in both syntactic positions, i.e., in the complement of *wh* and in the embedded subject position, such that R-expressions were read significantly longer than pronouns. We argued that this effect can be explained by two separate processes based on the syntactic position of the DP. Following work by Gordon and Hendrick (1998), we argued that the DP type effect in the complement of *wh* was reflective of R-expressions having greater discourse properties compared to pronouns when they are initially introduced. However, we suggested that the effect of R-expressions showing a greater processing cost compared to pronouns on the embedded subject position was reflective of the parser attempting to reconstruct the phrase containing the pronoun to a structural position below its antecedent. In doing so, the structure would yield a Principle B violation. We argued that the longer reading times observed on the R-expression embedded subject were reflective of the parser reanalyzing its parse from a binding relation to a co-reference interpretation at this position in the structure. Since a binding relation between the pronoun and the embedded subject R-expression is not possible, the parser must adopt a non-structural (co-reference) relation between the two DPs. As this is not the preferred parse of the sentence, a processing cost is incurred.

In Experiment 3, we replicated the DP results from Experiment 2 and also suggested that the results from both syntactic positions cannot be solely attributed to R-expressions carrying more discourse properties than pronouns. If the results could be solely attributed to R-expressions carrying more discourse properties than pronouns, we expected "familiar" R-expressions to incur the same processing cost as pronouns. In Experiment 3, we manipulated the familiarity of the R-expressions by using the names of famous people. We predicted that if our previous results could be attributed to discourse property effects in both syntactic positions, they should disappear when the names are familiar to participants. In other words, we expected familiar names and pronouns to be read similarly but unfamiliar names to still incur a processing cost. In Experiment 3, we found that R-expressions were overall more costly compared to pronouns and thus, the prediction that pronouns and familiar R-expressions should incur the same processing cost was not borne out. Familiar R-expressions showed less of a processing cost compared to unfamiliar ones but overall, they were still more costly than pronouns. As a result, we argued that attributing our previous effects to discourse properties alone would not be a sufficient explanation.

Crucially, since we did not observe the same effects with pronouns in Experiments 2 and 3 as we did with anaphors in Experiment 1, we argued that anaphors are fundamentally different than pronouns. We further suggested that our results are compatible with a reconstruction analysis of binding at LF. Anaphors must find their antecedents within the sentence and following a structural analysis (binding), the anaphor must be bound by its antecedent at LF. In our stimuli, the only way to adopt this analysis is for the phrase containing the anaphor to reconstruct to a position below its antecedent at LF (i.e., interpret the lower copy of the complement of *wh*). In contrast, pronouns cannot be locally bound by their antecedents. Consequently, if the phrase containing the pronoun reconstructs below its antecedent at LF, a Principle B violation is incurred. Since we found that participants still interpreted the two DPs as co-indexed, we suggested that the parser adopts a co-reference analysis between the two DPs at the embedded subject and this analysis is costly. This analysis explains why we observed a later cost (i.e., further to the right) in the pronoun conditions in Experiments 2 and 3. Our results from Experiments 1-3 are therefore compatible with a structural analysis of antecedent resolution.

4.1.1.2 Effects of verb type

Another important finding from Experiment 2 was that we found significant effects on the verb region, such that non-creation verbs were read longer than creation verbs. Assuming that additional operations, such as reconstruction, are costly, this effect was unexpected under the hypothesis that *many-x* needs to reconstruct for the *de dicto* interpretation when the embedded verb is a creation verb, but not when it is a non-creation verb. This effect was also surprising because we did not find a reliable effect of verb type in Experiment 1. As mentioned in the Chapter 3, the types of comprehension questions asked in Experiment 2 were different from those asked in Experiment 1. More precisely, some of the CQs in Experiment 2 directly asked about the event in the creation or non-creation verb, i.e., *Did the pictures already exist?* These questions therefore may have made participants more aware of the verb types and they may have quickly learned to pay attention to the verb because they could be asked a question about it. In other words, these types of questions may have encouraged participants to pay more attention to the verb region and they may have been processing this region more deeply. In the series of experiments reported in this chapter (Experiments 4a, 4b, 5, 6a, and 6b), we also asked deep processing questions about the relation between the anaphor/pronoun and its antecedent in an effort to encourage participants to interpret the DPs in the sentences.

4.1.1.3 No significant interaction between type of DP and type of verb

We did not observe an interaction between verb type and DP type in either Experiment 1 or 2. If creation verbs force the parser to reconstruct *many-x*, we expected to observe a facilitation effect when *many-x* contained an anaphor in Experiment 1 (since anaphors must be c-commanded by their antecedents). In Experiment 2, we expected to observe binding violations if the parser was forced to reconstruct *many-x* below the embedded subject at LF. If *many-x* contained a pronoun, we might predict a reanalysis effect after the creation verb since reconstruction would yield a Principle B violation. Crucially, the creation verb should also force the phrase containing the R-expression to reconstruct below the verb at LF which would yield a Principle C violation, if the two DPs are interpreted as co-referential. However, we did not observe any real-time effects of the creation verb in either experiment. The lack of a creation verb effect could be attributed to (at least) two explanations. One possibility is that participants were simply not processing the semantic distinction between the creation and non-creation verbs. However, we do not think this is the correct explanation because participants were fairly accurate on the comprehension questions that specifically asked about the creation or non-creation event. A second possibility is that participants are not sensitive to Principle C violations. We had predicted that the effect should be stronger in Experiment 2 because reconstruction would lead to an ungrammatical structure, in contrast to Experiment 1, where reconstruction creates a grammatical structure. However, if participants are not sensitive to Principle C effects, then we should not observe an effect at all. Our results from Experiment 4b might shed some light on this issue so we leave the discussion for later in the chapter. Note, however, that if this explanation is correct, we will need to explain why Kazanina et al. (2007) found significant effects of Principle C violations on the surface. Recall that they found that the parser did not consider a potential referent for a cataphoric pronoun if that referent would violate Principle C. Since the verb type manipulation did not reveal significant findings and since it complicates the design of our studies, we did not use this manipulation in the subsequent studies reported in this thesis. Moreover, since we did not observe reliable effects at the verb site in any of the previous studies, we remain agnostic about what parsing operations might take place at this region in the sentence. All of our stimuli consisted on embedded questions and thus, they all required syntactic integration of the complement of *wh* in the lowest copy position (object position of the verb). Since we did not observe any differences between conditions at this region in our stimuli, we cannot comment further on what form the parsing mechanism might take at this region.

4.1.2 Goal of the current experiments

The goal of the current series of experiments was to further tease apart whether the effects found in Experiments 1 and 2 can be explained by a reconstruction account of binding at LF or whether they would be better explained by the parser simply looking for the closest gender-matched antecedent in the sentence. Thus far, this thesis has assumed that the parser is sensitive to grammatical constraints on anaphoric and pronominal binding, i.e., the binding principles. Binding Theory is a structural account of anaphoric and pronominal binding and thus is sensitive to the syntactic positions of the pronominal referents and their antecedents. Following this view, we have assumed that if a pronominal referent is not in an appropriate position to be bound by its antecedent on the surface, it can reconstruct to a lower position at LF in order to be structurally bound. However, in all previous experiments, the stimulus sentences only contained one other DP in the sentence that matched the pronominal in gender. Crucially, this DP also appeared in a structural position in which it could c-command the pronominal at LF. Therefore, while such results are compatible with a reconstruction analysis of binding, this is not the only explanation of the results.

An alternative explanation is that the parser was simply looking for a feature-matched antecedent for the pronominal (i.e., matching gender and number features) but it does not care about the structural position of the antecedent. Under this alternative hypothesis, an anaphor does not need to find a structurally-appropriate antecedent; instead, it is sufficient to find a feature-matched antecedent somewhere in the sentence. Following much previous work on language processing, there is a preference to resolve open dependencies as soon as possible. Therefore, if this simple search hypothesis is the correct analysis, then the search for an antecedent for the anaphor should end as soon as a feature-matched antecedent is found in the sentence, irrespective of its syntactic position. In contrast, if the reconstruction account is on the right track, feature-matched DPs that are not in appropriate structural positions, i.e., c-command positions, should not be considered as potential antecedents for the anaphor. Our goal in the series of experiments reported in this chapter was to explore these types of questions more directly by manipulating the structural positions of potential antecedents for anaphors/pronouns and by manipulating whether these potential antecedents matched or did not match the pronominals in gender.

In this chapter, we also explore whether the parser can reconstruct the complement of *wh* to an intermediate reconstruction position. In Experiments 6a and 6b, we manipulated whether there was more than one potential antecedent for the pronominal in the sentence. In these experiments, the potential antecedent was always found in a subject position, meaning that it was always potential structural antecedent. Whether or not this subject DP could locally bind the pronominal, however, was dependent on whether the complement of *wh* could reconstruct to an intermediate copy position or whether it could only reconstruct to the lowest copy position (base-generated position). Thus, in the two versions of Experiment 6, we also explore whether the parser is sensitive to intermediate reconstruction positions.

4.2 Experiment 4a

4.2.1 Design

The experimental design was adapted from the previous experiments. In this version of the experiment, we manipulated the type of DP found in the complement of *wh* (anaphor or pronoun) but not the type of verb (all verbs were non-creation verbs). Additionally, we manipulated whether a feature-matched potential antecedent for the anaphor/pronoun appeared before (i.e., in a linearly closer position) the embedded subject. In our previous experiments, only one feature-matched antecedent appeared in the sentence. This DP also occurred in a structurally appropriate syntactic position to c-command the anaphor/pronoun under a reconstruction analysis. As a result, we do not know how the parser processes constructions in which there are two potential antecedents for a pronominal and only one of them occurs in a structurally appropriate position.[78] Stronger evidence for the hypothesis that the parser is sensitive to structural constraints on referential processing would be results demonstrating that the parser only considers potential antecedents for pronominals when they are in structurally appropriate positions and that it ignores potential antecedents occurring in structurally inappropriate positions, i.e., positions in which the antecedent could not bind the pronominal.

Following previous work on referential processing, we have argued in the previous

[78] By "structurally appropriate," we are always referring to whether the DP appears in a position from which it could c-command the pronominal at LF.

chapters that the parser prefers to find antecedents for anaphors/pronouns within the sentence, at least within an experimental setting (see e.g., Van Gompel and Liversedge, 2003). The previous experiments reported in this thesis all used stimuli where there was only one DP that matched the pronominal in relevant phi-features (i.e., gender, number). The goal of Experiment 4a was to determine precisely how the parser determines which antecedent is appropriate for a pronominal referent. Since our previous studies only had one feature-matched antecedent for the pronominal, this DP was the only potential antecedent the parser needed to consider. In addition, this antecedent was always in an appropriate structural position to satisfy Binding Principle A in the anaphor conditions. The pronoun conditions present a more complicated case, as argued in Chapter 3. Reconstruction of the phrase containing the pronoun to a position below the embedded subject would yield a Principle B violation at LF because the antecedent would c-command the pronoun. In such cases, the parser must adopt a co-reference interpretation of the two DPs in order for the parse to be interpretable.

In the series of experiments reported in the current chapter, we investigated how the parser finds an antecedent for a pronoun/anaphor when there are competing potential antecedents in the sentence. To investigate this question experimentally, we inserted intervening adjuncts in between the *wh*-filler phrase containing the anaphor/pronoun and the embedded subject. As in our previous studies, in Experiment 4a, the embedded subject always matched the anaphor/pronoun in gender, meaning that the parser could always find an antecedent for the pronominal in the embedded subject position. The critical manipulation was whether the intervening adjunct also contained a gender-matched potential antecedent or not. Sample stimuli are shown in (1). The intervening adjuncts, which were always prepositional phrases, contained a genitive R-expression, which either matched or did not match the pronoun or anaphor in gender, compare (1-a)-(1-c) to (1-b)-(1-d).

(1) The reporters wondered ...

a. how many lies about himself$_i$ in Sean's report you asked Jeff$_i$ to deny

b. how many lies about himself$_i$ in Shelly's report you asked Jeff$_i$ to deny

c. how many lies about him$_i$ in Sean's report you asked Jeff$_i$ to deny

d. how many lies about him$_i$ in Shelly's report you asked Jeff$_i$ to deny

... for the TV interview.

The intervening adjunct was always 3 words in length and appeared immediately following the anaphor/pronoun but before the subject of the asking event, which, as in Experiments 1-3, was always the pronoun *you*. Our critical assumption was that since the DP appeared within an adjunct phrase, it was never in an appropriate structural position to bind the pronominal. This assumption follows from a revised notion of c-command proposed by Bruening (2014), which he calls *phase-command* (p. 243). *Phase-command* is defined as in (2).[79]

(2) a. Phase command: X phase-commands Y iff there is no ZP, ZP a phasal node such that ZP dominates X but does not dominate Y.

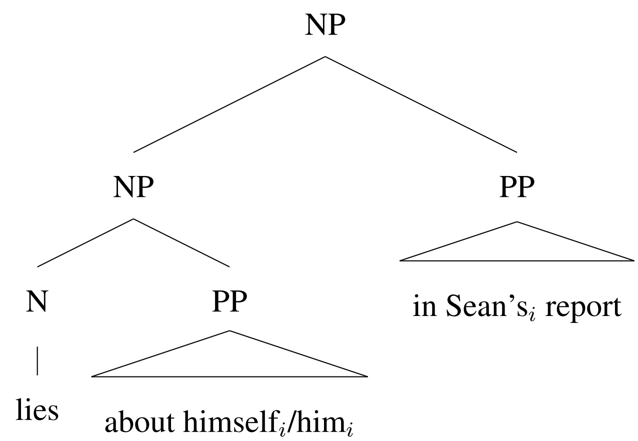
b. Phasal nodes: CP, vP, NP

Crucially for Bruening (2014)'s proposal, PP is not a phasal node. If we assume the adjunction structure in (3) for the complement of *wh* in (1), we can see that the anaphor/pronoun c-commands the R-expression. If the two DPs are co-indexed, this

[79] Precedence is also an essential part of Bruening (2014)'s proposal but this notion is not relevant to the current work.

structure should incur a Principle C violation. Thus, it should not be possible for the pronominal to be co-referential with the DP in the intervening adjunct phrase.

(3)



If we were to find speakers who accept a reading in which the pronominal and the DP in the adjunct are co-indexed, this might provide evidence for analyzing PPs as phasal nodes (see e.g., van Riemsdijk, 1978; Abels, 2012). If PPs are phasal nodes, then the pronominal would not phase command the PP adjunct.[80,81]

Generally, we assumed there was no syntactic position to which the parser could reconstruct the phrase containing the anaphor/pronoun within the adjunct phrase. In other words, there was no reconstruction position that would create a structure where the DP in the adjunct would c-command the anaphor/pronoun at LF.[82] An equal number of trials with

[80] This analysis would eliminate the Principle C violation but in anaphor conditions, we would still need to ensure that the anaphor can be bound by the antecedent at LF. Since there is no reconstruction position in the adjunct, it is unclear how the parser would satisfy Principle A in these environments. Reconstruction of the phrase containing the anaphor to a lower copy position, i.e., below the embedded subject would allow the anaphor to be locally bound by the embedded subject but it would then not be in the same local domain as the DP in the adjunct.

[81] It is also important to note that this analysis assumes that only the complement of *wh* would need to reconstruct. The complement of *wh* includes the picture DP and its PP complement (containing the pronominal). Adjuncts do not need to reconstruct along with the phrase. We remain agnostic about the specifics of this assumption since they are not relevant to the current work but a theory of Late Merge (see e.g., Fox, 2002) would account for this analysis.

[82] An additional consideration is that the DP in the adjunct was always genitive and therefore did not match the anaphor/pronoun in Case features. We put this issue aside for the time being and assume that gender and

male and female DPs occurring both in the adjunct phrase and as the embedded subject were found across the experiment. As in previous experiments, we ensured that the proper names used in the experiment were not gender neutral. We used the word-by-word, moving window self-paced reading methodology (Just et al., 1982) but in contrast to previous studies, the naturalness rating task occurred within the reading task, i.e., participants were asked to read each sentence word-by-word and also provide a naturalness judgement on the sentence they just read. The procedure will be explained in more detail after we outline the main predictions of this study.

4.2.1.1 Predictions

This version of the experiment allows us to directly investigate the structural versus non-structural hypotheses, as discussed in Chapter 3. By manipulating whether or not a competing potential antecedent was found in the intervening adjunct, we are able to investigate the parser's sensitivity to structural constraints on antecedent resolution, as proposed in the theoretical literature (see e.g., Chomsky, 1981, and many others). The two competing hypotheses make different predictions for the stimulus sentences in (1). We will go over these different hypotheses separately.

Non-structural analysis: Anaphors According to the *non-structural* account, as soon as the parser finds an anaphor in the sentence, it should look for an antecedent with matching features. Following this account, the structural position of the antecedent should not matter; what is important is that the DP bears appropriate features. Since the parser prefers the simplest possible parse of a sentence (see e.g., Frazier and Fodor, 1978), we might predict that it will prefer to satisfy open dependencies as soon as possible. Thus, there should

number features are more relevant, at least at an initial stage of processing.

be a general preference to find an antecedent with matching features in the linearly closest position.[83] At a general level, if this preference is on the right track, the parser should prefer to associate the pronominal with the first DP it finds in the string as long as the phi-features on that DP match the phi-features on the pronominal.

In the current experiment, the first DP that the parser finds is found in the intervening adjunct. Crucially, this DP is not in an appropriate position to locally bind the anaphor at LF. The only structurally appropriate antecedent for the anaphor is the embedded subject. If the non-structural account is on the right track, the structural position of the first potential antecedent should not matter since the parser should only care about finding a feature-matched antecedent for the anaphor. Following this account, once the parser has found the feature-matched DP in the intervening adjunct position, the parser should be satisfied that it has found an antecedent for the anaphor and it should stop looking for an antecedent further to the right in the sentence. In other words, any processing difficulty (i.e., longer reading times) associated with the parser looking for an antecedent for the anaphor should disappear after it has found the DP in the adjunct. Note that we manipulated whether the DP in the adjunct matched or did not match the anaphor. If the non-structural account is on the right track, we should observe that processing difficulty associated with the parser looking for an antecedent for the anaphor should disappear if the DP occurring in the intervening adjunct matches the anaphor in gender. If the DP does not match the anaphor in gender, the parser should continue looking for an antecedent further to the right in the linear string.

Structural analysis: Anaphors In contrast, following the structural account, the parser is sensitive to the syntactic constraints imposed on anaphors with respect to their

[83]This hypothesis is also supported by previous work on cataphoric processing: processing difficulty is found when the first potential antecedent site mismatches the cataphor in gender, suggesting that the parser had predicted that it would find the antecedent for the pronoun in this position.

antecedents, i.e., binding constraints. As a result, the parser should never consider DPs that occur in positions in which they cannot structurally bind the anaphor as potential antecedents for the anaphor. Since the DPs occurring in the intervening adjuncts are never able to structurally bind the anaphor at LF (since there is no reconstruction position within the adjunct), they should not be considered as potential antecedents for the anaphors. The parser may initially consider them as antecedents solely due to the fact that they match the anaphors in features. However, it should quickly determine that these DPs are not structurally appropriate and it should continue to look for an antecedent for the anaphor. If processing difficulty is associated with the parser looking for an appropriate match for the anaphor, we should observe this processing difficulty (reflected by longer reading times) until the parser finds the embedded subject further to the right in the string. Therefore, following this account, the anaphor conditions should be read similarly, irrespective of the gender of the DP found in the intervening adjunct. In both cases, the parser needs to find a structurally appropriate antecedent for the anaphor and the only DP that is in a structurally appropriate position is the embedded subject.

The prediction that we might observe processing difficulty on a DP in the adjunct if it matches the anaphor in gender might seem surprising if the parser is only concerned with the structural positions of potential antecedents. If the parser was to only consider structural positions of the DP, it should dismiss any DP that is not found in a position from which it could bind the anaphor, irrespective of its phi-features. We propose that the parser takes both morphological features and the syntactic position of the DP into account but not at the same time.[84,85] Under this view, upon finding an anaphor further to the left

[84]Thanks for Martin Hackl for discussing this idea with me.

[85]Note that this proposal differs from the results reported in Dillon et al. (2013), where the researchers did not find intrusion effects for antecedent resolution (see also Sturt, 2003). They argued that syntactic information is used when determining an antecedent for a reflexive pronoun (anaphor in our terms). However,

in the string, the parser looks for an antecedent and uses morphological features as an initial filter. For each DP that the parser encounters, the parser checks whether this DP matches the anaphor in relevant phi-features. If the two DPs do not match in features, the DP is disregarded because it cannot serve as an antecedent for the anaphor based on morphological features alone. However, if the two DPs match in relevant phi-features, the parser is then able to examine whether the DP is found in a structurally accessible position. Structural accessibility is determined by binding principles.

Thus, assuming that both morphological features and syntactic position are taken into consideration, an effect based on the gender features of the potential antecedent DP could be found in two ways. On the one hand, matching gender features could lead to a processing cost (slower reading times). This analysis would follow if the gender-matched DP is not a possible antecedent due to structural considerations. The parser is given a potential DP antecedent that matches the pronominal in gender but the structural configuration in which it appears is not suitable for binding. Thus, a gender-matched DP should incur a processing cost if it is not found in a structural appropriate position. On the other hand, mismatching gender features might incur a processing cost. This would follow if the DP could be a possible antecedent based on the structural configuration but the gender features on the DP mismatch the gender features on the pronominal. In this case, the initial filter of gender features might tell the parser that this DP is not an appropriate antecedent but since the DP still appears in a structural position that is appropriate for a structural antecedent (because it can bind the DP at LF, for example), the parser might have predicted that it would find an appropriate antecedent in this position. Since it is given conflicting information (features

Dillon et al. (2013) used number agreement and not gender and their potentially intrusive antecedents were always found in structurally inaccessible positions. In our Experiment 6, we investigated the potential effect of a second DP in a structurally accessible position, depending on the reconstruction position of the DP. Thus, the experimental set-ups may be too different to warrant a comparison.

vs. structural environment), a processing cost is incurred. Therefore, if the morphological features and syntactic structure associated with the DPs interact, we might expect to find effects of both gender-matched and gender-mismatched DPs. Crucially, whether gender matching or mismatching is predicted to show the effect depends on the structural position in which the DP appears. We would not expect to observe any processing effects if the DP matched the pronominal in gender and it was also found in an appropriate structural position; this would be a suitable antecedent. We also would not expect to find a processing cost if the DP mismatched the pronominal in gender and was not in a structural position from which it could bind the pronominal; this would not be a suitable antecedent based on both criteria.

Non-structural analysis: Pronouns The predictions are different for the conditions containing pronouns. Following the *non-structural* account, the parser should simply look for the linearly closest potential feature-matched antecedent for the pronoun, irrespective of its syntactic position within the sentence. In the current experiment, this means that the parser should associate the pronoun with the DP found in the intervening adjunct position when they match in gender. This DP is the linearly closest possible antecedent for the pronoun. If the DP in the adjunct does not match the pronoun in gender, the parser should continue looking for an antecedent further to the right in the sentence. In such cases, it should associate the pronoun with the embedded subject, similarly to what we found in Experiment 2.

Structural analysis: Pronouns The *structural* analysis makes slightly different predictions for the pronoun conditions. Following a structural account, the parser should prefer to adopt a binding over a co-reference relation whenever possible. Thus, there should be an overall preference for the parser to reconstruct the phrase containing the pronoun to a position structurally lower than its antecedent. The parser first encounters a potential antecedent in the intervening adjunct phrase. The parser cannot reconstruct the phrase containing the pronoun to a position that is structurally lower than the DP within the adjunct because there is no reconstruction position within the adjunct. Does the parser still consider this DP as a potential antecedent? Recall that the parser still has a preference to satisfy open dependencies as soon as possible. This preference might force the parser to adopt an analysis where the pronoun is co-indexed with the DP in the adjunct phrase, provided they match in gender features. The linearly closest position in which the parser could reconstruct the phrase containing the pronoun is following the embedded subject. Assuming successive cyclic movement (Chomsky, 1995), there are other intermediate positions for reconstruction but these positions would not be below feature-matched antecedents for the pronoun.

Note that there is no position in the sentence in which the phrase containing the pronoun would reconstruct to a position that is structurally lower than the DP in the adjunct phrase. If the parser reconstructs the phrase containing the pronoun to a position below the embedded subject DP and also co-indexes the pronoun and the embedded subject, this should incur a Principle B violation (as in Experiment 2). In cases in which the embedded subject is the only gender-matched antecedent in the sentence, the parser will need to reanalyze its parse and adopt a co-reference interpretation. However, in cases in which there is another gender-matched DP in the sentence, i.e., when the DP in the intervening adjunct matches the pronoun in gender, the parser may determine that this DP should be co-indexed with the pronoun. In this case, no reanalysis is needed because this parse does not lead to a binding violation. If the parser prefers to find its referent in the linearly closest position, i.e., the DP in the adjunct, this would suggest that there is a general preference for pronouns to be co-indexed with linearly closer antecedents instead of waiting to determine whether a structural analysis is possible. As discussed in Chapter 3, it is also possible that the parser does not consider binding as a possible analysis until it finds a potential antecedent further to the right in the string. Thus, processing difficulty may only arise when the parser finds a potential antecedent further to the right and it must determine if this DP should be interpreted as co-referential with the pronoun.

As argued in Chapter 3, if the *many-x* phrase containing a pronoun reconstructs to a syntactic position below the embedded subject, this should incur a Principle B violation. Thus, in conditions in which the embedded subject is the only possible within-sentence antecedent for the pronoun, i.e., when the DP in the adjunct mismatches the pronoun in gender, the parser will be forced to adopt a co-reference relation between the pronoun and the embedded subject DP. As argued in Chapter 3, this parse of the sentence is dispreferred since there would be no structural relation between the pronoun and its co-referent but it avoids a Principle B violation under a reconstruction analysis. Following the non-structural analysis, the parser should not have a preference for binding over co-reference relations. Thus, we should not observe a reanalysis effect at the embedded subject region when this DP is the only possible co-referent for the pronoun, i.e., in the conditions in which the DP in the adjunct mismatches the pronoun in gender. Overall, we predict that the parser should prefer to find a referent for the pronoun in the linearly closest position, provided the closest potential DP matches the pronoun in gender.

4.2.1.2 Materials

Thirty-two sentence templates were created following the paradigm shown in (1). Target sentences were counterbalanced across four lists of stimuli combined with 40 sentences from an unrelated experiment and 36 fillers (16 of the fillers were created to resemble the target sentences, and 20 fillers resembled the unrelated experiment's items). In total, participants read and judged 108 stimulus sentences. Comprehension questions were asked after each sentence to ensure that participants were paying attention and comprehending the sentences as intended.

As in Experiment 2, we asked comprehension questions targeting participants' interpretation of the pronominal (here, anaphor or pronoun). Since we manipulated whether or not there was a competing potential antecedent in the sentence, these types of comprehension questions were important in determining how participants were interpreting the anaphor/pronoun. More precisely, we were interested in whether or not participants were co-indexing the anaphor/pronoun with the DP in the intervening adjunct or with the embedded subject DP. When the sentence contained an anaphor, we expected participants to reject co-indexation between the anaphor and the DP in the adjunct if it has a preference to adopt a structural analysis between anaphors and their antecedents, regardless of the gender of the intervening antecedent. Following the structural analysis, the only possible interpretation of the anaphor conditions should be one in which the anaphor is co-indexed with the embedded subject (Principle A). In contrast, if the parser is not sensitive to structural constraints, the parser should prefer to co-index the anaphor with the first DP it finds with matching features in the sentence. This means that it should be possible for the parser to co-index the anaphor with the DP in the intervening adjunct, even though this DP is not in a structural position in which it could c-command the anaphor at LF.

For the pronoun conditions, we predict that the parser should always prefer to co-index the pronoun with the DP found in the intervening adjunct, provided they match in gender features. If they do not match in features, the parser will be forced to continue looking further to the right for an antecedent. If the parser prefers to adopt binding relations over co-reference relations, it should attempt to reconstruct the phrase containing the pronoun to a position below the embedded subject. If the embedded subject is the only DP with matching features, the parser should determine that a binding relation would yield a Principle B violation and reanalyze its parse to adopt a co-reference relation between the pronoun and the embedded subject. If, however, the DP found in the intervening adjunct matches the pronoun in gender, the parser can co-index the pronoun with this DP. There would be no Principle B violation but the parser would have also not adopted a parse in which binding is preferred over co-reference. The CQ results from this experiment will enable us to determine if the parser's preferences are weighted when determining the referents for pronominal elements. For example, does the parser prioritize finding a referent that is in a linearly closer position or does it prioritize adopting binding over co-reference relations?

In Experiment 4a, we included 8 questions asking about how participants were interpreting the pronoun or anaphor. Because we were not sure how participants would respond to these questions, all responses to these types of questions were accepted as "correct." Participants were not provided with feedback on their answers to the co-indexation questions. We also included questions asking about the sentence-final adjunct, the picture DP, the matrix subject and the embedded verb. The wide variety of questions helped to ensure that participants were paying attention and reading the whole sentence.

4.2.1.3 Participants

Sixty-four participants completed the combined self-paced reading and naturalness judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). Three participants were removed from the analysis because they scored less than 60% on the comprehension questions throughout the whole experiment.[86] Four additional participants were excluded because they scored less than 60% on the critical trials. In the results section, the data for the remaining 57 participants (24 males, 33 females) will be presented. All participants were self-reported native speakers of English, aged between 19 and 55 years of age (*M* = 34.1 years, *SD* = 9.25). Participants were paid $9.00 USD for their participation.

4.2.1.4 Procedure

The experimental procedure was the same as in the previous studies with the exception that once participants had read each sentence word-by-word, they were asked to rate the sentence they just read for its naturalness on a scale of 1 (extremely unnatural) to 5 (perfectly natural). After providing a naturalness rating, they were presented with a yes-no comprehension question about the sentence. As in our previous studies, line breaks appeared after the picture DP in each sentence, as shown in (4).

(4) The reporters wondered how many lies
about himself$_i$ in Sean's report you asked Jeff$_i$ to deny for the TV interview.

[86] We used a lower comprehension cut off score compared to previous experiments because comprehension scores were overall lower in Experiment 4a. Rejecting participants based on the 75% cut off used in other experiments would have eliminated too many participants from our analysis.

4.2.2 Results

4.2.2.1 Outliers

Items with an average comprehension score less than 50% were eliminated from the analysis.[87] In total, 2 test items and 4 fillers were removed from the analysis. Accuracy on the comprehension questions for the critical trials was 83.3 % and 82.8% for all trials, after trimming out the low scoring items and participants. In order to only investigate judgements on trials where participants were interpreting the sentence as intended, trials where CQs were answered incorrectly were removed from the analysis.[88]

4.2.2.2 Statistical analyses

We used the same statistical analyses as described in the previous experiments with the same trimming procedures. We report here the final trimmed models. All statistical analyses for the naturalness judgements were calculated on the z-scored data across all trials.

4.2.2.3 Naturalness judgements

We repeat the paradigm from (1) as (5) below. The relevant variables that we investigated in this experiment are DP type and Match. DP type refers to the DP found in the filler phrase: anaphor conditions, as in (5-a) and (5-b) and pronoun conditions, as in (5-c) and (5-d). Match refers to whether the DP in the intervening adjunct matched or did not match

[87] Note that in our previous experiments, items with an an average score less than 60% were trimmed out of the analysis. We used a lower cut off score in this experiment because keeping the cut off at 60% would have eliminated too many trials. Overall, comprehension scores were lower in this version of the experiment. We will discuss possible reasons for the lower overall accuracy in the Discussion section.

[88] We kept all trials in which the question asked about co-indexation. We investigate these results in detail below.

the pronominal in gender: gender-matched conditions are found in (5-a) and (5-c) and gender-mismatched conditions are found in (5-b) and (5-d).

(5) The reporters wondered ...

a. how many lies about himself$_i$ in Sean's report you asked Jeff$_i$ to deny

b. how many lies about himself$_i$ in Shelly's report you asked Jeff$_i$ to deny

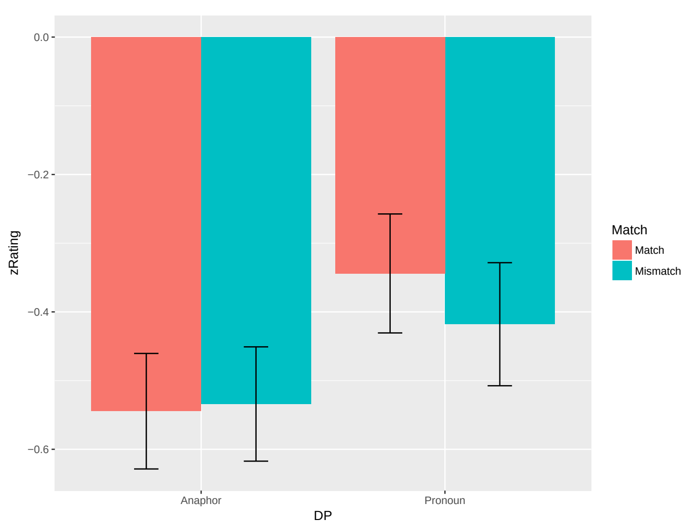
c. how many lies about him$_i$ in Sean's report you asked Jeff$_i$ to deny

d. how many lies about him$_i$ in Shelly's report you asked Jeff$_i$ to deny

... for the TV interview.

Mean *z*-scored naturalness judgements by condition are plotted in Figure 4.1. There was a reliable main effect of DP type such that conditions containing pronouns were more likely to be rated higher than conditions containing anaphors (β = 0.151, SE = 0.035, t = 4.372, $p < 0.001$). There was no reliable effect of Match nor was there a reliable interaction between DP type and Match. These results suggest that judgements were affected by the type of DP found in the filler phrase but that whether or not the pronominal found in the filler phrase matched the DP in the intervening adjunct in gender did not have an effect on participants' naturalness ratings. Recall that in this version of the experiment, the naturalness ratings were asked after participants had read each sentence word-by-word and before they had answered the comprehension questions. Therefore, unlike our previous naturalness judgement tasks where participants were presented with the whole sentence at once and then asked to rate that sentence, in this experiment, participants needed to engage in real-time processing before making a judgement. The DP type results suggest that participants are sensitive to the type of DP found in the filler phrase in both types of tasks, i.e., in cases where they need to make a judgement after having processed the whole sentence on a single screen and also in cases where they were required to read sentences word-by-word (and engage in incremental processing) before making this judgement. As in previous studies, conditions containing anaphors incur a greater processing cost than conditions containing other types of DPs (in this case, pronouns). However, the naturalness judgement task did not reveal any differences based on whether or not the DP in the intervening adjunct matched or did not match the pronominal in gender. Assuming that differences in ratings are reflective of differences in processing difficulty, this result might initially suggest that the gender of the intervening DP did not affect processing difficulty. The results from the reading version of the experiment will enable us to investigate whether this manipulation was effective or not.

Figure 4.1: Mean naturalness judgements (in z-scores) by DP type (anaphor, pronoun) and Match (match, mismatch) in Experiment 4a. The Match variable refers to whether or not the DP in the intervening adjunct matched or mismatched the pronominal in gender. Error bars represent 95% confidence intervals.



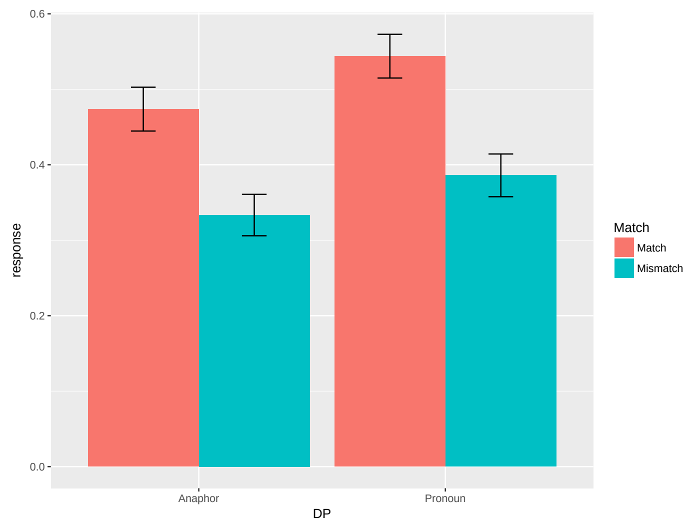
4.2.2.4 Comprehension questions

We were also interested in how participants were interpreting the sentences, which we probed with CQs asking about co-indexation between the pronominal and either the DP in the adjunct or the embedded subject DP. For example, for the sentence in (5-a), a CQ asking about co-indexation between the anaphor and the DP in the adjunct would have been: *Were the lies about Sean?* and a question asking about co-indexation between the anaphor and the embedded subject DP would have been: *Were the lies about Jeff?* Since we asked yes-no comprehension questions in this version of the experiment, we will present the results based on the syntactic position of the DP (intervening adjunct, embedded subject) separately.

Comprehension questions asking about the DP in the adjunct The mean percentage of *yes* responses to questions asking about whether the DP in the intervening adjunct and the pronominal were co-indexed are plotted in Figure 4.2. Across conditions, the mean percentage of *yes* responses to these questions was 43.4%, suggesting that participants were not interpreting the intervening DP as a potential referent for the anaphor/pronoun the majority of the time. However, this average is still fairly high considering that we were expecting participants to immediately reject this DP as a potential antecedent for the anaphor conditions.

Answers were coded as 1 if the participant answered *yes* to the comprehension question and 0 if they answered *no*, meaning that higher values represent a higher likelihood of a co-indexed interpretation. Since the data formed a binomial distribution (1 or 0), we ran logistic regression models on the data with random effects participants and items and random slopes by participant for DP type and Match, whenever the models converged.

Figure 4.2: Percentage of *yes* responses to the co-indexation CQs in Experiment 4a when the CQ asked about whether the DP in the intervening adjunct and the pronominal were co-indexed. Error bars represent 95% confidence intervals.

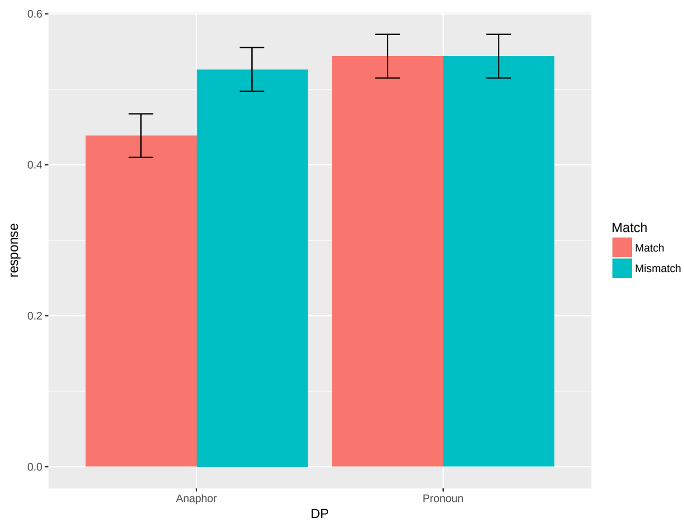


There was a significant main effect of Match such that participants were less likely to respond *yes* in mismatch conditions compared to match conditions (β = -1.29, SE = 0.09, z = -14.18, $p < 0.001$), as expected. There was also a significant main effect of DP type, indicating that participants were more likely to respond *yes* in pronoun conditions compared to anaphor conditions (β = 0.52, SE = 0.08, z = 6.22, $p < 0.001$). The interaction between the two factors did not reach significance at the 5% threshold. These results suggest that participants were sensitive to the gender of the DP in the adjunct and that they were less likely to consider a gender-mismatched DP as a potential referent for the pronominal, as expected. We also found a significant effect based on the type of DP found in the filler phrase, indicating that co-indexation was more likely for pronoun conditions compared to anaphor conditions. This result suggests that participants were sensitive to the fact that the DP in the adjunct was not an appropriate match for anaphor conditions and they were more likely to choose this DP as the antecedent for the pronoun conditions.

Comprehension questions asking about the embedded subject DP The mean percentage of *yes* responses to co-indexation questions that asked about whether the embedded subject DP and the pronominal were co-indexed are plotted in Figure 4.3. Note that we expected participants to answer *yes* the majority of the time for these questions, especially when the DP in the intervening adjunct mismatched the anaphor/pronoun in gender. In fact, when the *wh*-phrase contained an anaphor, we expected participants to answer *yes* 100% of the time because the embedded subject is the only structurally accessible antecedent for the anaphor. In the anaphor cases, the embedded subject is the only structurally appropriate antecedent, assuming that the parser reconstructs the phrase containing the anaphor to a position below the embedded subject. As discussed in the introduction, in the pronoun cases, the embedded subject is a possible referent under a co-reference interpretation. If the parser co-indexes the pronoun with the embedded subject DP and reconstructs the phrase containing the pronoun, this should incur a Principle B violation. However, it is still possible to co-index the pronoun with the embedded subject DP under a co-reference analysis. In fact, this is the only possible analysis when the embedded subject is the only gender-matched antecedent in the sentence, i.e., the DP in the adjunct mismatches the pronoun in gender.

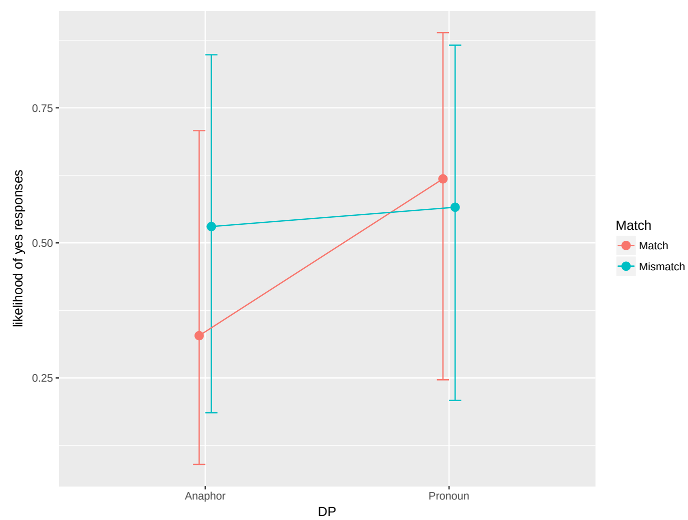
Across conditions, participants responded *yes* only 51.3% of the time to these questions, suggesting that the embedded subject was selected as the referent for the pronominal at chance. These results suggest that participants did not have a preference to co-index the pronominal with the embedded subject DP, contrary to our predictions at least for the anaphor cases. We again ran generalized logistic regression models on the binomial data. There was a significant main effect of DP type such that participants were more likely to respond *yes* to these CQs in pronoun conditions compared to anaphor conditions (β = 1.45, SE = 0.13, z = 11.04, $p < 0.001$, model not shown). However, there was no

Figure 4.3: Percentage of *yes* responses to the co-indexation CQs in Experiment 4a when the CQ asked about whether the pronominal and the embedded subject were co-indexed. Error bars represent 95% confidence intervals.



significant main effect of Match. We did find a significant interaction between DP type and Match (β = -1.06, SE = 0.18, z = -5.75, $p < 0.001$), as plotted in Figure 4.4. The interaction demonstrates that responses were similar between Match and Mismatch cases in pronoun conditions but that questions were more likely to be answered as *yes* when the DP in the adjunct mismatched the anaphor in gender compared to when it matched the anaphor in gender. These results therefore suggest that the matching intervening DP was interfering with how participants were processing the anaphor conditions but did not show an effect in the pronoun conditions. More precisely, the likelihood of participants selecting the embedded subject as the referent for the anaphor was higher when the intervening DP mismatched the anaphor in gender, compared to when it mismatched the anaphor in gender. We will discuss the consequences of these results after going over the reading time results.

Figure 4.4: Interaction between DP and Match for the CQs asking about co-indexation between the embedded subject and the pronominal in Experiment 4a. Error bars represent 95% confidence intervals.



Summary: Comprehension questions The comprehension questions that directly probed participants' interpretation of the relation between the anaphor/pronoun and one of the other feature-matched DPs in the sentence (DP in the adjunct or embedded subject DP) showed that participants were not always interpreting the pronominal referents as intended. These results actually indicate that participants accepted the DP in the intervening adjunct as a potential antecedent for the anaphor/pronoun approximately half the time, even when it mismatched the pronominal in gender. Participants were overall less likely to choose the gender-mismatched DP as the referent for the anaphor/pronoun but the mean percentages of *yes* responses was still quite high considering that this DP should be automatically disregarded based on morphological features alone. This result might suggest that participants were guessing the answers to the CQs. If they were truly paying attention, we would expect them to easily dismiss a gender-mismatched DP as a potential antecedent. Crucially, participants accepted the DP in the intervening adjunct as the antecedent for the anaphor just under 50% of the time when they matched in gender. This result is unexpected under a structural analysis of anaphoric binding. Under a reconstruction analysis, only the embedded subject DP should be considered as a referent for the anaphor. For the pronoun conditions, we found that participants were more likely to answer *yes* for the DPs occurring in both syntactic positions, compared to the anaphor conditions. Overall, these results seem to suggest that pronouns prefer to find referents within the sentence but there did not seem to be an overall preference for the pronoun to be co-indexed with the DP occurring in the adjunct or the embedded subject DP.

4.2.3 Self-paced reading experiment

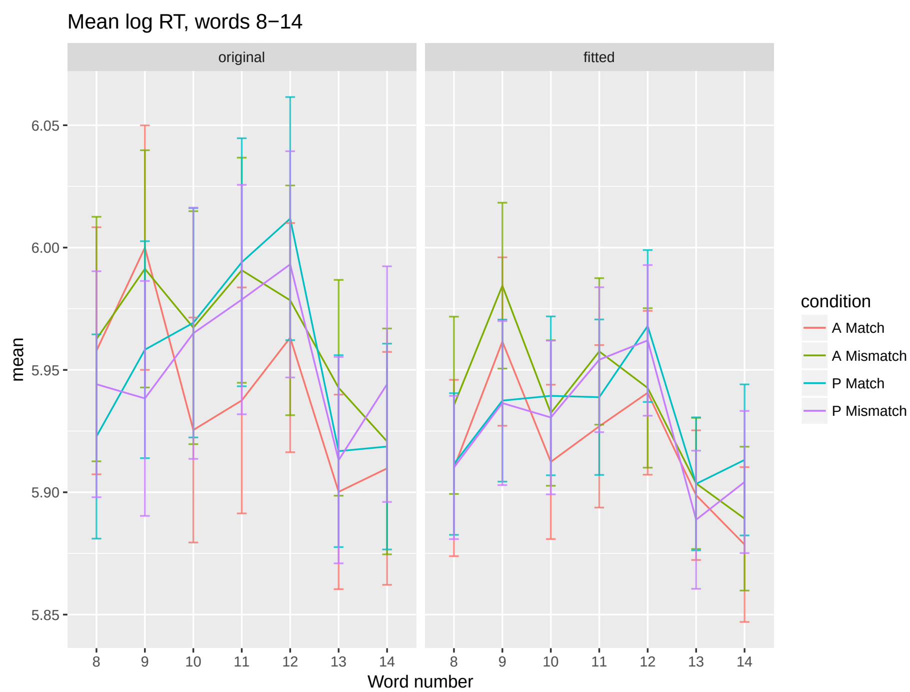
4.2.3.1 Outliers

We removed outliers in the same way as outlined in previous chapters. In this experiment, we also included participants' z-scored judgement on each sentence as a predictor in our statistical model. Overall, we found that when participants provided higher judgements, they read the sentence faster.

4.2.3.2 Reading time results

The mean log reading times for words 8-14 are plotted in Figure 4.5. The anaphor/pronoun in the filler phrase appeared on word 8, the DP in the intervening adjunct appeared on word 10 and the embedded subject DP appeared on word 14. The most reliable effect in this data set was found on word 9, one word following the anaphor/pronoun. In the log reading times, we found that pronoun conditions were read faster than anaphor conditions (β = -0.043, SE = 0.02, t = -2.139, p = 0.037, model not shown).[89] We did not find any effects of the matching or mismatching intervener nor did we observe an interaction between Match and DP type.

Figure 4.5: Mean log reading times on words 8-14 in Experiment 4a by DP Type and Match. Original means in the left panel and fitted means in the right panel. Error bars represent 95% confidence intervals.



4.2.3.3 Experiment 4a Discussion

The fact that reading times were not affected by the gender manipulation is surprising. Our comprehension results seemed to suggest that participants were sensitive to the gender

[89]This effect did not reach significance at the 5% threshold in the raw or residual reading times. We have no explanation for why this might be the case but we interpret this effect cautiously since it does not seem to be as strong as the previous effects we have observed at this region.

manipulation, at least when they were asked comprehension questions about the DP in the adjunct. Participants were less likely to respond *yes* when asked about whether the pronominal and the DP in the adjunct were co-indexed, but only if the gender of the DP in the intervening adjunct mismatched the pronominal in gender. Participants still accepted gender-mismatched DPs as antecedents for anaphors/pronouns, as shown by our comprehension results. We had predicted that these interpretations should be immediately disregarded by the parser, based on morphological constraints alone but also because the DP is not a structurally appropriate antecedent. In such cases, the only possible antecedent for the pronominal is the gender-matched DP in the sentence (the embedded subject).

Participants also did not seem to show a preference about which DP they preferred as the antecedent for the pronoun. One possible explanation for these results is that participants may not have been fully interpreting the anaphors and pronouns and assigning them an interpretation right away. If the parser does not immediately interpret the anaphor/pronoun with respect to its antecedent and instead waits until later in the sentence (i.e., further to the right) to determine the appropriate antecedent, this could have made the parser unaware of the matching and mismatching DPs in the adjunct phrases. In other words, the parser might wait until it has processed the entire sentence before it makes a commitment about how to interpret the anaphor/pronoun. If this is the case, it might not show sensitivity to the gender manipulation in incremental processing.[90] In Experiment 4b, we used a more sensitive experimental methodology, namely, eye-tracking while reading, to investigate how these types of sentences are processed when participants can read the whole sentence on a single line and have the opportunity to reread parts of the sentence that

[90] Of course we still need to explain why participants accept parses in which the pronominal is co-indexed with a DP that mismatches it in gender. Comprehension questions are asked after the whole sentence and therefore, once participants have processed the whole structure. One possibility is that our sentences might have been too difficult for participants to read and comprehend. They may have just been guessing on the CQs. We currently do not have a clear explanation for these results and put them aside for the time being.

they may find difficult. The eye-tracking results will enable us to determine whether these sentences were simply too difficult to process and participants stopped trying to process them properly. They will also allow us to examine if the self-paced reading methodology might not have been sensitive enough to show potential effects of our gender manipulation.

4.3 Experiment 4b

4.3.1 Design

In Experiment 4b, we used the same stimulus sentences as in Experiment 4a except that sentences were presented on one line, i.e., no line break appeared in the sentences. We used the eye-tracking while reading methodology which allows researchers to measure participants eye movements as they read sentences. Crucially, this methodology allows participants to re-read earlier parts of the sentence that they may have found difficult and thus, can be considered to be a methodology that captures more natural reading compared to the self-paced reading methodology. In this version of the experiment, we further manipulated the types of comprehension questions being asked after each sentence. In an effort to encourage participants to interpret the anaphor or pronoun, we only asked "deep processing" questions after each sentence (see e.g., Stewart et al., 2007 for results showing that the type of CQs asked in an experiment can have an influence on real-time results). In this version of the experiment, our CQs directly asked participants to choose which within-sentence referent the pronominal in the filler phrase referred to. Instead of asking yes-no questions, we provided participants with two names and asked them to select which name the anaphor/pronoun referred to. For example, *Who were the pictures of? Options: Mary,* *Jessica*.[91] By manipulating the types of answers, participants were required to associate the anaphor/pronoun to an antecedent within the sentence and they could not simply respond *yes* or *no*, as in the previous version of the experiment. The yes-no processing questions could be viewed as a measure of shallow processing since participants could answer the questions correctly without needing to interpret the anaphor/pronoun. By directly asking participants to select an antecedent for the anaphor/pronoun, we hypothesized that they would be forced to interpret the pronominal and they would not be able to engage in shallow processing.

4.3.1.1 Predictions

The predictions were the same as in Experiment 4a. We will reiterate our predictions based on the gender manipulation. If the parser takes into consideration both morphological features and structural accessibility, we should observe an effect of gender matching if the DP is not in a structurally accessible position yet it matches the pronominal in gender. In this case, the morphological features on the DP tell the parser that it is a suitable antecedent but it is not in a structural position from which it can c-command the pronominal. In contrast, we should observe an effect of gender mismatch if the DP is in a structurally accessible position but it mismatches the DP in gender. In this latter case, the morphological features tell the parser that this is not an appropriate antecedent but its structural position still makes it accessible for binding.

[91] Note that asking participants to select the antecedent for the anaphor/pronoun was done indirectly because we did not want to draw attention to the DP manipulation in our stimulus sentences.

4.3.1.2 Materials

The critical sentences were the same thirty-two templates as those used in Experiment 4a. One hundred and eight filler sentences which were either from unrelated experiments or they were filler sentences designed to resemble the experimental items were also used. In total, participants read one hundred and forty stimulus sentences. The CQs asked after the filler sentences were similar in structure to the CQs asked after the critical sentences.

4.3.1.3 Participants

Forty-three participants completed the eye-tracking while reading study. Twelve participants were excluded from the analysis: 7 because they were non-native speakers, 2 due to technical difficulties, and 3 because they did not complete the experiment. The remaining thirty-one participants were self-reported native speakers of English with normal or corrected-to-normal vision and were recruited from the University of Toronto community. Participants were paid $10.00 CAD for their participation.[92]

4.3.1.4 Procedure

Participants were tested individually in a private room. While participants were completing the eye-tracking experiment, the experimenter was in the room but separated from the participant via a dividing wall. Participants wore the Eye-Link II (SR Research, Mississauga, Canada), a head-mounted eye-tracker with two cameras that used an infrared light to track eye-movements. The eye-tracker was connected to a PC computer that recorded the positions of the eye-movements on the screen at a rate of 500Hz. While participants were able to view the sentences with both eyes, only the eye-movements of the

[92]Thanks to Meg Grant and her research assistants at the University of Toronto for running this experiment.

right eye were recorded.

Once participants had provided informed consent, their right eye was calibrated to the eye-tracker. Participants were instructed to read the sentences presented to them on the screen at a natural pace. The order in which the sentences was presented to participants was randomized. To ensure that participants' eye-movements were being tracked by the eye-tracker as accurately as possible, prior to each sentence being shown on the screen, a black box appeared at the left-hand side of the screen. Participants were asked to look at this black box at the beginning of each trial and once their eye movement had been detected, the sentence appeared. This procedure ensured that the eye-tracker was tracking participants' eye-movements at the appropriate location on the screen. Once they had finished reading each sentence, they were instructed to select a button on a game controller to move on to the comprehension question. The comprehension questions always had two possible answers, which participants responded to on the game controller. Upon answering the comprehension question, participants were presented with the next sentence and comprehension question and the process continued as described above. The experiment took approximately an hour to complete.

4.3.2 Results

Reading times in the critical sentences were investigated by region. We isolated 8 regions in our critical sentences, as shown in (6). In the results section, we focus on regions 3 to 6 (from the complement of *wh* to the embedded subject) as these are the regions where we had made specific predictions about expected reading times.

(6) R1: The reporters wondered

R2: how many lies

R3: about himself/him

R4: in Sean/Shelly's report

R5: you asked

R6: Jeff

R7: to deny

R8: for the interview.

4.3.2.1 Dependent measures

We investigated the results of seven different eye-movement measures, which have been argued to measure different levels of processing (see Staub and Rayner, 2007 for an overview). *First fixation time* is the time spent on the region during the first fixation, provided the region was fixated on at all. *First pass reading time* is the sum of all fixations on a region during the first pass, i.e., does not include any re-reading time. *Second pass reading time* is the sum of all fixations on a region during the second pass, i.e., from entering the region for the second time during reading. Note that when the region is not refixated, then fixation times of 0 are indicated. For our purposes, when analyzing second pass times, we removed any items in which the region of interest was not refixated before doing additional data trimming. *Go-past time* is the sum of all fixations in a region from first entering the region until leaving it to the right. This measure therefore also includes any fixations made on the region when regressions were made back to earlier parts of the sentence. We also removed any items in which go-past times were 0 before doing any further data trimming. *Total reading time* is the sum of all fixations made in the region, including forward and regressive eye-movements. *Regressions-in* is the probability of regressing into a region (scored as a binomial distribution, either 1 or 0). *Regressions-out* is the probability of regressing out of a region (also scored as a binomial distribution, either 1 or 0).

4.3.2.2 Outliers

The critical region in our stimuli was the embedded subject (R6). We considered this region to be the critical region because if participants did not fixate on the embedded subject, then we would not be able to determine how participants are interpreting the embedded subject and whether they are considering it as a potential antecedent for the anaphor or pronoun. Items in which a blink occurred on the embedded subject or in which the embedded subject was not fixated at all were thus removed from the analysis. This resulted in a loss of 7.46% of possible data. Data was also trimmed according to outlying reading times for each eye-movement measure. The trimming criteria and percentage of data lost for each measure are shown in Table 4.1.

Table 4.1: Criteria used to eliminate outliers in eye-movement measures in Experiment 4b.

Measure	Range of data excluded	% of data lost
First fixation	$>$ 100 ms & $<$ 600 ms	3.5 %
First pass	$>$ 100 ms & $<$ 600 ms	2.7 %
Second pass	$<$ 1500 ms	4.7 %
Go past	$<$ 4500 ms	2.2 %
Total time	$>$ 100 ms & $<$ 2500 ms	2.1 %

4.3.2.3 Statistical analyses

We ran the same statistical analysis procedures as described in the previous experiments on all eye-movement measures. Statistical models were run both on the raw and log reading times. We report the log reading times because they better control for the non-normal distributions of reading time data.

4.3.3 Results

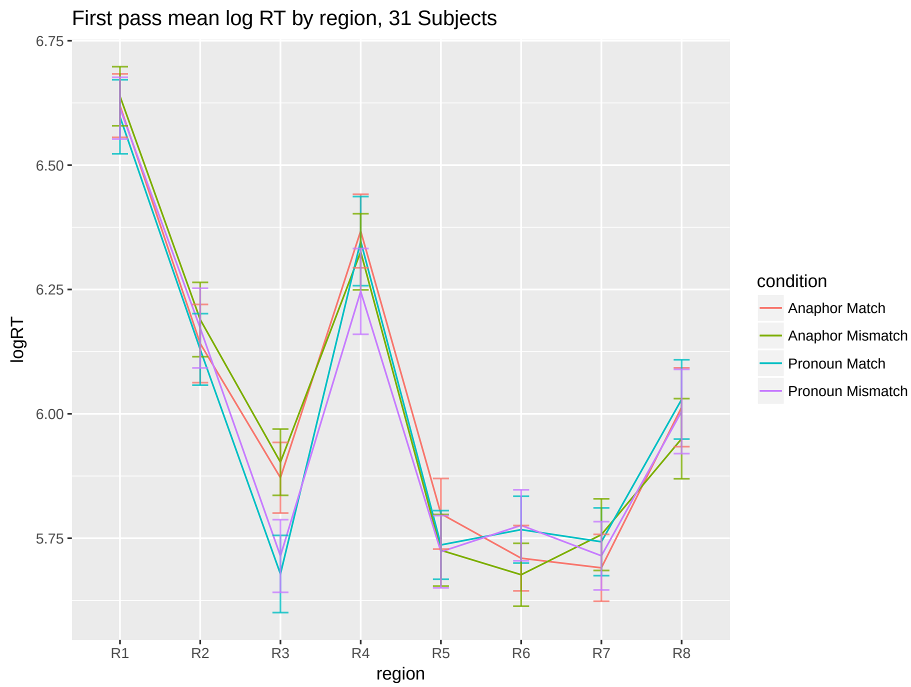
4.3.3.1 First pass times

No significant effects were found in first fixation times. The first fixation time measure has been characterized as an early measure of processing (see e.g., Clifton et al., 2007). On the very first fixation of a word/region, participants may not yet be interpreting the region. Since the stimulus sentences used in this experiments are long and complex, it is possible that participants do not start interpreting the words/regions during the very first fixations.

The mean log reading times for the first pass reading time measure are plotted in Figure 4.6. In the first pass reading times, which include the first fixations as well as any other fixations on that region before leaving it to the right or left, we found that pronoun conditions were read faster than anaphor conditions in region 3, i.e., when the pronoun or anaphor is introduced (β = -0.18, SE = 0.03, t = -5.8, $p < 0.001$, model not shown). This effect is also marginally significant at region 4, i.e., the adjunct region (β = -0.07, SE = 0.03, t = -1.9, p = 0.056, model not shown). We interpret this effect as reflecting a processing cost for anaphors compared to pronouns. Since a cost of anaphor conditions was found in Experiment 1 (Chapter 2) as well as in Experiment 4a, this effect is not particularly surprising.

On region 5, i.e., *you asked*, one region following the adjunct containing the gender-matched or gender-mismatched DP, we found a significant effect of Match such that conditions in which the DP in the adjunct mismatched the anaphor/pronoun in gender were read faster than conditions in which the two DPs matched in gender (β = -0.07, SE = 0.03, t = -2.45, p = 0.01, model not shown). At first glance, this effect might seem

Figure 4.6: Mean log first pass reading times by region in Experiment 4b by DP Type and Match. Error bars represent 95% confidence intervals.



surprising considering that previous studies investigating gender mismatch in pronominal dependencies has found that conditions in which both DPs *matched* in gender incurred less of a processing cost, i.e., faster reading times (Van Gompel and Liversedge, 2003; Kazanina et al., 2007). One possible explanation for why we observe different effects in the current experiment compared to those found in previous studies is that in our stimuli, the first potential antecedent for the anaphor/pronoun is found in the intervening adjunct. Since the structure of this adjunct is likely more difficult to process (because the parser needs to determine where to attach it in the structure), it also needs to determine if the DP is a viable antecedent for the pronominal. If the DP mismatches the anaphor/pronoun in gender, the parser can easily dismiss this DP as a potential antecedent. Since the parser is able to quickly disregard this DP due to gender mismatching, this condition incurs less of a processing cost compared to conditions in which the parser must seriously consider the DP as a potential antecedent. As previously discussed, if the two DPs match in gender, the parser faces a conflict. Morphological features tell the parser that this DP is an appropriate antecedent, i.e., because the two DPs have matching phi-features (at least for gender and number). The parser then needs to look at the structural configuration. For anaphor conditions, the parser determines that the matching DP in the adjunct is not structurally appropriate since there is no position in the adjunct to which the phrase can reconstruct. As a result, we observe an effect of gender matching in this position.

The effect of gender matching was a main effect and there was no main effect of the type of DP on this region. This means that first pass reading times were longer on this region if the DP in the adjunct matched the pronominal in gender irrespective of the type of DP found in the complement of *wh*, i.e., the effect was found in both anaphor and pronoun conditions. In the pronoun condition, we had predicted that it should be possible for the pronoun to co-refer with the DP in the intervening adjunct because such an analysis would not incur a binding violation. An analysis in which the pronoun is co-indexed with the embedded subject DP would incur a Principle B violation under a reconstruction analysis at LF. It is thus unclear to us why there is an effect of gender match in the pronoun condition if this DP can serve as an antecedent for the pronoun. One possibility is that the parser was not expecting to find a potential antecedent for the pronoun in this syntactic position. A processing cost is incurred when the parser must consider whether this DP is a possible antecedent for the pronoun. Our main prediction at this region, however, was not borne out since we did not observe differences in gender match effects based on the type of DP in the complement of *wh*.

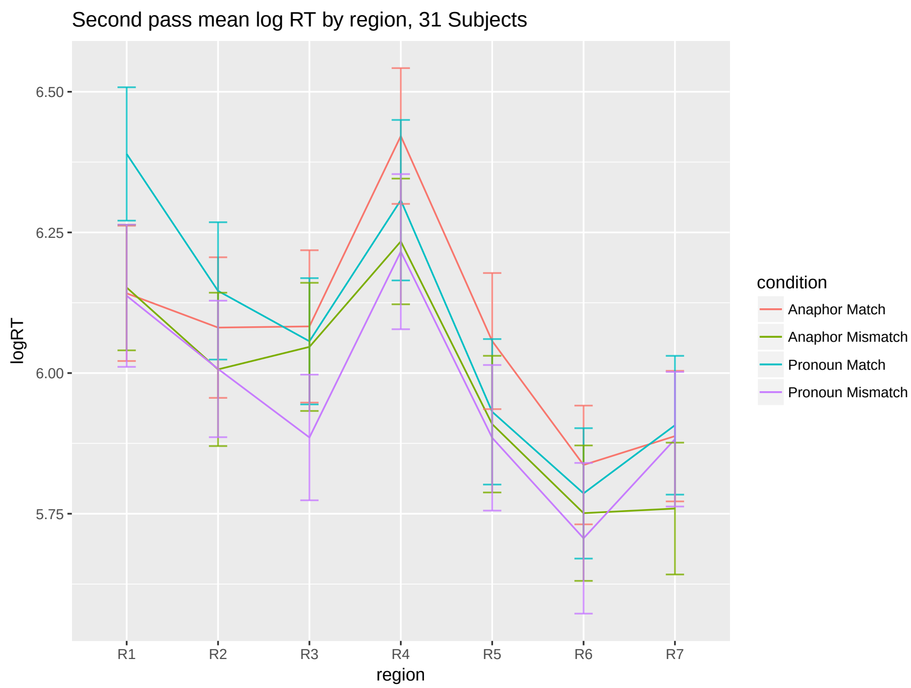
Finally, on region 6, i.e., the embedded subject, pronoun conditions were read longer than anaphor conditions (β = 0.06, SE = 0.03, t = 2.06, p = 0.04, model not shown). This effect replicates the effects found in Experiments 1 and 2. In Experiment 1, we did not observe any reading time differences between the anaphor and R-expression conditions on the embedded subject. We argued that this effect demonstrated that the increased processing difficulty associated with anaphors is alleviated once the antecedent for the anaphor has been found. In Experiment 2, we found that conditions containing a pronoun in the *wh*-phrase incurred a greater processing cost on the embedded subject, which was an R-expression. We argued that this effect reflected the parser attempting to reconstruct the complement of *wh*, containing a pronoun, to a position structurally lower than the embedded subject. Upon reaching the embedded subject, the parser determines that this reconstruction operation incurs a Binding Principle B violation, assuming that the two DPs are co-indexed. As a result, the parser must reanalyze its original parse and adopt a co-reference interpretation. The eye-tracking results found in the current experiment provide further support for this interpretation of our previous results.

Overall, the first pass reading time results suggest that the parser is able to make a distinction between the two types of DPs found in the filler phrase and the gender of the DP found in the intervening adjunct phrase from a very early stage of processing, i.e., during the very first pass of reading the sentence. It remains unclear to us why we did not observe differences in gender match effects on the region following the adjunct based on the type of DP in the complement of *wh*. This result is unexpected if it is possible for the pronoun to co-refer with the gender-matched DP in the adjunct and if such an analysis is not possible in anaphor conditions.

4.3.3.2 Second pass times

The mean log second pass reading times are plotted in Figure 4.7. Effects of both DP type and Match were found on region 3 (*of him/himself*) in the second pass reading times demonstrating that pronoun conditions were read marginally faster than anaphor conditions (β = -0.11, SE = 0.06, t = -1.9, p = 0.058, model not shown) and mismatched conditions were read significantly faster than matched conditions (β = -0.12, SE = 0.06, t = -2.02, p = 0.04, model not shown). At region 4, i.e., the intervening adjunct region, we did not find any significant effects of DP type but mismatched conditions were again read significantly faster than matched conditions (β = -0.16, SE = 0.06, t = -2.66, p = 0.008, model not shown). No other regions showed significant effects of either factor. The second pass reading time measure reflects participants' second pass of reading through the sentence, once they have initially processed the regions. These results show that there is still processing difficulty associated with anaphors during the second pass of the sentence. The greater processing load associated with Match conditions, i.e., conditions in which the DP in the adjunct matched the anaphor/pronoun in gender, is still significant during the second pass of the sentence and it is also found in the filler region, i.e., where the anaphor/pronoun is found. These results therefore suggest that the difficulty associated with gender-matched conditions persists for a long period of time and the parser is even sensitive to this effect at the filler region in second pass reading times. This effect suggests that the parser actively considers the gender-matched DP when reading the filler phrase during second pass reading times. We interpret this effect as reflecting the parser's attempt to determine whether the DP in the adjunct is a suitable antecedent for the anaphor/pronoun. As previously mentioned, the morphological features on this DP match the pronominal and as a result, it seems to be an appropriate antecedent based on this factor alone. Once the parser has determined that the features match, it must then consider the structural configuration of the DPs. Thus, the Match effects might reflect the parser initially considering this DP as a potential antecedent based on phi-features alone.

Figure 4.7: Mean log second pass reading times by region in Experiment 4b by DP Type and Match. Error bars represent 95% confidence intervals.



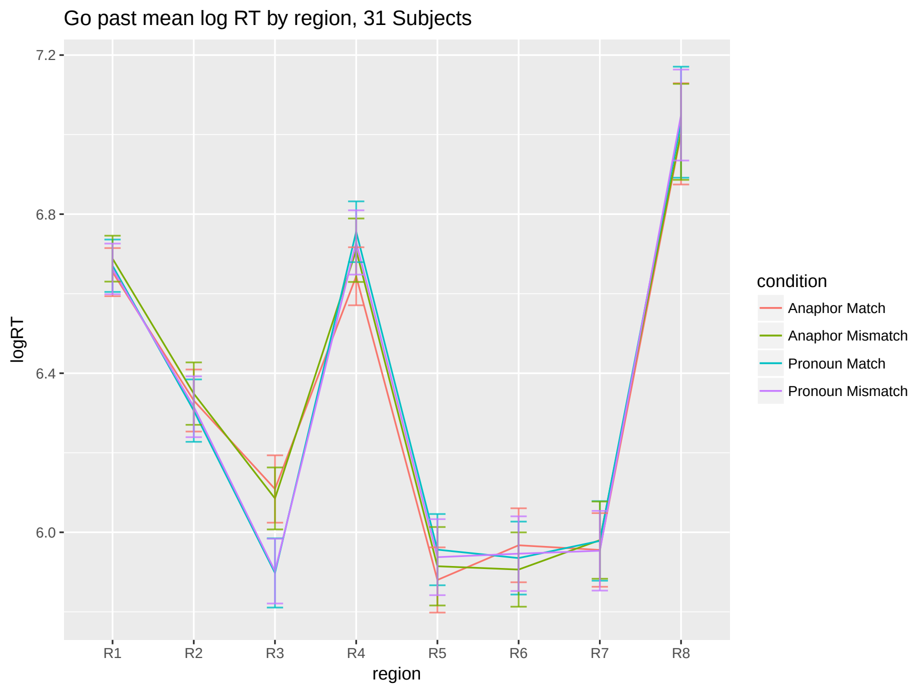
4.3.3.3 Go-past times

The mean log go-past reading times are plotted in Figure 4.8. Similarly to first pass reading times, the go-past reading time measure can also be understood as a measure of early processing. However, unlike first pass reading times, it also include re-reading times from previous regions. Thus, longer reading times in this measure can be understood as reflecting processing difficulty that requires re-reading until the parser can proceed to the next region in the sentence. There was a significant main effect of DP type on region 3 (β = -0.21, SE

= 0.03, t = -6.11, $p < 0.001$, model not shown), demonstrating that conditions containing pronouns were read significantly faster than conditions containing anaphors. At region 4, i.e., the intervening adjunct region, the opposite effect was found: conditions containing pronouns were read significantly longer than conditions containing anaphors (β = 0.09, SE = 0.03, t = 2.65, p = 0.008, model not shown). No other effect reached significance at the 5% threshold. We interpret the processing cost of the pronoun at the intervening adjunct region as reflecting the parser's attempt to associate the pronoun with the DP in the adjunct.[93] Crucially, the cost of anaphors is only found when the DP is initially introduced and disappears on the following region, where the adjunct is found. We interpret this effect as suggesting that the parser does not consider the DP in the adjunct as a potential antecedent for the anaphor. The regression results (which we report below) also support this interpretation of the results.

[93]We might predict that we should observe an interaction between DP type and Match at this region. More precisely, if the parser is attempting to associate the DP in the adjunct with the pronoun, we should find longer reading times in pronoun match conditions. Unfortunately, there was no reliable interaction at this region. However, since we do find a reliable main effect of pronouns being more costly than anaphors at the intervening adjunct region, we argue that this effect suggests that the parser at least considers the intervening DP as a potential antecedent and this causes processing difficulty. Why the results do not seem to depend on the gender is the intervening DP is unclear to us but it might suggest that participants were not yet processing these DPs at a deep level of processing.

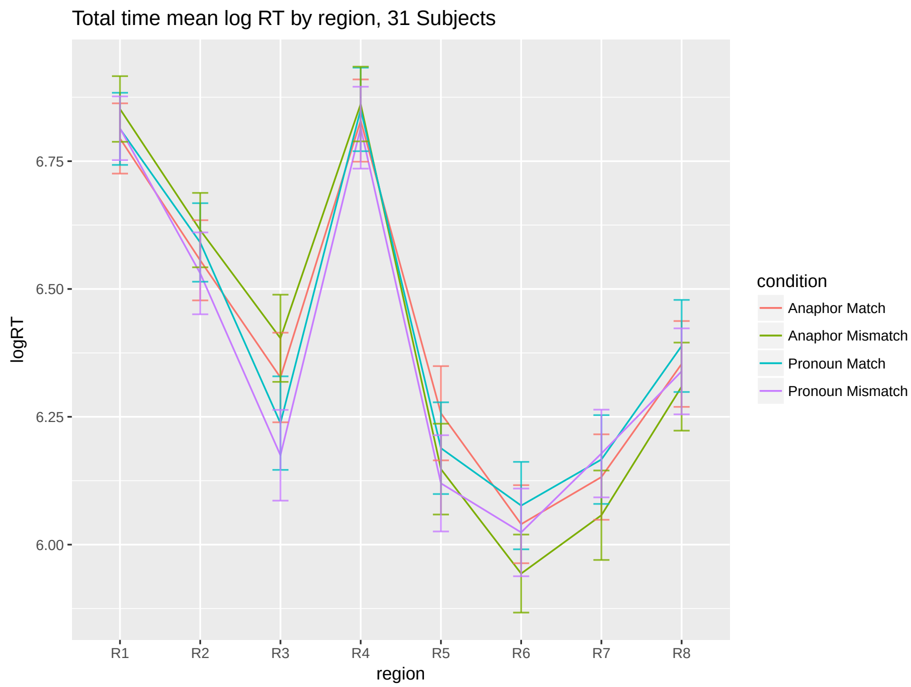
Figure 4.8: Mean log go-past reading times by region in Experiment 4b by DP Type and Match. Error bars represent 95% confidence intervals.



4.3.3.4 Total reading times

The mean log total reading times are plotted in Figure 4.9. The total reading time measure can be interpreted as an overall processing measure because it includes the sum of all reading times on a particular region, including any re-reading times after the participant has read further to the right or has re-read previous regions in the sentence. Processing difficulty in this measure, as indicated by longer reading times, can therefore be understood as a general processing cost on that region, even during a later stage of processing. In total reading times, we found a significant main effect of DP type on region 3 (β = -0.18, SE = 0.04, t = -4.48, $p < 0.001$, model not shown), demonstrating that conditions containing pronouns were read significantly faster than conditions containing anaphors. The total reading time results for the Match variable did not reach significance at the 5% threshold at this region. These results therefore suggest that anaphor conditions generally incur a processing cost, even at a later stage of processing once the parser has had the opportunity to re-read previous regions in the sentence. Since we do not observe a significant effect of the Match variable at this region in total reading times, the effect found in second pass reading times does not persist to an even later stage of processing. The effect observed in second pass reading times, therefore, can be understood as a reanalysis effect during the second reading of the sentence only. At region 5, one region following the intervening adjunct, we observed a significant effect of Match (β = -0.11, SE = 0.04, t = -2.77, p = 0.006, model not shown), showing that conditions in which the DP in the adjunct mismatched the anaphor/pronoun in gender were read significantly faster than conditions in which both DPs matched in gender. Since this effect is found on the region following the intervening adjunct (and therefore, the gender-matched or mismatched DP), we interpret it as reflecting the parser's attempt to determine if the gender-matched DP in the previous region can be associated with the anaphor/pronoun. Since gender-mismatched conditions do not incur this processing cost, this suggests that they are not considered as potential referents for the anaphor/pronoun, as expected. No other effects reached significance at the 5% threshold.

Figure 4.9: Mean log total reading times by region in Experiment 4b by DP Type and Match. Error bars represent 95% confidence intervals.



4.3.3.5 Regressions in

Regression data form a binomial distribution. Data are coded as 1 if there was a regression into a region and 0 if there was not a regression into that region. Since the data form a binomial distribution, a generalized logistic regression model was run on the regression data by region. We interpreted regressions as a reflex of the parser looking for an antecedent. Thus, regressions into a regression are reflective of the parser searching for an antecedent in that region. Results indicate that participants were more likely to regress into region 3 in gender-mismatched conditions compared to gender-matched conditions (β = 0.29, SE = 0.15, z = 1.98, p = 0.048, model not shown). Region 3 is the filler phrase, i.e., when the anaphor/pronoun is introduced.

We suggest that the parser regresses into the anaphor/pronoun region when the DP in the adjunct mismatches the pronominal in gender in an attempt to find an antecedent for the pronominal. Based on morphological features alone, the DP in the adjunct cannot be an appropriate associate for the pronominal if it does not match the pronominal in gender. If we interpret regressions as reflecting the parser's attempt to associate the DP with an antecedent, this result is rather puzzling. These regressions could reflect the parser's attempt to associate the pronominal with a DP antecedent and more regressions are found when the DP in the adjunct mismatches the pronominal in gender. However, note that this effect is the opposite of what we found in the reading time results. In reading times, we found that there was a cost when the DP in the adjunct matched the pronominal in gender. We had interpreted this effect as demonstrating that the parser considers both morphological features and syntactic structure when determining antecedents for pronominals. We do not have a clear explanation for the difference in the pattern of results. One possible explanation is that the two different measures (regressions vs. reading times) reflect different processes. Regressions are a measure of whether participants were more likely to leave a region and re-read another region but reading times reflect an overall cost of reading a particular region. Since examining the differences between these two measures was not the goal of the current thesis, we leave this question aside for the time being.

Participants were also more likely to regress into region 5 in anaphor conditions compared to pronoun conditions (β = -0.53, SE = 0.16, t = -3.42, $p < 0.001$, model not shown). Region 5 is where the subject of the asking event is found. This subject was always the pronoun, *you*, and was therefore never an appropriate match for either the anaphor or pronoun. This region is the first region where the parser might find a structurally appropriate antecedent for the anaphor. Following Principle A, the anaphor requires an antecedent that can c-command it at LF. Regressions into this region might be reflective of the parser attempting to find a structurally appropriate antecedent for the anaphor in the linearly closest syntactic position. The parser regresses more often into the subject of the asking event region in anaphor conditions in an attempt to find a structurally appropriate antecedent for the anaphor. Pronoun conditions do not show this same effect, suggesting that the parser is sensitive to the different structural constraints on anaphors and pronouns.[94] No other effects reached significance at the 5% threshold.

4.3.3.6 Regressions out

We also analyzed the results based on how likely participants were to regress *out* of a region. We interpret regressions out of a region as reflecting the parser's attempt to associate two DPs. The parser regresses out of a region to re-read a previous region in the sentence. The region from which it regresses out causes processing difficulty requiring that it re-read a previous region to process it again. We found no significant effects at region 3 (the filler phrase).

At region 4 (the intervening adjunct region), there were significant effects of both DP type and Match. The effect of DP type reveals that participants were more likely to regress out of a pronoun condition compared to an anaphor condition (β = 0.51, SE = 0.16, t = 3.08, p = 0.002, model not shown). We interpret this effect as reflecting the parser's attempt to associate the DP in the adjunct with the pronoun. Upon reaching the DP in

[94] This effect could potentially be viewed as an "active search" effect. Since the parser must find an antecedent for the anaphor further to the right, it looks for a DP that could serve as an antecedent on each subsequent word/region. However, if this was the case, we might expect to find that there were more regressions into all regions following the anaphor/pronoun until the antecedent was reached further to the right. Since the regression effect only reaches significance on region 5 (the subject of the asking event), this might suggest that the parser only looks for an antecedent for the anaphor in structurally accessible positions. In region 5, the subject of the asking event is structurally accessible but it does not bear appropriate morphological features to bind the anaphor. Thus, these results might be taken as evidence that the parser targets specific regions when trying to resolve anaphoric dependencies.

the adjunct, it must now determine if this DP is a possible antecedent for the pronoun or not. Consequently, more regressions out of this region are found as the parser reconsiders the pronoun in the previous region. The effect of Match reveals that participants were more likely to regress out of mismatch condition compared to a match condition (β = 0.45, SE = 0.16, t = 2.7, p = 0.007, model not shown), replicating the effect found in the regressions in measure. Since we found main effects of both DP type and Match at this region, this result might be interpreted as demonstrating that the parser considers the DP in the adjunct as a potential antecedent for the pronoun when it matches the pronoun in gender. Regressions out of pronoun conditions and out of gender-mismatched conditions may reflect this process.[95,96]

At region 6, i.e., the embedded subject region, we found a significant effect of DP type demonstrating that participants were less likely to regress out of a pronoun condition compared to an anaphor condition (β = -0.41, SE = 0.21, t = -1.98, p = 0.048, model not shown). We interpret this result as an effect of reconstruction. More regressions out of the embedded subject region in anaphor conditions suggests that the parser had reconstructed the complement of *wh* containing an anaphor to the embedded subject position and that it associates the anaphor and the embedded subject. In the pronoun conditions, reconstruction to the lower copy position would yield a Principle B violation. Note that this effect should not be viewed as showing that the parser never associates the pronoun and the embedded

[95]Of course, the gender-matched effect collapses the anaphor and pronoun conditions. If this effect was being driven by the pronoun match conditions, we might expect an interaction between these two factors at this region. However, the interaction between DP type and Match did not reach significance at this region, suggesting that there were overall more regressions out of adjunct region in gender-matched conditions, irrespective of the type of DP in the complement of *wh*.

[96]Another possibility for the effect of gender matching is that the parser regresses out of the region in order to double check if the pronominal from the previous region actually mismatches the DP in gender. If the two DPs mismatch in gender, the parser is able to quickly disregard the DP in the adjunct as a potential antecedent for the pronominal. However, this explanation of the results would be able to explain any regression effects, i.e., regressions happen to double check that the gender of the two DPs match or mismatch. A follow-up study would need to be conducted to directly test if there is any validity to this possible explanation.

subject. In fact, when the DP in the adjunct mismatches the pronoun in gender, the pronoun must find its antecedent in the embedded subject position. Since there are fewer regressions out of the embedded subject position in pronoun conditions, we can interpret this result as showing that co-reference between the pronoun and embedded subject is possible but the pronoun can also find its antecedent elsewhere. The same is not true for anaphors which must find their antecedents in the embedded subject position. No other effects reached significance at the 5% threshold.

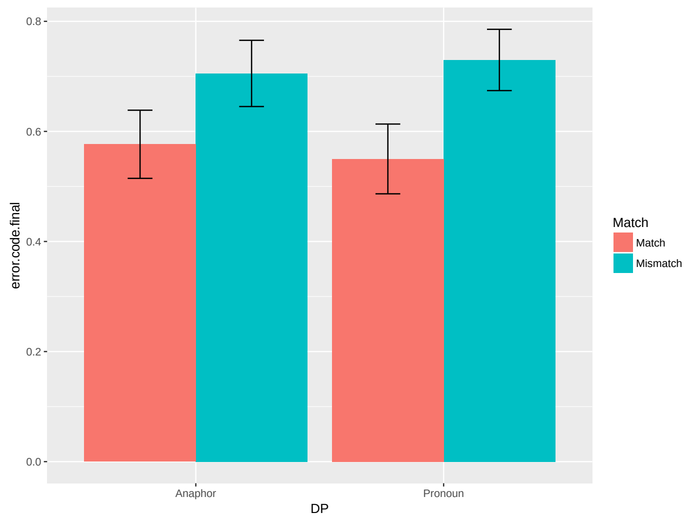
4.3.3.7 Comprehension questions

In this version of the experiment, we manipulated the types of answers presented to participants in the comprehension questions. Instead of providing them with yes-no comprehension questions, we specifically asked participants to select a referent for the anaphor or pronoun. CQs were of the form *Who were the pictures of?* and participants had to select one of the two names provided, e.g., *Sean* or *Shelly*. We hoped that asking participants to explicitly select one of the names would force them to interpret the pronominal and therefore, engage in a deeper level of processing. In our previous experiments, we asked participants yes-no comprehension questions and they could interpret the sentences at a more shallow level of processing. The names that participants were asked to select from were always the same names as those used in the experimental sentences. In other words, when the sentence contained a gender-mismatched DP in the intervening adjunct, the possible answers to the CQs also contained one gender-matched name and one gender-mismatched name. The order in which the names were presented was counter-balanced across the experiment, i.e., in half of the trials, the names were presented in the same order as the order in which they appeared in the experimental sentence but in the other half of the trials, they were presented in the opposite order.

The mean percentage of responses to the CQs asked in Experiment 4b are presented in Figure 4.10.[97] The plot depicts the percentage of responses in which participants selected the embedded subject as the antecedent for the pronoun or anaphor. In cases in which the DP in the adjunct mismatched the anaphor/pronoun in gender, the embedded subject was selected as the antecedent for the pronominal roughly 70% of the time (70.5% of the time for anaphors and 73% of the time for pronouns). We might have expected these numbers to be higher because in such cases, the other DP mismatches the anaphor/pronoun in gender. When the DP in the adjunct matched the anaphor/pronoun in gender, participants selected the embedded subject as the antecedent for the pronominal at just above chance (57.7% for anaphors and 55% for pronouns). Following the structural analysis, this result is surprising for anaphor conditions since the only structural antecedent in the sentences is the embedded subject. The DP in the adjunct should not be a possible antecedent for anaphors. The results for the pronoun conditions suggest that participants were just as likely to select the DP in the intervening adjunct as they were to select the embedded subject, provided they both matched the pronoun in gender. Assuming that pronouns do not need to find a structural antecedent and therefore do not need to reconstruct below their antecedent, this result is not surprising. We will discuss some possible explanations for these results in the General Discussion section.

[97] Due to a programming error, one pronoun gender-matched condition and three anaphor gender-mismatched conditions were removed from the analysis.

Figure 4.10: Percentage of responses to the comprehension questions in which participants interpreted the pronominal and the embedded subject as co-indexed in Experiment 4b. Error bars represent 95% confidence intervals.



4.3.4 General Discussion: Experiments 4a and 4b

In Experiment 4b, we used the same stimuli as in Experiment 4a to investigate whether a more sensitive experimental methodology (namely, eye-tracking while reading) would enable us to determine whether a DP intervening in between a pronominal referent and a potential antecedent further to the right (embedded subject) has an effect on how the parser determines the appropriate antecedent for an anaphor or pronoun. In the previous experiments reported in this thesis, the only DP in the sentence that matched the anaphor/pronoun in gender was the embedded subject and thus, there was only one possible antecedent for the anaphor/pronoun. In addition, this DP was always found in a position in which it could c-command the pronominal at LF, provided the phrase containing the anaphor/pronoun reconstructed at LF. The goal of Experiments 4a and 4b was to investigate how the parser processes constructions in which there is more than one possible gender-matched antecedent within the sentence, only one of which occurs in an appropriate structural position to c-command the anaphor/pronoun at LF. As in our previous experiments, the embedded subject in Experiments 4a and 4b always matched the anaphor/pronoun in gender. In these studies, we added an intervening adjunct (which was always a prepositional phrase) in between the surface position of the anaphor/pronoun and the embedded subject. The intervening adjunct either contained a DP that matched the anaphor/pronoun in gender or it contained a gender-mismatched DP. Crucially, since the gender-matched/mismatched DP was embedded in an adjunct, it was not in a structural position in which it could bind the pronominal at LF, even assuming a reconstruction analysis. More precisely, in these constructions, there was no syntactic position structurally below the DP in the adjunct to which the anaphor/pronoun could reconstruct in order to be bound by this DP.

4.3.4.1 Experiment 4a summary

In the self-paced reading version of the experiment (Experiment 4a), we only found that anaphor conditions were read longer than pronoun conditions on the word following the DP itself. We did not find any significant effects of the DP in the intervening adjunct nor did we observe any later effects based on the type of DP found in the filler phrase. When discussing the results from Experiment 4a, we suggested that the lack of these effects might be attributed to the fact that participants were not interpreting the anaphor/pronoun as intended. This speculation is further supported by their answers to the CQs, which were more or less at chance, even when there was only one gender-matched DP in the sentence. It is therefore possible that participants may not have been always interpreting the anaphor/pronoun in our stimulus sentences. The CQ results did reveal that participants were less likely to co-index the pronominal with the DP in the adjunct when it mismatched the pronominal in gender, suggesting that they were sensitive to the gender manipulation on at least some of the trials (recall that we had only a few trials with these types of CQs). We also found that co-indexation was more likely in pronoun conditions compared to anaphor conditions in both DP positions. This effect can be easily explained for the DP appearing in the intervening adjunct position. Following the structural analysis, this DP is only a possible antecedent for pronouns but not for anaphors, thus explaining why participants were more likely to consider this DP as an antecedent for pronouns. It is puzzling why we find this same effect at the embedded subject site, which should be the only possible antecedent for anaphors, irrespective of the gender of the DP in the intervening adjunct. One possible way to explain these effects is that the type of CQs asked might still not be forcing participants to engage in deep processing. In Experiment 4a, we asked participants CQs targeting their interpretation of the anaphor/pronoun but we still used yes-no questions, e.g., *Were the lies about Jeff?* It is possible that these types of CQs still did not force participants to process the sentences at a deep level and still allowed them to adopt a superficial interpretation. Following Ferreira et al. (2002)'s work on *good enough* processing, comprehenders do not always create fully specified syntactic representations in processing. If participants engage in a superficial level of processing, they may not fully interpret the pronoun in our sentences. For example, it would be sufficient to interpret *her* as *a female referent who this sentence is about*, but not actually determine which real-world referent *her* refers to. In Experiment 4b, we used different types of CQs which aimed to access a deeper level of processing.

4.3.4.2 Experiment 4b summary

In the eye-tracking version of the experiment (Experiment 4b), we found significant effects using the same stimuli as in Experiment 4a. As previously mentioned, in this version, we also manipulated the CQs such that participants were forced to engage in a deeper level of processing at least with respect to antecedent resolution for the anaphors and pronouns. More precisely, we asked questions of the type, *Who were the lies about?* and gave participants two names as possible answers, e.g., *Sean, Jeff*. By forcing participants to select one of the names, we hoped to access a deeper level of processing in the critical sentences. In this experiment, participants were much more accurate in selecting the embedded subject as the antecedent for the pronominal when the DP in the adjunct mismatched the anaphor/pronoun in gender ($\sim$ 70% of the time). However, they only selected the embedded subject above chance when the DP in the adjunct mismatched the pronominal in gender. The comprehension results therefore suggest that the gender-matched intervening DP affected antecedent resolution for both anaphors and pronouns.

During initial processing (first pass reading times), we found that pronoun conditions were read faster than anaphor conditions. Since we have previously found that anaphor conditions incur a greater processing cost (although our previous comparison was with R-expressions), this effect was not surprising. We also observed a Match effect during the initial stage of processing which demonstrated that conditions in which the DP in the adjunct *matched* the anaphor/pronoun in gender were read longer than conditions in which it mismatched the pronominal in gender. This effect was puzzling considering that the opposite effect was found in previous work on referential processing, i.e., previous work has found that gender-matched conditions incur less of a processing cost compared to gender-mismatched conditions. However, in previous studies, the DP of interest was always found in a subject position and could be reliably predicted as an antecedent for the pronominal. Upon finding a cataphor earlier in the sentence, the parser could reliably predict that it would find the antecedent further to the right. Since it was expecting a gender-matched DP, finding a gender-mismatched DP violates the parser's expectations, incurring a processing cost. In the current study, the DP in the intervening adjunct cannot be reliably predicted. The parser is looking for an antecedent for the pronominal but might expect to find it further to the right in the string of words. Upon reaching a (unexpected) DP that matches the pronominal in gender, the parser must now consider whether this DP might be a match for the pronominal. This incurs a processing cost. When the DP mismatches the pronominal in gender, the parser can easily dismiss it as a potential antecedent and no processing difficulty is incurred.

The most interesting results from Experiment 4b were found in the regression results. We interpreted regressions as demonstrating an effect of the parser attempting to associate two DPs in our sentences. While we observed that gender-matched conditions were read longer than gender-mismatched conditions in the reading time measures, we found more regressions into the complement of *wh* in gender-mismatched conditions, compared to gender-matched conditions. We currently do not have an explanation for the difference in results between the two experiments. We also find more regressions into the subject of the asking event region when the filler phrase contained an anaphor. This effect, although not explicitly predicted, can be easily explained by the fact that anaphors must find structural antecedents and the linearly closest potential position for the anaphor to find a structural antecedent is in this region. This effect suggests that the parser makes predictions about the types of words/phrase it will find further to the right in the string of words and crucially, that it is sensitive to the syntactic structure of these strings.

Participants were more likely to regress out of the adjunct region if the sentence contained a pronoun, suggesting that the parser seems to be considering the DP in the intervening adjunct as a potential antecedent for pronouns and not for anaphors. We also found more regressions out of the adjunct when the DP mismatched the pronominal in gender. Since we found an effect of both DP type and Match at the intervening adjunct region, we suggested that the parser attempts to associate the pronoun with the DP in the adjunct phrase. Processing difficulty is incurred when this DP mismatches the pronoun in gender since it cannot be associated with the pronoun due to mismatching morphological features. We also found more regressions out of the embedded subject region in anaphor conditions compared to pronoun conditions, which we interpret as a reconstruction effect. While the parser must reconstruct the complement of *wh* below the embedded subject in anaphor conditions, this is not the case for pronoun conditions. Since fewer regressions were found in pronoun conditions, this suggests that the parser can find an antecedent for the pronoun both in the embedded subject position and elsewhere, e.g., the intervening adjunct DP.

4.3.4.3 Experiment 4a vs. Experiment 4b results

Overall, we argue that the results from Experiment 4b provide evidence in favour of a structural analysis of antecedent resolution. However, we are still puzzled as to why we did not find the same effects in both the self-paced reading experiment and the eye-tracking experiment, even though we used the same stimulus items. As previously mentioned, eye-tracking is a much more sensitive methodology than self-paced reading because it allows researchers to measure participants' eye-movements during several different stages of processing due to different eye-movement measures, e.g., first pass reading times, go-past times, total times, etc. Participants are also free to re-read the sentences more than once, allowing for a more natural reading level. Thus, we might have observed differences in the effects due to differences in methodology. It is also possible that the stimulus sentences used in both versions of Experiment 4 were too complex to result in any interpretable data. Our CQ data shows that participants found the questions very difficult, even when the two DPs mismatched in gender. It is thus possible that participants were not able to properly process these sentences, due to their complexity. In any case, it is difficult to compare the results observed in Experiment 4b to the results reported in previous chapters, which all used the self-paced reading methodology.[98] We therefore ran follow-up experiments investigating the question of gender-mismatch and manipulated the syntactic position of the gender-matched or mismatched DP.

4.4 Experiment 5

4.4.1 Design

The results found in Experiment 4b suggest that the parser is sensitive to a potential antecedent DP intervening between the anaphor/pronoun and the embedded subject. We were also able to find some results suggesting that the intervening DP played a role in

[98] It is also important to note that the participants who completed Experiment 4a were recruited on Amazon Mechanical Turk while those who completed Experiment 4b were undergraduates from the University of Toronto community. The latter group could have contained more proficient readers and they are all from a similar demographic. This fact seems supported by the fact that participants were more accurate on the CQs in Experiment 4b, compared to Experiment 4a. In Experiment 4a, the accuracy was at about chance, even when the potential antecedent mismatched the pronominal in gender. The low accuracy in Experiment 4a could thus be explained by lower proficiency readers. In addition, since Experiment 4b used the eye-tracking methodology, participants were able to re-read earlier parts of the sentence. This is not an option in self-paced reading. Thus, even if there were less proficient readers who completed Experiment 4b, the methodology allowed them to re-read the sentences for comprehension. This could explain the higher accuracy in this version of the experiment.

our CQ results from Experiment 4a but we were unable to find real-time effects of the intervener. We have attributed the differences found between the two experiments to the fact that we used two different experimental methodologies (self-paced reading and eye-tracking). The goal of Experiment 5 was to investigate whether we can find a gender manipulation effect using the self-paced reading methodology. This would enable us to compare the results to those reported in previous chapters since our previous experiments also used self-paced reading.

In the two versions of Experiment 4, the intervening DP was embedded in a prepositional phrase adjunct appearing in between the *wh*-filler phrase and the embedded subject. This PP adjunct might have seemed unnatural to participants and might have caused processing difficulty on its own, before the participants had even tried to process the DP appearing within the adjunct phrase. Since we did not find any reading time effects after the word following the DP in the filler phrase (word 9) in Experiment 4a, participants may have found the sentences too difficult to process and may have simply stopped trying to interpret the sentences.[99] The CQ results from this experiment also suggest that participants were not engaging in a deep level of processing.

In an attempt to remedy the overall processing difficulty associated with the adjuncts in Experiment 4a, in Experiment 5, we used a different type of adjunct phrase, namely a gerund, which modified the picture DP. As a result, the syntactic attachment site of the adjunct was clear and we hoped that this would alleviate processing difficulty, i.e., because participants would not be trying to figure out where the phrase needs to attach in the structure. Sample stimuli are shown in (7). The intervening adjuncts, which were

[99]If participants found the sentences too complex to parse, this would mean our results from Experiment 4a (and maybe Experiment 4b) would not be interpretable. If we were to replicate some of the effects found in Experiments 4a-4b in our follow-up studies, this would suggest that the earlier effects are actually interpretable.

always gerund phrases, contained a genitive R-expression, which either matched or did not match the pronoun or anaphor in gender, compare (7-a)-(7-c) to (7-b)-(7-d). The sentence templates were otherwise similar to those used in Experiments 4a-4b.

(7) The reporters wondered ...

a. how many lies about himself$_i$ crashing Sean's car you asked Jeff$_i$ to deny

b. how many lies about himself$_i$ crashing Shelly's car you asked Jeff$_i$ to deny

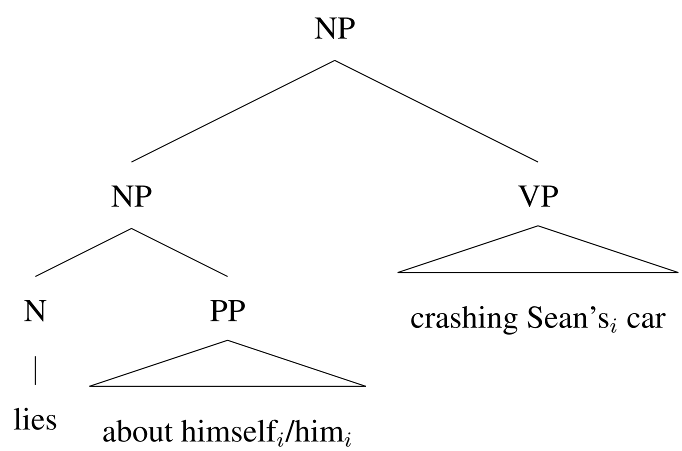
c. how many lies about him$_i$ crashing Sean's car you asked Jeff$_i$ to deny

d. how many lies about him$_i$ crashing Shelly's car you asked Jeff$_i$ to deny

... for the TV interview.

Since we changed the type of adjunct, the DP consisted of a different structure, as shown in (8). For simplicity, we will assume that the gerund phrase is a VP but it could also be analyzed as a NP. In either case, the gerund could now form its own phasal node, following Bruening (2014).

(8)



Consider again the definition of *phase-command* repeated below as (9).

(9) a. Phase command: X phase-commands Y iff there is no ZP, ZP a phasal node such that ZP dominates X but does not dominate Y.

b. Phasal nodes: CP, vP, NP

In the structure in (8), if the anaphor/pronoun is co-indexed with the R-expression in the gerundial adjunct, we still predict a Principle C violation, even though the R-expression is found in its own phasal domain. This is because the pronominal and the R-expression are still within the same larger phrase (here, NP). Consequently, the anaphor/pronoun still phase commands the R-expression and this should incur a Principle C violation if the two DPs are co-indexed. Thus, even though we have changed the structure of the intervening adjunct, the pronominal and the R-expression are still in the same structural relationship and our predictions remain the same.

4.4.1.1 Predictions

The predictions are the same as in Experiment 4a. In the intervening adjunct position, if the parser is sensitive to both morphological features and syntactic position, we predicted an effect of gender matching at least in the anaphor condition. While the DP in the adjunct matches the anaphor in gender, it is not found in a structural position from which it could bind the anaphor, even under a reconstruction analysis. For pronoun conditions, we predicted that the parser might consider the DP in the intervening adjunct as a potential antecedent when it matched the pronoun in gender.[100] Even if the parser reconstructs the phrase containing the pronoun to a structurally lower position, this DP would not c-command the pronoun at LF and therefore, not violate Principle B. If the embedded subject

[100]Crucially, for this analysis to be possible, the complement of *wh* containing a pronoun must reconstruct to a structural position that is (at least) below the embedded subject. If the pronoun were to be interpreted in its surface position and co-indexed with the DP in the adjunct, this would incur a Principle C violation.

is the only gender-matched DP in the sentence and if the parser prefers binding relations over co-reference relations, it should reconstruct the phrase containing the pronoun to a structural position below the embedded subject. Co-indexation should lead to a Principle B violation, forcing the parser to adopt a co-reference over a binding relation.[101]

4.4.1.2 Materials

Thirty-two sentence templates were created following the paradigm shown in (7). Target sentences were counterbalanced across four lists of stimuli combined with 32 filler sentences. The filler sentences were constructed to resemble the target sentences. In total, participants read and judged 64 stimulus sentences. Comprehension questions were asked after each sentence to ensure that participants were paying attention and comprehending the sentences as intended. As in the eye-tracking study (Experiment 4b), in order to encourage participants to interpret the anaphor/pronoun, we asked comprehension questions directly targeting participants' interpretation of the anaphor or pronoun. More precisely, we asked questions of the type, *Who were the lies about?* and presented

[101] Our results are compatible with this analysis but we are aware of a potential counterexample:

(i) ?The reporters wondered how many lies about him$_i$ crashing Shelly's car you asked every student$_i$ to deny.

It seems possible for the pronoun to be co-referential with the quantified phrase in the embedded subject position. As a result, a variable binding analysis is possible in this sentence. To be able to derive this example, the complement of *wh* containing the pronoun must be able to reconstruct below the embedded subject at LF. Since we did not use quantified phrases in our studies, we are unable to comment further on this example. One potential explanation, suggested by Martin Hackl, is that the pronoun might act as a specified subject in the DP constituent, as in (ii). This would mean that the DP could be analyzed as its own clause. If the phrase reconstructed below the embedded subject, we would no longer incur a Principle B violation since the pronoun and the embedded subject would be in separate clauses.

(ii) lies about [him crashing Shelly's car]

An interesting follow up study would be to investigate our stimulus sentences but using quantified phrases as embedded subjects, instead of R-expressions.

participants with two names as possible answers, e.g., *Sean, Jeff*. We counterbalanced the order in which the options were presented to participants: in half of the trials, the names were presented in the same order as the order in which they appeared in the experimental sentence and in the other half of the trials, the names were presented in the opposite order. Twenty comprehension questions directly targeted participants' interpretation of the anaphor/pronoun, six questions asked about the possessor of the DP in the adjunct and six questions asked about the embedded subject.

4.4.1.3 Participants

Ninety-nine participants completed the combined self-paced reading and naturalness judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). One participant was excluded because they did not complete the task as instructed and another participant was excluded because their demographic questionnaire revealed that they were not of the age of majority. Eleven participants were excluded because they scored less than 60% on the comprehension questions across the whole experiment. The questions that asked participants about their interpretation of the pronominal in the filler phrase were not counted towards participants' comprehension accuracy since the answers to these questions depended on how participants interpreted the anaphor/pronoun. Since we had so many of these types of questions, we did not trim based on accuracy on the critical trials alone. Two participants were excluded because they lost more than 20% of their reading time data after we trimmed out outliers. We report the data for the remaining eighty-four participants (47 males, 37 females). All participants were self-reported native speakers of English, aged between 20 and 60 years of age (M = 33.7 years, SD = 8.9). Participants were paid $4.50 USD for their participation.

4.4.1.4 Procedure

The experimental procedure was the same as in Experiment 4a. Participants were given feedback after responding to the comprehension questions.[102]

4.4.2 Naturalness judgement results

4.4.2.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 5 fillers were removed from the analysis: 4 due to low accuracy and one due to a programming error. Accuracy on the comprehension questions for the test trials was reasonable at 85.2% and 81.2% for all trials, after trimming.[103] In order to only investigate judgements on trials where participants were interpreting the sentence as intended, trials where CQs were answered incorrectly were removed from the analysis.[104]

4.4.2.2 Statistical analyses

We used the same statistical analyses as described for the previous experiments with the same trimming procedures. We report here the final trimmed models. All statistical analyses for the rating data were calculated on the z-scored data across all trials.

Mean z-scored naturalness judgements are plotted in Figure 4.11. There was a reliable

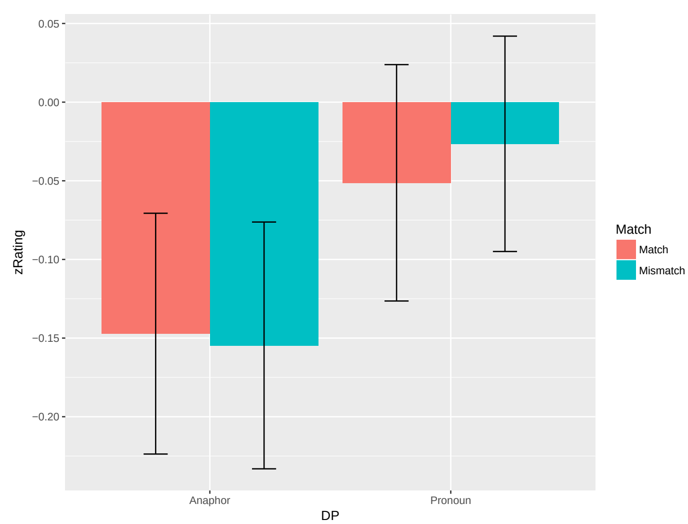
[102] Note that we had only coded the embedded subject as the correct answer to the binding questions, thus encouraging participants to reconstruct the complement of *wh* to this position. We had originally assumed that the DP in the adjunct could not be an appropriate antecedent for either pronominal, especially in cases where it mismatched the pronominal in gender. Based on the results of the comprehension questions, it does not seem like this feedback influenced participants' answers to these questions.

[103] Note that we must proceed cautiously when considering the accuracy for the test items because this score only includes 12 items. The items in which we asked about the interpretation of the anaphor/pronoun were not included when calculating accuracy.

[104] We kept all trials in which the question asked about co-indexation. We investigate these results in detail below.

main effect of DP type such that conditions containing pronouns were more likely to be rated higher than conditions containing anaphors (β = 0.101, SE = 0.038, t = 2.664, $p <$ 0.01), replicating effects found in previous studies. There was no reliable effect of Match nor was there a reliable interaction between DP and Match.

Figure 4.11: Mean naturalness judgements (in z-scores) for DP type (anaphor, pronoun) and Match (match, mismatch) in Experiment 5. Error bars represent 95% confidence intervals.



4.4.3 Comprehension questions

As in the other experiments, we were also interested in how participants were interpreting the anaphor or pronoun. We probed their interpretation of the pronominals using comprehension questions that directly asked about who the DP described in the picture DP referred to, e.g., *Who were the lies about?* The mean percentage of responses where participants selected the embedded subject DP as the referent for the anaphor/pronoun is plotted in Figure 4.12. Across anaphor and pronoun conditions, participants selected the embedded subject as the antecedent for the pronominal 53.9% of the time. We found no significant effects of either DP type or Match, suggesting that the type of DP found in the filler phrase did not affect how participants answered the CQs nor did the gender manipulation in the intervening adjunct.

We did, however, find a significant interaction between DP type and Match (β = 0.56, SE = 0.056, z = 9.998, $p < 0.001$), as plotted in Figure 4.13. We interpret the interaction as demonstrating that while participants were just as likely to choose the embedded subject as the antecedent for the anaphor in both gender-matched and gender-mismatched conditions, participants were more likely to choose the embedded subject as the antecedent for the pronoun in gender-mismatched conditions, compared to gender-matched conditions. This suggests that participants were more sensitive to the gender manipulation in pronoun conditions, compared to anaphor conditions. When the DP in the adjunct mismatches the pronominal in gender, the embedded subject should be the only possible within-sentence antecedent. The interaction shows that accuracy was higher in pronoun mismatch conditions, suggesting that participants were more accurate in this condition. Nevertheless, participants did not seem to show a preference for selecting a particular antecedent for either the anaphor or pronoun and the means are low overall. We will first present the results from the reading portion of the experiment before providing possible explanations for these results.

Figure 4.12: Percentage of responses in which participants selected the embedded subject DP as the antecedent by DP type (anaphor, pronoun) and Match (match, mismatch). The Match variable refers to whether or not the DP in the adjunct matched or mismatched the anaphor/pronoun in gender. Error bars represent 95% confidence intervals.

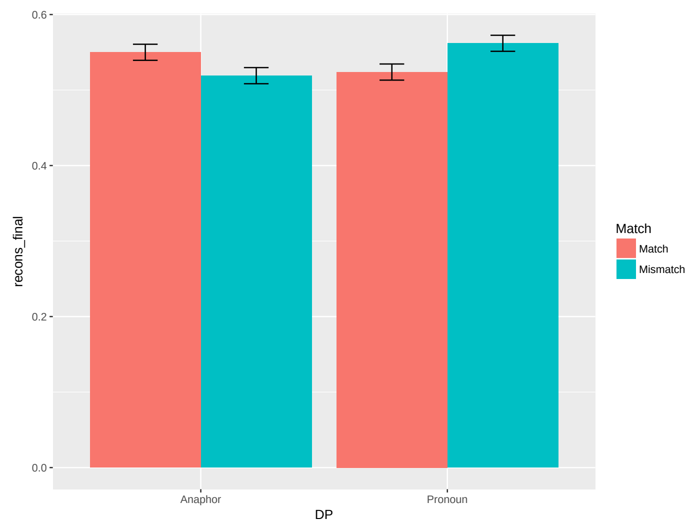
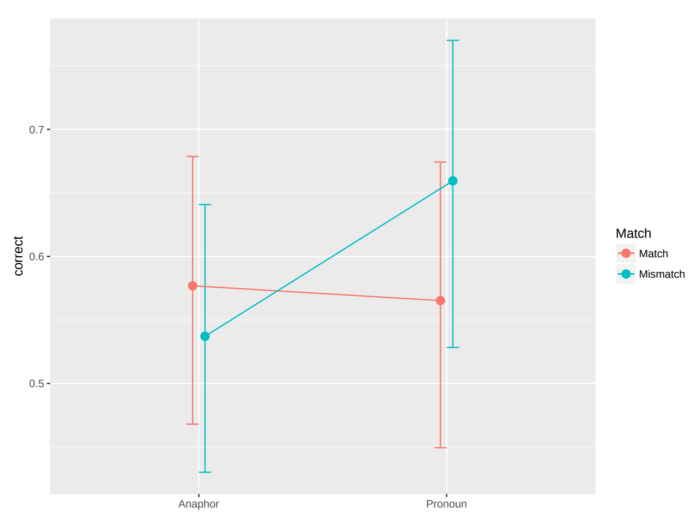


Figure 4.13: Interaction between DP type and Match in Experiment 5. The Match variable refers to whether or not the DP in the adjunct matched or mismatched the anaphor/pronoun in gender. Error bars represent 95% confidence intervals.



4.4.4 Self-paced reading experiment

4.4.4.1 Outliers

We removed outliers in the same way as outlined in previous experiments. As in previous studies, data points that were +/- 2 SDs away from the mean by participants were removed (less than 6% of data across all transformations). We also only investigated trials where participants answered the CQs correctly. Since we were interested in their interpretation of the anaphor/pronoun, when the CQ asked about co-indexation, both answers were accepted.

4.4.4.2 Reading time results

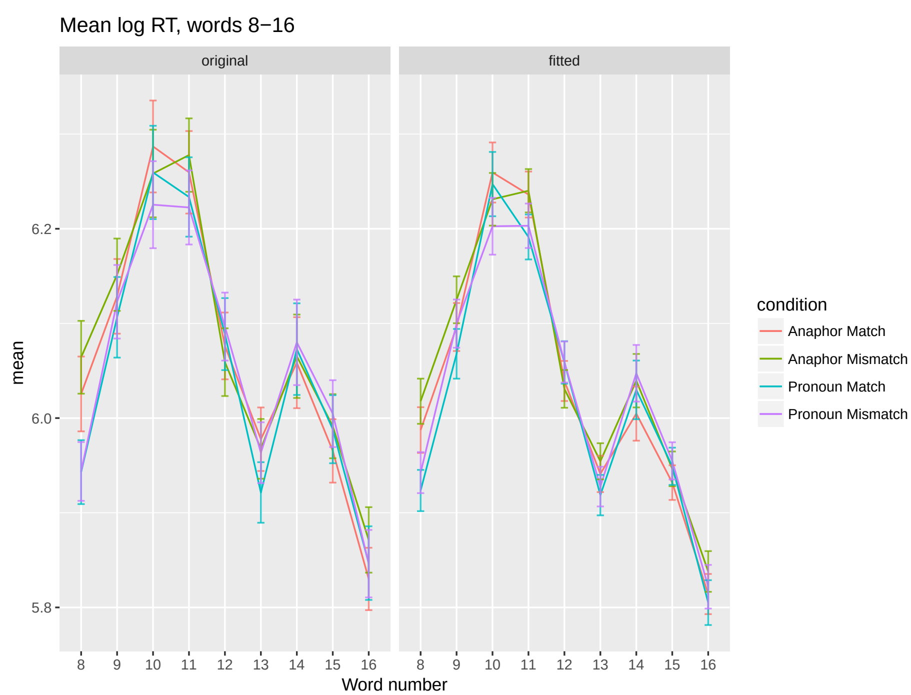
In this experiment, the DP in the filler phrase occurred on word 8, the intervening DP in the adjunct appeared on word 10 and the embedded subject appeared on word 14. Mean log reading times for words 8-16 are plotted in Figure 4.14.[105] There were significant effects of the type of DP at word 9,[106] on the word after the pronoun/anaphor in the raw (β = -14.29, SE = 6.93, t = -2.062, p = 0.043) residual (β = -13.32, SE = 6.57, t = -2.029, p = 0.043) and log reading times (β = -0.03, SE = 0.13, t = -2.482, p = 0.013, models not shown). At word 10, when the DP in the adjunct is introduced, we found significant effects of Match such that mismatch conditions were read faster than match conditions (raw: β = -39.25, SE = 15.86, t = -2.475, p = 0.02, residual: β = -36.57, SE = 15.95, t = -2.293, p = 0.025, log (marginal): β = -0.037, SE = 0.02, t = -1.839, p = 0.069, models not shown). However, there was no significant effect of DP type at word 10, suggesting that participants are sensitive to

[105] Models on words 8-9 were run only on the DP variable since the Match variable does not come into play until word 10. Models were run using both variables as predictors from word 10 onwards.

[106] Upon inspection of the graph, the type of DP also appears to show a strong effect. This effect is not reliable, however, since it is not significant in the residual reading times, which take into consideration the number of characters in the word. This suggests that this effect is dependent on length of the word and not on the type of DP.

whether the DP in the adjunct matches or mismatches the anaphor/pronoun or gender but that the Match effect does not differ based on the type of DP found in the filler phrase. Since we find an effect based on whether or not the DP in the adjunct matches the pronominal in gender (i.e., gender matching incurs a cost), this suggests that the parser is sensitive to the structural configuration in which the intervening DP appears. The DP matches the pronominal in gender which passes the first filter. It is an appropriate antecedent based on morphological features. However, it is not structurally appropriate, which is a factor the parser might only consider after it has deemed the DP appropriate morphologically. Since it has passed the morphological filter but not the syntactic filter, a processing cost is incurred on gender-matched DPs.

Figure 4.14: Mean log reading times on words 8-16 in Experiment 5 by DP Type and Match. Original means are plotted in the left panel and fitted values are plotted in the right panel. Error bars represent 95% confidence intervals.



We found no significant main effects of either DP type or Match on words 11 to 13 in the raw or residual reading times. In the log reading times, we found significant effects of DP type on word 11, one word following the DP in the intervening adjunct such that pronoun conditions were read faster than anaphor conditions (β = -0.035, SE = 0.022, t = -2.387, p = 0.017). On word 12, the effect of DP type reaches marginal significance, demonstrating that pronoun conditions were read longer than anaphor conditions (β = 0.022, SE = 0.013, t = 1.765, p = 0.079).

In the log reading times, on word 13, one word before the embedded subject, we found that pronoun conditions were again read faster than anaphor conditions (β = -0.027, SE = 0.011, t = -2.424, p = 0.016). We also found a significant interaction between DP type and Match at this region (raw: β = 18.47, SE = 9.75, t = -1.89, p = 0.06, residual: β = 19.27, SE = 9.68, t = -1.99, p = 0.05, models not shown). The model for the log reading times is shown in Table 4.2 and the interaction at this region is plotted in Figure 4.15. As we can see in the plot, while reading times for anaphor conditions were similar for both gender-matched and gender-mismatched conditions, pronoun conditions were read longer in the gender-mismatched conditions, compared to the gender-matched conditions. We interpret this effect as showing that participants were more sensitive to the Match manipulation in pronoun conditions, which we had also found in our comprehension data. Recall that the Match manipulation affected the DP in the intervening adjunct phrase. Since this DP is never a structurally appropriate antecedent for the anaphor, it follows that this manipulation might not affect reading times for the anaphor conditions. In contrast, since this DP might be considered a potential antecedent for pronouns, gender-mismatched conditions might have been read longer because the parser was expecting a gender-matched DP in this position. An important question is why this effect appears so late. The DP in the intervening adjunct is found on word 10 but this effect appears three words afterwards. One possible reason for why the effect might appear further to the right is that these constructions are difficult to process and the parser might still be interpreting the adjunct phrase for several words after it is initially read. A related question we might ask is why it shows up so *early*, i.e., *before* the embedded subject. Since the DP in the adjunct did not match the pronoun in gender, the parser might have been anticipating that the embedded subject would match the pronoun in gender. Note that this effect could also be interpreted as reflecting the parser's preference to adopt a binding relation over co-reference, again replicating the effects found in Experiment 2. What seems to be most relevant is that the interaction is driven by the pronoun conditions and not by the anaphor conditions, suggesting that the pronoun conditions are more affected by the gender of the DP found in the intervening adjunct, even at a later stage of processing.

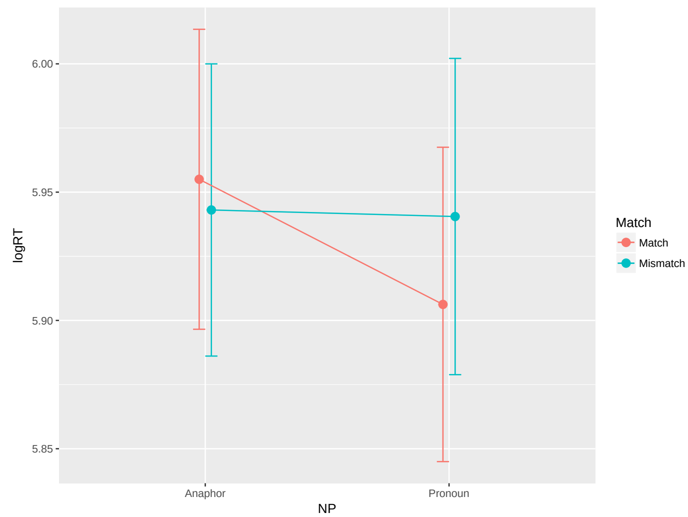
Table 4.2: Final mixed-effects model for log reading times at word 13, one word before the embedded subject (N = 2452 before trimming, 2398 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.955	0.298	199.77	0.000
Pronoun	-0.049	0.0165	-2.964	0.003
Mismatch	-0.012	0.018	-0.0656	0.514
Pronoun*Mismatch	0.0462	0.023	2.049	0.043

Random effects: by-participant intercept (SD = 0.251), by-item intercept (SD = 0.023), by-participant random slope for DP type (SD = 0.052), the correlation between by-participant random intercept and slope for DP type (r = 0.17), by-participant random slope for Mismatch (SD = 0.09), the correlation between by-participant random intercept and slope for Mismatch (r = -0.26), by-participant random slope for DP type * Mismatch (SD = 0.58) and the correlation between random intercept and slope for the interaction (SD = 0.27).

We did not find any reliable effects on the embedded subject itself (word 14) or on the word following the embedded subject (word 15). As a result, we did not replicate

Figure 4.15: Interaction between DP type and Match at word 13 in the log reading times in Experiment 5. Error bars represent 95% confidence intervals.



the effects found in Experiments 1-2 in this version of the experiment. On word 16, two words following the embedded subject, there was a significant interaction between DP and Match in the residual reading times (β = -18.07, SE = 9.15, t = -1.975, p = 0.048, model not shown), showing that while pronoun conditions were read similarly in both Match and Mismatch conditions, anaphor conditions were read longer in Mismatch conditions compared to Match conditions. We do not find this effect in any other transformations and as a result, do not deem it to be a reliable effect.[107] We found no other reliable effects at the 5% threshold in this experiment.

[107]It is also unclear how to interpret this effect. In Mismatch conditions, the embedded subject is the only feature-matched DP in the sentence. This effect therefore might be interpreted as demonstrating that the parser has reconstructed the complement of *wh* so that the embedded subject can structurally bind the anaphor. The question is why we do not also see this effect in Match conditions. If the parser abides by Principle A, the embedded subject should be the only structurally accessible antecedent in the sentence. It is also unclear why this effect is only found in the residual reading times and not in any of the other transformations. We therefore put this effect aside for the purposes of this thesis.

4.4.5 Experiment 5 Discussion

In Experiment 5, we aimed to investigate whether we could find effects of gender-matched DPs intervening between a pronominal embedded in the filler phrase and the embedded subject further to the right in the sentence using the self-paced reading methodology. To make the sentences easier to process, we used gerundial adjunct phrases with embedded gender-matched and gender-mismatched DPs instead of prepositional phrase adjuncts.

4.4.5.1 Comprehension Questions

We also manipulated the types of comprehension questions asked after participants had read the critical sentences. More precisely, in the majority of the critical trials (20/32), participants were directly asked about their interpretation of the anaphor/pronoun. The possible answers to these comprehension questions were the names of the two DPs in the sentences, thus forcing participants to make a decision about how they are interpreting the pronominal. Unfortunately, our CQ results did not reveal any significant differences based on the type of DP found in the complement of *wh* or on the gender of the DP found in the intervening adjunct. Overall, participants chose the embedded subject as the antecedent for the pronominal roughly half of the time (see Figure 4.12), even in cases where it was the only gender-matched antecedent within the sentence. This result is surprising, especially in the gender-mismatched cases. We did find an interaction between DP type and Match, suggesting that the pronoun conditions were more affected by the gender manipulation. More precisely, the results showed that participants were more likely to choose the embedded subject as the antecedent for the pronoun when the DP in the adjunct mismatched the pronoun in gender (see Figure 4.13). In contrast, the likelihood of selecting the embedded subject as the antecedent for the anaphor was not affected by the gender of the DP in the adjunct. At first glance, this might seem to support the hypothesis that anaphors always need to find structural antecedents and in our stimuli, the only structurally appropriate antecedent was the embedded subject. However, the overall means of selecting the embedded subject as the antecedent were quite low, suggesting that participants do not show a preference with respect to which DP should serve as the antecedent for the pronominal. Our reading time results suggest that participants were sensitive the Match variable used in this experiment so it is currently unclear why we do not also observe such effects in the comprehension questions. In Experiments 6a and 6b, we also investigate potential gender-matching effects when DPs are found in subject positions. After reporting those results in the next section, we will discuss possible explanations for the CQ results observed in Experiment 5.

4.4.5.2 Reading time results

Crucially, we did observe reading time effects in this version of the experiment, suggesting that the gerundial adjunct phrases made the sentences easier to process compared to the PP adjuncts. We replicated the DP effect found one word following the anaphor/pronoun, demonstrating that anaphor conditions incur a greater processing cost compared to pronoun conditions in this region. We also found a Match effect on the DP in the intervening adjunct showing that gender-matched conditions were read longer than gender-mismatched conditions. This effect replicates the gender-match effect observed in Experiment 4b. This effect again suggests that the parser is able to easily dismiss the gender-mismatched DP as a potential antecedent for the pronominal but it considers the gender-matched DP as a potential antecedent. An effect of gender-matching is expected if the parser finds a DP that matches the pronominal in gender but it is not found in structurally accessible position.

However, we did not observe differences at this word based on the type of DP found in the filler phrase. This suggests that both types of DPs were affected by the gender-matching effect.

We did observe a DP type effect on the word following the DP in the adjunct (word 11), showing that pronoun conditions were read faster than anaphor conditions.[108] One possible explanation for this effect is that in the pronoun conditions, the DP in the adjunct is a possible antecedent. Thus, there is less of a processing cost in the pronoun conditions once the DP in the adjunct has been found. However, in the anaphor conditions, the DP in the adjunct is not an appropriate antecedent and thus, the parser must keep looking for an antecedent for the anaphor further to the right in the string. Since the parser has not yet found the antecedent, a processing cost is incurred.

It is also worth mentioning that this same DP effect (i.e., pronoun conditions read faster than anaphor conditions) is also found on word 13 but we also found a significant interaction between DP type and Match at this word. This result shows that while pronoun conditions were generally less costly than anaphor conditions, the pronoun conditions show a difference between the gender-matched and gender-mismatched conditions at this word region. More precisely, the gender-matched conditions were read faster than the gender-mismatched conditions, Overall, this suggests that the gender-matched conditions incurred *less* of a processing cost for the pronouns. We interpret this finding as demonstrating that the parser was able to find an antecedent for the pronoun in the intervening adjunct when the DP matched the pronoun in gender and that this process alleviated the processing load in these conditions. However, since the anaphor conditions were read similarly for both gender-matched and gender-mismatched conditions, this suggests that there is no reliable

[108]Note that this effect only reached significance in the log reading times and we are therefore cautious about our interpretation of these results

difference in processing load in gender-matched versus gender-mismatched conditions when the complement of *wh* contained an anaphor. This could be taken as evidence that the parser adopts a structural account of antecedent resolution for anaphors.

4.4.5.3 Summary: Experiments 4a, 4b and 5

To sum up, in Experiments 4a, 4b and 5, we investigated whether a gender-matched or gender-mismatched potential antecedent DP embedded inside an adjunct intervening between the pronominal and its embedded subject would have an effect on how the parser finds antecedents for anaphors and pronouns. Overall, the results suggest that the gender-matched intervening DP had more of an effect in the pronoun conditions, compared to the anaphor conditions. We interpret this as suggesting that the parser considered this DP as a potential antecedent for the pronoun but not for anaphors. Thus, the parser does seem to be sensitive to the structural restrictions imposed on anaphors with respect to their antecedents. Unfortunately, the results from our comprehension questions directly targeting participants' interpretation of the anaphor/pronoun do not clearly show whether the parser has a preference about which DP it prefers to associate with the anaphor/pronoun. We currently do not have an explanation for this finding.[109] We further investigate how participants interpret the pronominals in Experiments 6a and 6b.

[109] One possibility is that our stimuli were too difficult to process. Since our stimulus sentences were being presented to participants word-by-word, participants needed to keep the previous words and previously built structure in memory as they continued to process the sentence. This process would require quite a few memory resources and may have made it difficult for participants to keep track of the different possible referents in the sentence. A second related possibility is that participants were simply not able to comprehend the sentences as we intended. Since their comprehension of our critical sentences was poor, as shown in their comprehension scores, participants were unable to determine which DP was the appropriate antecedent for the anaphor/pronoun. This likely means that they were not fully interpreting the anaphor/pronoun. If they were not fully interpreting the anaphor/pronoun, we would not expect to find differences in reading times based on this variable. Thanks to Ivona Kučerová for bringing up this point.

4.5 Experiment 6

4.5.1 Design

In Experiments 4a, 4b and 5, we investigated how the parser processes constructions in which there are two DPs in the sentence that could be potential antecedents for anaphors/pronouns. As in our previous studies, the embedded subject always matched the pronominal in relevant phi-features. The second potential antecedent DP always occurred in a linearly closer syntactic position and it either matched or did not match the anaphor/pronoun in gender. Crucially, this DP was always embedded in an adjunct phrase that could not structurally bind the anaphor/pronoun, even under a reconstruction analysis. The only DP that could therefore be in a binding relation with the anaphor/pronoun was the embedded subject DP.

Thus far, we have ignored the fact that there is in fact another DP in our stimulus sentences that is found in a position in which it could structurally bind the anaphor/pronoun under a reconstruction account. This DP is the subject of the asking event. In all of the previous experiments reported in this thesis, the subject of the asking event was always the pronoun *you* and it was therefore not an appropriate match for the third person pronominals. The goal of Experiment 6 was to investigate whether participants consider this DP as a potential antecedent for the pronoun or anaphor when it bears appropriate gender features. Assuming successive cyclic movement of *wh*-phrases (Chomsky, 1995), there is an intermediate reconstruction position that is structurally below the subject of the asking event. Thus, if the parser reconstructs to this position, it should be possible for the pronoun/anaphor to enter a binding relation with this DP.

In the two versions of Experiment 6, we manipulated whether the subject of the asking event matched or did not match the anaphor/pronoun in gender features. We used this manipulation because both DPs, i.e., the subject of the asking event and the embedded subject, could both c-command the phrase containing the anaphor/pronoun, provided the parser can reconstruct to a position that is structurally lower than the DP. If the parser reconstructs to the intermediate copy position, the subject of the asking event could bind the pronominal at LF. However, if the parser reconstructs to the base-generated position (i.e., the lowest copy position), the embedded subject could bind the pronominal at LF. We ran two versions of Experiment 6 to investigate whether the parser can reconstruct to an intermediate position. We will briefly explain the two versions of this experiment and the manipulations used in each version before reporting our results.

4.5.1.1 Experiment 6a

In Experiment 6a, we manipulated the gender of the subject DP of the asking event, as shown in (10). As in the other experiments reported in this chapter, we included intervening adjuncts appearing in between the picture DP and the asking event. However, unlike the other experiments, the intervening adjunct in the two versions of Experiment 6 did not contain a DP that the parser might consider as a potential antecedent for the pronominal. In Experiment 6a, the subject DP of the asking event either matched the pronominal in gender, as in (19)-(21), or did not match the pronominal in gender, as in (20)-(22). In this version of the experiment, the embedded subject DP always matched the pronominal in gender.

(10) The broadcasters wondered ...

a. how many lies about himself$_i$ crashing a Rolls-Royce Sean had asked Jeff$_i$ to deny

b. how many lies about himself$_i$ crashing a Rolls-Royce Shelly had asked Jeff$_i$ to deny

c. how many lies about him$_i$ crashing a Rolls-Royce Sean had asked Jeff$_i$ to deny

d. how many lies about him$_i$ crashing a Rolls-Royce Shelly had asked Jeff$_i$ to deny

... for the TV interview.

4.5.1.2 Experiment 6b

In Experiment 6b, we manipulated the gender of both subject DPs in our stimulus sentences, meaning that there was only ever one DP in the sentence that matched the pronominal in gender. Sample stimuli are shown in (11). Either the subject of the asking event matched the pronominal in gender and the embedded subject did not, as in (11-a)-(11-c), or the subject of the asking event mismatched the pronominal in gender and the embedded subject matched the pronominal in gender, as in (11-b)-(11-d). Thus, in this version of the experiment, there was only gender-matched antecedent for the pronominal within each sentence. If the parser must reconstruct to the lower position, i.e., below the embedded subject, it should not be possible for the anaphor to co-refer with the subject of the asking event. However, if the parser can reconstruct to an intermediate position, the subject of the asking event should be a potential antecedent for the anaphor, provided they match in gender.

(11) The broadcasters wondered ...

a. how many lies about himself$_i$ crashing a Rolls-Royce Sean$_i$ had asked Shelly to deny

b. how many lies about himself$_i$ crashing a Rolls-Royce Shelly had asked Jeff$_i$ to deny

c. how many lies about him$_i$ crashing a Rolls-Royce Sean$_i$ had asked Shelly to deny

d. how many lies about him$_i$ crashing a Rolls-Royce Shelly had asked Jeff$_i$ to deny

... for the TV interview.

4.5.1.3 Predictions

The predictions are different for the two versions of Experiment 6. In Experiment 6a, the embedded subject DP always matched the anaphor/pronoun in gender (see (10)). The relevant manipulation was whether the subject of the asking event also matched the pronominal in gender. We predicted that if the parser could reconstruct to an intermediate position, it should be possible to associate the anaphor with the subject DP of the asking event, provided they match in gender. If there is a reconstruction below that is structurally lower than the subject DP of the asking event, the anaphor would be structurally bound by the DP at LF. Assuming that phrases containing pronouns also reconstruct, it should not be possible for the pronoun to be bound by this DP because reconstruction to the intermediate position should incur a Principle B violation. In such cases, it should be possible for the parser to reconstruct the phrase to the lower syntactic position, i.e., a position that is structurally lower than the embedded DP subject, and for the pronoun to still be co-referential with the subject DP of the asking event. Another possibility is that the parser reconstructs to the lower syntactic position, i.e., to a position that is structurally lower than the embedded subject. Under a structural analysis of binding, this should be the only option for the conditions in which the embedded subject is the only gender-matched DP in the sentence and the filler phrase contains an anaphor, e.g., (20). However, this should also be a possibility when the sentence contains two gender-matched DPs, provided the lower reconstruction position is available to the parser.

For the pronoun conditions, we should observe different effects of Principle B depending on the reconstruction position of the phrase containing the pronoun. If the parser can reconstruct to the intermediate position and there is a gender-matched antecedent in this position, this should incur a Principle B violation. Thus, we predict a reanalysis effect since the parser would need to adopt a co-reference interpretation over a binding interpretation. However, if the parser reconstructs to the lower position, i.e., structurally below the embedded subject, it can still enter a binding relation with the subject DP of the asking event and this parse would not incur a Principle B violation (since they would not be within the same local domain). In cases where the only gender-matched DP in the sentence is the embedded subject, we predict the same effect as what we found in previous studies. In such cases, if the parser reconstructs the phrase containing the pronoun to a position that is structurally lower than the embedded subject, this should incur a Principle B violation. Thus, we predict a reanalysis effect at this lower position since the parser must adopt a co-reference interpretation over a binding relation between the two DPs. The gender manipulation used in this experiment therefore allows us to investigate the question of which reconstruction positions are available to the parser in a direct way.

The predictions for Experiment 6b are similar to those from Experiment 6a except that in Experiment 6b, there is only one gender-matched DP in each sentence (see (11)). Consequently, there is only one DP in each sentence that bears the same gender features as the pronominals. If the parser can reconstruct to intermediate positions when the DP in that position bears the correct morphological features, it should be possible for the anaphor conditions to always find structural antecedents within the stimulus sentences. Similarly, if the parser prefers binding relations over co-reference relations and therefore reconstructs to a structural position in which the DP can bind the pronoun, we predict reanalysis effects for our pronoun conditions, depending on the position to which the parser reconstructs. For example, in (11-c), if the parser reconstructs to the intermediate position structurally below the subject of the asking event (*Sean*), this should incur a Principle B violation. The parser can either reconstruct the phrase further to the right so that the pronoun and the antecedent are not in the same local domain or it can adopt a co-reference relation between the two DPs. Similarly, in (11-d), if the parser reconstructs the phrase containing the pronoun below the embedded subject, this should incur a Principle B violation and it will need to adopt a co-reference relation over a binding relation. This should incur a processing cost.

4.5.1.4 Materials

Thirty-two sentence templates were created for both versions of the experiment following the paradigms shown in (10) and (11). Target sentences were counterbalanced across four lists of stimuli combined with 32 filler sentences. The filler sentences were constructed to resemble the target sentences. In total, participants read and judged 64 stimulus sentences. Comprehension questions were asked after each sentence to ensure that participants were paying attention and comprehending the sentences as intended. As in Experiment 5, in order to encourage participants to interpret the anaphor/pronoun, we asked comprehension questions directly targeting participants' interpretation of the anaphor or pronoun. More precisely, we asked questions of the type, *Who were the lies about?* and presented participants with two names as possible answers, e.g., *Sean,* *Jeff*. Twenty comprehension questions directly targeted participants' interpretation of the anaphor/pronoun, six questions asked about the matrix subject, i.e., *broadcasters*, and six questions asked about the embedded subject. In Experiment 6a, comprehension questions that asked about the DP in the filler phrase in sentences in which both DPs matched the pronominal in gender were not coded as correct or incorrect since participants' answers to these questions depended on how they interpreted the sentence. If only one of the DPs matched the pronominal in gender, that DP was coded as the correct answer to the comprehension question. Thus, in Experiment 6b, all comprehension questions were assigned a correct answer. We will go through the results for each version of the experiment separately and then summarize the results from both experiments.

4.5.2 Experiment 6a

4.5.2.1 Participants

Eighty participants completed the combined self-paced reading and naturalness judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). One participant was excluded because they did not complete the task as instructed. Nine participants were excluded because they scored less than 60% on the comprehension questions across the whole experiment. Two additional participants were excluded because they scored less than 55% on the test trials alone.[110] Two participants were excluded because they lost more than 20% of their reading time data after we trimmed out outliers. We report the data for the remaining sixty-six participants

[110] This cut off might seem arbitrary but a higher cut off of 60% would have eliminated 3 participants who scored 59% on the test trials. In this version of the experiment, we had comprehension questions which did not have correct answers and therefore there were several questions that were not included in participants' overall score.

(29 males, 37 females). All participants were self-reported native speakers of English, aged between 18 and 58 years of age (*M* = 33.3 years, *SD* = 9.03). Participants were paid $4.50 USD for their participation.

4.5.2.2 Procedure

The experimental procedure was the same as in Experiments 4a and 5. Participants were given feedback for the co-indexation questions only when there was only one DP in the sentence that matched the pronominal in gender. If both DPs matched in gender, either answer was accepted.

4.5.3 Results

4.5.3.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 5 fillers were removed from the analysis. Accuracy on the comprehension questions for the test trials was reasonable at 80.4% and 80.6% for all trials, after trimming.[111]

4.5.3.2 Statistical analyses

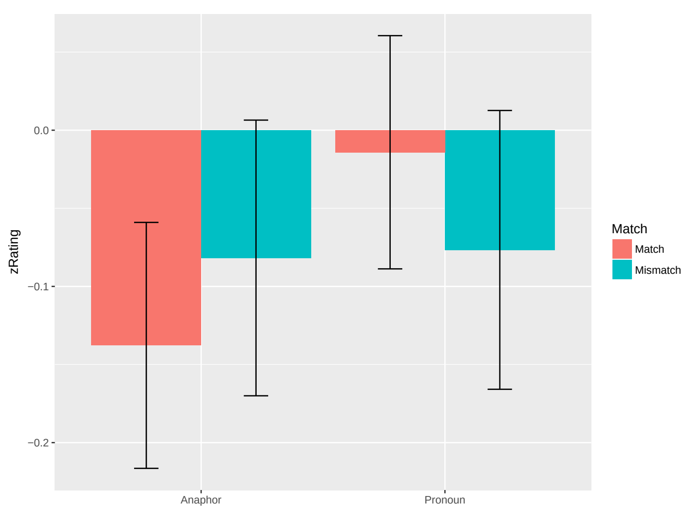
We used the same statistical analyses as described in the previous experiments with the same trimming procedures. We report here the final trimmed models. All statistical analyses for the rating data were calculated on the *z*-scored data across all trials.

[111] If the sentence contained two gender-matched DPs and the CQ asked about co-indexation, both answers were accepted. However, if only one of the DPs matched the pronominal in gender, only the gender-matched DP was coded as correct. The critical results remain the same if we accepted both answers as correct, even in cases where one of the DPs mismatched the pronominal in gender. As a result, we report the results based on the more conservative data set.

4.5.3.3 Naturalness judgements

Mean z-scored naturalness judgements are plotted in Figure 4.16. Conditions containing pronouns were marginally more likely to be rated higher than conditions containing anaphors (β = 0.07, SE = 0.039, t = 1.99, p = 0.051). There was no reliable effect of Match nor was there a reliable interaction between DP and Match. These results suggest that in Experiment 6a, the type of DP found in the filler phrase had an effect on participants' naturalness ratings such that they found anaphor conditions more difficult to process compared to pronoun conditions but that whether or not the DP subject of the asking event matched or did not match the pronominal in gender did not affect how they rated the sentences. Assuming that lower ratings reflect a higher processing cost, these results suggest that the gender manipulation did not affect the processing cost of these sentences.

Figure 4.16: Mean naturalness judgements (in z-scores) for DP type (anaphor, pronoun) and Match (match, mismatch) in Experiment 6a. In this experiment, the Match variable refers to whether or not the DP subject of the asking event matched or did not match the pronominal in gender. Error bars represent 95% confidence intervals.



4.5.3.4 Comprehension questions

We also asked comprehension questions targeting participants' interpretation of the anaphor/pronoun. In cases where there was only one gender-matched DP in the sentence (which was always the embedded subject in this version of Experiment 6), we coded that DP as the correct answer. Figure 4.17 shows the percentage of responses in which participants selected the antecedent for the pronominal that would abide by structural constraints, e.g., binding theory. To be explicit, we assumed that the parser can always reconstruct the phrase containing the pronominal to the position below the embedded subject. In order to abide by Principle A, it is always possible for the parser to choose the embedded subject as the antecedent for the anaphor, even in cases where there is another gender-matched DP in the sentence. If the parser is also able to reconstruct to an intermediate copy position, below the subject of the asking event, it might also be possible for the parser to reconstruct to this intermediate position and satisfy Principle A, provided the anaphor and the subject of the asking event match in gender. However, since we did not know if the parser would be sensitive to the intermediate reconstruction site, we coded co-indexation with the embedded subject as the correct answer. To abide by Principle B, the parser should choose the DP subject of the asking event when it matches the pronoun in gender. If it chooses the embedded subject as the antecedent for the pronoun and reconstructs the complement of *wh* to adopt a binding relation between the two DPs, this would violate Principle B since the two DPs would be within the same local domain.[112] In cases where there is only one gender-matched DP in the sentence (the embedded subject), the parser will have no choice but to select this DP as the antecedent for the pronoun. Under a reconstruction analysis, this should violate Principle B so the parser will need to adopt a

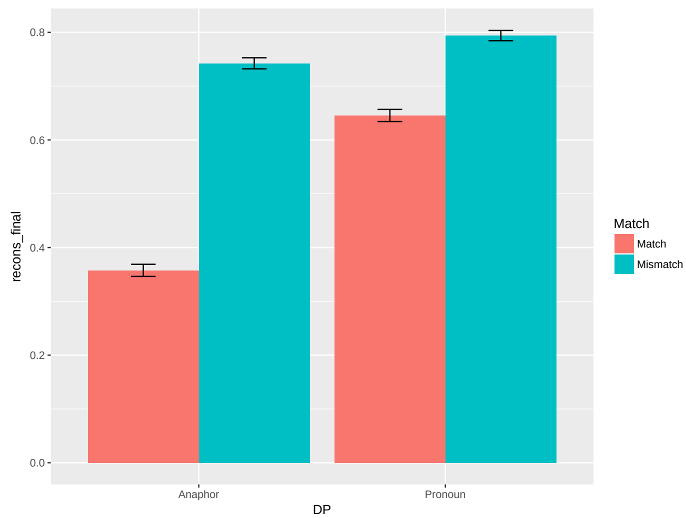
[112]This parse would be fine if the parser adopts a co-reference relation between the two DPs but such an analysis violates the parser's preferences and should thus only be pursued as a last resort.

co-reference relation over a binding relation.

Upon visual inspection of the mean responses, accuracy on the CQs was high in the gender-mismatching conditions. This result suggests that participants were sensitive to the gender manipulation and associated the pronominal with the gender-matching DP in the sentence the majority of the time. When both the subject of the asking event DP and the embedded subject DP matched the pronominal in gender, the results suggest that in pronoun conditions, there is a preference for co-indexing the pronoun with the subject of the asking event. In such cases, it still seems possible for the parser to adopt an interpretation in which the pronoun and the embedded subject DP are co-indexed but this is not the preferred interpretation. Interestingly, the results also suggest that the parser can adopt an analysis in which the anaphor and the subject of the asking event are co-indexed. In fact, this seems to be the preferred interpretation in gender-matched conditions.

We ran generalized logistic regression models on the comprehension data. The model revealed main effects of both DP type and Match. The main effect of DP type revealed that participants were more likely to select the structurally appropriate antecedent DP in pronoun conditions compared to anaphor conditions (β = 1.73, SE = 0.47, t = 3.71, $p <$ 0.001). The main effect of Match revealed that participants were more likely to select the structurally appropriate antecedent DP in mismatch conditions compared to match conditions (β = 3.42, SE = 0.48, t = 7.2, $p < 0.001$). These results suggest that participants were aware of the gender mismatch manipulation and that they were more likely to choose the DP that matched the anaphor/pronoun in gender as its antecedent, as expected. The DP results suggest that participants were more likely to choose the subject of the asking event as the antecedent for the pronoun when both DPs in the sentence matched the pronoun in gender. This is the parse predicted by the structural account of binding at LF since

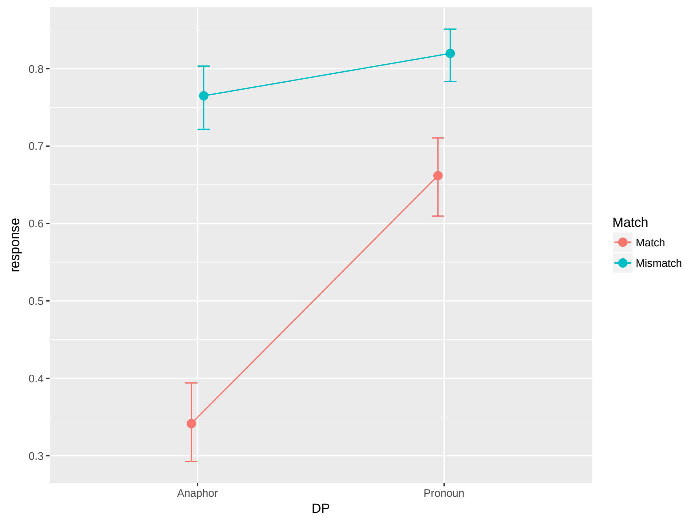
Figure 4.17: Percentage of responses in which participants selected the structurally appropriate DP as the antecedent for the anaphor/pronoun in Experiment 6a. For anaphor conditions, this means selecting the embedded subject in both gender-matched and gender-mismatched conditions. For pronoun conditions, this means selecting the subject of the asking event in gender-matched conditions and the embedded subject in gender-mismatched conditions. In the pronoun conditions, it is also possible to choose the embedded subject in the gender-matched conditions but since this analysis would violate the parser's preference for binding, we assumed that it was not the preferred parse of the sentence. Error bars represent 95% confidence intervals.



the subject of the asking event and the pronoun would be in separate binding domains at LF. Although these results support our predictions for the pronoun cases, the anaphor conditions suggest that the parser only abided by structural constraints when there was only one gender-matched DP in the sentence. When both the subject of the asking event and the embedded DP subject matched the anaphor in gender, participants preferred to co-index the anaphor with the subject of the asking event. This parse is not expected if the parser must reconstruct the phrase containing the anaphor to the lower reconstruction position, i.e., below the embedded subject, but it is expected if it is possible for the parser to reconstruct the phrase containing the anaphor to an intermediate position below the subject of the asking event.

We also observed a significant interaction between DP type and Match, as plotted in Figure 4.18. The interaction demonstrates that while participants were more likely to choose the structurally appropriate DP in the mismatch conditions (i.e., the DP that appeared in the embedded subject position) compared to the match conditions, the difference between match and mismatch conditions was stronger in the anaphor conditions compared to the pronoun conditions (β = -0.99, SE = 0.06, t = -17.45, $p < 0.001$, model not shown). This means that the anaphor match conditions seem to be driving this effect. More precisely, this result suggests that participants are able to associate the anaphor with the subject of the asking event when the two DPs match in phi-features. Assuming that the parser abides by structural constraints, this result provides support for an intermediate reconstruction site that is structurally lower than the subject of the asking event. If the parser reconstructs to an intermediate copy position below the subject of the asking event, the intermediate subject would c-command the complement of *wh* and this would satisfy Binding Principle A.

Figure 4.18: Interaction plot showing the effect of DP type by Match in the comprehension questions asking about co-indexation in Experiment 6a. Error bars represent 95% confidence intervals.



It is important to note that these results suggest that participants generally preferred to co-index the pronominal with the first DP in the sentence, i.e., the subject of the asking event, provided this DP matched the pronominal in gender. It is possible that this preference is simply due to the fact that this is the first DP that the parser finds in the sentence and that this effect is not due to structural considerations. Note this effect is similar to the effects reported by Frazier et al. (1996) where they also found that participants preferred to interpret an anaphor in a *wh*-phrase as referring to the matrix subject, instead of the embedded subject. The results from the reading portion of the task will enable us to investigate whether the comprehension results point to a structural explanation or whether they may be attributable to the parser choosing the linearly closest DP as the antecedent for the pronoun/anaphor.

4.5.4 Self-paced reading experiment

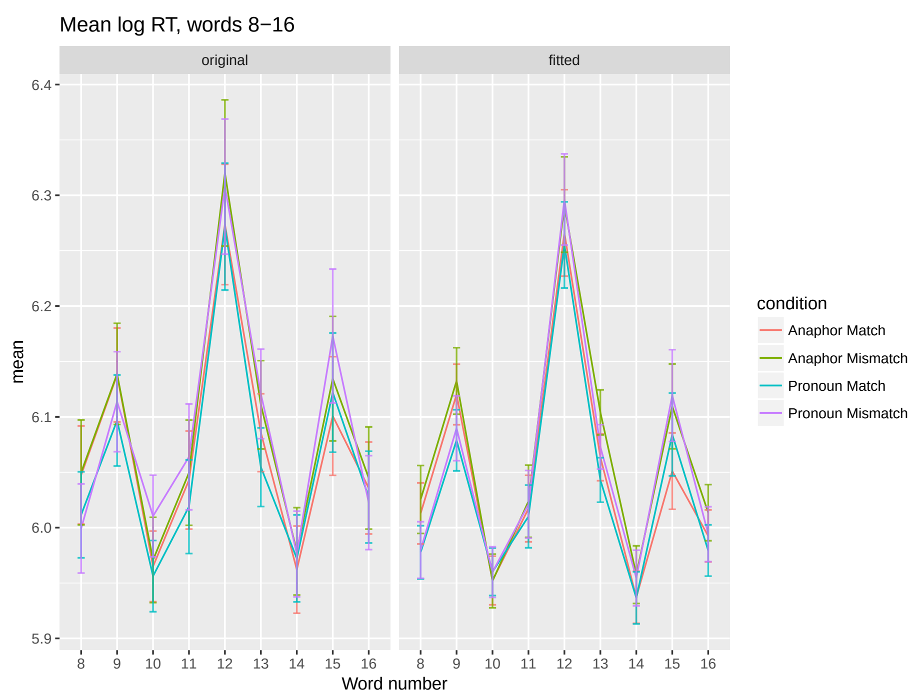
4.5.4.1 Outliers

We removed outliers in the same way as outlined in previous experiments. Trimming data points that were +/- 2 SDs away from the mean by participants removed less than 6% of data across all transformations.

4.5.4.2 Reading time results

In this experiment, the DP in the filler phrase occurred at word 8, the subject of the asking event appeared at word 12 and the embedded subject appeared at word 15. The mean log reading times on words 8-16 by DP type and Match are plotted in Figure 4.19.

Figure 4.19: Mean log reading times on words 8-16 in Experiment 6a by DP Type and Match. Error bars represent 95% confidence intervals.



There were significant or marginal effects of the type of DP at word 9, the word after the pronoun/anaphor in the raw (marginal: β = -16.26, SE = 8.55, t = -1.901, p = 0.062), residual (marginal: β = -17.16, SE = 9.18, t = -1.87, p = 0.066) and log reading times (β = -0.04, SE = 0.17, t = -2.41, p = 0.019, models not shown), such that pronouns were read faster than anaphors. At word 13, one word following the subject of the asking event, we found significant effects of both DP type (raw (marginal): β = -11.7, SE = 6.08, t = -1.926, p = 0.054)[113] and Match (raw: β = 15.56, SE = 7.65, t = 2.034, p = 0.046, residual: β = 14.32, SE = 7.806, t = 1.834, p = 0.07), indicating that pronoun conditions were read faster than anaphor conditions but that mismatch conditions were read longer than match conditions. Table 4.3 shows these effects for the log reading times at this word. Note that this Match effect is the opposite of what we found in Experiments 4b and 5. One reason for why this could be the case is because in the current experiment, the subject of the asking event could be a potential antecedent for the pronominal but in Experiments 4b and 5, the DP in the adjunct could not be a potential antecedent for the anaphor. If there is an intermediate reconstruction position below the subject of the asking event, this might explain why mismatch conditions were read longer than match conditions. The parser was expecting to find an antecedent for the pronominal in this position but instead finds a DP that mismatches the pronominal in gender. The parser must therefore continue looking further to the right in the string for an appropriate antecedent. No other main effects were significant at the 5% threshold. We also did not find any significant interactions between DP Type and Match.

[113]This effect does not reach significance in the residual reading times.

Table 4.3: Final mixed-effects model for log reading times at word 13, one word after the subject of the asking event (N = 1766 before trimming, 1721 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	6.064	0.316	192.13	0.000
Pronoun	-0.028	0.0133	-2.115	0.035
Mismatch	0.031	0.016	1.989	0.051

Random effects: by-participant intercept (SD = 0.236), by-item intercept (SD = 0.027), by-participant random slope for Mismatch (SD = 0.06), the correlation between by-participant random intercept and slope for Mismatch (r = -0.39).

4.5.5 Experiment 6b

4.5.5.1 Participants

Eighty-one participants completed the combined self-paced reading and naturalness judgement study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). Two participants were excluded because they did not complete the task as instructed. Four participants were excluded because they scored less than 60% on the comprehension questions across the whole experiment. Four additional participants were excluded because they scored less than 60% on the test trials alone.[114] We report the data for the remaining seventy-one participants (33 males, 38 females). All participants were self-reported native speakers of English, aged between 18 and 68 years of age (*M* = 34.2 years, *SD* = 10.2). Participants were paid $4.50 USD for their participation.

4.5.5.2 Procedure

The experimental procedure was the same as in Experiments 4a, 5 and 6a.

[114] In this experiment, there was only one gender-matched DP in each stimulus sentence. Consequently, for the co-indexation questions, we coded the answer corresponding the gender-matched DP as the correct answer. There was thus a correct answer for each sentence.

4.5.6 Results

4.5.6.1 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 5 fillers were removed from the analysis. Accuracy on the comprehension questions for the test trials was reasonable at 82.5% and 80.1% for all trials, after trimming.

4.5.6.2 Statistical analyses

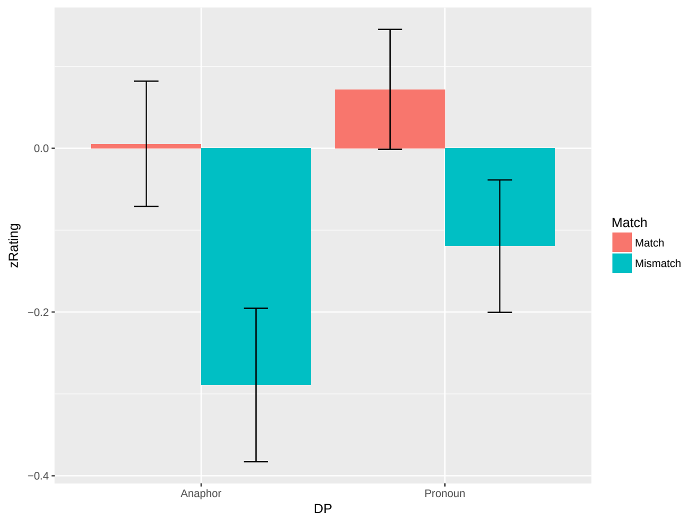
We used the same statistical analyses as described in the previous experiments with the same trimming procedures. We report here the final trimmed models. All statistical analyses for the rating data were calculated on the z-scored data across all trials.

4.5.6.3 Naturalness judgements

Mean z-scored naturalness judgements are plotted in Figure 4.20. In this version of the experiment, we found significant effects of both DP type and Match. Conditions were coded as matched if the subject of the asking event matched the pronominal in gender. Conditions containing pronouns were marginally more likely to be rated higher than conditions containing anaphors (β = 0.08, SE = 0.041, t = 1.85, p = 0.069). Conditions in which the subject of the asking event mismatched the pronominal in gender were more likely to be rated lower than conditions in which this DP matched the pronominal in gender (β = -0.22, SE = 0.053, t = -4.17, $p < 0.001$). There was no reliable interaction between the two factors. As in other experiments, these results suggest that conditions containing anaphors were more difficult to process than conditions containing pronouns. The Match effect suggests that participants were sensitive to the gender manipulation used in the experiment and more precisely, they found that conditions in which the subject of the asking event mismatched the anaphor/pronoun in gender were more difficult to process than conditions in which it matched the anaphor/pronoun in gender. Thus, participants seem to prefer constructions in which the subject of the asking event is the antecedent for the pronominal.

What is interesting about these results is that we used this same manipulation in Experiment 6a but did not find a significant effect of Match in our naturalness ratings. The difference between the two experiments was the gender of the embedded subject. In Experiment 6a, the embedded subject always matched the anaphor/pronoun in gender but in Experiment 6b, we also manipulated the gender of the embedded subject. One possible explanation for the difference in results is that even though finding a gender-matched DP as the subject of the asking event conformed to the parser's expectations in Experiment 6a, upon reaching another gender-matched DP in the embedded subject position, the parser needed to make a decision about which DP is the appropriate antecedent. In contrast, in Experiment 6b, there is only ever one gender-matched DP in the sentence so the parser is able to easily determine which DP is the appropriate antecedent for the anaphor/pronoun. Due to this lack of competition effects, the Match effect is more easily detectable in these results. We will return to the differences between the two experiments in the General Discussion section.

Figure 4.20: Mean naturalness judgements (in z-scores) for DP type (anaphor, pronoun) and Match (match, mismatch) in Experiment 6b. Error bars represent 95% confidence intervals.



4.5.6.4 Comprehension questions

In this experiment, we also asked comprehension questions targeting participants' interpretation of the pronominal in the filler phrase. In this version of the experiment, there was only one gender-matched DP in the gender, which either occurred as the subject of the asking event or as the embedded subject. The mean percentage of correct responses to the questions that directly asked about participants' interpretation of the pronominals are plotted in Figure 4.21. Since there was only one DP in the sentence that matched the pronominal in gender, a correct response meant that participants associated the pronominal with the gender-matched DP. For gender-matched conditions, it is the intermediate subject (subject of the asking event) that matches the pronominal but for the gender-mismatched conditions, it is the embedded subject. Accuracy on these questions in this version of the experiment was quite high. For the match conditions, i.e., conditions in which the subject of the asking event matched the pronominal in gender, participants answered the CQ correctly

87.5% of the time, compared to 75.5% of the time when the embedded subject DP was the only possible referent for the pronominal. Therefore, participants were fairly accurate in responding to these CQs. However, since there was only ever one gender-matched DP in the sentence in this version of the experiment, there was only one possible answer to each CQ. Since the match conditions had higher accuracy than the mismatch conditions, it seems as though the parser prefers to associate the anaphor/pronoun with the first DP it finds in the sentence. The accuracy for the mismatch conditions was 75%, indicating that participants answered 25% of the questions incorrectly, even though there was only one possible gender-matched antecedent within the sentence.

We ran generalized logistic regression models on the comprehension data. The mean comprehension scores by condition are plotted in Figure 4.21. There was a reliable main effect of Match (β = -1.7, SE = 0.05, t = -37.28, $p < 0.001$), indicating that participants were less likely to answer the CQ correctly if the pronominal and the DP subject of the asking event mismatched in gender. There was no reliable effect of DP type. There was also a reliable interaction between DP type and Match (β = 0.47, SE = 0.09, t = 5.47, $p < 0.001$), as plotted in Figure 4.22. The interaction shows that while CQs asked in Mismatch conditions were more likely to be answered incorrectly than those asked in Match conditions, this effect seemed stronger for pronoun conditions compared to anaphor conditions. This suggests that the gender manipulation affected the processing of the pronoun conditions more so than the anaphor conditions meaning that when the DP subject of the asking event mismatched the pronominal in gender, CQ scores were lower for pronouns than for anaphors. Similar CQ scores were found for both types of DPs when the subject of the asking event matched the pronominal in gender. One possible reason for this is that the parser had a stronger expectation of finding a gender-matched DP in the linearly closest subject position when the filler phrase contained a pronoun compared to an anaphor. We will discuss some possible reasons for why this might have been the case after we report the reading time data.

Figure 4.21: Percentage of correct responses to the CQs in Experiment 6b. The Match variable refers to the gender of the subject of the asking event. Error bars represent 95% confidence intervals.

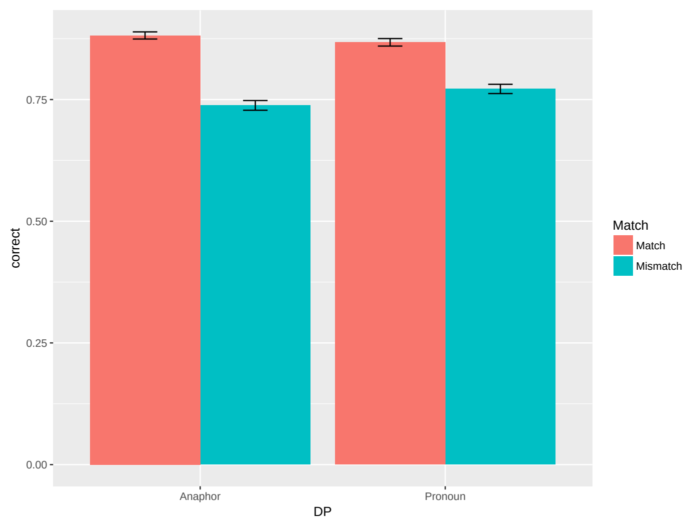
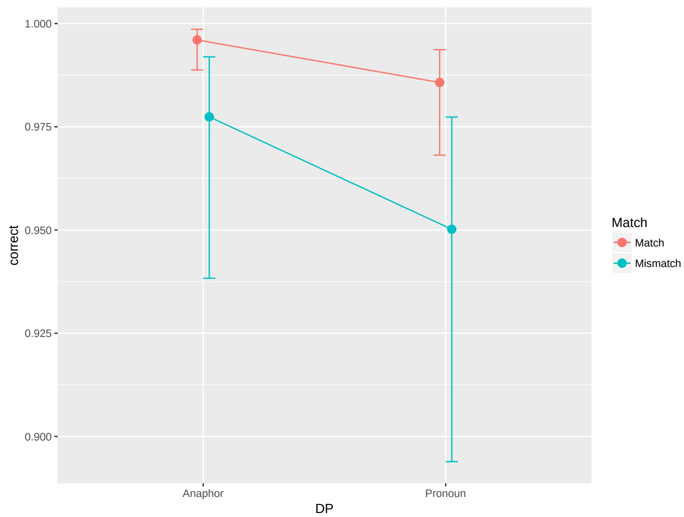


Figure 4.22: Interaction between DP type and Match in the comprehension data for Experiment 6b. Error bars represent 95% confidence intervals.



4.5.7 Self-paced reading experiment

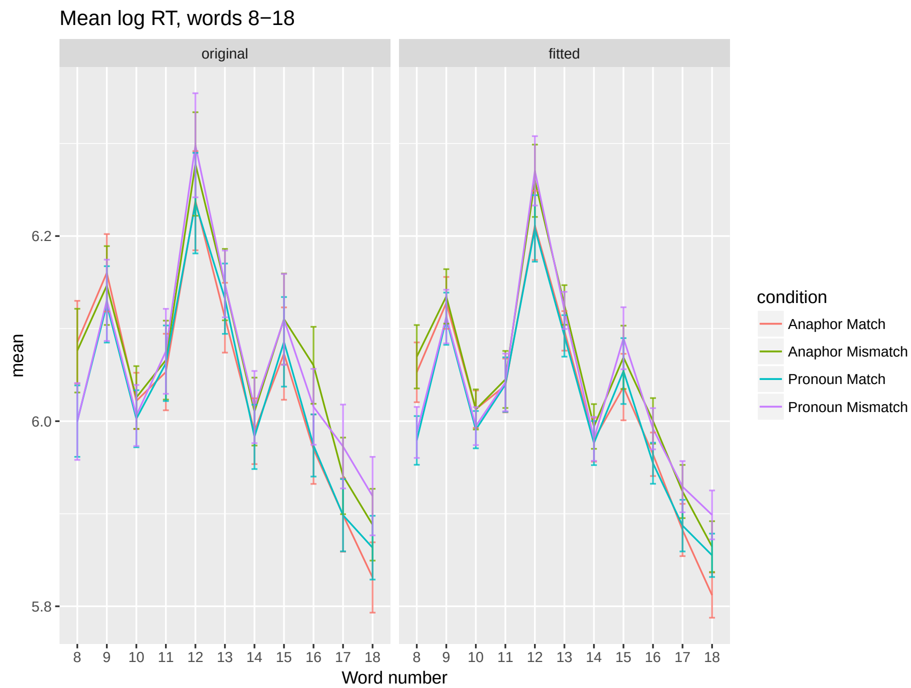
4.5.7.1 Outliers

We removed outliers in the same way as outlined in previous experiments. As in other experiments, reading times that were +/- 2 SDs from the mean by subject were removed (less than 5.5% of data across transformations).

4.5.7.2 Reading time results

In this experiment, the DP in the filler phrase occurred on word 8, the subject of the asking event appeared at word 12 and the embedded subject appeared on word 15. The mean log reading times on words 8-18 are plotted in Figure 4.23.

Figure 4.23: Mean log reading times on words 8-18 in Experiment 6b by DP Type and Match. Error bars represent 95% confidence intervals.



We found no reliable effects on words 8-11.[115] On word 12, when the subject of the asking event is introduced, mismatch conditions were read longer than match conditions (raw: β = 41.26, SE = 14.22, t = 2.902, p = 0.005, residual: β = 37.09, SE = 13.49, t = 2.75, p = 0.008, log: β = 0.05, SE = 0.02, t = 2.51, p = 0.015, models not shown). No reliable effects were found on words 13-15. On word 16, one word following the embedded subject, we again found that mismatch conditions were read longer than match conditions (raw: β = 24.72, SE = 10.04, t = 2.461, p = 0.017, residual: β = 12.89, SE = 5.51, t = 2.34, p = 0.02, log: β = 0.034, SE = 0.15, t = 2.23, p = 0.03, models not shown). Note that in this case, the embedded subject *matches* the pronominal in gender (it is the subject of the

[115] Upon inspection of the graph, one might think that there is an effect of DP type at word 8. As in other experiments, this effect is not significant in the residual reading times, suggesting that the number of characters in the word (pronoun vs anaphor) is driving this effect.

asking event that matches the pronominal in gender). This same effect was also found on word 17, two words following the embedded subject (raw: β = 14.96, SE = 5.71, t = 2.62, p = 0.01, residual: β = 11.59, SE = 5.41, t = 2.14, p = 0.032, log: β = 0.04, SE = 0.14, t = 2.51, p = 0.014, models not shown). On word 18, three words following the embedded subject, we again find this effect (raw: β = 17.81, SE = 6.74, t = 2.64, p = 0.01, residual: β = 17.16, SE = 6.94, t = 2.47, p = 0.016, models not shown). We also find a marginal effect of DP type, showing that pronoun conditions were read longer than anaphor conditions (raw: β = 10.26, SE = 5.71, t = 1.8, p = 0.077, residual: β = 11089, SE = 5.69, t = 1.95, p = 0.056, models not shown). The model for the log data on word 18 is found in Table 4.4. These effects disappeared at word 19. No interactions between DP type and Match were significant in this experiment.

Table 4.4: Final mixed-effects model for log reading times at word 18, three words after the embedded subject (N = 1827 before trimming, 1774 after trimming), reported as the regression coefficient estimates, standard errors, t-values and p-values.

Variable	Coefficient estimate	Std. error	t-value	p-value
Intercept	5.81	0.035	167.91	0.000
Pronoun	0.038	0.015	2.499	0.015
Mismatch	0.039	0.016	12.462	0.016

Random effects: by-participant intercept (SD = 0.276), by-item intercept (SD = 0.02), by-participant random slope for DP type (SD = 0.07), the correlation between by-participant random intercept and slope for DP type (r = -0.22), by-participant random slope for Mismatch (SD = 0.08), and the correlation between by-participant random intercept and slope for Mismatch (r = 0.01).

4.5.8 Discussion: Experiment 6

In Experiments 6a and 6b, we investigated how the parser processes constructions in which there are two DPs that could act as structural antecedents for pronominals. In the previous experiments reported in this chapter, we had investigated how the parser processes sentences in which there are two potential antecedents for pronominals but we ensured that one of these DPs was not found in a structural position in which it could bind the pronominal, even under a reconstruction analysis. In Experiments 6a and 6b, we used DPs that were both found in subject positions and could therefore bind the pronominal at LF, assuming that there is an intermediate reconstruction position that is structurally lower than the intermediate subject, i.e, the subject of the asking event. Crucially, the two DPs were found in two linearly distinct positions. The subject of the asking event was found in a position that was linearly closer to the pronominal compared to the embedded subject position.

4.5.8.1 Experiment 6a

In Experiment 6a, the embedded subject always matched the pronominal in gender. The relevant manipulation was whether the subject of the asking event also matched the pronominal in gender. The results from our comprehension task revealed that participants were more likely to choose the structurally appropriate antecedent when the subject of the asking event mismatched the anaphor/pronoun in gender, compared to when it matched the pronoun in gender. While these results suggest that the parser was sensitive to the gender manipulation, when the subject of the asking event mismatched the anaphor/pronoun in gender, there was only one gender-matched antecedent in the sentence (i.e., the embedded subject). These results therefore show that participants were sensitive to the gender manipulation used in the experiment but they do not tell us if the parser was abiding by structural constraints or not. We also found that participants were more likely to choose the structurally appropriate antecedent when the filler phrase contained a pronoun. While this result does support our hypothesis that the parser abides by structural constraints, if participants were more likely to choose the structurally appropriate antecedent for pronouns, this means that they were more likely to choose the first DP antecedent (i.e., the subject of the asking event) when both DPs matched the pronoun in gender. This result is compatible with a structural analysis but it is also compatible with the view that the parser prefers to satisfy open dependencies as soon as possible (see also Frazier et al., 1996). The reading time results on the word following the subject of the asking event (word 13) support a structural analysis in which the parser is expecting to find the antecedent for the pronoun in the first subject position. Overall, pronoun conditions are read faster since the parser was expecting to find the antecedent in this position and resolve the open dependency. Longer reading times are incurred when the intermediate subject DP mismatches the pronominal in gender since the parser must now continue looking for an antecedent and this violates the parser's expectations.

These results are also compatible with our interpretation of our Experiment 2 results. In Experiment 2, we argued that if the phi-features on the embedded subject match the phi-features on the pronoun, the parser attempts to adopt a binding relation between the two DPs. If binding is not possible (because adopting such a relation would incur a Principle B violation), the parser must adopt a co-reference analysis, which incurs a processing cost. In both versions of Experiment 6, a binding analysis between the pronoun and the intermediate is subject is possible, provided they match in gender. The parser can determine that this analysis is possible as soon as it reaches the verb *asked*. Since the parser can adopt its preferred analysis of the sentence (i.e., binding > co-reference), these conditions are read faster. A relevant question is why we observe longer reading times on the intermediate subject in anaphor conditions in cases where the intermediate subject matches the anaphor in gender. In these conditions, participants seem to be adopting an analysis in which the intermediate subject binds the anaphor, similarly to the analysis adopted in the pronoun conditions. However, pronoun conditions are read faster than anaphor conditions. One reason why this might be the case is because the parser must reconstruct the complement of *wh* containing an anaphor to a position that is structurally below the intermediate subject but still local to the intermediate subject (in order to abide by Binding Principle A). In contrast, in order to abide by Binding Principle B, the parser must reconstruct the complement of *wh* containing a pronoun to a structural position that is *not* local to the intermediate subject position. Crucially, the effect of DP type on the intermediate subject is only significant in Experiment 6a, where the embedded subject always matched the pronominal in gender and we manipulated whether the intermediate subject also matched the pronominal in gender. We did not replicate the DP type effect in Experiment 6b where there was only one gender-matching DP in the sentence, which either occurred in the intermediate or embedded subject position. One reason why this might be the case is because the structure of our sentences was very predictable in Experiment 6b. Thus, participants might have figured out that when the intermediate subject matched the anaphor in gender, the embedded subject would not be an appropriate antecedent. Overall, the results of both experiments seem to suggest that participants still have a preference to adopt an analysis in which the anaphor reconstructs to the embedded subject position and when it finds a matching DP in the intermediate subject position, a processing cost is incurred. Note that the effect of mismatch conditions incurring a greater processing cost compared to match conditions replicates previous results on cataphoric processing (see e.g., Van Gompel and Liversedge, 2003).

4.5.8.2 Experiment 6b

In Experiment 6b, we manipulated not only the gender of the DP subject of the asking event but also the embedded subject. Thus, for each sentence, there was only one gender-matched DP in the sentence. In our comprehension data, we found that participants were less likely to answer the question correctly in gender-mismatched conditions compared to gender-matched conditions. Recall that the gender-matched manipulation affected the subject of the asking event, i.e., the first possible DP antecedent in the sentence. This result suggests that the parser was expecting to find an antecedent for the pronominal in this position. We also observed an interaction between DP and Match in the comprehension data, showing that pronoun conditions were more affected by the Match manipulation than anaphor conditions. More precisely, participants were more likely to answer the CQ correctly in pronoun conditions when the subject of the asking event matched the pronoun in gender. This results suggests that the parser had a strong expectation about finding an antecedent for the pronoun in the first subject position. The reading time data showed that when the subject of the asking event was introduced (word 12), mismatch conditions were read longer than match conditions, suggesting that mismatch conditions incurred a greater processing cost. If the parser was expecting to find an antecedent for the pronominal in this position, this result is in line with that hypothesis. When the subject of the asking event mismatches the pronominal in gender, it is not an appropriate antecedent and the parser must continue looking for an antecedent, incurring a processing cost. Interestingly, we also found a gender effect later in the sentence, when the embedded subject was introduced. This effect again indicated that mismatched conditions were read longer than matched conditions. Note that at the embedded subject in Experiment 6b, in gender mismatched conditions, the embedded subject matches the pronominal in gender and is therefore the appropriate antecedent. We interpret longer reading times at this position as reflecting the parser's attempt to associate the pronominal with the embedded subject DP as its antecedent. These results seem to suggest that the parser has a general preference to find antecedents in linearly closer positions. When this is not possible, due to gender mismatch, a processing cost is incurred when an appropriate antecedent is located further to the right. The longer reading times in mismatch conditions in Experiment 6b could therefore be viewed as an effect of reconstruction. This effect replicates the effect found in Experiment 2, where we found that conditions containing a pronoun in the complement of *wh* were costly at the embedded subject. We have not previously observed such a cost when the complement of *wh* contains an anaphor. However, this was the first experiment where the embedded subject did not always match the anaphor in gender. Thus, the greater processing cost at the embedded subject might be reflective of reconstruction operations for binding.

4.5.8.3 Different effects in Experiments 6a and 6b

Why did we observe different effects in the two versions of Experiment 6, even though they both manipulated whether the subject of the asking event matched or did not match the anaphor/pronoun in gender? In Experiment 6a, the embedded subject always matched the pronominal in gender and participants could have learned this pattern and this might explain why we did not observe any differences in reading times at the embedded subject position. In contrast, in Experiment 6b, we also manipulated the gender of the embedded subject DP. Since the embedded subject DP could also mismatch the pronominal in gender, we were able to observe reading time differences based on this manipulation. The effect found in Experiment 6b suggests that processing difficulty is incurred later in the sentence, i.e., at the embedded subject site, when it matched the pronominal in gender. We attribute this effect to the parser establishing a co-referential relation between the pronominal and the embedded subject DP. We were not able to observe this effect in Experiment 6a because the embedded subject always matched the pronominal in gender.

4.5.8.4 Experiment 6 takeaway

Overall, the results from these experiments suggest that upon reaching a pronominal early in the linear string, i.e., in the filler phrase, the parser makes predictions about when it will find the antecedent for that pronominal further to the right in the string. If the first DP subject mismatches the pronominal in gender, processing difficulty is incurred. The parser must wait until it has built more structure to be able to associate the pronominal with an antecedent. These results are compatible with a structural account of referent resolution. Assuming that anaphors must always be found below their antecedents at LF, it follows that pronoun conditions should be processed faster when a gender-matched antecedent is found in the intermediate subject position (Experiment 6a). Phrases containing pronouns do not need to reconstruct but those containing anaphors must reconstruct below the antecedent, incurring a processing cost. We were unable to replicate this same effect in Experiment 6b, which might suggest that participants were able to predict where the gender-matched DP would be found in the sentence since there was only ever one gender-matched antecedent in the stimulus sentences in Experiment 6b.[116]

[116]Note that our CQ results from Experiment 6b still point to a difference between anaphors and pronouns at the subject of the asking event. We currently have no explanation for why we did not also observe these effects in the reading time data.

4.6 General Discussion and chapter conclusions

We began Chapter 4 asking how the parser might process constructions in which there is a second potential antecedent in the sentence. In the previous experiments reported in this thesis (Chapters 2 and 3), the stimulus sentences had only one DP that could be co-referential with the pronominals. Our goal in this chapter was to investigate how the parser finds an antecedent for a pronominal when there is another DP in the sentence that could be co-referential with the anaphor/pronoun.

4.6.1 Relevant manipulations: intervening DP in adjunct or subject position

In Experiments 4a, 4b and 5, we inserted potential antecedents DPs in adjunct phrases that intervened in between the filler phrase and the embedded subject. The DPs in the adjuncts were not in appropriate positions to be able to structurally bind the pronominals. Therefore, we predicted that this manipulation would enable us to investigate if the parser is sensitive to structural constraints on binding and co-reference, assuming that the DP in the adjunct is never a structurally appropriate antecedent for the pronominal. In Experiments 4a-4b, we used prepositional phrase adjuncts and in Experiment 5, we used gerundial adjuncts. In all cases, possessive DPs that either matched or did not match the pronominal in gender appeared in the adjunct phrase. Since these DPs were embedded in adjuncts, they were not found in appropriate structural positions to be able to bind the anaphor/pronoun at LF, even under a reconstruction analysis of binding. In other words, there was no reconstruction position that would place the phrase containing the anaphor/pronoun below the antecedent. In Experiments 6a and 6b, we also used the gender manipulation but we manipulated whether the gender of the subject of the asking event matched or mismatched the gender of the pronominal. These experiments presented us with new questions about how the parser determines antecedents for pronominals because the subject of the asking event could c-command the anaphor/pronoun at LF, provided the phrase containing the pronominal can reconstruct to an intermediate position.

4.6.2 Match effect depends on syntactic position of DP

We found different match effects based on the syntactic position of the DP. When the DP was found in an intervening adjunct, gender-matched DPs were read longer than gender-mismatched DPs. We interpret this effect as reflecting a violation of the parser's expectations. The parser was looking for an antecedent for the pronominal but was not expecting to find a possible antecedent in the intervening adjunct. When it finds a gender-matched DP, it must now consider whether this DP is a possible antecedent for the pronominal and this incurs a processing cost. However, when the DP occurred in a subject position, we found that gender-mismatched DPs were read longer than gender-matched DPs. We again interpret this effect as reflecting a violation of the parser's expectations. In this case, the parser was expecting to find a possible antecedent for the pronominal. Processing difficulty is incurred when the DP mismatches the anaphor/pronoun in gender. Note that this latter effect replicates the effects found in other gender manipulation studies with cataphors (see e.g., Van Gompel and Liversedge, 2003; Kazanina et al., 2007). Since the gender effects are different in the two sets of experiments and the difference between the two experiments is the syntactic position of the relevant DP, we argue that these effects reflect the parser's sensitivity to structural considerations. More precisely, the parser seems to be sensitive to the syntactic position of a potential antecedent DP and whether it occurs in a structural position in which it could serve as an antecedent for a pronominal (i.e., subject DP) or whether it cannot (i.e., because it is embedded in an adjunct).

4.6.3 How to account for the reported data: complexity metrics

The experiments reported thus far in this thesis suggest that different types of DPs embedded in *wh*-fillers are subject to different grammatical constraints. A consistent effect we have shown is that when DPs are initially introduced, they immediately incur different processing costs. More precisely, anaphors incur a greater processing cost than both pronouns and R-expressions and R-expressions incur a greater processing cost than pronouns.[117] The gender effects reported in this chapter suggest that the parser is also sensitive to the syntactic position of an antecedent DP. In the Match experiments, we did not observe that the different types of DPs (in these cases, anaphors and pronouns) show different processing costs across several different words in the sentence. We do not have a clear explanation for why this might be the case except that the gender effects might have been stronger than any effects of the DPs looking for antecedents.

In Chapter 5, we propose a series of complexity metrics to account for the results found in the processing studies reported in this thesis. These complexity metrics are all based on research in theoretical literature in the domain of referent resolution. More precisely, they are based on how different types of DPs (anaphors, pronouns and R-expressions) receive their interpretation in a given sentence. In addition, since all of our stimulus sentences were embedded *wh*-questions, we introduce complexity metrics to explain how the *wh*-filler phrase gets interpreted in real-time processing (as first discussed in Chapter 1). Moreover, we also consider that the left-to-right parser has certain general processing constraints that

[117] We ran an experiment to specifically investigate the cost associated with the DPs themselves (Experiment 7) which we report in Chapter 5.

may guide how it interprets sentences incrementally. We will use these complexity metrics to explain the processing data presented in this thesis. Since these complexity metrics are based on grammatical constraints, the left-to-right parser will be argued to be constrained by principles of the grammar, as proposed in the theoretical linguistics literature. We will show both where these metrics make the right predictions and where they fail to account for the observed data. The consequences of this proposal will be discussed in the next chapter.

Chapter 5

Accounting for the data: Complexity metrics

5.1 Introduction

In Chapters 2-4, we reported a series of psycholinguistic experiments designed to systematically investigate how the parser finds antecedents for different types of pronominals (anaphors, pronouns) when they linearly precede their antecedents. The different types of DPs were embedded in *wh*-phrases and the (potential) antecedent was always found further to the right in the string. We manipulated the syntactic position of the DPs in order to test for whether the parser is sensitive to a mismatch between the string of words it receives as input (roughly, the PF representation) and the LF structure it must derive in order for the string to be interpreted (*PF-LF mismatch*). We used R-expressions (proper names) as our baseline condition since they do not need to find linguistic antecedents in order to be interpreted. More precisely, we assumed that upon finding an R-expression in the complement of *wh*, the parser is able to assign the complement phrase a semantic denotation. In contrast, when the phrase is embedded with an anaphor, it cannot be assigned a semantic denotation since the interpretation of an an anaphor depends on its antecedent. If the phrase contains a pronoun, whether or not it can be interpreted in its surface position depends on the assignment function of the pronoun. If the pronoun is interpreted as a bound variable, its interpretation depends on its antecedent, meaning that it cannot be assigned a semantic denotation when it is encountered. However, if the pronoun is referential (a free pronoun), it receives its interpretation from the context. As a result, it can be assigned a semantic denotation and interpreted when it is encountered. Throughout the thesis, we have been working with the assumption that the parser is sensitive to the grammar's preferences and that it attempts to abide by such preferences whenever possible. Consequently, we have pursued a structural explanation of our results, arguing that when the grammar's preferences cannot be met, processing costs are incurred.

Following much previous work in psycholinguistics, we have interpreted longer reading times at particular regions in the string of words as reflecting a greater processing cost at that region. In the previous chapters, we have argued for different types of processing costs either arising due to a general processing preference (i.e., resolving open dependencies as soon as possible) or due to theoretical linguistic principles (i.e., binding theory constraints on the interpretation of different types of DPs). We have observed interesting interactions between these different processing costs. Generally, such interactions arise when the parser must choose between two different preferences or two different grammatical constraints. To explain these interactions, we have developed post-hoc arguments suggesting that the parser prioritizes certain constraints over others. In this chapter, we aim to further refine these costs and specifically show how the different types of processing costs can interact with each other. To achieve this goal, we propose a series of complexity metrics in this chapter. These complexity metrics are designed to be a first step towards a processing model that can account for effects of interpretation in real-time processing. Since the data reported in this thesis are complex and we have examined many different variables in our experiments, the metrics that we will propose in this chapter will not be able to account for all of the collected data. We will point out both where the metrics make the right predictions and where they fall short.

It is important to note that this is not the first attempt to develop a processing model to account for psycholinguistic data. However, the focus of previous processing models has been to explain effects of syntactic complexity (for instance, subject versus object relative clauses) in real-time processing. We will briefly discuss two of these models and show how they might be extended to the current work and where they make the wrong predictions or do not make any predictions for the type of data reported in this thesis.

5.2 Previous processing models

There have been two influential families of processing models previously proposed in the psycholinguistic literature: i) distance-based models (Dependency Locality Theory, Gibson, 1998, 2000) and ii) expectation-based models (Hale, 2001; Levy, 2008). We limit our discussion to these two families because they used similar experimental methodologies to test their predictions and thus can be easily compared to the results of the current thesis. These processing models also assume that costs are incurred on each word in the sentence as it is read from left-to-right, which is also one of our assumptions based on a left-to-right incremental parser. It is important to note that these previous models were developed to account for differences in structural complexity arising in the surface form of the sentence, i.e., at the PF level of representation. The goal of these models was to explain why certain syntactic constructions incur a greater processing cost at certain regions of the sentence but they were not concerned with how the parser arrives at an appropriate interpretation for such constructions. These models all take relative clause structure as their baseline and as such, are concerned with how the parser finds the gap site (or integration position) for a filler.[118] For example, consider the relative clause structures shown in (1).

(1) a. *Subject relative clause*
The man who$_i$ t_i saw the woman was very tall.
b. *Object relative clause*
The man who$_i$ the woman saw t_i was very tall.

The subject relative clause structure, found in (1-a), is argued to be easier to process compared to the object relative clause structure, shown in (1-b). One prevalent explanation for this effect is that the gap position for the relative pronoun, *who* is linearly closer in the subject relative clause structure compared to the object relative clause structure. Thus, the distance between the filler and its gap position have been argued to contribute to processing difficulty. The different models differ based on *where* they expect processing difficulty to arise. After briefly going over the relevant previous processing models and explaining how they can or cannot be extended to account for the current results, we will introduce a series of complexity metrics designed to account for the results observed in the previous chapters.

[118] Of course, these structures necessarily also involve a mismatch between the PF and LF forms of the sentences since the parser must locate a gap site for the filler to be able to integrate the filler into the structure. The reason we are suggesting that the effects of structural complexity arise in the surface form of the sentence is because the difference in complexity depends on the order of the words in their surface form, as we will discuss following the example.

5.2.1 Dependency Locality Theory

Dependency Locality Theory (DLT), developed by Gibson (1998, 2000), predicts greater processing difficulty (greater cost) to arise based on the number of *new* discourse referents that intervene between a filler and its gap site, or integration position. Upon finding a *wh*-filler, the parser must find its gap site in order to integrate that filler. A crucial notion for DLT is the *distance* between a filler and its gap site: the longer the parser has to wait to find a gap site and satisfy an open dependency, the greater the cost. Crucially, this theory predicts that the cost should arise on the discourse referents.[119] The more discourse referents found between the filler and its gap site, the greater the processing cost. In this theory, both verbs and nouns are considered discourse referents and thus, the theory predicts increased costs on each verb or noun the parser encounters in between the filler and its gap site.

In the domain of relative clause processing, DLT is able to account for why object relative clauses incur a greater cost compared to subject relative clauses. In the subject relative clause construction, there is no discourse referent intervening between the filler and its gap site. In contrast, in the object relative clause structure, there are two discourse referents appearing in between the filler and its gap site (*the woman* and *saw*). Due to these intervening discourse referents, object relative clauses are predicted to be more difficult to

[119] Gibson (1998) suggests that while all new words encountered in a particular string incur some cost, new discourse referents incur a greater cost. He argues that new discourse referents incur a greater cost than other words because once they have been introduced, they can be referred to later in the discourse using an anaphor, i.e., pronoun for DP and tense on verbs. As a result, discourse referents incur greater computational complexity because the parser must build a structure for that referent so that it can be referred to later in the discourse. Thus, Gibson (1998)'s model does take the semantics of discourse referents into account, which we also build into the proposed processing model. Even though DLT takes some semantic effects into account, the model does not consider semantic effects that might arise due to the parser not being able to interpret the structure and thus does not make correct predictions for our findings. As we will show in section 5.4, the complexity metrics proposed in this chapter are able to account for the effects predicted by DLT but they also predict some of the other effects observed in this thesis, which would not be accounted for under Gibson (1998)'s theory.

process compared to their subject counterparts. DLT predicts the processing cost associated with object relative clauses to arise at the embedded verb in the object relative clause structure, i.e., at the end of the relative clause.

At first glance, DLT might be extendable to account for some of the results reported in this thesis, specifically the parser needing to find a lower copy position for an anaphor to be interpreted. Since DLT only considers surface syntactic relations, it does not make predictions that differ based on the syntactic complexity of the filler phrase. DLT would only predict a difference in processing difficulty in the filler phrase if one type of DP introduced a new discourse referent but another did not. For example, if the filler phrase contained an anaphor that was co-referential with a previously introduced DP, it should be less costly than a new R-expression in the filler phrase. DLT does make straightforward predictions about where we would predict processing difficulty to arise.

Crucially, this theory predicts that we should observe a processing cost on new discourse referents that intervene between a filler and its gap position. In our stimuli, the gap position is found after the embedded verb; thus, each new discourse referent between the filler and the embedded verb should incur a processing cost. In our stimuli, we would only expect an increased cost on different types of DPs, provided they are new discourse referents, and on the verb *ask/asking*. DLT can therefore be used as a way to test for whether the parser considers given DPs as introducing new discourse referents or not. If the parser considers a given DP as introducing a new discourse referent, this DP cannot be in a binding or co-reference relation with the pronominal because if the two DPs were co-indexed, the DP would not introduce a new discourse reference.

In Experiments 4-6, we also observed differences based on the gender of the intervening DP. While DLT does not make a distinction between different types of DPs based on their gender, the phi-features on the DP can be used a way to control for whether or not that DP introduces a new discourse referent, i.e., if the phi-features do not match, the DP must introduce a new referent and it should incur a processing cost. We will further discuss where DLT makes the right predictions for our data and where it falls short when we discuss our proposed complexity metrics (section 5.4).

5.2.2 Expectation-based models

While there are several expectation-based models available in the processing literature, we will limit our discussion to two of them here: Hale (2001)'s surprisal theory and Levy (2008)'s expectation-based model. These models argue that greater processing difficulty arises when the incoming structure does not match the parser's expectations. These theories assume that as the parser receives more input (words), it revises its predictions about the upcoming structure. If the parser is only given a partial input, possible structural continuations are ranked in parallel. The difficulty of each incoming word depends on how many resources it takes for the parser to show that new word's effect on the current ranking of possible structures. If a subsequent word disconfirms a structure that had high probability (i.e., it was more expected), increased processing difficulty is expected (longer reading times). According to Levy (2008), as the sentence continues and the parser is given more input, the expectation for a particular word may increase. In such cases, processing difficulty should be alleviated and we would predict less of a processing load (faster reading times).

Evidence for these models has been shown using relative clause processing.[120] Following expectation-based models, processing difficulty in object relatives is predicted to arise at the

[120] Evidence has also been shown using many other constructions but we will limit our discussion to relative clause processing for the purposes of the thesis.

subject of the object relative clause because this is where the unexpected structure is found. Upon finding a relative pronoun, the parser expects to find a verb on the following word (corresponding to the subject relative clause structure). When it finds a DP instead of a verb, processing difficulty is incurred because this structure violates the parser's expectations. However, once this cost is incurred, the parser has received more input and it is able to make stronger predictions about the structure of the rest of the sentence. In other words, it is now clear to the parser that this structure is an object relative clause structure. Thus, in an object relative construction, the expectation-based model predicts less of a processing cost at the verb site because the parser will have a strong expectation of finding a verb in that position once it has encountered the unexpected DP earlier in the string (i.e., the subject of the relative clause). Thus, even though object relative clause constructions are first predicted to incur a greater processing cost when the expected structure is found (i.e., at the subject of the object relative clause), the parser is then able to make clearer and more specific predictions about the rest of the structure. This expectation is predicted to alleviate any processing difficulty associated with object relatives later (i.e., further to the right) in the string of words.[121]

Can expectation-based models explain our results? The answer is rather unclear but we will briefly discuss some possibilities. Whether or not expectation-based models can explain the results reported in the current thesis depend on the parser's predictions about a particular syntactic environment. As the parser is given more input (words), it revises its expectations about what it expects to find further to the right in the sentence. Let us assume that the introduction of an anaphor introduces a syntactic environment in which the parser knows that it must find an antecedent further to the right (following Principle A of

[121]Note that Hale (2001)'s surprisal theory does not predict that object relatives will incur less of a processing cost later in the sentence after the unexpected structure was previously encountered. This prediction is solely expected under Levy (2008)'s account.

the Binding Theory). As the number of words following the anaphor increases, so does the parser's expectation for finding an antecedent.

Processing difficulty in surprisal theory depends on the frequency of a particular word in a given syntactic environment. If a given word is frequently found in a given syntactic environment, the expectation for that type of word in that particular environment will be high and we predict a low processing cost. In contrast, if a given word is not frequently found in that particular syntactic environment, the expectation for that particular word will be fairly low and we would predict a processing cost. If we consider the syntactic environment used in the current experiments, we might expect that it is unexpected to find an anaphor embedded in a *wh*-phrase that linearly precedes its antecedent and that it might be slightly more frequent to find a pronoun in this position (i.e., cataphoric pronouns).[122] Thus, expectation-based models might be able to explain the main effect of anaphors incurring a cost compared to pronouns but it is unclear what this model would predict with respect to how the parser finds antecedents for the pronominals. Both anaphors and pronouns need to find antecedents, preferably within the sentence, unless the pronoun is interpreted as referential which we have found is not the preferred interpretation. However, only anaphors need to find *local* antecedents. There are no such grammatical constraints encoded in expectation-based models. Processing difficulty depends on the frequency of particular words in particular syntactic environments. Upon finding an anaphor/pronoun early in the string of words (i.e., further to the left), the parser would predict that an antecedent should be found further to the right. As more words are read, the expectation for that antecedent will increase, depending on the syntactic context. However, whether the antecedent is or is not found in a local syntactic position would not come into play

[122] To tell the whole story, one would need to conduct a systematic corpora search but this is beyond the scope of the current thesis.

in expectation-based models.[123] It is also rather unclear how these models would explain our gender effects, which differ based on the syntactic position of the intervening DP (i.e., whether it is in a subject or adjunct position).[124] There does not seem to be a straightforward way to extend the predictions of expectation-based models to the results observed in this thesis and we leave it to future work to examine this question more carefully.

5.3 Complexity metrics

In this section, we will go over each of the proposed complexity metrics in detail. These complexity metrics are a first step towards the development of a processing model to account for the effects reported in this thesis. We will show that the most interesting effects arise when these costs interact with each other. After each cost has been explained, we will show where these costs make correct predictions about our real-time processing data and where they make the wrong predictions or fail to account for some of the collected data. In cases where the proposed metrics do not make the right predictions, we will suggest possible explanations for the results as well as avenues for future research. Since stimulus sentences were presented to participants word-by-word, we will also assume that complexity costs are calculated on each word from left to right as the sentence is read.

[123] Another factor to consider is the linguistic experience of the speaker. It is possible that a particular speaker's experience might lead them to expect local antecedents when they are given a particular syntactic environment.

[124] More broadly, these theories also do not make a distinction between arguments and adjuncts since they do not take the structural positions of phrases into account. In the metrics proposed in this chapter, we make different gender match/mismatch predictions based on whether or not the DP appears in a structural position from which it can enter a binding relation with a pronominal.

5.3.1 Cost 1: interpretation of copies

The first complexity metric is based on what we assume to be a general processing preference to interpret copies in their surface (PF) positions, as first discussed in Chapter 1. We propose that it is simpler (more economical) to interpret a copy in the position in which it appears on the surface (i.e., where it is pronounced) rather than interpreting it in a distinct structural position. This proposal is based on the idea that open dependencies incur a processing cost until they are satisfied. The longer the parser has to maintain an open dependency, the greater the processing cost. We propose that if a DP can be assigned a semantic denotation when it is first encountered, it will be easier (less costly) to maintain in memory compared to a DP that cannot be assigned a semantic denotation when it is first encountered. For example, if the DP can be assigned a semantic denotation, it will be interpreted as an individual (i.e., when it is an R-expression or a referential pronoun). This means that the parser must keep the meaning of the individual in mind until it finds the gap position further to the right. However, if the DP cannot be assigned a semantic denotation when it is first encountered, i.e., when it contains an anaphor or an anaphoric pronoun, the parser must look for an antecedent for that DP to be interpreted. In this case, the pronominal is interpreted as a bound variable that can only be fully interpreted when its binder (antecedent) is found further to the right. This means that the parser must keep a partial function in mind, not only looking for the gap site but also for the antecedent for the pronominal. Since what the parser needs to keep in memory is more complex when it must look for an antecedent for the pronominal, we propose that it is more costly to maintain this linguistic object in memory, compared to when the DP can be fully interpreted upon first encounter.[125]

[125] We are again putting aside syntactic integration of the filler at the gap site. For us, if the copy can be assigned a semantic denotation in the higher copy position, it is assigned an interpretation in this position.

In order to make clear predictions for the interpretation of moved phrases in real-time processing, we have been adopting the copy theory of movement (Chomsky, 1993). Under copy theory, if a phrase moves in the syntax (before the derivation reaches a Spell-Out domain), a copy of that phrase is found both in its base-generated position and in its surface (or PF) position. Following successive cyclic movement (Chomsky, 1995), if the phrase must also stop off at any intermediate positions before reaching its final position, a copy of the phrase will also be found in these intermediate positions. Crucially, we assume that copies are created in the syntax before the derivation is spelled out. Where the phrase gets pronounced is determined at the PF level of representation. What is relevant for the current experiments is where the phrase gets interpreted at LF. Since we assume that the copies are created in the syntax, they are also present when the derivation is sent to the LF interface. We argue that LF has access to the whole chain and thus, that it can interpret the phrase in either the high position (i.e., where the phrase is pronounced) or in a lower position (either the base-generated position or an intermediate position). LF's access to the whole chain is particularly relevant in the domain of *wh*-dependencies. We have been following the proposal that *wh* must always be interpreted high in order to interpret the structure as an interrogative,[126] but that the complement of *wh* can be interpreted in (one of) the lower copy position(s). Following our economy constraint, the parser should prefer to interpret the whole phrase (*wh* and its complement) in its surface position, whenever possible. If it is not possible to interpret the whole phrase in its surface position (due to the phrase not having a semantic denotation, for example), the complement of *wh* can reconstruct to a

The parser will also need to locate the gap site further to the right to be able to syntactically integrate the filler in the lower copy position and assign the phrase its thematic role in the sentence. However, since this operation is required in all of stimuli, irrespective of the type of DP in the filler phrase, we do not make any predictions based on filler-gap dependencies at the gap position and we also do not make any claims about parsing for syntactic integration of the filler in the lowest copy position.

[126] See Chapter 1 for further details.

lower copy position to be assigned a semantic denotation and be interpreted. We consider reconstruction to be a last resort operation and that the parser will only reconstruct when it is not possible to interpret the whole phrase further to the left.[127]

We propose that a copy that cannot be interpreted in its surface position incurs a cost of 1 unit (henceforth, +1 CU, *uninterpretable copy*). If a copy cannot be interpreted in its surface position, this means that the phrase cannot be assigned a semantic denotation when it is encountered. If the phrase cannot be assigned a semantic denotation when it is encountered, we assume that the parser creates a partial denotation for that phrase which it must keep in memory until it is able to integrate the phrase into the structure (for example, when the phrase contains an anaphor, upon finding its antecedent). Since we predict a cost to arise until the phrase can be integrated (and thus, interpreted), this cost should not only be incurred in the higher copy position but also on each subsequent word until the parser finds an integration position (i.e., a lower copy). Upon finding a lower copy position, the processing cost should disappear because the phrase can now be assigned a semantic denotation and be interpreted.[128]

[127] Since all of our stimuli were *wh*-questions, there was always a base-generated position to which the parser could reconstruct the complement of *wh*. If the parser was insensitive to other structural properties in our sentences, we might have predicted that it would adopt a simple analysis and always reconstruct the complement of *wh* to this lower position. Generally, in the base-generated position, it would be able to assign the phrase a semantic denotation and syntactically integrate it in the structure. This would enable the phrase to be interpreted in the base-generated position and also be assigned a thematic role in this position. However, in our experiments, we failed to find parsing effects at the base-generated position, but we did find effects at intermediate copy positions, suggesting that the parser is sensitive to the fact that reconstruction can target intermediate copies, at least for interpretation reasons.

[128] We are not assuming that the parser predicts the syntactic position where it might find a copy, however, such an analysis might be possible if we assume that the parser is sensitive to the syntactic structure. Work on filler-gap dependencies (Crain and Fodor, 1985; Stowe, 1986) has shown that the parser engages in an active search for a gap position in the linearly closest syntactic position where such a gap is syntactically licit. In the current studies, the parser would not only need to predict where lower copies are found but also where appropriate antecedents are found. We remain agnostic about this question and leave it to future work to examine further.

5.3.1.1 Complement of *wh* containing R-expressions

Let us consider the relevant examples that we have been investigating in this thesis. Recall that we are following Sauerland and Elbourne (2002) and others in assuming that *wh* must always be interpreted high (in its surface position) but the complement of *wh* can be interpreted in a lower structural position (reconstructed reading). In general, it is more economical to interpret the complement of *wh* in its surface position compared to its reconstructed position. The simplest case would be one where the *wh*-phrase is embedded with an R-expression, as in (2).

(2) [*wh* R-expression$_i$] ... embedded subject$_j$ ... [~~R-expression$_i$~~]

In (2), the parser can assign the complement of *wh* a semantic denotation as soon as it is encountered. The R-expression is referential and its interpretation does not depend on another DP in the structure. As a result, the complement of *wh* can be interpreted in the higher copy position. Thus, we can view this example as the baseline condition, at least with respect to antecedent resolution.

5.3.1.2 Complement of *wh* containing anaphors

We expect a cost to arise if the parser needs to find an antecedent for a DP further to the right in the sentence. Consider the case where the *wh*-phrase contains an anaphor, as in (3).

(3) [*wh* anaphor$_i$] ... antecedent$_i$ ... [~~anaphor$_i$~~]

In (3), the complement of *wh* containing an anaphor cannot be assigned a semantic denotation when it is encountered. This is because the interpretation of the anaphor depends on its antecedent. We assume that to be able to interpret a phrase, it must carry a semantic denotation. Since there is no appropriate antecedent in the preceding linguistic context, the parser must look further to the right for an antecedent to be able to interpret the anaphor. Assuming that the parser is sensitive to Binding Principle A, it will need to locate a local antecedent in a position further to the right. Therefore, in order for the anaphor to be assigned a semantic denotation and integrated within the structure, the parser must reconstruct the complement of *wh*, which contains the anaphor, to a position below the antecedent (in this case, the embedded subject).[129]

5.3.1.3 Complement of *wh* containing pronouns

The situation is more complicated when the complement of *wh* contains a pronoun, as in (4).

(4) [*wh* pronoun$_i$] ... antecedent$_{i/j}$ ... [~~pronoun$_i$~~]

The pronoun can be interpreted in two ways, depending on its assignment function. If the pronoun is to be interpreted as referential (a free pronoun), the complement of *wh* can be assigned a semantic denotation in its surface position and be interpreted. This parse would allow the parser to satisfy its preference to interpret phrases as soon as they are encountered.

However, following previous processing work (e.g., Van Gompel and Liversedge, 2003), there is a preference to find antecedents for pronominals within the sentence. Following this preference, the interpretation of the pronoun depends on its within-sentence antecedent. Upon finding a pronoun in the complement of *wh*, the parser might prefer

[129] Note that the lower copy position to which the parser reconstructs the complement of *wh* may or may not correspond to the lower copy position where the *wh*-filler must be integrated, i.e., the embedded object position. Following successive cyclic movement, there is an intermediate copy position that is structurally lower than the antecedent but higher than object position. It is possible that the complement of *wh* reconstructs to this position in order to bind the anaphor with the antecedent. Unfortunately, the results of our experiments are compatible with both possibilities.

to find an antecedent for that pronoun within the sentence. Since there is no appropriate antecedent for the pronoun in the preceding linguistic context (the experimental sentence), the parser must also locate an antecedent further to the right. If the parser is sensitive to structural constraints, it should be sensitive to Binding Principle B in this condition. In order to satisfy this constraint, the complement of *wh* cannot reconstruct to a position that is structurally lower than the antecedent. We have also been assuming that there is a preference for binding over co-reference interpretations (Reinhart, 1983). Following this preference, if the complement of *wh* contains a pronoun, the parser will have a preference to adopt a binding relation between that pronoun and its antecedent, compared to a co-reference relation. However, in its surface position (i.e., in the complement of *wh*), the pronoun cannot (yet) enter a binding relation with another DP since there is no linguistic antecedent at that point in the string. Binding may only become an option for the parser when it finds an antecedent further to the right in the structure. As a result, we may only observe effects of the pronoun once the antecedent is reached. Note that if the semantic denotation of the pronoun depends on its antecedent, the parser is able to satisfy the preference to find within-sentence antecedents for pronouns but it is not able to satisfy the preference to interpret phrases when they are encountered. Thus, results from the pronoun condition may help us understand if these preferences are assigned different weights by the parser.

5.3.1.4 Summary: interpretation of copies containing different types of DPs

In terms of interpreting copies, the simplest case is when the *wh*-complement contains an R-expression. The parser can assign the phrase a semantic denotation as soon as it is encountered and thus, interpret the phrase in the higher copy position. The most costly case is when the *wh*-complement contains an anaphor. In this case, in order to assign the phrase a semantic denotation and satisfy Binding Principle A, the parser must reconstruct the *wh*-complement to a syntactically lower position. Since the higher copy is uninterpretable when the complement of *wh* contains an anaphor, a cost should be incurred until a lower copy is found further to the right in the linear string. This cost is predicted to arise on each word until the copy is located. When the *wh*-complement contains a pronoun, the parser can adopt two distinct analyses. Either the complement of *wh* can be interpreted in its surface position, just like R-expressions, and the pronoun is interpreted as free, or the complement of *wh* be can be interpreted in a lower copy position and the pronoun is co-referential with another DP in the sentence, just like anaphors. The parser may not adopt a binding analysis with pronouns until it finds an antecedent further to the right in the sentence and binding becomes an option. Upon reaching a potential antecedent, the parser may attempt to adopt a binding relation between the two DPs. In order to do so, the complement of *wh* containing a pronoun will need to reconstruct to a lower copy position. Doing so may incur binding violations (specifically, Binding Principle B) and the parser will need to reanalyze is parse.

5.3.2 Cost 2: Cost of introducing a discourse referent

We propose the second processing cost arises when a new discourse referent must be added to the common ground.[130] With all three types of DPs (anaphors, pronouns and R-expressions), the referent of the DP needs to be supported by the common ground.[131] The DPs must be part of the common ground in order for speakers and listeners to be able

[130]Note that this cost is similar to Gibson (1998)'s proposal that new discourse referents incur a cost. The proposed cost arises when a new discourse referent is introduced but our other proposed costs are not dependent on the number of discourse referents that have been introduced, as outlined in DLT.

[131]We acknowledge that in our experiments, the DPs would all be considered "out of the blue." If we assume that the discourse in an experimental setting is the experimental sentence, the referents in common ground will be reset for each sentence.

to keep track of the discourse. One semantic theory that can be easily extended to real-time processing predictions is Heim (1983)'s File Change Semantics (see also Kamp, 1981). Following this proposal, listeners can keep track of what is happening in the discourse using file cards. Different file cards indicate which referents have been introduced in the discourse. There is a new file card introduced for each new discourse referent.

We propose that there is a cost for introducing a new file card, which will add the referent to the common ground. In our system, introducing a new file card incurs a cost of 1 cost unit (+1 CU, *new*). For our experiments, we propose that the introduction of each new R-expression (or proper name) will incur a cost of 1. This is because R-expressions necessarily introduce a new discourse referent and thus require a new file card.

What happens with pronouns and anaphors? Whether or not pronominals incur a processing cost depends on when they are introduced in the linear string. If they are introduced before their antecedent, then they introduce a new discourse referent, requiring a new file card, and incur a cost of 1 CU. However, if the anaphor/pronoun is co-referential with another DP that is already part of the common ground, it will not incur a cost attributed to the introduction of a new discourse referent. Instead, there will be a cost for adding information to the previously introduced file card. We will discuss this cost in the next sub-section.

5.3.3 Cost 3: Cost of co-indexation

The third processing cost arises when the parser must co-index a DP with a previously introduced DP. This cost is incurred because more information has been collected about the previously introduced DP and the file card must be updated with this new information. In our experiments, this cost is incurred when the parser encounters a pronoun or anaphor further to the right in the linear string. Since the pronoun/anaphor cannot be interpreted unless it is co-indexed with another DP in the sentence, the parser must determine which referent is the appropriate one for each pronoun/anaphor and this incurs a processing cost. The cost is incurred when the parser has found a co-referent within the sentence. Which DP incurs the cost, i.e., antecedent or pronominal, depends on the preceding linguistic context. If the pronominal (anaphor/pronoun) linearly precedes its antecedent, the cost will be incurred on the antecedent because it is at this point that the parser must co-index the pronominal with its antecedent.[132] If the pronominal linearly follows its antecedent, the cost is incurred on the pronominal. For simplicity, we will call this cost a cost of co-indexation and assume that it is a cost of 1 whenever the parser must co-index a DP with a co-referent (+1 CU, *existing*).[133]

5.3.4 Cost 4: Cost of indices

The cost of introducing a new discourse referent and of co-indexation can be incurred on all three types of DPs. The DPs all have in common that they need to be part of the common ground, which is encoded using file cards. How do the different types of DPs differ? Following previous work in semantic theory, we argue that they differ based on whether or not they enter the derivation with an index. If a DP has an index, it needs to be bound by another DP. If the DP does not have an index, it is referential. Pronouns can

[132] At first glance, this prediction might seem inconsistent with our results from Experiment 1, where we did not observe a cost on the embedded subject, which was the antecedent for the anaphor. As we will see shortly, we predict that the different costs can interact with one another, meaning that sometimes there is more than one cost reflected on a particular word. If the cost of an anaphor in a particular syntactic environment is the same as another DP, e.g., an R-expression, in that same syntactic environment, then we do not expect to find a processing cost associated with the anaphor.

[133] This cost might also be viewed as a reintegration cost, i.e., the parser needs to reintegrate the previously introduced DP into the linguistic context. In file change semantics, the parser needs to add to the already introduced file card and thus, it must relocate that file card in order to add the new information.

both enter the derivation with an index or without an index. Anaphors necessarily enter the derivation with an index.

Following work by Heim (1998), Roelofsen (2010, 2011), we assume that in order for two DPs to enter a binding relation, they must both bear an index. Anaphors and pronouns inherently carry indices because they are always referentially dependent on another DP in order to be interpreted. However, R-expressions do not inherently enter the derivation with an index. In order for an R-expression to be in a binding relation with another DP and serve as a binder, an index must be introduced on the R-expression. In order to be assigned an index, a DP must undergo movement (Roelofsen, 2010, 2011; Heim, 1998). This work assumes that an index is introduced if the DP undergoes quantifier raising (QR), as schematized in (5).

(5) $[_{TP}$ X $[_{DP}$ Q] Y] $\rightarrow$ $[_{TP}$ X $[_{DP}$ Q$]^{n}$ $[_{TP}$ X t_n Y]]

Since an additional operation (i.e., introduction of an index on the DP) must occur with R-expressions but not with anaphors/pronouns, we assume that it incurs a processing cost.

How does this work in left-to-right parsing? Upon encountering an R-expression, the parser will not know whether this DP will be in a binding relation with another DP in the sentence. However, since it is possible for each R-expression to enter a binding relation and serve as binders and since R-expressions require indices in order to enter binding relations, we will assume that the parser will always create a binder index for each R-expression it encounters. Consequently, R-expressions will always incur this processing cost (+1 CU, *index*).[134]

[134] For simplicity, we assume that this cost is incurred on each R-expression. In order for two DPs to enter a binding relation, they both need to bear an index. For the parser to be able to consider whether an R-expression can serve as a binder for another DP in the sentence, it must introduce a binder index on each R-expression, even if the R-expression does not enter a binding relation in that particular sentence.

5.3.5 Cost 5: Cost of co-reference over binding

The fifth cost we propose is that there is a general cost to adopt a co-reference relation over a binding relation (Reinhart, 1983). Following the discussion from Chapter 1, we assume that there is a general grammatical preference to adopt a structural relation between two DPs whenever possible (binding). We further assume that the parser is sensitive to the grammar's preference and will attempt to adopt whichever parse follows the grammar's preferences. If the parser cannot adopt a binding interpretation (due to structural constraints such as binding theory) and must adopt a co-reference interpretation, this incurs a cost. The parser begins interpreting the string according to the grammar's preferences and attempts to adopt a binding relation between two DPs. When such an interpretation is not possible, the parser must reanalyze its parse and adopt a different interpretation, incurring a processing cost. We will call this cost *coref* and assume it contributes a cost of 1 CU.

5.3.6 Cost 6: Cost of (mis)matching features

As previously mentioned in Chapter 4, we predict that both a gender-matched and a gender-mismatched DP can incur a cost, depending on its structural position. A DP that matches a pronominal in gender will incur a cost if it is not found in a structurally accessible position from which it can bind the pronominal. In contrast, a DP that mismatches a pronominal in gender will incur a cost if it is found in a structurally accessible position. In the latter case, the syntactic structure tells the parser that this DP should be an appropriate antecedent but the phi-features on the DP do not match the pronominal. This cost therefore depends on the parser checking two different aspects of the DP. As a first filter, it may check the phi-features and determine that the DP does match the pronominal in gender and thus deem it to be an appropriate antecedent. As a second filter, it determines that it is not found in an appropriate syntactic configuration, incurring a processing cost (+1 CU, *gender match*). In the other case, the parser may initially determine that the DP does not match the pronominal in gender and it could thus disregard this DP as a potential antecedent. However, if the parser is sensitive to the structural configuration in which the DP appears, it might determine that this DP is in an appropriate structural position for binding and this incurs a processing cost (+1 CU, *gender mismatch*).[135]

5.3.7 Summary

The processing costs that we propose are summarized in Table 5.1. For simplicity, we assume that each of these costs incurs a cost of 1 CU. At this point, we do not have any reason to suggest that any of these costs are more costly than others but we do not discount the possibility that some costs might incur a greater burden on the parsing system than others.

Table 5.1: Proposed processing costs

Cost	Description	Weight
Unint copy	preference to interpret copy in surface position	1
New	new discourse referent	1
Existing	co-indexation	1
Index	introduction of index on R-exp	1
Coref	coreference $>$ binding	1
Gender match	gender-matched referent	1
Gender mismatch	gender-mismatched referent	1

[135]This proposed cost assumes that we should only observe effects of gender-(mis)matching if the parser is actively looking for an antecedent for a pronominal. If an association between the DP and a previously introduced pronominal was not being considered, we would not expect to observe any effects based on whether the DP matches or mismatches the pronominal in gender. Thanks to Susana Béjar for bringing up this point.

5.4 Applying the complexity metrics

In this section, we will discuss some of the interactions between the metrics that arise when different processing costs are incurred on a single region in the experiments. Since we are assuming a left-to-right incremental parser, we assume that the costs are incurred on each word as the sentence is processed left-to-right. We predict that the greater the cost (i.e., the more cost units that are incurred due to the various complexity metrics proposed in the previous section), the longer that word/region should be read. For space reasons, we will go over the major predictions that our metrics make across experiments rather than detailing the predictions of each metric on each word in the experiments. For transparency, we present schematics demonstrating word-by-word predictions for each experiment when we discuss the major predictions.

5.4.1 Type of DP found in the filler phrase

The complexity metrics introduced in the previous section make specific predictions about the costs of different types of DPs. Across experiments, different types of DPs were presented in positions in which they linearly preceded their antecedents. In Experiment 1 (anaphors vs R-expressions), we found that anaphor conditions were more costly than R-expression conditions on the words following the DP in the filler phrase. This effect is predicted by the *uninterpretable copy* complexity metric. Upon finding the anaphor in the *wh*-filler phrase, the parser can only interpret *wh* in the surface position. The complement of *wh* must reconstruct to a position in which the anaphor can be bound by a c-commanding antecedent. We interpret the longer reading times on the words following the filler phrase in anaphor conditions as reflecting this *uninterpretable copy* cost.

5.4.1.1 Applying the metrics: Experiment 1

Interactions between the different complexity metrics arise on the DP itself in the filler phrase. What we mean by this is that when different metrics apply in a single region, they make specific predictions for our reading time data. We will first consider Experiment 1. The relevant examples are shown in (6).

(6) *Experiment 1 stimuli refresher*

The relatives wondered...

a. how many pictures of **herself**$_i$ you are asking **Mary**$_i$ to bring

b. how many pictures of **Paul** you are asking **Mary** to bring

...for the photo album.

In (7)-(8), we show the predicted computational costs on each word in the experimental items. We will concentrate on the costs that apply on the DP in the complement of *wh*.

(7) *Applying the metrics: anaphor condition*

how many pictures of herself <u>you are asking</u> Mary to bring

↓	↓	↓
+1 new +1 unint copy	+1 new + 3 unint copy	+1 existing +1 index

Total cost: 8 CUs

(8) *Applying the metrics: R-expression condition*

how many pictures of Paul you are asking Mary to bring

$\downarrow \quad \downarrow \quad \downarrow$

$\boxed{\begin{array}{l}\text{+1 new}\\ \text{+1 index}\end{array}} \quad \boxed{\text{+1 new}} \quad \boxed{\begin{array}{l}\text{+1 new}\\ \text{+ 1 index}\end{array}}$

Total cost: 5 CUs

DP in the complement of *wh* When the DP is introduced, whether it is an anaphor or an R-expression, a cost of 2 CUs is incurred. In both cases, this DP introduces a new discourse referent. R-expressions are always new discourse referents and the anaphor is a new discourse referent because there is no antecedent for the anaphor in the preceding linguistic context (the sentence). When the DP is an anaphor, there is a cost of an uninterpretable copy since the complement of *wh* cannot be assigned a semantic denotation and must be interpreted in a lower copy position (reconstruction).[136] In contrast, when the DP is an R-expression, the complement of *wh* can be assigned a semantic denotation in its surface position and thus, it can be interpreted. Nevertheless, there is a cost of introducing a binder index since this DP could serve as a binder for a pronominal found further to the right in the string.

Even though these complexity metrics are different, in both cases, we are predicting a cost of 2 CUs when the DP is introduced in the complement of *wh* position.[137] The

[136]This complexity metric assumes that the parser is sensitive to structural requirements. An alternative is that the parser simply looks further to the right in search of an antecedent, as discussed in Chapter 2.

[137]In the current thesis, we assume that the different types of costs incur the same number of CUs, namely, 1 CU per cost. It is possible that different types of costs may incur different numbers of cost units. We would need to conduct more research in order to determine if certain costs are more costly than others but we do not dismiss this possibility here.

difference between reading times at the DP in the anaphor and R-expression conditions in Experiment 1 was not statistically reliable which might suggest that this prediction was borne out. However, it is often the case that processing costs are observed in the spill-over region, i.e., on the words following the costly region. We did, in fact, find that anaphor conditions were more costly than R-expression conditions two and three words following the DP (*are, asking*). We argue that this effect can be explained as resulting from the parser looking for a lower copy position to which it can reconstruct the phrase containing the anaphor. The phrase containing the anaphor cannot be interpreted when it is encountered so the parser must look further to the right for a position where the phrase can be interpreted. We suggest that this cost persists until the lower copy position is found. Why this cost seems to only become active two words after the anaphor was introduced is currently not clear to us. One possibility is that since the sentences are presented one word at a time, the parser does not compute that the phrase containing the anaphor cannot be interpreted right away. Thus, the cost associated with anaphors is not a lexical effect of anaphors being inherently harder than pronouns but it reflects structural operations. This analysis is supported by a follow up study, which will be reported in section 5.4.2.

We additionally predict that the cost of an uninterpretable copy in the anaphor conditions should persist until the embedded subject is found further to the right in the string. We have suggested that this cost should be found on each word until the embedded subject is reached. As a result, we are predicting an increasing cost on each word until the embedded subject is found. This effect was not borne out. While we did find a cost of anaphor conditions on two and three words following the anaphor, this cost did not increase as more words were processed. We do not have an explanation for why this is the case. Even though we predict this increase in complexity costs when anaphors precede their antecedents in subsequent experiments, we also fail to find this effect. This might suggest that we are making the wrong prediction altogether or that our experimental method is not sensitive enough to detect this effect.

Embedded subject DP position Interesting interactions between the different complexity metrics are also predicted to occur at the embedded subject region. Again, in both conditions, our metrics predict a cost of 2 CUs. In both cases, the embedded subject is an R-expression and thus, a binder index must be introduced, incurring a cost of 1 CU. There is also a cost for adding this DP to a file card in the anaphor condition. However, in the R-expression condition, the cost is due to the parser needing to introduce a new file card for this DP. Thus, even though the costs are attributable to different complexity metrics, in both cases, we predict a cost of 2 CUs on this region. If our complexity metrics are on the right track, we should not observe reading time differences between the two conditions at this region. This prediction was borne out: the difference between reading times on the embedded subject in the two conditions was not statistically reliable.[138]

5.4.1.2 Contrasting the proposed metrics to the predictions of DLT: Experiment 1

Recall that DLT predicts processing costs to arise on *new* discourse referents in between a filler and its gap site. Thus, in our anaphor condition in Experiment 1, DLT would predict a cost on *asking* since a new discourse referent is introduced in this position (the event of *asking*. DLT would not predict a cost on the deictic pronoun, *you*, since such pronouns

[138]It is important to note however that we did find longer reading times on the embedded subject in the raw and residual reading times in the anaphor condition compared to the R-expression condition. One possible explanation is that this region is roughly the intermediate copy position for the *wh*-phrase. This effect could either be attributed to a spill over effect, i.e., there is still a cost of an uninterpretable copy at this position or, ii) a reconstruction effect, i.e., the copy is able to receive an interpretation at this position and this operation incurs a cost. However, since this effect was not found across all transformations, it is unclear whether it is a reliable effect and proceed cautiously in our interpretation of this effect.

are argued to not introduce discourse referents in the theory nor would it predict a cost on *are* since there is no discourse referent introduced by the auxiliary. As a result, DLT is unable to account for the effects observed in Experiment 1. It would only predict a cost on *asking*, which is borne out, but it would fail to account for the cost we observed on *are*. If we compare the pronoun and R-expression conditions of Experiment 1, DLT would also predict a cost on the embedded subject, *Mary*, in the R-expression conditions. This is because *Mary* introduces a new discourse referent. In the anaphor condition, *Mary* is the antecedent for the anaphor and thus, does not introduce a new discourse referent. We did not observe a cost on the embedded subject in the R-expression condition. Thus, DLT again makes the wrong prediction for our Experiment 1.

5.4.1.3 Applying the metrics: Experiment 2

We also predict relevant interactions between the complexity metrics when the *wh*-filler is embedded with a pronoun, which was investigated in Experiment 2. Sample stimuli are shown in (9).

(9) *Experiment 2 stimuli refresher*

The reporters wondered...

a. how many rumors discrediting **her**$_i$ as a witness you are asking Mary$_i$ to deny

b. how many rumors discrediting **Mary** as a witness you are asking her$_i$ to deny

...for the TV interview.

The predicted processing costs are shown on each word in (10)-(11).

(10) *Applying the metrics: pronoun ... embedded subject R-expression*

how many rumors discrediting her as a witness you are asking Mary to deny

+1 new

+2 new

+1 new?/existing?
+1 co-reference
+ 1 index

Total cost: 6 CUs

(11) *Applying the metrics: R-expression ... embedded subject pronoun*

how many rumors discrediting Mary as a witness you are asking her to deny

+1 new
+1 index

+2 new

+1 new?/existing?

Total cost: 5 CUs

DP 1 region Again we will concentrate on the DP regions. Our metrics predict that pronouns should be less costly than R-expressions when they appear in the complement of *wh*. The pronoun incurs a cost of 1 CU because it introduces a new discourse referent. This cost also applies when the complement of *wh* contains an R-expression but in the R-expression condition, there is also a cost incurred for introducing a binder index. Since there are 2 CUs incurred in the R-expression condition but only 1 CU incurred in the pronoun condition, we predict that R-expressions should be more costly. This prediction was borne out in Experiment 2 in the spill over region. This effect was found on the word following the DP in the *wh*-phrase.[139]

DP 2 region We also predicted interactions between the different complexity metrics at the embedded subject region in Experiment 2. In this experiment, the embedded subject was either a pronoun or an R-expression. If the embedded subject was an R-expression (see (10)), we predicted a cost of 3 CUs at this region. Even though the pronoun in the filler phrase does not create an open dependency (unlike the anaphor condition which must find a local antecedent), we assume that the parser still prefers to find referents for pronominals within the sentence (Van Gompel and Liversedge, 2003). Thus, in this condition, the parser must determine whether the embedded subject should be associated with the previously introduced pronoun or whether the two DPs should be analyzed as separate referents and each introduce a new file card. The first cost at the embedded subject region therefore arises due to referent resolution. The parser must determine if the R-expression should introduce a new file card or whether it needs to be added to the file card previously introduced by the pronoun. At this point in the derivation, the parser faces two conflicting alternatives. Following the general processing preference to find referents for pronominals within the sentence, it should associate the pronoun with the embedded subject. Furthermore, it should attempt to bind the pronoun with the R-expression because there is a preference for structural binding over co-reference interpretations. However, in order to abide by structural constraints (binding theory), the parser must conclude that the R-expression introduces a new discourse referent. If the parser attempts to establish a binding relation

[139]Note that since pronouns contain significantly less characters than R-expressions, any effects found on the DP itself would be considered unreliable because they could have arisen due to word length, i.e., longer words are read for a longer amount of time. One of our data transformation was residual reading times which take into consideration the number of characters in each word. Indeed, we found that the effects on the DP itself were not statistically reliable in the residual reading times, even though they came out significant in the raw and log reading times.

between the R-expression and the previously introduced pronoun, this parse would violate Binding Principle C. The conflict between these two alternative parses has been indicated under one cost in (10)-(11), labelled as *+1 new?/existing?*.[140]

Based on the results of our comprehension questions, participants preferred to associate the pronoun with the R-expression found in the sentence. Since we assume that the parser is sensitive to grammatical constraints, this result suggests that a co-reference interpretation is adopted in these stimuli.[141] Co-reference interpretations are more costly than binding interpretations. Consequently, adopting a co-reference interpretation between the two DPs incurs a cost of 1 CU. Finally, since *Mary* is an R-expression and since the parser prefers binding interpretations over co-reference interpretations, it must create a binder index on *Mary* which is also costly.[142]

In contrast, if the embedded subject is a pronoun (see (11)), we only predicted a cost of 1 CU. Similarly to the condition in which the embedded subject is an R-expression, the parser must determine if the pronoun should be added to an existing file card or whether it introduces a new file card. In this condition, the parser can associate the pronoun with

[140] It is also possible that these conflicting parses lead to a greater processing cost, compared to cases where it is clear that the DP should be added to a new file card or that it introduces its own file card. Unfortunately, the results of our experiments do not allow us to differentiate between these alternatives. We will thus assume the simplest analysis that there is a cost of 1 CU introduced at this region.

[141] This would mean that certain preferences are more preferred than others. Even though binding is preferred over co-reference, such an interpretation is not possible if it would violate structural constraints.

[142] Again, this does not seem to be the most economical possible parse of this sentence. Since the parser does not know which R-expressions will enter into binding relations with other DPs in the sentence, it must assume that all R-expressions require an index. Another possibility is that the parser will not posit indices on R-expressions until it determines that they will be required in the sentence. In the latter case, we would predict that R-expressions would incur a cost only when the parser has determined that the DP is in a binding relation with another DP in the sentence and thus requires an index. Since the current experiments all use self-paced reading (with the exception of Experiment 4b), it is not possible to test this hypothesis. If the R-expression is presented further to the left in the sentence, the parser will not know that that DP requires an index until it reaches a pronominal further to the right. As a result, we would predict a processing cost of the R-expression later in the sentence, after the pronominal has been read, which may take the form of re-reading the earlier part of the sentence. As our design does not allow us to investigate this question directly, we leave it aside in the current work.

another DP in the sentence, provided it adopts a co-reference analysis. The R-expression cannot bind the pronoun since it does not c-command the pronoun. The pronoun can also be interpreted as referring to a referent outside of the sentence. Our comprehension results from Experiment 2 show that participants have a preference for interpreting the two DPs as co-referential, suggesting that they adopt a co-reference analysis. One might ask why there is no cost for co-reference in this condition. If we assume that the parser does not even consider a binding analysis until it is faced with an environment in which such an analysis would be possible, it should not even consider binding until it gets to the embedded subject, which is the pronoun in this experiment. Upon reaching the pronoun, the parser might consider a binding analysis but it should quickly determine that such an analysis is not possible since there is no potential binder in the sentence. As a result, it can adopt a co-reference analysis. Since it does not need to reanalyze its parse in this case, no cost is incurred. Overall, in Experiment 2, we predicted a cost for R-expressions over pronouns, whether they occurred in the *wh*-phrase or as the embedded subject but importantly, the reasons these DPs incurred a cost depended on their syntactic position in the sentence.

Our predictions were borne out. We found longer reading times when the embedded subject was an R-expression compared to when it was a pronoun.[143] This effect was also found on the word following the embedded subject (*to*). We interpret this effect as a spill over effect from the previous region.[144] Overall, regions containing R-expressions were more costly than regions containing pronouns. We have argued that the effects can be explained differently depending on the syntactic position of the DP. When the R-expression is initially introduced in the *wh*-phrase, we have argued that the increased cost can be attributed to the requirement that the DP have an index to enter a binding relation with

[143] We are hesitant to consider the effect on the embedded subject itself as a reliable effect due to the word length issues discussed earlier.

[144] Importantly, this effect is not confounded by word length effects.

another DP in the sentence. In contrast, longer reading times on the R-expression embedded subject can be explained as an effect of the parser adopting a co-reference interpretation instead of the preferred binding interpretation.

For space reasons, we will not list the complexity metrics for Experiment 3, which was designed as a follow-up study to Experiment 2 but used R-expressions that were familiar to participants (famous names) as well as unfamiliar names (made-up names matching the familiar names in number of syllables). The results of this experiment replicated those found in Experiment 2. At the first DP region (in the *wh*-phrase), the effect was stronger when the R-expression was unfamiliar compared to familiar. In contrast, at the embedded subject region, the familiarity variable was not a reliable predictor of the reading time data. We interpreted this effect as demonstrating that simple lexical effects are insufficient to explain our results, i.e., R-expressions are not simply read longer because they carry more discourse properties compared to pronouns. How can we explain the familiarity effects using our complexity metrics? One possibility is introducing a cost for unfamiliar R-expressions but no cost for familiar R-expressions. In the case of the familiar R-expressions, it is simple to map the R-expression onto a real-world referent but it is not possible to do so for an unfamiliar R-expression. As this question is beyond the scope of the current thesis and as Experiment 3 was conducted as a control experiment for Experiment 2, we leave it aside for the time being.

5.4.1.4 Contrasting the proposed metrics to the predictions of DLT: Experiment 2

DLT would not be able to account for the effects observed on the first DP region in Experiment 2. Since both the pronoun and the R-expression in the complement of *wh* introduce a new discourse referent, we would not expect to observe a difference in processing cost. Furthermore, since DLT predicts processing costs to arise in the region in between the filler and its gap, it is unclear if the theory would even predict a cost of the DP if it is found within the filler itself. Assuming that the pronoun and R-expression are co-indexed in our stimuli (as suggested by the answers to the comprehension questions), on the second DP region, DLT would not predict a cost on either the embedded subject pronoun or R-expression. In both cases, the DP does not introduce a new discourse referent because it is co-referential with a previously introduced DP. Consequently, DLT is unable to account for the collected data from our Experiment 2.

5.4.2 Follow-up study: Experiment 7

The way that the different complexity metrics interact allows us to make further predictions about how the parser should process the different types of DPs when they are presented in the same syntactic environment. Thus far, we have compared the processing of two different types of DPs in a single experiment but we have not yet investigated the costs associated with processing all three types of DPs in the same experiment. To fill this gap, we conducted Experiment 7.

The main goal of Experiment 7 was to investigate if our complexity metrics make the right predictions when all three types of DPs are presented in the same syntactic environment. Thus far, we have been examining the different types of DPs when they are embedded in *wh*-fillers that linearly appear further to the left than their antecedents (for anaphors and pronouns). Doing so enabled us to investigate how the parser finds antecedents for pronominals that appear in syntactic positions in which they cannot be interpreted since they could not be assigned a semantic denotation and thus could not be integrated into the structure. Experiment 7 aimed to investigate if our complexity metrics accurately predict which type of DP will be more costly than another. As a result, it was important to examine the different DPs in syntactic positions in which all structural considerations were satisfied in the surface representation. In other words, it was important to directly compare the processing of the three types of DPs in surface positions in which they could be interpreted and the parser did not need to look elsewhere in the string to find an interpretation position. This would ensure that any effects found on the DPs could be solely attributed to the processing of those DPs and not to other factors, such as whether or not the parser must search for an antecedent for a pronominal or a gap site for a *wh*-phrase.

Experiment 7 also enabled us to investigate an alternative explanation of our DP results from previous experiments. In our earlier experiments, we observed processing differences based on the type of DP found in the complement of *wh* and we argued that these effects can be explained by grammatical principles. An alternative explanation is that these processing differences arise due to inherent lexical properties of the DPs themselves. For example, perhaps anaphors are inherently more costly than R-expressions due to their abstract nature, i.e., *himself* must refer to another DP in the sentence whereas *John* refers to a real-world entity.

5.4.2.1 Design

In Experiment 7, we used simple declarative sentences that were adapted from our previous stimulus sentences, as shown in (12).

(12) The reporters realized that ...

a. you had asked Alexa to deny three lies about herself

b. you had asked Alexa to deny three lies about her

c. you had asked Alexa to deny three lies about Sean

... for yesterday's live television interview.

The different types of DPs were still embedded in a picture DP but the sentences did not involve movement and consequently, the picture DP containing the anaphor, pronoun, or R-expression appeared in canonical object position in the linear string. This means that when the parser encounters the DP in the string of words, the DP already appears in an interpretation position. In other words, the parser can interpret the DP in the position in which it is found in the linear string. As previously argued, a more economical parse is one in which the PF and LF positions match.[145] As in our other studies, the matrix subject was a plural DP that could not serve as an antecedent for the third person pronominal found further to the right in the string. The subject of the asking event was the pronoun *you* which was also not a suitable antecedent for a later pronominal. Thus, the only possible within-sentence antecedent was the subject of the embedded verb which always matched the anaphor/pronoun in gender. Crucially, this DP c-commanded the anaphor at LF, satisfying Binding Principle A. As a result, the anaphor could be assigned a semantic denotation and interpreted in the position in which it appeared in the surface representation. This also means that the embedded subject c-commanded the pronoun, which should lead to a Principle B violation if the two DPs are co-indexed. We will elaborate further on this prediction in the next section. In this experiment, we used three conditions, corresponding to the three different types of DPs. The picture DP either contained an anaphor, as in (12-a), a pronoun, as in (12-b), or an R-expression, as in (12-c).

[145] Of course, the PF representation is linear and not hierarchical so this is not exact. However, since the phrase appeared in canonical object position, the DPs were found in positions where they could be interpreted at LF without the need for additional operations, like reconstruction. This representation is simpler than one in which the phrase is found in a position where it cannot be interpreted on the surface and must be interpreted in a distinct position at LF.

5.4.2.2 Predictions

Following our complexity metrics, we predict that the three DPs should be processed similarly when they appear in canonical object position. When an anaphor is found in this position, we predict a cost of 1 CU for co-indexation. The anaphor must be co-indexed with a previously introduced DP in order to be interpreted. In other words, the anaphor must be added to an existing file card. In the pronoun condition, the cost that is incurred depends on how the pronoun is interpreted. Upon encountering a pronoun in this syntactic position, the parser must determine if the pronoun should be added to an existing file card or if it introduces a new file card. If it introduces a new file card, there is only a cost of 1 CU. If the pronoun is interpreted as co-indexed with another DP in the sentence, there will be a cost for adding that pronoun to an existing file card. Crucially, if the co-indexation parse is adopted, the parser will be unable to satisfy its preference to adopt a binding relation between the two DPs and will be forced to adopt a co-reference relation. Adopting a structural relation between the two DPs would incur a Principle B violation because the embedded subject would locally c-command the pronoun at LF. Therefore, if the co-indexation analysis is adopted, we predict a cost of 2 CUs.[146] Finally, in the R-expression condition, we predict

[146]It is currently puzzling to us why the co-reference analysis seems to be possible at all. If *Alexa* and *her* are co-indexed, as shown in (i-a), the sentence is marginal for some speakers but perfectly acceptable for others, despite the fact that this structure clearly violates Principle B. However, as shown in (i-b), a quantified phrase cannot bind the pronoun in the same syntactic environment. These examples suggest that a coreference analysis is possible in this environment. This is puzzling because co-reference is predicted to only be possible when binding is not possible or when it would lead to a distinct interpretation (Grodzinsky and Reinhart, 1993's Rule I).

(i) a. ?The reporters realized that you had asked Alexa$_i$ to deny three lies about her$_i$.
b. *The reporters realized that you had asked every student$_i$ to deny three lies about her$_i$.

Interestingly, we seem to lose this contrast in finite environments, where co-indexation between the R-expression subject and the pronoun is unacceptable, as shown in (ii). The contrast shows up again in non-finite environments, as shown in (iii). Thanks to Martin Hackl for these examples.

(ii) a. *Alexa$_i$ denied three lies about her$_i$.

a cost of 2 CUs. The first cost arises because the DP introduces a new discourse referent and therefore, a new file card. The second cost arises because the parser must introduce a binder index for the R-expression. The R-expression can serve as a potential binder and thus requires an index. Thus, following our complexity metrics, R-expressions should be more difficult to process compared to anaphors and pronouns, if the parser creates a new file card for the pronoun. If the parser attempts to co-index the pronoun with the embedded subject, anaphors should be less costly than both pronouns and R-expressions.

We have argued that if our complexity metrics make the right predictions, they also provide support for a structural analysis, suggesting that the parser is sensitive to structural properties. However, there is also an alternative non-structural hypothesis which is that the DPs differ based on inherent lexical properties and these properties affect how they are processed in real-time. If our previous results showing a difference between the processing of anaphors and R-expressions can be explained by inherent lexical properties, we should be able to replicate them when the DPs appear in any syntactic position. Following this analysis, the difficulty associated with anaphors (Experiment 1) arises because their inherent lexical properties make them more difficult to process compared to R-expressions. In contrast, if these previous effects can be explained by structural restrictions, we predict that R-expressions should be more costly than anaphors in Experiment 7. Note that this is the opposite of what we saw in Experiment 1. Crucially, in Experiment 7, the anaphor is locally bound in its surface position.

---

b. *Every girl$_i$ denied three lies about her$_i$.

(iii) a. ?Alexa$_i$ had to deny three lies about her$_i$.
b. *Every girl$_i$ had to deny three lies about her$_i$.

These examples suggest that co-reference is possible in non-finite environments, when Binding Principle B is violated. We currently do not know of or have an explanation for this contrast. The comprehension results in Experiment 7, however, will help us to better understand if the judgement in (i-a) is real, i.e., whether participants judge the embedded subject and the pronoun as co-referential.

What does the lexical (non-structural) hypothesis predict for the pronoun conditions? If we presume that pronouns could also be construed as abstract since they also need to have antecedents, we might expect that the two types of pronominals should be read similarly. In contrast, if anaphors are more abstract than pronouns, and therefore, inherently more difficult, anaphors should incur a greater processing cost. Note that the lexical hypothesis thus makes two different predictions depending on how the pronouns are analyzed, similarly to our complexity metrics which made two different predictions based on how the parser interprets the pronoun. If pronouns are processed similarly to anaphors and this is due to their lexical properties, we predict that the pronominals should incur a greater processing cost compared to R-expressions. Crucially, if we were to find that R-expressions are more difficult than both anaphors and pronouns, this would suggest that our complexity metrics are on the right track.[147]

5.4.2.3 Materials

Thirty sentence templates were created following the paradigm shown in (12). Target sentences were counterbalanced across three lists of stimuli combined with 30 filler sentences. The filler sentences were constructed to resemble the target sentences. In total, participants read 60 stimulus sentences. Comprehension questions were asked after each sentence to ensure that participants were paying attention and comprehending the sentences as intended. We asked comprehension questions directly targeting participants' interpretation of the anaphor or pronoun to ensure that participants were fully interpreting the anaphor/pronoun. To be consistent with our previous studies, we used Yes-No comprehension questions in this experiment. For the questions that asked about participants' interpretation of the critical

[147]Note that this result would suggest that the parser adopts a co-reference relation between the pronoun and the embedded subject, incurring only a cost of 1 CU. Our metrics predict that anaphor conditions should only incur a cost of 1 CU but that R-expressions incur a cost of 2 CUs.

DP in the picture DP, we asked questions of the type *Was Alexa asked to deny three lies about herself/Sean for the interview?* There was a correct answer for the anaphor and R-expression trials but the pronoun trials could be answered either way. As a result, we accepted both answers for the pronoun conditions. Six comprehension questions directly targeted participants' interpretation of the anaphor/pronoun, six questions asked about the matrix subject, six questions asked about the picture DP itself, six questions asked about the adjunct phrase, and six questions asked about the complement of the picture DP.

5.4.2.4 Participants

Sixty-four participants completed the self-paced reading study on Amazon Mechanical Turk using the Ibex farm software (Alex Drummond, http://spellout.net/ibexfarm/). Note that this was a self-paced reading study with CQs after each sentence but it did not include a naturalness rating task. Eleven participants were excluded because they scored less than 75% on the comprehension questions across the whole experiment. One participant was excluded due to a data collection error. We report the data for the remaining fifty-two participants (22 males, 30 females). All participants were self-reported native speakers of English, aged between 23 and 56 years of age (M = 35.7 years, SD = 8.9). Participants were paid $5.40 USD for their participation.

5.4.2.5 Procedure

Participants were asked to read each sentence word-by-word for comprehension and answer a yes-no CQ about the sentence they had just read.

5.4.2.6 Outliers

Items with an average comprehension score less than 60% were eliminated from the analysis. In total, 1 test trial and 6 fillers were removed from the analysis. Accuracy on the comprehension questions for the test trials was high at 87.7% and 86.7% for all trials, after trimming. In order to only investigate reading times on trials where participants were interpreting the sentence as intended, trials where CQs were answered incorrectly were removed from the analysis.[148]

5.4.2.7 Statistical analyses

We used the same statistical analyses as described for the previous experiments with the same trimming procedures. Trimming data points that were +/- 2 SDs away from the mean by participants resulted in a loss of less than 6.5% of data across all transformations. We report here the final trimmed models.

5.4.2.8 Reading time results

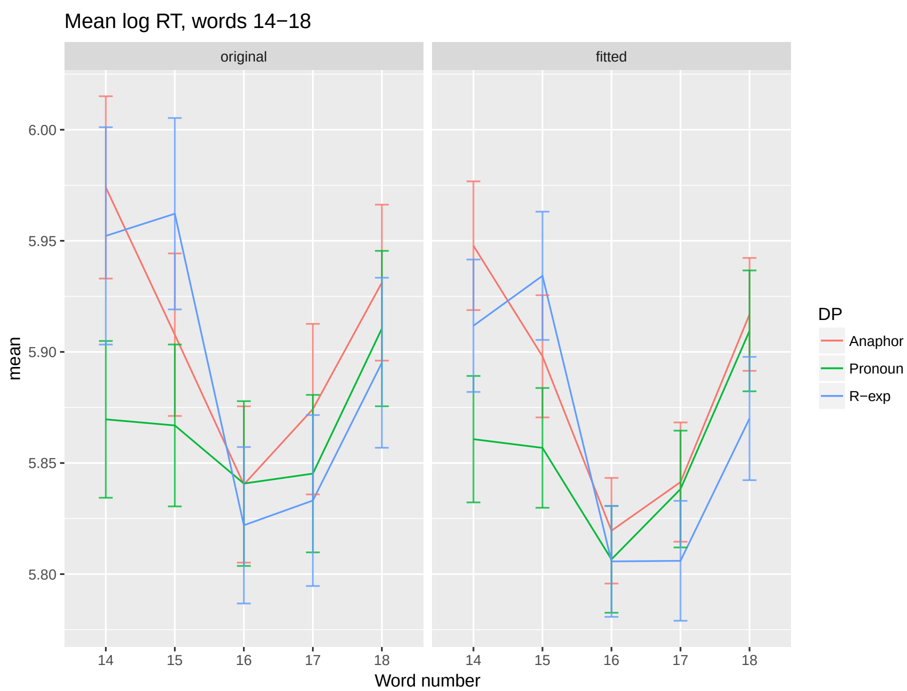
In this experiment, the DP in the picture DP occurred on word 14. We investigated reading times on the DP itself as well as three words following it. Mean log reading times for words 14-18 are plotted in Figure 5.1. There were significant effects of the type of DP at word 14, when the DP is introduced but these effects were only found in the raw and log reading times and not in the residual reading times, which take word length into consideration. Consequently, we deem the results on the DP itself to be unreliable since they seem to be driven by the length of the word, i.e., longer words are read for a longer amount of time.

A significant effect of the type of DP was found on word 15, one word following the

[148] We kept all trials in which there was no clear answer, i.e., the pronoun trials.

DP in the picture DP. Pronoun conditions were read faster than anaphor (raw: β = 18.48, SE = 8.63, t = 2.141, p = 0.04, residual: β = 18.35, SE = 8.81, t = 2.083, p = 0.043, log: β = 0.039, SE = 0.016, t = 2.441, p = 0.015, models not shown) and R-expression conditions (β = 29.46, SE = 8.79, t = 3.352, p = 0.002, residual: β = 33.91, SE = 9.71, t = 3.493, p = 0.001, log: β = -0.08, SE = 0.016, t = 5.024, $p < 0.001$, models not shown). Crucially, while R-expression conditions were read numerically longer than anaphor conditions, this effect did not reach significance at the 5% threshold in the raw and residual reading times. In the log reading times, this effect was marginal (β = 0.05, SE = 0.03, t = 1.824, p = 0.07). No effects reach significance on word 16, two words following the DP. Interestingly, on word 17, three words following the DP, R-expression conditions were read significantly faster than anaphor conditions (raw: β = -13.44, SE = 6.19, t = -2.169, p = 0.03, residual: β = -14.8, SE = 6.79, t = -2.180, p = 0.03, log: β = -0.04, SE = 0.02, t = -2.214, p = 0.03) and pronoun conditions (raw (marginal): β = -10.48, SE = 6.14, t = -1.705, p = 0.09, residual: β = -15.04, SE = 7.11, t = -2.114, p = 0.04, log: β = -0.403, SE = 0.02, t = -1.976, p = 0.05). These effects were also found on word 18, except that the difference between R-expressions and pronouns did not reach significance at the 5% threshold.

Figure 5.1: Mean log reading times on words 14-18 in Experiment 7 by DP Type. Original means are plotted in the left panel and fitted values are plotted in the right panel. Error bars represent 95% confidence intervals.



5.4.2.9 Discussion

R-expressions vs pronouns The effects found in Experiment 7 replicate some of the previous effects found in the studies reported in this thesis. On the word following the DP, we found that R-expression conditions were more costly (read longer) than pronoun conditions. We interpret this effect as demonstrating that when they are first introduced, R-expressions carry greater discourse properties compared to pronouns. These additional discourse properties incur a processing cost. Note that this effect replicates the effects found in Experiments 2 and 3. This effect is predicted by our complexity metrics under the assumption that the parser assigns a new file card to the pronoun when it appears in canonical object position. This effect would not be predicted by the lexical hypothesis under the assumption that pronouns are inherently more abstract compared to R-expressions.

Anaphors vs pronouns We also found that pronoun conditions were read faster than anaphor conditions, which suggests that anaphor conditions incur a greater processing cost. This effect is not predicted by our complexity metrics. Our metrics either predict that the two types of pronominals should be processed similarly (both incurring a cost of 1 CU) but only if the pronoun is interpreted as introducing a new file card or that pronouns should be more costly than anaphors (due to binding > co-reference). We see two possible explanations for this effect in the current study. One possibility supports the lexical hypothesis: the lexical properties associated with anaphors are more costly than those associated with pronouns. Another possibility is to appeal to structural analyses. In the anaphor condition, the parser must establish a binding relation between the anaphor and its antecedent, which is found further to the left in the string. Even though the parser has already built the structure containing the antecedent, it still needs to locate the antecedent and associate it with the anaphor. We suggest that in order to do so, the parser must backwards search the already built structure and locate the antecedent. In contrast, pronouns do not need to be structurally bound to antecedents. In fact, if our complexity metrics are on the right track, the parser adopts a co-reference analysis between the pronoun and the antecedent in these syntactic environments. Since pronouns do not need to be bound, the parser does not need to conduct a backwards search to find their antecedents.

An open question is how the parser still associates the earlier DP with the pronoun. Even under a co-reference analysis, the two DPs are interpreted as referring to the same referent. The results from our CQs that targeted participants' interpretation of the pronoun support this analysis. As in previous experiments, we found that there is a strong preference to interpret the pronoun and the R-expression as referring to the same referent. Participants responded *yes* to these CQs, confirming that they interpreted the pronoun and R-expression as referring to the same entity, 85.6% of the time. These results therefore suggest that participants were more likely to interpret the pronoun as referring to the previously introduced R-expression, even though it is possible for the pronoun to find its referent outside this sentence. This result once again provides support for Van Gompel and Liversedge (2003)'s argument that the parser prefers to find within-sentence antecedents. At this point, both a structural and a non-structural analysis could explain the difference between pronouns and anaphors in this experiment. Our complexity metrics do not adequately predict this effect but we can see how it might follow from structural properties of the DPs. Our metrics do not consider the syntactic position of the DP in the sentence and as a result, they fail to account for certain effects, such as this difference between anaphors and pronouns in Experiment 7. It is currently unclear to us how we might encode the syntactic position of the DP into our complexity metrics. We will discuss this issue in greater detail in the next sub-section as well as in Chapter 6.

It is important to address a possible discrepancy between the pronoun results found in Experiment 7 and those found in Experiment 2. In Experiment 2, we argued that the reading time difference between pronouns and R-expressions found at the embedded subject could be explained as a reanalysis effect, i.e., the parser prefers binding relations over co-reference interpretations. When a binding relation between two DPs is not possible because it would violate a structural constraint, the parser is forced to reanalyze its parse and adopt a co-reference interpretation, incurring a processing cost. In Experiment 7, the syntactic environment is the same as it was in Experiment 2, except that it does not involve movement. As a result, the parser should again not be able to adopt a binding analysis in the pronoun conditions. In this case, we have suggested that the parser automatically adopts a co-reference relation and this parse is less costly than one in which it attempts to associate the two DPs via binding. One possible explanation is that these results show a difference between anaphoric and cataphoric processing. In Experiment 2, the results were always found on cataphoric conditions, i.e., the pronoun always linearly preceded its antecedent. In contrast, in Experiment 7, we only used anaphoric conditions. Finding a cataphor early in the sentence (further to the left) sets up an expectation for binding, i.e., the parser expects to find an antecedent for that pronoun further to the right in the linear string. However, there is no such expectation in the anaphoric case since the antecedent has already been encountered further to the left. Unfortunately, it is beyond the scope of the current thesis to investigate this issue further and we leave it to future work to provide us with a better understanding of these findings.

R-expressions vs anaphors Crucially, the difference between the anaphor and R-expression conditions on the word following the DP did not reach significance at the 5% threshold. This result suggests that our previous effects from Experiment 1 cannot simply be explained by a lexical difference between anaphors and R-expressions. If our previous findings could be attributed to a lexical effect, we would expect to find the same effect when the DPs are found in canonical object position. Since the effect disappears when the DPs are found further to the right in the linear string, we argue that our result from Experiment 2 cannot be explained as a simple lexical difference between anaphors and R-expressions but that it reflects a processing difference associated with anaphors linearly preceding their antecedents. It is important to note that our complexity metrics actually predict that R-expression conditions should be read longer than anaphor conditions and this is not borne out. We do not have a clear explanation for this result. One possibility is that the cost of a binder index on R-expressions is not required when the DP appears in canonical object position. At this point in the sentence, especially within the experimental setting where all the sentences follow a similar structure and length, the parser might determine that this DP will not serve as a binder and therefore, not introduce an index for that DP. If this was the case, both anaphors and R-expressions should only incur a cost of 1 CU. Again, this explanation takes into consideration the syntactic position of the critical DP, which our complexity metrics do not yet consider.

We found some puzzling results on words 17-18, three and four words following the critical DP, showing that R-expression conditions incurred less of a processing cost compared to anaphor and pronoun conditions. Since this effect appears so late in the sentence, it is not clear to us how to interpret this effect. One possible explanation is that this effect reflects the parser establishing an interpretation for the anaphors/pronouns, i.e., co-indexing the DPs with the antecedent found further to the left. Since the R-expression cannot be co-indexed with another DP in the sentence, it incurs less of a processing cost because it can be interpreted as soon as it is encountered.

To summarize, the results of Experiment 7 argue against a lexical explanation for our results from Experiment 1. If the difference between R-expressions and anaphors could be explained by lexical effects, we would expect to observe the same reading time effects when the DPs are found in canonical object position. Since we did not replicate our previous results and did not find a reliable difference between anaphors and R-expressions in Experiment 7, we argue that our Experiment 1 results can be explained by a structural account of antecedent resolution.

5.4.3 Match-Mismatch effects

In this sub-section, we will consider the gender match/mismatch effects that our complexity metrics predict. We will concentrate on Experiments 4-6 but only consider the results that are relevant for our gender predictions. Our complexity metrics predict a cost of finding a DP that matches a previously introduced pronominal in gender if that DP is in a structural position from which is cannot structurally bind the pronominal. However, we predict a cost of gender mismatch if the DP does not match the pronominal in gender but it is found in a structural position from which it could bind the pronominal.

In general, this thesis has assumed that the parser is sensitive to the structural configuration in which the DP appears in the sentence. We have argued in favour of several processing preferences which are based on grammatical constraints. These preferences favour analyses in which structural constraints are not violated. The cost of gender match or mismatch also relies on the parser's sensitivity to structure. If, however, we consider an alternative in which the parser is not sensitive to the structural environment in which a DP appears, we might predict that a gender-matched DP incurs a cost, irrespective of where it appears in the sentence. Gender matching would incur a cost because the parser must consider whether this DP must be associated with a previously introduced pronominal. As a result, this cost allows us to investigate if a non-structural, pure gender-matching account could explain our effects. Since our other complexity metrics make use of structural constraints, we predict that if our metrics are on the right track, they would provide evidence for a structural analysis of our results. In contrast, if the non-structural analysis is on the right track, the gender-matching cost should not interact with our other complexity metrics and our proposed model should fail at predicting the results.

For space reasons, we will group together the experiments based on the syntactic position of the intervening DP which had the gender manipulation (either gender match or gender mismatch). The critical DP either appeared embedded in an adjunct phrase (Experiments 4a, 4b, and 5) or as a subject (Experiments 6a-6b). Crucially, when the DP appeared embedded in an adjunct, it was unable to c-command the anaphor/pronoun at LF. In contrast, if the DP appeared in subject position, it was able to c-command the pronominal at LF, depending on the reconstruction position of the complement of *wh*, i.e., either an intermediate copy position or the base-generated position.

5.4.3.1 Intervening DP in an adjunct phrase

In Experiments 4a, 4b, and 5, the DP appeared in an adjunct phrase. In the two versions of Experiment 4, the adjunct was a PP but in Experiment 5, it was a gerund phrase, as shown in (13).

(13) *Experiments 4-5 stimuli refresher*

The reporters wondered...

a. how many lies about **himself**$_i$ {in **Sean's** report} / {crashing **Sean's** car} you asked Jeff$_i$ to deny

b. how many lies about **himself**$_i$ {in **Shelly's** report} / {crashing **Shelly's** car} you asked Jeff$_i$ to deny

c. how many lies about **him**$_i$ {in **Sean's** report} / {crashing **Sean's** car} you asked Jeff$_i$ to deny

d. how many lies about **him**$_i$ {in **Shelly's** report} / {crashing **Shelly's** car} you asked Jeff$_i$ to deny

...for the TV interview.

The predicted processing costs are shown on each word in (14)-(17).[149]

(14) *Applying the metrics: anaphor, gender-matched intervener*

how many lies about himself in Sean's report you asked Jeff to deny

↓	↓	↓
+1 new +1 unint copy	+1 new +1 index +1 gender match +5 unint copy	+1 existing +1 index

Total cost: 12 CUs

(15) *Applying the metrics: anaphor, gender-mismatched intervener*

how many lies about himself in Shelly's report you asked Jeff to deny

↓	↓	↓
+1 new +1 unint copy	+1 new +1 index +5 unint copy	+1 existing +1 index

Total cost: 11 CUs

[149] Here we show the costs on the stimulus sentences from Experiments 4a-4b but the predicted costs are the same for Experiment 5.

(16) *Applying the metrics: pronoun, gender-matched intervener*

how many lies about him <u>in Sean's report you asked</u> Jeff to deny

↓	↓	↓
+1 new	+1 existing?/new? +1 index +1 gender match +1 co-reference?	+1 new?/existing? +1 gender match +1 co-reference? +1 index

Total cost: 9 CUs

(17) *Applying the metrics: pronoun, gender-mismatched intervener*

how many lies about him <u>in Shelly's report you asked</u> Jeff to deny

↓	↓	↓
+1 new	+1 new +1 index	+1 new?/existing? +1 gender match +1 co-reference? +1 index

Total cost: 7 CUs

The critical region is the adjunct region where the intervening DP is introduced. The DP in the adjunct introduces a new discourse referent but this is the case both for gender-matching and gender-mismatching DPs. The DP in the adjunct is not in an appropriate position to bind the pronominal. Thus, we predict a cost of gender-matching in this position, e.g., (15) and (17), compared to gender-mismatching, e.g., (14) and (16). The parser finds a gender-matched DP which passes the morphological filter for binding but since this DP is not found in a position from which it can structurally bind the pronominal, it is not a suitable antecedent.

We will consider the anaphor and pronoun conditions separately since they make distinct predictions. In the anaphor conditions, if the intervening DP matches the anaphor in gender, we predict a cost of gender matching. This DP is not in an appropriate position from which it can bind the anaphor, even under a reconstruction analysis. There will also be a cost for introducing a new file card for this DP and creation of a binder index. Since the complement of *wh* containing an anaphor cannot be interpreted in its surface position, there is also the cost of finding the lower copy. Thus, in the intervening DP region, there is a cost of 4 CUs when the DP matches the anaphor in gender. In contrast, if the DP does not match the anaphor in gender, there is a predicted cost of 3 CUs. Since the DP does not match the anaphor in gender and it also is not in an appropriate structural position to bind the anaphor, it can be easily dismissed as a potential antecedent.

In the pronoun conditions, upon reaching the intervening DP in the adjunct, the parser must determine if this DP can be co-referential with the pronoun. Following the binding over co-reference preference, this DP could not bind the pronoun. However, it still might be possible for the parser to adopt a co-reference relation between the pronoun and the intervening DP. In this case, we would also predict a cost of co-reference.[150] There is also a cost for introducing a binder index for the DP. Thus, when there is a DP embedded in an intervening adjunct, there should be a greater cost for gender matching in both anaphor and pronoun conditions but we do not predict a difference between the anaphor and pronoun conditions at this position since they both incur a cost of 4 CUs, see (14)-(16).

[150] An alternative is that there is a greater cost for determining whether the DP is a new or existing referent. This cost could have a greater weight than cases where it is evident that the DP refers to a new discourse referent. If this alternative cost is on the right track, we would not need the gender match cost for pronoun conditions. The data do not allow us to distinguish between these two alternative complexity costs. For simplicity, we assume that gender matching DPs in structurally illicit positions automatically incur a cost and that determining whether the DP is a new or existing referent only incurs 1 CU.

These metrics also make different predictions between anaphor and pronoun conditions at the embedded subject region. Specifically, pronoun conditions are predicted to be more costly than anaphor conditions. The DP embedded subject matches the pronoun in gender but it is not structurally appropriate since Principle B would be violated under a reconstruction analysis. In contrast, in the anaphor condition, the gender-matched DP embedded subject is structurally appropriate since reconstruction to a copy position below the embedded subject would satisfy Principle A.

Metrics fail: Experiment 4a There were no significant effects of the intervener in Experiment 4a. We are currently puzzled by this result. As argued in Chapter 4, one possible way to explain these results is that participants were not fully interpreting the anaphor/pronoun, as suggested by our comprehension questions showing that participants accepted gender-mismatched DPs as antecedents for the pronominals. We will return to this issue when we discuss some of the open questions in Chapter 6.

Metrics make the right predictions: Experiments 4b and 5 In Experiment 4b, when used the same stimuli but a more sensitive methodology, namely, eye-tracking while reading. We were able to find significant effects based on the gender of the intervening DP in the adjunct region. We found a gender-matching effect on the region directly following the adjunct (*you asked*). Conditions containing DPs that mismatched the pronominal in gender incurred less of a processing cost (read faster) compared to conditions containing DPs that matched the pronominal in gender. This effect was predicted by our complexity metrics. This effect was found in the first pass reading times, suggesting that the parser is sensitive to gender effects during initial parsing. This effect was also found in the second pass reading times and total reading times, again suggesting that it is a prolonged effect that continues past the initial stages of parsing.[151] This gender effect was replicated in Experiment 5 on the intervening DP when we used gerundial adjuncts.

In Experiment 4b (eye-tracking), pronoun conditions incurred a greater cost compared to anaphor conditions on the embedded subject. This effect is predicted by our complexity metrics because the embedded subject matches the pronoun in gender but it is not structurally appropriate.[152,153] It is important to note that while our complexity metrics seem to make the right predictions at the embedded subject, this effect would suggest that the parser always associates the pronoun with the embedded subject DP. However, our comprehension results from Experiments 4b and 5 do not support this analysis. Using a more sensitive methodology, we were able to find reading time effects that are in line with the proposed complexity metrics at the intervener DP.[154]

[151] We also found that pronoun conditions incurred a greater processing cost compared to anaphor conditions in go-past times. This effect was not predicted by our complexity metrics. In Chapter 4, we suggested that this effect arises because the parser considers the DP in the adjunct as a potential referent for the pronoun but not for the anaphor. Our complexity metrics do not make such a prediction because they assume the self-paced reading methodology in which reading times are calculated on each word in a left-to-right fashion. Eye-tracking enables us to not only investigate regions in the sentence but it also allows us to look at different reading time measures which track different stages of processing. Since the go-past measure also takes into consideration re-reading times and our complexity metrics were not designed with re-reading times in mind, we leave it to future work to investigate this effect further.

[152] In Chapter 4, we suggested that this effect arose due to competition between the embedded subject and the DP in the adjunct since both of these DPs can serve as the antecedent for the pronoun. Since there are two possible antecedents for the pronoun, upon reaching the embedded subject, the parser must decide which DP is the appropriate antecedent. Crucially, competition effects should only arise when both DPs match the pronoun in gender but the effect shows a main effect of DP type but not of match. However, across conditions, the embedded subject always matched the pronoun in gender and thus, could always be associated with the pronoun. Since this effect was found in the first pass reading times, a very early stage of processing, it suggests that the parser was sensitive to the possibility that the embedded subject could be associated with the pronoun. We are unable to explain why we do not also observe a significant effect of Match at this region, which would be predicted if the pronoun effect was due to competition effects.

[153] We leave aside the regression results since they are not comparable to the results of our self-paced reading experiments. As a reminder, regressions reflect a re-analysis of a previous region. Since self-paced reading does not measure re-reading times, we are unable to compare our results from these two different experimental tasks using our complexity metrics.

[154] We note that some effects may have also arisen because we asked deep processing questions in this version of the experiment. These CQs forced participants to assign a referent to the anaphor/pronoun and therefore, forced participants to actually interpret the anaphor/pronoun. In Experiment 4a, participants could

Metrics fail: Experiments 4b and 5 We note that not all effects predicted by our complexity metrics are borne out in this experiment. Specifically, since the anaphor is searching for an antecedent, we predicted a cost on the words following the anaphor which was not borne out. One possible explanation is that the intervening region has a high processing load and this masks any potential DP effects in this region. The intervening region is an adjunct which likely increases the processing load. Assuming a left-to-right incremental parser, upon finding an adjunct phrase, the parser must determine where the adjunct should be syntactically integrated in the hierarchical structure. The adjunct phrase introduces its own clause which must be built before it can be integrated into the larger structure that the parser is building. In addition, the adjunct contains a DP that either matches or mismatches the previously introduced pronominal in gender. Even if this DP cannot serve as an antecedent (in the case of anaphors), it still competes as a potential antecedent, disrupting processing.

As discussed in Chapter 4, we also found an interaction between the type of DP and whether the intervener matched or did not match the pronominal in gender, one word before the embedded subject (word 13) in Experiment 5. This effect was driven by the pronoun conditions. More precisely, pronoun conditions incurred a greater cost (read longer) in the gender-mismatched conditions compared to the gender-matched conditions. The anaphor conditions were read similarly, irrespective of whether the intervener matched or did not match the anaphor in gender. The proposed complexity metrics do not directly predict this effect. In the previous chapter, we argued that the anaphor conditions do not show a difference based on the gender of the intervener because this DP is never an appropriate antecedent for the anaphor. However, the intervener could be an appropriate antecedent for the pronoun, but only if it matches the pronoun in gender. If the parser expects a

have scored highly on the CQs without actually interpreting the pronominals.

gender-matched antecedent in the pronoun conditions (because only a gender-matched DP would be an appropriate antecedent) upon finding a gender-mismatched antecedent, a cost is incurred. This result suggests that we need to revise our complexity metrics to be able to account for this effect. One possibility is that there is a cost associated with the parser needing to continue searching for an antecedent for the pronoun. This would violate the parser's preference to resolve open dependencies in the linearly closest position. The intervening DP could serve as the antecedent for the pronoun but only if it matches the pronoun in gender. If it mismatches the pronoun in gender, it is no longer an appropriate match. Nevertheless, this DP is still found in the linearly closest syntactic position in which the pronoun could find its antecedent, violating the parser's preference and this is costly.

We also acknowledge that our CQ results in both Experiments 4b and 5 do not show a clear preference for whether the pronominal refers to the DP in the adjunct or the embedded subject DP, especially in gender-matched conditions. When participants were specifically asked to select a referent for the pronominal, they were not consistent with respect to which DP they chose. Our metrics do not make any predictions about these results and assume that the parser attempts to abide by structural constraints, even though the comprehension question results do not show these same effects.

DLT predictions It is again unclear what predictions DLT would make for our Experiment 4-5 stimuli. If processing costs are only calculated after the filler and before the gap position, DLT would not predict any processing effects on the intervening DP in the adjunct since the adjunct is part of the filler phrase. If DLT did predict that processing costs could be incurred in the adjunct, these costs would depend on whether or not the DP in the adjunct introduces a new discourse referent. Crucially, we observed a cost of gender-matching. If the pronoun can be co-indexed with the DP in the adjunct, this means that the intervening

DP does not introduce a new discourse referent and should not incur a processing cost. Our effects at this position are exactly the opposite of what DLT might predict. If the embedded subject is co-referential with the pronominal (anaphor or pronoun), DLT would not predict a processing cost in this position since in both cases, the DP does not introduce a new discourse referent. However, in the pronoun condition, if the embedded subject is a new discourse referent, then DLT would predict a cost of pronoun conditions at this region. This latter analysis is potentially compatible with our results at the embedded subject position in Experiment 4b.

5.4.3.2 Intervening DP in subject position

In Experiments 6a and 6b, we explored how the parser finds antecedents for pronominals when there is a second possible antecedent for the pronominal in the sentence which is also found in an appropriate structural position. In Experiment 6a, we manipulated whether the subject of the asking event matched or did not match the pronominal in gender. In Experiment 6b, we manipulated the gender of both subject DPs. Therefore, in Experiment 6b, there was only ever one possible gender-matched antecedent for the pronominal. It was either found in the subject of the asking event position or in the embedded subject position. Sample stimuli are shown in (18).

(18) *Experiment 6a-6b stimuli refresher*

The broadcasters wondered...

a. how many lies about **himself**$_i$ crashing a Rolls-Royce **Sean** had asked Jeff$_i$ / Shelly$_i$ to deny

b. how many lies about **himself**$_i$ crashing a Rolls-Royce **Shelly** had asked Jeff$_i$ to deny

c. how many lies about **him**$_i$ crashing a Rolls-Royce **Sean** had asked Jeff$_i$ / Shelly$_i$ to deny

d. how many lies about **him**$_i$ crashing a Rolls-Royce **Shelly** had asked Jeff$_i$ to deny

...for the TV interview.

The proposed complexity metrics on each word for both experiments are shown in (19)-(22).

(19) *Applying the metrics: anaphor, gender-matched subject of asking event*

... himself crashing a Rolls-Royce **Sean** had asked Jeff / Shelly to deny

+1 new +1 unint copy	+1 existing?/new? +1 index +6 unint copy	+1 new?/existing? +1 index	/	+1 new +1 index +1 gender mismatch

Total cost: Experiment 6a: 13 CUs; Experiment 6b: 14 CUs

(20) *Applying the metrics: anaphor, gender-mismatched subject of asking event*

... himself crashing a Rolls-Royce **Shelly** had asked Jeff to deny

+1 new +1 unint copy	+1 new +1 index + 1 gender mismatch +6 unint copy	+1 new?/existing? +1 index

Total cost: 13 CUs

(21) *Applying the metrics: pronoun, gender-matched subject of asking event*

... him crashing a Rolls-Royce **Sean** had asked Jeff / Shelly to deny

him	Sean	Jeff	/ Shelly
+1 new	+1 existing?/new? +1 index +1 gender match	+1 new?/existing? +1 co-reference? +1 index	+1 new +1 gender mismatch +1 index

Total cost: Experiment 6a: 7 CUs; Experiment 6b: 7 CUs

(22) *Applying the metrics: pronoun, gender-mismatched subject of asking event*

... him crashing a Rolls-Royce **Shelly** had asked Jeff to deny

him	Shelly	Jeff
+1 new	+1 new +1 gender mismatch +1 index	+1 new?/existing? +1 co-reference? +1 index

Total cost: 7 CUs

In this version of the experiment, we made different predictions based on which reconstruction positions are available to the parser. Since we manipulated whether the subject of the asking event matched or did not match the pronominal in gender, it was possible for the parser to find a structural antecedent for the pronominal in this syntactic position. If there is an intermediate reconstruction position which is structurally below the subject of the asking event, the parser should find an antecedent for the anaphor when it reaches a gender-matched DP at the subject of the asking event. However, if the subject of the asking event does not match the anaphor in gender, the parser will need to keep looking further to the right for an antecedent for the anaphor.

The predictions differ for the pronoun conditions due to Principle B. If there is an intermediate reconstruction position and if the DP at this intermediate position (subject of the asking event) is an appropriate match, reconstructing to this position should be costly because co-indexing the pronoun and the DP subject of the asking event would incur a Binding Principle B violation. If the parser adopts an interpretation in which the pronoun and the DP subject of the asking event are co-indexed and if it abides by structural considerations, it will need to adopt a co-reference interpretation. This interpretation should be costly because binding interpretations are preferred over co-reference interpretations. However, if the parser reconstructs the phrase containing the pronoun to a position below the embedded subject, it would be possible to co-index the pronoun and the DP subject of the asking event and not incur a Principle B violation (since the pronoun and its antecedent would be in separate binding domains).

Where the metrics make some of the right predictions We concentrate on the effects found at the relevant DP regions but we acknowledge that our complexity metrics do not make the right predictions for other regions in the sentence. We discussed some of these issues in Chapter 4 so we leave them aside here. On the word following the subject of the asking event, we found a significant effect of Match in Experiment 6b. The Match effect shows that conditions in which the DP subject of the asking event mismatched the anaphor/pronoun in gender were read longer than conditions in which it matched the pronominal in gender. We replicated this effect on the subject of the asking event in Experiment 6b. The effect of gender-mismatch on the subject of the asking event is predicted by our complexity metrics. We predict a cost of gender mismatch when the DP is found in a structurally appropriate position but its morphological features do not match the morphological features on the pronominal.

Where the metrics make the wrong predictions On the word following the subject of the asking event, we found also found a significant effect of DP type in Experiment 6a. The DP type effect demonstrates that pronoun conditions were read faster than anaphor conditions. This effect would be predicted if the parser still has not found an antecedent for the anaphor but has found an appropriate match for the pronoun, i.e., this is a cost of an uninterpretable copy. Our comprehension question results from this experiment, however, suggest that the parser can find an antecedent for the anaphor in the subject of the asking event position. If the DP type effect reflected a cost because the anaphor cannot be interpreted, it should not be possible for participants to interpret the intervening DP and the anaphor as co-indexed. In fact in gender-matched conditions (when both DPs matched the anaphor in gender), the parser preferred to interpret the subject of the asking event and the anaphor as co-indexed. Thus, our complexity metric at this region does not straightforwardly predict the DP effect.[155]

We found a gender match effect on the embedded subject region in Experiment 6b. Note that at this region, the embedded subject matches the pronominal in gender. We found this effect on the three words following the embedded subject, suggesting that it is a robust effect. This effect is not predicted by our complexity metrics since we predict a cost for finding a DP that mismatches the pronominal in gender when the DP is in a structural position from which is could bind a pronominal. In Chapter 4, we suggested that this result might reflect an effect of reconstruction. This DP is the only gender-matched DP in the sentence and thus it is the only DP that can be associated with the pronominal. We may not observe gender mismatch effects in this position in Experiment 6b because there is only

[155] Another potential way to explain the processing cost associated with anaphors at this position would be to argue that this is an effect of reconstruction to the intermediate copy position. However, such an effect would only be predicted in gender-matched effects and we observed a general main effect of DP type in this position.

ever one gender-matched DP in the sentence.

Experiments 6a-6b had also aimed to investigate whether there are two possible reconstruction positions in our stimulus sentences: i) below the subject of the asking event and ii) below the embedded subject. Whether we can find real-time evidence for two distinct reconstruction positions remains an open question. We will discuss this issue in Chapter 6.

DLT predictions for our Experiments 6a-6b DLT makes the same predictions on the DPs as the current theory does. On the subject of the asking event, DLT would predict a higher cost for a new discourse referent. Thus, if this DP mismatches the pronominal in gender, DLT would predict a cost. This was borne out in our experiment. At the embedded subject region, DLT would again predict a processing cost for a new discourse referent, i.e., a gender-mismatching DP. We found that gender-matched DPs incurred a processing cost at the embedded subject region in Experiment 6b. DLT would not predict this effect but neither did our complexity metrics.

Summary: match/mismatch effects Importantly, we made different gender match/mismatch predictions based on the syntactic position of the potential antecedents in Experiments 4-6. We predicted a cost of gender matching if the DP was found in an inappropriate structural position (for binding) but a cost of gender mismatching if the DP was found in a structurally appropriate position. We made the right predictions for our observed data at the intervening adjunct position in Experiments 4b and 5, where we found a cost of gender match when the DP was embedded in an adjunct. We also predicted a cost of gender mismatch at the intermediate subject position in Experiments 6a-6b. However, we did not predict a cost of gender matching at the embedded subject position in Experiment 6b. We have provided some possible explanations for these effects in the previous sections.

5.5 Chapter summary and conclusions

In this chapter, we proposed a series of complexity metrics to account for the real-time processing data reported in Chapters 2-4. We also reported the results of a new experiment aimed at testing the proposed metrics based on the type of DP in the complement of *wh*. These complexity metrics were designed to be a first step towards a real-time processing model account for effects of interpretation. Since we used the self-paced reading methodology, we proposed that costs are calculated on each word as the sentence is read from left-to-right. We predicted that the greater the cost on a particular word/region, the longer that word/region should be read. We proposed 6 complexity metrics to predict the processing data. We also assumed that the grammar has particular preferences, i.e., binding $>$ co-reference, structural considerations, principles of economy, etc. We assumed that the left-to-right parser is sensitive to the grammar's preferences and will attempt to build a structure that abides by these preferences. In general, the parser prefers the simplest possible parse for a given string of words.

The proposed complexity metrics were able to account for some of the data collected in our real-time experiments but they were unable to account for all of the collected data. In general, when the *wh*-phrase contains an anaphor, processing difficulty is incurred on the words following the anaphor. We have attributed this effect to the cost of an uninterpretable copy. Since there is no linguistic antecedent earlier in the sentence, only *wh* can be interpreted in the higher copy position. The complement of *wh*, which contains the anaphor, cannot be interpreted high and must reconstruct to a lower copy position. This lower position must be one from which the anaphor can be structurally bound by an appropriate antecedent. Our complexity metric predicts that this cost should arise on each intervening word in between the anaphor and the lower copy position. Such an effect was only partly borne out in Experiment 1. While we observed a cost of anaphors in our later experiments, it was not a continuous effect found on each word until the lower copy position was encountered. We can only speculate about why this is the case. One possibility is that the later experiments (Experiments 4-6) were all more complex than Experiment 1, involving adjuncts or additional subject positions intervening between the anaphor and the embedded subject. Since the parser needed to process and build the structure for the adjunct, this might have made processing more difficult and masked any effects of the parser searching for the lower copy (or more generally, antecedent).

The costs of a new discourse referent and of co-reference, both related to file card semantics (Heim, 1983), seemed to make accurate predictions for our data. These costs were generally combined with other costs, thus predicting greater processing effects depending on the number of metrics appearing on a particular word in the sentence. We most clearly saw interactions of different metrics when we examined effects arising based on the type of DP found in a particular syntactic environment.

We also introduced a cost of a binder index on R-expressions (proper names), assuming that these DPs need to QR in order to introduce an index and serve as binders (Heim, 1998; Roelofsen, 2010, 2011). While these costs seemed to account for the collected data, it is unclear if this metric is the most explanatory. An alternative is that the discourse properties associated with proper names are more costly than those associated with pronominals, especially when the R-expressions are introduced out of the blue in an experiment. This cost could be encoded in a more highly weighted cost of introducing a new file card for R-expressions. Our cost of a binder index is also predicted to arise on each R-expression, even if that DP never serves as a binder. While this is not the most economical option, it is unclear how else we might encode such a cost. If the parser fails to create an index for an R-expression that must later serve as a binder, the derivation will crash since a pronominal would not be associated with an antecedent. Since our experiments were not designed to tease apart these alternatives, we are not able to comment on which explanation would better predict our data.

Our proposed cost of gender-(mis)matching makes accurate predictions for most of the collected data. When a DP is found in a structurally inappropriate position (adjunct), we observed a cost of gender matching. When a DP is found in a structurally appropriate position, we observe a cost of gender mismatching. In cases where there was an earlier cost of gender mismatching, we find a cost of gender matching on a later subject position. Our metrics currently do not predict this effect but we have provided some possible explanations to account for these effects. Crucially, the gender (mis)match results show different effects depending on the structural position of the intervening DP, suggesting that the parser is sensitive to the structural configuration in which a DP appears.

In this chapter, we did not propose complexity metrics to account for the creation verb manipulations discussed in Chapters 2-3. Since our manipulations failed to show any interpretable results, we do not make any claims about potential processing effects based on these different verb types (and their potential interaction with different types of DPs).

To conclude, the complexity metrics proposed in this chapter are a first step towards a better understanding of how the parser finds referents for pronominals that linearly precede their antecedents. These metrics assume that a structural analysis is simpler than a non-structural one and thus, that the parser is sensitive to the grammar's preferences, at least in the domain of referential resolution. In the next chapter, we will summarize the main findings of each chapter and go over some open questions that have arisen in this thesis. We will then briefly discuss the theoretical consequences of this work.

Chapter 6

Conclusion and open questions

6.1 Thesis summary

This thesis began by asking whether the left-to-right parser is sensitive to effects of interpretation in real-time processing. More precisely, we examined whether the parser is sensitive to the syntactic structure it needs to build in order to interpret (and assign meaning to) a given string of words (LF structure). To directly investigate this question, we investigated whether the parser is sensitive to binding constraints (Chomsky, 1981) imposed on LF structure. In real-time processing, the parser is given a linear string of words as input (roughly, the PF representation). To interpret that string of words, the parser must build a hierarchical structure that can be interpreted by the semantic component (LF representation). We have assumed that building an LF representation based on the PF form of a sentence occurs via the syntactic module.

Following proposals in theoretical linguistics (see e.g., Lebeaux, 1990; Heycock, 1995; Fox, 1998, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004, among others), the LF representation of a string of words does not necessarily directly correspond to its

PF representation. This means that the interpretation position of a given phrase in the LF representation does not always correspond to the position in which it appears in the pronounced form of the sentence. We assume that it is simpler to interpret a given phrase if the position in which it appears on the surface roughly corresponds to the position in which it can be interpreted at LF. In some cases, the LF representation can be derived based on the order in which the words appear in the PF representation. We assumed that when the LF representation could be built as soon as the words in the string were processed, i.e., the parser was able to build the LF structure and integrate the words into that structure as soon as the words were presented (PF-LF match), this would be more economical than if the parser needed to interpret a phrase in a distinct position from where it appeared on the surface (PF-LF mismatch).

To directly investigate the question of whether the parser is sensitive to these two levels of representation, we investigated constructions whose PF and LF forms have been argued to be distinct. Our strategy was to use constraints on the interpretation of different types of English DPs to investigate this question. In particular, we investigated constructions in which binding constraints seemed to be violated at the PF level of representation. Assuming that binding constraints must hold at LF (Fox, 2000; Nissenbaum, 2000; Fox and Nissenbaum, 2004), we examined whether the parser is sensitive to such constraints. If so, these results would provide evidence that the parser is sensitive to LF structure. In order to satisfy binding constraints at LF, theoretical linguistics have argued that the DP can *reconstruct* to structurally lower position at LF, i.e., below its antecedent, in order to be interpreted (see e.g., Lebeaux, 1990; Heycock, 1995, among others). Such cases can be viewed as a *PF-LF mismatch*: a given phrase is pronounced in a distinct position from where it is interpreted.

Wh-constructions are seemingly obvious candidates to investigate PF-LF mismatches in real-time processing. In a *wh*-construction, the *wh*-phrase is pronounced at the beginning of the sentence but it must be integrated into the structure in a position that is further to the right in the linear string. Much previous work in psycholinguistics has investigated how the parser processes these types of constructions and they are known under the label of *filler-gap dependencies* (see e.g., Crain and Fodor, 1985; Stowe, 1986, among many others). Filler-gap dependencies can be understood as reflecting one type of mismatch between PF and LF. We argued that the current work adds to the large body of work on filler-gap dependencies but addresses whether the parser is sensitive to PF-LF mismatches using a different empirical domain, referential dependencies. More precisely, we examined whether the complement of *wh* itself can be interpreted in its surface position or whether it must be interpreted in a distinct position at LF.

The experimental design used in the current thesis was built based on previous work in the filler-gap domain but specifically investigated how the complement of *wh* is assigned an interpretation in real-time. In order to investigate whether the parser is sensitive to LF effects, our design used two main components: 1) reconstruction and 2) *wh*-fillers embedded with different types of DPs. Reconstruction was used as a way to control for whether the complement of *wh* could be interpreted in its surface position (PF) or whether it needed to be interpreted in a distinct position from where it appears on the surface. If the phrase could be interpreted in its surface position, we assumed that there was a match between the PF and LF representations and thus that this parse was simpler (more economical) than a parse in which the LF and PF positions of a given phrase were completely distinct. We embedded the *wh*-fillers with different types of DPs because the DPs differ in how they are interpreted (assigned a semantic denotation). The type of

DP allowed us to control for whether the parser could interpret the complement of *wh* in its surface position or whether it needed to reconstruct the phrase containing the DP to a lower copy position at LF. We adopted a general economy principle that the parser prefers the simplest possible parse of a sentence. Extending this idea to LF processing and to the question of interest in the current thesis, it should only be possible to interpret the complement of *wh* in the lower (reconstructed) position when the phrase cannot be interpreted in its surface position (high position, where it is pronounced). In other words, this parse is a last resort since it is more complicated (and therefore, less economical) to derive this interpretation of the sentence.

We also investigated whether different motivations for reconstruction might interact in processing. *How-many* questions have been shown to be ambiguous between a *de re* and a *de dicto* interpretation (Obenauer, 1984). In order to derive the *de dicto* interpretation, the complement of *wh* must reconstruct below a scope bearing operator. However, in *how many* question containing creation verbs, only the *de dicto* interpretation is possible, meaning that reconstruction is obligatory (Heycock, 1995). In such cases, the parser will not know that it must reconstruct the complement of *wh* until it has found the creation verb further to the right in the sentence. If the complement of *wh* contains an anaphor, the parser is given an early indication that it must reconstruct further to the right in the string in order to interpret the DP. If the parser has already reconstructed a phrase to a lower position, we predicted that it should be easier for the parser to engage in a second reconstruction operation. Building on a paradigm introduced in Hackl et al. (2012), we predicted that reconstruction for binding would facilitate reconstruction for scope.

In a series of psycholinguistic experiments, we investigated how the parser assigns a semantic interpretation to the complement of *wh* when the complement phrase cannot be interpreted in the position in which it appears on the surface. More precisely, by using embedded *wh*-constructions, we presented different types of DPs in positions in which they linearly preceded their antecedents (for anaphors and pronouns). In Experiments 1-2, we also investigated whether the parser was sensitive to reconstruction for the *de dicto* interpretations of creation verbs. We will briefly summarize the main findings of each chapter. We will then outline some open questions and avenues for future research. We will conclude this chapter and the thesis by discussing the theoretical consequences of this work.

6.1.1 Chapter-by-chapter summaries

In Chapter 2, we investigated whether the parser is sensitive to a PF-LF mismatch using constraints on the processing of anaphors. In Experiment 1, we presented anaphors in *wh*-fillers in positions in which they linearly preceded their antecedents. We compared reading times on anaphors and R-expressions in the same syntactic position. When anaphors are first encountered in the complement of *wh*, they are uninterpretable since they cannot be assigned a semantic denotation in this position; their interpretation depends on their antecedent. Following Binding Principle A, anaphors must be bound by local antecedents. To be able to interpret the phrase containing the anaphor, the parser would need to reconstruct the phrase to a lower copy position, below the antecedent (embedded subject). We also manipulated the type of verb found in the embedded clause. If the verb was a creation verb, reconstruction for the *de dicto* interpretation was obligatory. However, the parser would only find out that the embedded verb is a verb of creation upon finding the verb site. Since we used two different motivations for reconstruction, we predicted that an early signal for reconstruction for binding (an anaphor in the complement of *wh*) would facilitate reconstruction for the *de dicto* interpretation of creation verbs. Results indicated that while anaphors and R-expressions were processed similarly when they were first encountered (i.e., in the *wh*-phrase), anaphor conditions incurred a greater processing cost (longer reading times) on the words following the DP until the antecedent was found further to the right in the string. We interpreted this effect as providing evidence for a reconstruction analysis of anaphoric binding. Upon finding an anaphor that does not yet have an antecedent, the parser is given a signal that it must reconstruct the phrase containing the anaphor (complement of *wh*) to a lower syntactic copy position in which it will be bound by its antecedent and receive a semantic denotation. These results are also compatible with a non-structural account of antecedent resolution following which the parser simply searches for an antecedent that matches the anaphor in morphological features irrespective of its syntactic position. However, we did not find a reliable effect of the verb type. Thus, we make no claims about parsing at the embedded verb position or about the parser's (in)sensitivity to reconstruction for *de dicto* interpretations.

In Chapter 3, we examined whether we could find stronger evidence for a structural analysis of our results from Experiment 1. In Experiments 2 and 3, we examined how the parser finds antecedents for pronouns both when they linearly precede and linearly follow their antecedents. We compared the processing of pronouns and R-expressions presented in *wh*-fillers that appeared at the beginning of the sentence. Since pronouns can also serve as embedded subjects, we also manipulated whether the embedded subject was a pronoun or R-expression. Results indicated that conditions containing R-expressions incurred a greater processing cost in both syntactic positions. We argued that the increased processing difficulty associated with R-expressions when they appear inside *wh*-fillers could be explained by a greater number of discourse properties on R-expressions compared to pronouns. In contrast, we suggested that the effect found in the embedded subject region could be explained as a reanalysis effect. When the embedded subject was an R-expression, the DP in the *wh*-filler was a pronoun. If the parser respects the grammar's preference to adopt binding over co-reference interpretations, it will attempt to reconstruct the pronoun in the *wh*-filler to a syntactically lower position. However, in our stimulus sentences, reconstructing the *wh*-phrase containing a pronoun to a position below the embedded subject would incur a Principle B violation. As a result, the parser is forced to adopt a co-reference interpretation between the pronoun and its antecedent and doing so incurs a processing cost because it is not the preferred parse. Our argument that the effects on the two different regions reflect different processing costs was supported by different effect sizes found on the two positions and an interaction between DP type and its position in the sentence. In Experiment 2, we again used the creation verb manipulation, predicting that if reconstruction for *de dicto* interpretations was obligatory, we would observe Principle B and C violations when the embedded verb was a creation verb. We did not find such effects, again making it unclear if the parser is sensitive to reconstruction for *de dicto* interpretations. We did find a main effect of Verb type in Experiment 2, showing that non-creation verbs incurred a greater cost than creation verbs. We suggested that this effect arose due to the non-creation verb sentences seeming more unnatural than the creation verb ones. Since we failed to find effects of reconstruction for creation verbs in both Experiments 1-2, we dropped this manipulation in subsequent studies.

In Experiment 3, we additionally investigated whether the effects found in Experiment 2 could be attributed solely to R-expressions carrying more discourse properties than pronouns. In this follow-up study, we used both familiar (names of famous people) and unfamiliar R-expressions (made-up names with matching number of syllables). We replicated the effects found in Experiment 2 in both syntactic positions. The familiarity manipulation was significant in the first DP position (i.e., in the *wh*-filler phrase), demonstrating that familiar R-expressions incurred less of a processing cost compared to unfamiliar R-expressions. Crucially, R-expressions still incurred a greater processing cost compared to pronouns, despite the familiarity manipulation, suggesting that there is a general cost associated with R-expressions that cannot simply be explained as them being unfamiliar to participants. If our previous effect could be solely explained as resulting from the R-expressions being unfamiliar to participants, we would have expected familiar R-expressions to behave like pronouns in our Experiment 3. At the second DP region (i.e., the embedded subject), the familiarity manipulation did not reach significance, suggesting that our Experiment 2 results at this region cannot be explained as resulting from R-expressions carrying more discourse properties than pronouns.

In Chapter 4, we investigated how the parser processes constructions in which there are two possible antecedents for a pronominal found further to the left in the string of words. In Experiments 1-3, there was only one possible antecedent for the pronominals and it always occurred in the embedded subject position from which the pronominals could be structurally bound under a reconstruction analysis at LF. In Experiments 4a, 4b and 5, a second potential antecedent appeared in an adjunct phrase that linearly followed the *wh*-filler phrase containing the pronominal but linearly preceding the embedded subject. In these experiments, we compared the processing of anaphors and pronouns embedded in *wh*-filler phrases. The second potential antecedent DP in the adjunct phrase either matched or did not match the pronominal in gender. We used adjuncts to investigate whether the parser considers potential antecedents for pronominals even when they do not appear in appropriate structural positions. Since the DPs were embedded in adjuncts, they were not found in structural positions from which they could bind the pronominals at LF (i.e., they would not c-command the pronominal, even under a reconstruction analysis). In Experiments 6a-6b, the stimulus sentences again contained two potential antecedents for the anaphor or pronoun. However, in this case, the second potential antecedent appeared as the subject of the asking event and was thus in an appropriate position to bind the anaphor or pronoun at LF, depending on the reconstruction position of the complement of *wh*. We again manipulated the gender of the second potential antecedent.

Crucially, we observed different gender effects based on the syntactic position of the intervening DP. When the second potential antecedent DP appeared in adjunct position, DPs that matched the pronominals in gender incurred a processing cost compared to DPs that mismatched the pronominals in gender. We suggested that this effect arises because the parser is given conflicting information based on morphological and syntactic properties. If the DP is embedded in an adjunct, it is not in a structural position from which it can bind the pronominal, even under a reconstruction analysis. If, however, the morphological features are appropriate, the DP has passed the first filter (morphological features) but then the syntactic configuration is inappropriate and the parser must reject this DP as a potential antecedent. In contrast, when the potential antecedent appeared in subject position, we found the opposite effect. In this case, DPs that mismatched pronominals in gender incurred a greater processing cost compared to DPs that matched the pronominals in gender. In this case, since the DP in subject position can serve as an antecedent for a pronominal found further to the left, the parser was expecting to find a DP that matched the pronominal in morphological features. When the DP instead mismatches the pronominal in gender, the parser is again given conflicting information and it must revise its parse. Thus, the results from the experiments reported in Chapter 4 demonstrate that not only is the parser sensitive to gender effects associated with potential antecedents but it is also sensitive to their syntactic positions.

In Chapter 5, we proposed a series of complexity metrics to account for the results reported in Chapters 2-4. These complexity metrics were designed as a first step towards a model of LF processing. The proposed metrics are based on constraints proposed in semantic theory as well as general processing preferences. The interaction between different complexity metrics was most relevant for our discussion. The complexity metrics were shown to correctly predict many of the effects arising based on the type of DP (anaphor, pronoun, R-expression). Crucially it was the combination of various complexity metrics that correctly predicted some of the collected data. We also pointed out where our metrics failed to make the right predictions for the collected data. We also saw that the current version of the complexity metrics predict some but not all of the gender effects reported in Chapter 4. The experiments reported in Chapters 2-4 were complex and involved many different variables. Future work should continue to fine tune these experiments in an effort to develop more precise complexity metrics that predict all of the collected data.

6.2 Open questions

In the following sub-sections, we briefly discuss some of the open questions/issues that have arisen while writing this thesis. These questions present possible avenues for future research in the domain of LF processing. After presenting some of the open questions, we will discuss the theoretical contribution of this thesis and provide some concluding remarks.

6.2.1 Multiple referents

We have been assuming that it is simpler (more economical) if there is a match between the PF and LF representations of a string of words, i.e., if it is possible to build the LF structure from the linearized string of words. It is thus less economical to build an LF structure that is completely distinct from the PF representation of the sentence. However, there seems to be another way to investigate this question that we have generally ignored in this thesis.[156] Consider the following example:

(1) Joanne had no idea which picture of herself Lady Gaga put on the cover of the album.

(Brian Dillon, PC)

This sentence contains a *wh*-phrase which is embedded with an anaphor. The anaphor needs to find a structurally appropriate (c-commanding) antecedent within the sentence. (1) differs from our experimental stimuli because in this sentence, the matrix subject can seemingly serve as the antecedent for the anaphor. The matrix subject, *Joanne*, bears appropriate gender and number features to serve as the antecedent for the anaphor in the *wh*-phrase. Since it is structurally higher than the anaphor, it is also appears to be in an appropriate structural position to be able to c-command and, therefore bind, the anaphor at LF. In this thesis, we have been assuming a strictly structural definition of Binding Theory following which the anaphor must be locally bound by its antecedent.[157] For simplicity, we have assumed that the local binding domain is the clause. This means that the anaphor and its antecedent must appear in the same clause and the antecedent must c-command

[156] Thanks to Brian Dillon for bringing this to our attention.

[157] Non-structural analyses of binding have been generally put aside in this thesis but see Pollard and Sag (1992); Reinhart and Reuland (1993) for further details.

the anaphor in order to abide by Binding Principle A (Chomsky, 1981, 1986, see also Charnavel and Sportiche, 2016 for a recent proposal arguing for a revised, more restricted definition of Principle A). Notice that in (1), the matrix subject and the anaphor do not appear in the same binding domain because they are not part of the same clause. The matrix subject, *Joanne* is part of the matrix clause, whereas the anaphor, *herself*, is contained in the embedded clause. Thus, if the structural analysis of binding theory is on the right track, it should not be possible for the matrix subject to bind the anaphor in this example.

There is another DP in the sentence that can serve as the antecedent for the anaphor: *Lady Gaga*. In order to derive this interpretation, the parser must reconstruct the phrase containing the anaphor to a position that is structurally lower than *Lady Gaga*. In our experimental stimuli, this was the only interpretation we could derive (either under a structural or non-structural analysis of binding) because the matrix subject never matched the anaphor in morphological features (it was always a plural DP). The example in (1) therefore presents the parser with a puzzle. In order to adopt a structural analysis of binding at LF, it must reconstruct the phrase containing the anaphor (complement of *wh*). Since this parse involves a PF-LF mismatch, it is less economical than one in which the PF and LF representation match. To have a match between the LF and PF representations, the parser would need to adopt a non-structural analysis of binding in which the anaphor is associated with the matrix subject, *Joanne*. However, this parse would violate the structural analysis of binding.

The parser is therefore faced with two alternatives, both of which violate one of its preferences. If it is the case that reconstructed readings are strongly dispreferred, there should be a general preference to interpret *herself* as referring to *Joanne* and it should not be possible for the anaphor to refer to *Lady Gaga*. No reconstruction is required if the anaphor refers to the matrix subject because the anaphor can be interpreted as soon as it is encountered. We could view this as a backwards search mechanism. Upon reaching the anaphor, the parser reviews the previous linguistic discourse and finds a match (*Joanne*). It can therefore interpret the anaphor and does not need to continue looking further to the right for an antecedent. However, this parse would violate the grammar's preference to adopt structural relations between DPs since this parse would violate Binding Principle A.

If the preference to adopt structural relations between DPs is more highly weighted than the preference to interpret phrases as soon as they encountered (PF-LF match), the parser should reconstruct the phrase containing the anaphor and only the reading in which the anaphor is co-indexed with *Lady Gaga* should be possible. This prediction seems to be borne out: it seems much more natural to interpret *herself* as referring to *Lady Gaga* in this sentence, even though this interpretation requires reconstruction of the complement of *wh*, as shown in (2).

(2) Joanne had no idea [which picture of herself] Lady Gaga put ~~[picture of herself]~~ on the cover of the album.

Unfortunately, we did not conduct experiments where the matrix subject of the sentence matched the pronominal in gender and number features. To investigate whether the parser is sensitive to the fact that the matrix subject is in a separate binding domain and therefore cannot locally bind the anaphor, we would need to run an experiment with this additional manipulation. If the parser abides by the locality restriction on Principle A, we would predict that the parser should not consider the matrix subject as a potential binder for the anaphor and that it should adopt the reconstruction analysis.

A related question is what happens if there are two potential antecedents in the sentence and both are in the same binding domain as the anaphor? Consider (3).

(3) Joanne had no idea which picture of herself Lady Gaga thought that Tori Amos would put on the cover of the album.

(Brian Dillon, PC)

Assuming that anaphors need to find local antecedents and that the matrix subject and the anaphor are in separate binding domains, the matrix subject cannot serve as the antecedent for the anaphor. The parser must therefore reconstruct the complement of *wh* to a lower structural position so that the anaphor can find a local antecedent. In (3), there are two possible antecedents, depending on the reconstruction position of the complement of *wh*. If the parser reconstructs the phrase containing the anaphor to the lower gap position (following the verb, *put*), the only local antecedent would be *Tori Amos*. The anaphor could not be associated with *Lady Gaga* because they are in separate binding domains. However, assuming successive cyclic movement, the *wh*-phrase must stop off at the edge of each intermediate clause before reaching its final position. Consequently, it is also possible for the anaphor to be associated with *Lady Gaga* if there is a reconstruction position at the edge of the intermediate clause. This example is similar to the stimulus sentences used in Experiments 6a-6b where we observed that participants prefer to associate the anaphor in the *wh*-filler with the closest potential antecedent, i.e., *Lady Gaga*. It is unclear if this is truly a preference or if our results might reflect a working memory effect. The sentences that are being presented to participants are long and are always presented word by word. As each new word is presented, the previous word disappears. As a result, the parser must keep all previously encountered words in memory as it continues to build the structure. It is possible that the parser associates the anaphor with the linearly closest gender-matched DP to alleviate working memory resources. Even though it would still need to keep the previously built structure in active memory, it would be able to resolve the open dependency, alleviating memory resources. One way to test this would be to conduct a study using a methodology that allows us to present the whole sentence to participants at once, such as eye-tracking, instead of presenting the sentences to participants in a word-by-word fashion.

6.2.2 Types of comprehension questions

We have briefly discussed the issue of the types of comprehension questions used in our experiments but we have not yet discussed the significance of the findings. Crucially, we have shown that we can obtain different results in reading times when different types of comprehension questions are used in our experiments (see also Stewart et al., 2007; Swets et al., 2008). When surface processing questions are used and participants are able to answer questions correctly if they simply retain the sentence in memory, we are able to capture a very basic understanding of how participants interpret our critical sentences. Surface processing questions do not require participants to interpret the different DPs in the sentence. As a result, we have no way of knowing if participants are actually interpreting the anaphor/pronoun or whether they are assigning the pronominal a very surface level meaning. When we used deep processing questions, in which participants were forced to interpret the anaphor/pronoun because they were required to choose a referent for the pronominal, we found different reading time results. These results suggest that the types of questions participants are asked has an influence on their real-time processing and how they interpret the critical sentences. In real-time processing studies, comprehension questions tend to be used to verify that participants were paying attention to the sentences they were reading. Higher comprehension scores are interpreted as evidence that participants were paying attention and fully reading the critical sentences. However, if we obtain different results based on the types of comprehension questions asked, researchers should be paying attention to the types of questions they are using in their experiments, especially when investigating questions about interpretation.[158]

6.2.3 Reconstruction for scope vs reconstruction for binding

In Experiments 1 and 2, in addition to investigating how the parser finds referents for pronominals found further to the left in the linear string, we also investigated whether the parser is sensitive to scope reconstruction effects in real-time processing. Our specific research question was whether finding an early signal for reconstruction would make a second reconstruction operation further to the right easier to process. We used reconstruction for binding as our early signal. The assumption was that upon finding an anaphor in the *wh*-phrase at the beginning of the sentence, the parser would be prepared to reconstruct the complement of *wh* to a lower structural position. Since the parser was already given an early signal about reconstruction of the complement of *wh*, we predicted that it would be less costly for the parser to perform a second reconstruction operation later in the sentence (i.e., further to the right). We targeted the second reconstruction operation with creation verbs which force *de dicto* readings of *how-many* questions. In order to derive the *de dicto* reading, the *many-x* phrase must be interpreted lower in the structure, structurally below the scope-bearing operator (in our cases, the verb *ask*). We predicted that finding an anaphor early in the sentence would signal to the parser

[158] A related point is that the type of experimental methodology might impact the results. See Grant (2016) for experimental data on ambiguity resolution demonstrating differences in the results obtained in eye-tracking versus self-paced reading studies.

that the phrase containing the anaphor must reconstruct to a structurally lower position. Upon encountering the creation verb, the parser would still need to perform another reconstruction operation but it should be easier to do so when it had been given an early signal about reconstruction for binding, compared to a construction in which the parser was not given an early signal about reconstruction for binding but still encounters a creation verb further to the right in the sentence. The crucial prediction was that we should observe an interaction between the type of DP found in the *wh*-filler and the type of verb found in the embedded clause. Reconstruction for scope should be less costly when the filler contains an anaphor compared to when it contains an R-expression.

Our results from Experiments 1 and 2 showed effects of the type of DP found in the filler phrase and we have argued that these effects are compatible with a reconstruction account of binding at LF. However, we did not find effects of the type of verb used in the sentences and we crucially did not find an interaction between the type of DP and the type of verb. Thus, we did not find evidence that an early signal for reconstruction for binding facilitates later reconstruction for scope. However, since we failed to find effects of the type of verb in our experiments,[159] it is unclear if our methodology was not sensitive enough to find such an effect, if the parser is not sensitive to reconstruction for scope or if reconstruction for scope and reconstruction for binding are fundamentally different, at least in the domain of real-time processing. We have argued that our other results are compatible with a reconstruction account of binding at LF. Thus, the parser seems to be sensitive to reconstruction effects in real-time processing.

Why we were not able to also find reconstruction effects for scope at LF? One

[159]Note that we did find an effect of the type of verb in Experiment 2 demonstrating that non-creation verbs were read significantly longer than creation verbs. This result was unexpected as we expected creation verbs to be more costly than non-creation verbs. As discussed in Chapter 3, we asked CQs directly targeting participants' interpretation of the event in Experiment 2 which may have drawn their attention to the semantics of the verbs. No such questions were asked in Experiment 1.

possibility is that the difference between the readings associated with creation and non-creation verbs is quite subtle and such readings are not able to be computed in rapid online parsing. This does not seem to be the right approach since our CQs targeting participants' interpretation of the verb types revealed that they interpreted creation and non-creation verbs differently. Another possibility is that participants were deriving the *de dicto* interpretation across conditions. Creation verbs force the *de dicto* interpretation of *many-x*; the *de re* reading is unavailable. Both readings are possible with non-creation verbs. As a starting point, we assumed that the *de re* reading was the simpler representation since *many-x* can be interpreted in its surface position. However, maybe this assumption is not on the right track and participants prefer to adopt a *de dicto* interpretation. One way to investigate this question further would be to test constructions that are only compatible with a *de re* interpretation. If our assumption that the *de re* reading is the simpler parse is not on the right track, then parses that are only compatible with the *de re* reading should be more costly than those that are compatible with both *de re* and *de dicto* readings. It is also possible that *de re* and *de dicto* readings do not differ based on structure and thus, we would not expect to observe differences between the two readings in processing. See the discussion in Section 2.5.3 (Chapter 2). We leave it to future work to investigate these questions further.

6.3 Theoretical contribution and thesis conclusions

The results of the current thesis demonstrate that the parser is sensitive to PF-LF mismatches in real-time processing. Foundational work on filler-gap dependencies (Stowe, 1986; Crain and Fodor, 1985) has investigated whether the left-to-right parser is sensitive to syntactic constraints on *wh*-movement. This work has shown that upon finding a *wh*-phrase at the beginning of the linear string, the parser predicts that it will find a gap further to the right in the string. It must find this gap to satisfy syntactic constraints on *wh*-movement. When the closest linear position in which a gap could be found is filled with an overt phrase, processing difficulty is incurred and this is when we find the filled-gap effect. Work on filler-gap dependencies can be understood as demonstrating one type of PF-LF mismatch in processing. Novel to the current thesis is that the parser is also sensitive to whether a phrase can be interpreted in its surface position (higher copy position) or whether it must be interpreted in a distinct position (lower copy position) at LF. We assumed that to be able to interpret a phrase, it must have a semantic denotation. In general, our results from a series of experiments are compatible with a reconstruction account of referential processing at LF. These results suggest that the parser is sensitive to PF-LF mismatches in the domain of interpretation. To our knowledge, this thesis is the first attempt to directly investigate whether the parser is sensitive to reconstruction operations required for semantic interpretation in real-time processing. As discussed in the previous sub-sections, we are still left with many unanswered questions which open many avenues for future research. We will conclude the thesis with a discussion of the main theoretical contribution of this work.

Recall that we began this thesis assuming Chomsky (1995)'s Y-model. We assumed that the phonological form of the sentence (PF) is the parser's input and the parser's goal is to derive a structure that is readable by the semantic component (LF). To investigate this question, we used constructions argued to have distinct PF and LF representations. We assumed that if the parser was sensitive to LF, it would be sensitive to PF-LF mismatches in processing. As a result, we argued that real-time processing studies provide us with a tool to directly investigate whether the parser is sensitive to the LF component of the grammar and if so, how it builds structures that are readable by the semantic component. The results presented in Chapters 2-4 suggest that there is mapping between the PF and LF interfaces via Spell-Out (syntax). If there was no mapping between the two interfaces, the parser would not be able to derive an LF structure from the PF input. Generally, the results of our experiments suggest that the parser is sensitive to the LF structure it needs to build in real-time processing. We assume that the parser must be able to access the form of the derivation at one of the Spell-Out domains. After reaching Spell-Out, the derivation is sent to the interfaces and no further syntactic operations on that domain can take place. Since the parser needs to have access to the Spell-Out form of the sentence but its input is the PF form of the sentence, the parser must be able to take the linearized form of the sentence and reconstruct its Spell-Out form. Once the parser has access to the Spell-Out form of the sentence, it can build an appropriate LF structure. Since copies appear in (at least) two distinct positions at Spell-Out, the parser is able to interpret the phrase in any of these copy positions, as long as it follows the grammar's preferences and structural constraints. Thus, our results suggest not only that the parser is sensitive to LF but also that it is able to (re)build the Spell-Out form of the sentence in order to derive an interpretable LF, i.e., a structure that is readable by the semantic component.

If the parser was not sensitive to LF structure, we would not expect to observe reading time differences based on the structural position of different types of DPs. The finding that the parser builds interpretable LF structures is an important one because very little work in psycholinguistics has investigated potential effects of LF structure even though they have been extensively studied in the theoretical linguistics literature. In real-time processing studies, participants are presented with the PF representation of the sentence. Previous results demonstrating that the parser is sensitive to syntactic structure in processing, namely, work on syntactic ambiguity resolution (Bever, 1970; Frazier and Fodor, 1978), has relied on cases in which the hierarchical structure could be built based on the order in which the words are presented on the surface, i.e., PF-LF match. Such results do not demonstrate if the parser is sensitive to a LF representation that is distinct from the pronounced form of the sentence (PF-LF mismatch). Previous work on semantic ambiguities (Tunstall, 1998; Anderson, 2004; Conroy, 2008) has shown that the parser is sensitive to LF when there is more than one possible semantic interpretation of a given string. In such cases, a given phrase could either be interpreted in its surface position (PF-LF match) or in a distinct structural position at LF (PF-LF mismatch). These two parses lead to two different interpretations of the sentence. In the current work, we investigated cases where a given phrase could not be interpreted in its surface position and needed to be interpreted in a distinct position at LF (see also Hackl et al., 2012) By systematically manipulating the PF and LF positions of phrases, building on previous work in the filler-gap domain (Crain and Fodor, 1985; Stowe, 1986), our results provide real-time evidence that the parser is sensitive to two different levels of representations. In general, the parser must take the PF input and build an LF representation that can be interpreted (see also Conroy, 2008). While we are left with many unanswered questions about how exactly this is accomplished, our experimental results provide a promising avenue for future research in the domain of interpretation.

References

Abels, Klaus. 2012. *Phases: An essay on cyclicity*. Berlin: Mouton de Gruyter.

Altmann, Gerry, and Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30:191–238.

Anderson, Catherine. 2004. The structure and real-time comprehension of quantifier scope ambiguity. Doctoral Dissertation, Northwestern University.

Baayen, R Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R Harald, Douglas J Davidson, and Douglas M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.

Baayen, R Harald, and Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3:12–28.

Badecker, William, and Kathleen Straub. 2002. The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:748.

Barr, Dale J, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, D.M., and D. Sarkar. 2007. The lme4 library. URL `https://CRAN.R-project.org/package=lme4`.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015a. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967* .

Bates, Douglas, Martin Mäechler, Ben Bolker, and Steve Walker. 2015b. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In *Cognition and the development of language*, ed. J.R. Hayes, 279–362. New York: John Wiley & Sons.

Boland, Julie E, Michael K Tanenhaus, and Susan M Garnsey. 1990. Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language* 29:413–432.

Bruening, Benjamin. 2014. Precede-and-command revisited. *Language* 90:342–388.

Charnavel, Isabelle, and Dominique Sportiche. 2016. Anaphor binding: What French inanimate anaphors show. *Linguistic Inquiry* 47:37–87.

Chomsky, Noam. 1981. *Government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001a. *Beyond explanatory adequacy*, volume 20. MITWPL.

Chomsky, Noam. 2001b. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Clifton, Charles, Shelia M Kennison, and Jason E Albrecht. 1997. Reading the words *her, his, him*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language* 36:276–292.

Clifton, Charles, Adrian Staub, and Keith Rayner. 2007. Eye movements in reading words and sentences. *Eye movements: A window on mind and brain* 341–372.

Conroy, Anastasia Marie. 2008. The role of verification strategies in semantic ambiguity resolution in children and adults. Doctoral Dissertation, University of Maryland.

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

Crain, Stephen, and Janet D Fodor. 1985. How can grammars help parsers? In *Natural Language Parsing: Psychological, Computational, and Theoretical perspectives*, ed. D.R. Dowty, L. Karttunen, and A.M. Zwicky, 94–128. Cambridge: Cambridge University Press.

Cunnings, Ian, Clare Patterson, and Claudia Felser. 2014. Variable binding and coreference in sentence comprehension: evidence from eye movements. *Journal of Memory and Language* 71:39–56.

Davies, Mark. 2013. Corpus of News on the Web (now): 3+ billion words from 20 countries, updated every day. URL `http://corpus.byu.edu/now`.

Dillon, Brian, Alan Mishler, Shayne Sloggett, and Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69:85–103.

Dowty, David. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.

Ferreira, Fernanda, Karl GD Bailey, and Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current directions in psychological science* 11:11–15.

Ferreira, Fernanda, and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25:348–368.

Ferreira, Fernanda, and John M Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30:725–745.

Fox, Daniel. 1998. Economy and semantic interpretation: A study of scope and variable binding. Doctoral Dissertation, Massachusetts Institute of Technology.

Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3:283–341.

Fox, Danny. 1999. Reconstruction, binding theory, and the interpretation of chains. *Linguistic Inquiry* 30:157–196.

Fox, Danny. 2000. *Economy and semantic interpretation*. Cambridge, MA: MIT Press.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Fox, Danny, and Jon Nissenbaum. 2004. Condition A and scope reconstruction. *Linguistic Inquiry* 35:475–485.

Frazier, Lyn. 1987. Sentence processing: A tutorial review. In *Attention and performance XII: The psychology of reading*, ed. M. Coltheart, 559–586. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frazier, Lyn, and Janet Dean Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6:291–325.

Frazier, Lyn, Bernadette Plunkett, and Charles Clifton. 1996. Reconstruction and scope. *University of Massachusetts Occasional Papers in Linguistics* 19:239–260.

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

Frazier, Lyn, and Keith Rayner. 1990. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language* 29:181–200.

Garnsey, Susan M, Michael K Tanenhaus, and Robert M Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18:51–60.

Gernsbacher, Morton Ann. 1990. *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gernsbacher, Morton Ann, and Joerg D Jescheniak. 1995. Cataphoric devices in spoken discourse. *Cognitive Psychology* 29:24–58.

Gernsbacher, Morton Ann, and Suzanne Shroyer. 1989. The cataphoric use of the indefinite this in spoken narratives. *Memory & Cognition* 17:536–540.

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

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain*, ed. Alec Marantz, Yasushi Miyashita, and Wayne A. O'Neil, 95–126. MIT Press.

Gordon, Peter C, Barbara J Grosz, and Laura A Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17:311–347.

Gordon, Peter C, and Randall Hendrick. 1998. The representation and processing of coreference in discourse. *Cognitive Science* 22:389–424.

Grant, Margaret. 2016. Ambiguity and incrementality in sentence processing. Talk presented at McMaster University's Cognitive Science of Language Lecture Series, January 2016.

Grodzinsky, Yosef, and Tanya Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24:69–101.

Hackl, Martin, Jorie Koster-Hale, and Jason Varvoutis. 2012. Quantification and ACD: Evidence from real-time sentence processing. *Journal of Semantics* 29:145–206.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 1–8. Association for Computational Linguistics.

Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Doctoral Dissertation, University of Massachusetts, Amherst.

Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. In *Meaning, use and interpretation of language*, ed. R. Baüerle, Ch. Schwarze, and A. von Stechow. Berlin: Mouton de Gruyter.

Heim, Irene. 1998. Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. *MIT Working Papers in Linguistics* 25:205–246.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Heycock, Caroline. 1995. Asymmetries in reconstruction. *Linguistic Inquiry* 26:547–570.

Jescheniak, Jörg D. 2000. The cataphoric use of spoken stress in narratives. *Psychological Research* 63:14–21.

Just, Marcel A, Patricia A Carpenter, and Jacqueline D Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111:228.

Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, ed. J.A.G. Groenendijk, T.M.V. Janssen, and B.J. Stokhof. Amsterdam: Mathematisch Centrum, University of Amsterdam.

Kazanina, Nina, Ellen F Lau, Moti Lieberman, Masaya Yoshida, and Colin Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56:384–409.

Kennison, Shelia M. 2003. Comprehending the pronouns *her, him, and his*: Implications for theories of referential processing. *Journal of Memory and Language* 49:335–352.

Kennison, Shelia M, Elaine C Fernandez, and J Michael Bowers. 2009. Processing differences for anaphoric and cataphoric pronouns: Implications for theories of discourse processing. *Discourse Processes* 46:25–45.

Kennison, Shelia M, and Peter C Gordon. 1997. Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes* 24:229–252.

Kratzer, Angelika. 1994. The event argument and the semantics of voice. Ms., University of Massachusetts, Amherst.

Krifka, Manfred. 1989. Nominalreferenz und zeitkonstitution: Studien zur theoretischen linguistik. *München: Wilhelm Fink* .

Lebeaux, David. 1990. Relative clauses, licensing, and the nature of the derivation. In *Proceedings of NELS 20*, ed. Juli Carter, Rose-Marie Déchaine, Bill Philip, and Tim Sherer, 318–332. Amherst, MA: GLSA.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106:1126–1177.

Marslen-Wilson, William. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244.

Marslen-Wilson, William. 1975. Sentence perception as an interactive parallel process. *Science* 189:226–228.

May, Robert. 1977. The grammar of quantification. Doctoral Dissertation, Massachusetts Institute of Technology.

Nicol, Janet, and David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18:5–19.

Nissenbaum, Jon. 2000. Investigations of covert movement. Doctoral Dissertation, Massachusetts Institute of Technology.

Obenauer, H-G. 1984. On the identification of empty categories. *Linguistic Review* 4:153–202.

Omaki, Akira. 2010. Commitment and flexibility in the developing parser. Doctoral Dissertation, University of Maryland, College Park.

Phillips, Colin. 1996. Order and structure. Doctoral Dissertation, Massachusetts Institute of Technology.

Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82:795–823.

Pinheiro, José C., and Douglas M. Bates. 2000. *Mixed effects models in S and S-PLUS*. New York: Springer.

Plunkett, Bernadette. 1991. The parsing of anaphor binding and levels of representation. *University of Massachusetts Occasional Papers in Linguistics* 15:205–275.

Pollard, Carl, and Ivan A Sag. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23:261–303.

Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. London: Croom Helm.

Reinhart, Tanya. 2006. *Interface strategies*. Cambridge, MA: MIT Press.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.

Reinhart, Tanya Miriam. 1976. The syntactic domain of anaphora. Doctoral Dissertation, Massachusetts Institute of Technology.

Reuland, Eric J. 1983. Governing-ing. *Linguistic Inquiry* 14:101–136.

van Riemsdijk, Henk. 1978. *A case study in syntactic markedness: The binding nature of prepositional phrases*. Dordrecht: Foris.

Roelofsen, Floris. 2010. Condition B effects in two simple steps. *Natural Language Semantics* 18:115–140.

Roelofsen, Floris. 2011. Free variable economy. *Linguistic Inquiry* 42:682–697.

Ross, John R. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Rullmann, Hotze. 1993. Scope ambiguities in how many-questions. Unpublished manuscript, University of Massachusetts.

Sag, Ivan A. 1976. Deletion and logical form. Doctoral Dissertation, Massachusetts Institute of Technology.

Sauerland, Uli. 1999. Erasability and interpretation. *Syntax* 2:161–188.

Sauerland, Uli, and Paul Elbourne. 2002. Total reconstruction, PF movement, and derivational order. *Linguistic Inquiry* 33:283–319.

Schütze, Carson T, and Jon Sprouse. 2014. Judgment data. *Research methods in linguistics* 27.

Sekerina, Irina A, Karin Stromswold, and Arild Hestvik. 2004. How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* 31:123–152.

Staub, Adrian, and Keith Rayner. 2007. Eye movements and on-line comprehension processes. In *The Oxford Handbook of Psycholinguistics*, 327–342. Oxford University Press.

von Stechow, Arnim. 2001. Temporally opaque arguments in verbs of creation. *Semantic Interfaces: Reference, Anaphora, and Aspect, CSLI* 278–319.

Stewart, Andrew J, Judith Holler, and Evan Kidd. 2007. Shallow processing of ambiguous pronouns: Evidence for delay. *The Quarterly Journal of Experimental Psychology* 60:1680–1696.

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

Stowe, Laurie A, Michael K Tanenhaus, and Greg Carlson. 1991. Filling gaps on-line: use of lexical and semantic information in sentence processing. *Language and Speech* 34:319–40.

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

Swaab, Tamara Y, C Christine Camblin, and Peter C Gordon. 2004. Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience* 16:715–726.

Swets, Benjamin, Timothy Desmet, Charles Clifton, and Fernanda Ferreira. 2008. Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition* 36:201–216.

Tanenhaus, Michael, Greg Carlson, and John C. Trueswell. 1989. The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes* 4:211–34.

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

Tunstall, Susanne Lynn. 1998. The interpretation of quantifiers: Semantics & processing. Doctoral Dissertation, University of Massachusetts, Amherst.

Van Gompel, Roger PG, and Simon P Liversedge. 2003. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:128.

Wagers, Matt, and Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45:395–433.

Zepter, Alex. 2000. Verbs of creation: Cause become or become cause. On Arnim von Stechow (1999): Temporally Opaque Arguments. Rutgers University: Unpublished Term Paper.

Appendices

Appendix A

Experiment 1 stimuli

(1) a. The editors wondered how many cartoons about herself you are asking Jessica to sketch for the upcoming edition of the student newspaper.
Was Jessica being asked to sketch cartoons about herself for the newspaper? (Yes)

b. The editors wondered how many cartoons about herself you are asking Jessica to approve for the upcoming edition of the student newspaper.
Was Jessica being asked to approve cartoons about herself for the newspaper? (Yes)

c. The editors wondered how many cartoons about Joseph you are asking Jessica to sketch for the upcoming edition of the student newspaper.
Was Jessica being asked to sketch cartoons about herself for the newspaper? (No)

d. The editors wondered how many cartoons about Joseph you are asking Jessica to approve for the upcoming edition of the student newspaper.
Was Jessica being asked to approve cartoons about herself for the newspaper? (No)

(2) a. The reporters wondered how many lies about herself you are asking Alexa to invent for tomorrow's live television interview.
Was it the police who wondered how many lies about herself Alexa was being asked to invent? (No)

b. The reporters wondered how many lies about herself you are asking Alexa to deny for tomorrow's live television interview.
Was it the police who wondered how many lies about herself Alexa was being asked to deny? (No)

c. The reporters wondered how many lies about Sean you are asking Alexa to invent for tomorrow's live television interview.
Was it the reporters who wondered how many lies about Sean Alexa was being asked to invent? (Yes)

d. The reporters wondered how many lies about Sean you are asking Alexa to deny for tomorrow's

live television interview.
Was it the reporters who wondered how many lies about Sean Alexa was being asked to deny? (Yes)

(3) a. The teachers wondered how many sketches of himself you are asking Stuart to draw for the art school application portfolio.
Did Stuart need to make some sketches of himself for the portfolio? (Yes)

b. The teachers wondered how many sketches of himself you are asking Stuart to adjust for the art school application portfolio.
Did Stuart need to fix some sketches of himself for the portfolio? (Yes)

c. The teachers wondered how many sketches of Lucy you are asking Stuart to draw for the art school application portfolio.
Did Stuart need to destroy some sketches of Lucy for the portfolio? (No)

d. The teachers wondered how many sketches of Lucy you are asking Stuart to adjust for the art school application portfolio.
Did Stuart need to destroy some sketches of Lucy for the portfolio? (No)

(4) a. The producers wondered how many stories about herself you are asking Larissa to draft for the behind-the-scenes special broadcast interview.
Was Larissa being asked to draft letters about herself for the interview? (No)

b. The producers wondered how many stories about herself you are asking Larissa to proofread for the behind-the-scenes special broadcast interview.
Was Larissa being asked to proofread letters about herself for the interview? (No)

c. The producers wondered how many stories about Henry you are asking Larissa to draft for the behind-the-scenes special broadcast interview.
Was Larissa being asked to draft stories about Henry for the interview? (Yes)

d. The producers wondered how many stories about Henry you are asking Larissa to proofread for the behind-the-scenes special broadcast interview.
Was Larissa being asked to draft stories about Henry for the interview? (Yes)

(5) a. The agents wondered how many songs about herself you are asking Heather to compose for the upcoming North American tour album.
Had Heather been required to compose some songs about herself for a fundraiser? (No)

b. The agents wondered how many songs about herself you are asking Heather to perform for the upcoming North American tour album.
Had Heather been required to perform some songs about herself for a fundraiser? (No)

c. The agents wondered how many songs about Bryan you are asking Heather to compose for the

upcoming North American tour album.
Had Heather been required to compose some songs about Bryan for an album? (Yes)

d. The agents wondered how many songs about Bryan you are asking Heather to perform for the upcoming North American tour album.
Had Heather been required to perform some songs about Bryan for an album? (Yes)

(6) a. The organizers wondered how many toy models of himself you are asking Charles to build for next weekend's superhero convention.
Was Charles being asked to build toy models of Clare for the convention? (No)

b. The organizers wondered how many toy models of himself you are asking Charles to buy at next weekend's superhero convention.
Was Charles being asked to buy toy models of Clare for the convention? (No)

c. The organizers wondered how many toy models of Clare you are asking Charles to build for next weekend's superhero convention.
Was Charles being asked to build toy models of Clare for the convention? (Yes)

d. The organizers wondered how many toy models of Clare you are asking Charles to buy at next weekend's superhero convention.
Was Charles being asked to buy toy models of Clare for the convention? (Yes)

(7) a. The curators wondered how many sculptures of herself you are asking Megan to carve for the museum's end of season art show.
Was Megan being asked to demolish some sculptures of herself for an art show? (No)

b. The curators wondered how many sculptures of herself you are asking Megan to critique for the museum's end of season art show.
Was Megan being asked to demolish some sculptures of herself for an art show? (No)

c. The curators wondered how many sculptures of Benjamin you are asking Megan to carve for the museum's end of season art show.
Was Megan being asked to make some sculptures of Benjamin for an art show? (Yes)

d. The curators wondered how many sculptures of Benjamin you are asking Megan to critique for the museum's end of season art show.
Was Megan being asked to judge some sculptures of Benjamin for an art show? (Yes)

(8) a. The network wondered how many documentaries about herself you are asking Betty to commission for the celebration of her ninetieth birthday.
Was it the network who wondered how many documentaries about herself Betty was being asked to commission? (Yes)

b. The network wondered how many documentaries about herself you are asking Betty to watch

at the celebration of her ninetieth birthday.
Was it the network who wondered how many documentaries about herself Betty was being asked to watch? (Yes)

c. The network wondered how many documentaries about Allen you are asking Betty to commission for the celebration of her ninetieth birthday.
Was it the fans who wondered how many documentaries about Allen Betty was being asked to commission? (No)

d. The network wondered how many documentaries about Allen you are asking Betty to watch at the celebration of her ninetieth birthday.
Was it the fans who wondered how many documentaries about Allen Betty was being asked to watch? (No)

(9) a. The museum wondered how many portraits of himself you are asking Jasper to paint for next week's grand opening exhibit.
Was Jasper being asked to paint pictures of himself for the exhibit? (Yes)

b. The museum wondered how many portraits of himself you are asking Jasper to judge at next week's grand opening exhibit.
Was Jasper being asked to judge pictures of himself for the exhibit? (Yes)

c. The museum wondered how many portraits of Celina you are asking Jasper to paint for next week's grand opening exhibit.
Was Jasper being asked to paint figurines of Celina for the exhibit? (No)

d. The museum wondered how many portraits of Celina you are asking Jasper to judge for next week's grand opening exhibit.
Was Jasper being asked to judge figurines of Celina for the exhibit? (No)

(10) a. The buyers wondered how many drawings of himself you are asking Brad to sketch for the worldwide release of his new book.
Had Brad been required to sketch some drawings of himself for a magazine? (No)

b. The buyers wondered how many drawings of himself you are asking Brad to dig up for the worldwide release of his new book.
Had Brad been required to dig up some drawings of himself for a magazine? (No)

c. The buyers wondered how many drawings of Julia you are asking Brad to sketch for the worldwide release of his new book.
Had Brad been required to sketch some drawings of Julia for a book? (Yes)

d. The buyers wondered how many drawings of Julia you are asking Brad to dig up for the worldwide release of his new book.
Had Brad been required to dig up some drawings of Julia for a book? (Yes)

(11) a. The comedians wondered how many jokes about herself you are asking Carly to make up for this weekend's sold out stand-up show.
Did Carly need to make up jokes about herself for a stand-up show? (Yes)

b. The comedians wondered how many jokes about herself you are asking Carly to practice for this weekend's sold out stand-up show.
Did Carly need to practice jokes about herself for a stand-up show? (Yes)

c. The comedians wondered how many jokes about Jared you are asking Carly to make up for this weekend's sold out stand-up show.
Did Carly need to make up jokes about herself for a stand-up show? (No)

d. The comedians wondered how many jokes about Jared you are asking Carly to practice for this weekend's sold out stand-up show.
Did Carly need to practice jokes about herself for a stand-up show? (No)

(12) a. The hosts wondered how many photographs of herself you are asking Christina to take for the online promotional advertisement.
Was Christina being asked to shoot photographs of herself for an advertisement? (Yes)

b. The hosts wondered how many photographs of herself you are asking Christina to copy for the online promotional advertisement.
Was Christina being asked to duplicate photographs of herself for an advertisement? (Yes)

c. The hosts wondered how many photographs of Kevin you are asking Christina to take for the online promotional advertisement.
Was Christina being asked to destroy photographs of Kevin for an advertisement? (No)

d. The hosts wondered how many photographs of Kevin you are asking Christina to copy for the online promotional advertisement.
Was Christina being asked to destroy photographs of Kevin for an advertisement? (No)

(13) a. The blogger-community wondered how many rumors about himself you are asking Derek to concoct for the social media campaign.
Was it the viewers who wondered how many rumors about himself Derek was being asked to concoct? (No)

b. The blogger-community wondered how many rumors about himself you are asking Derek to deny in the social media campaign.
Was it the viewers who wondered how many rumors about himself Derek was being asked to deny? (No)

c. The blogger-community wondered how many rumors about Sarah you are asking Derek to concoct for the social media campaign.
Was it the blogger-community who wondered how many rumors about Sarah Derek was being asked to concoct? (Yes)

d. The blogger-community wondered how many rumors about Sarah you are asking Derek to deny in the social media campaign.
Was it the blogger-community who wondered how many rumors about Sarah Derek was being asked to deny? (Yes)

(14) a. The viewers wondered how many rumors about herself you are asking Laura to make up for the upcoming reality TV show.
Was Laura being asked to make up some jokes about herself for the TV show? (No)

b. The viewers wondered how many rumors about herself you are asking Laura to discredit for the upcoming reality TV show.
Was Laura being asked to discredit some jokes about herself for the TV show? (No)

c. The viewers wondered how many rumors about Tom you are asking Laura to make up for the upcoming reality TV show.
Was Laura being asked to make up some gossip about Tom for the TV show? (Yes)

d. The viewers wondered how many rumors about Tom you are asking Laura to discredit for the upcoming reality TV show.
Was Laura being asked to discredit some gossip about Tom for the TV show? (Yes)

(15) a. The publicists wondered how many posts about himself you are asking Roger to create for the upcoming album release.
Was Roger being asked to create posts about himself for an album release? (Yes)

b. The publicists wondered how many posts about himself you are asking Roger to disregard for the upcoming album release.
Was Roger being asked to disregard posts about himself for an album release? (Yes)

c. The publicists wondered how many posts about Stephanie you are asking Roger to create for the upcoming album release.
Was Roger being asked to create posts about Stephanie for an awards ceremony? (No)

d. The publicists wondered how many posts about Stephanie you are asking Roger to disregard for the upcoming album release.
Was Roger being asked to disregard posts about Stephanie for an awards ceremony? (No)

(16) a. The directors wondered how many segments about herself you are asking Marisa to draft for the special edition broadcast.
Had Marisa been required to draft segments about Andrew for the broadcast? (No)

b. The directors wondered how many segments about herself you are asking Marisa to edit for the special edition broadcast.
Had Marisa been required to edit segments about Andrew for the broadcast? (No)

c. The directors wondered how many segments about Andrew you are asking Marisa to draft for the special edition broadcast.
Had Marisa been required to draft segments about Andrew for the broadcast? (Yes)

d. The directors wondered how many segments about Andrew you are asking Marisa to edit for the special edition broadcast.
Had Marisa been required to edit segments about Andrew for the broadcast? (Yes)

(17) a. The parents wondered how many columns about himself you are asking Joey to write for the drama club yearbook.
Was Joey being asked to delete columns about himself from the yearbook? (No)

b. The parents wondered how many columns about himself you are asking Joey to shorten for the drama club yearbook.
Was Joey being asked to delete columns about himself from the yearbook? (No)

c. The parents wondered how many columns about Susie you are asking Joey to write for the drama club yearbook.
Was Joey being asked to compose columns about Susie for the yearbook? (Yes)

d. The parents wondered how many columns about Susie you are asking Joey to shorten for the drama club yearbook.
Was Joey being asked to condense columns about Susie for the yearbook? (Yes)

(18) a. The instructors wondered how many pictures of herself you are asking Heidi to take for the art class scrapbook.
Was it the instructors who wondered how many pictures of herself Heidi was being asked to take? (Yes)

b. The instructors wondered how many pictures of herself you are asking Heidi to organize for the art class scrapbook.
Was it the instructors who wondered how many pictures of herself Heidi was being asked to organize? (Yes)

c. The instructors wondered how many pictures of Nicholas you are asking Heidi to take for the art class scrapbook.
Was it the photographers who wondered how many pictures of Nicholas Heidi was being asked to take? (No)

d. The instructors wondered how many pictures of Nicholas you are asking Heidi to organize for the art class scrapbook.
Was it the photographers who wondered how many pictures of Nicholas Heidi was being asked to organize? (No)

(19) a. The fans wondered how many tweets about himself you are asking Jimmy to post for that evening's late night TV episode.
Was Jimmy being asked to post some tweets about himself for a TV episode? (Yes)

b. The fans wondered how many tweets about himself you are asking Jimmy to retweet for that evening's late night TV episode.
Was Jimmy being asked to retweet some tweets about himself for a TV episode? (Yes)

c. The fans wondered how many tweets about Molly you are asking Jimmy to post for that evening's late night TV episode.
Was Jimmy being asked to post some tweets about himself for a TV episode? (No)

d. The fans wondered how many tweets about Molly you are asking Jimmy to retweet for that evening's late night TV episode.
Was Jimmy being asked to retweet some tweets about himself for a TV episode? (No)

(20) a. The journalists wondered how many stories about herself you are asking Ellen to concoct for the upcoming live TV interview.
Had Ellen been required to concoct stories about herself for an article? (No)

b. The journalists wondered how many stories about herself you are asking Ellen to verify for the upcoming live TV interview.
Had Ellen been required to verify stories about herself for an article? (No)

c. The journalists wondered how many stories about Bobby you are asking Ellen to concoct for the upcoming live TV interview.
Had Ellen been required to concoct stories about Bobby for an interview? (Yes)

d. The journalists wondered how many stories about Bobby you are asking Ellen to verify for the upcoming live TV interview.
Had Ellen been required to verify stories about Bobby for an interview? (Yes)

(21) a. The organizers wondered how many websites about himself you are asking Daniel to develop for next year's art exhibition.
Did Daniel need to create websites about himself for an exhibition? (Yes)

b. The organizers wondered how many websites about himself you are asking Daniel to revise for next year's art exhibition.
Did Daniel need to edit websites about himself for an exhibition? (Yes)

c. The organizers wondered how many websites about Sabrina you are asking Daniel to develop for next year's art exhibition.
Did Daniel need to take down websites about Sabrina for an exhibition? (No)

d. The organizers wondered how many websites about Sabrina you are asking Daniel to revise for next year's art exhibition.

Did Daniel need to take down websites about Sabrina for an exhibition? (No)

(22) a. The exhibitors wondered how many caricatures of himself you are asking Bob to draw for the upcoming artist showcase.
Was Bob being asked to draw pictures of himself for the showcase? (Yes)

b. The exhibitors wondered how many caricatures of himself you are asking Bob to display for the upcoming artist showcase.
Was Bob being asked to display pictures of himself for the showcase? (Yes)

c. The exhibitors wondered how many caricatures of Emily you are asking Bob to draw for the upcoming artist showcase.
Was Bob being asked to draw comics about Emily for the showcase? (No)

d. The exhibitors wondered how many caricatures of Emily you are asking Bob to display for the upcoming artist showcase.
Was Bob being asked to display comics about Emily for the showcase? (No)

(23) a. The columnists wondered how many images of himself you are asking Ryan to create for the latest issue of the scuba diving magazine.
Was it the readers who wondered how many images of himself Ryan was being asked to create? (No)

b. The columnists wondered how many images of himself you are asking Ryan to print for the latest issue of the scuba diving magazine.
Was it the readers who wondered how many images of himself Ryan was being asked to print? (No)

c. The columnists wondered how many images of Diane you are asking Ryan to create for the latest issue of the scuba diving magazine.
Was it the columnists who wondered how many images of Diane Ryan was being asked to create? (Yes)

d. The columnists wondered how many images of Diane you are asking Ryan to print for the latest issue of the scuba diving magazine.
Was it the columnists who wondered how many images of Diane Ryan was being asked to print? (Yes)

(24) a. The students wondered how many statues of himself you are asking Patrick to construct for the end of the year project.
Was Patrick being asked to construct websites about himself for the project? (No)

b. The students wondered how many statues of himself you are asking Patrick to comment on for the end of the year project.
Was Patrick being asked to comment on websites about himself for the project? (No)

c. The students wondered how many statues of Sophia you are asking Patrick to construct for the end of the year project.
Was Patrick being asked to construct sculptures of Sophia for the project? (Yes)

d. The students wondered how many statues of Sophia you are asking Patrick to comment on for the end of the year project.
Was Patrick being asked to comment on sculptures of Sophia for the project? (Yes)

(25) a. The actors wondered how many exposés about herself you are asking Jennifer to produce for the latest movie promotion.
Did Jennifer need to produce exposés about herself for a movie promotion? (Yes)

b. The actors wondered how many exposés about herself you are asking Jennifer to endorse for the latest movie promotion.
Did Jennifer need to endorse exposés about herself for a movie promotion? (Yes)

c. The actors wondered how many exposés about Liam you are asking Jennifer to produce for the latest movie promotion.
Did Jennifer need to produce exposés about Liam for a book release? (No)

d. The actors wondered how many exposés about Liam you are asking Jennifer to endorse for the latest movie promotion.
Did Jennifer need to endorse exposés about Liam for a book release? (No)

(26) a. The readers wondered how many articles about himself you are asking Lance to outline for the special issue of the hiking magazine.
Did Lance need to outline articles about Cindy for the magazine? (No)

b. The readers wondered how many articles about himself you are asking Lance to discredit in the special issue of the hiking magazine.
Did Lance need to discredit articles about Cindy in the magazine? (No)

c. The readers wondered how many articles about Cindy you are asking Lance to outline for the special issue of the hiking magazine.
Did Lance need to outline articles about Cindy for the magazine? (Yes)

d. The readers wondered how many articles about Cindy you are asking Lance to discredit in the special issue of the hiking magazine.
Did Lance need to discredit articles about Cindy in the magazine? (Yes)

(27) a. The administrators wondered how many photographs of herself you are asking Olivia to shoot for the final class project.
Was Olivia being asked to discard some photographs of herself for the project? (No)

b. The administrators wondered how many photographs of herself you are asking Olivia to touch up for the final class project.
Was Olivia being asked to discard some photographs of herself for the project? (No)

c. The administrators wondered how many photographs of Chris you are asking Olivia to shoot for the final class project.
Was Olivia being asked to take some photographs of Chris for the project? (Yes)

d. The administrators wondered how many photographs of Chris you are asking Olivia to touch up for the final class project.
Was Olivia being asked to fix some photographs of Chris for the project? (Yes)

(28) a. The broadcasters wondered how many videos of himself you are asking Robert to shoot for the latest promotional advertisement.
Was it the broadcasters who wondered how many videos of himself Robert was being asked to shoot? (Yes)

b. The broadcasters wondered how many videos of himself you are asking Robert to cut for the latest promotional advertisement.
Was it the broadcasters who wondered how many videos of himself Robert was being asked to cut? (Yes)

c. The broadcasters wondered how many videos of Amanda you are asking Robert to shoot for the latest promotional advertisement.
Was it the critics who wondered how many videos of Amanda Robert was being asked to shoot? (No)

d. The broadcasters wondered how many videos of Amanda you are asking Robert to cut for the latest promotional advertisement.
Was it the critics who wondered how many videos of Amanda Robert was being asked to cut? (No)

(29) a. The publishers wondered how many articles about herself you are asking Audrey to compose for the autobiographical collection.
Was Audrey being asked to compose articles about herself for a collection? (Yes)

b. The publishers wondered how many articles about herself you are asking Audrey to edit for the autobiographical collection.
Was Audrey being asked to edit articles about herself for a collection? (Yes)

c. The publishers wondered how many articles about Nathan you are asking Audrey to compose for the autobiographical collection.
Was Audrey being asked to compose songs about Nathan for a collection? (No)

d. The publishers wondered how many articles about Nathan you are asking Audrey to edit for the autobiographical collection.

Was Audrey being asked to edit songs about Nathan for a collection? (No)

(30) a. The roommates wondered how many figurines of himself you are asking Steven to make for the mid-semester art project.
Was Steven being asked to make figurines of himself for the project? (Yes)

b. The roommates wondered how many figurines of himself you are asking Steven to photograph for the mid-semester art project.
Was Steven being asked to photograph figurines of himself for the project? (Yes)

c. The roommates wondered how many figurines of Kimberley you are asking Steven to make for the mid-semester art project.
Was Steven being asked to make figurines of himself for the project? (No)

d. The roommates wondered how many figurines of Kimberley you are asking Steven to photograph for the mid-semester art project.
Was Steven being asked to photograph figurines of himself for the project? (No)

(31) a. The bloggers wondered how many videoclips of herself you are asking Chelsea to film for the special live television broadcast.
Did Chelsea need to record videoclips of herself for the broadcast? (Yes)

b. The bloggers wondered how many videoclips of herself you are asking Chelsea to watch for the special live television broadcast.
Did Chelsea need to view videoclips of herself for the broadcast? (Yes)

c. The bloggers wondered how many videoclips of Marc you are asking Chelsea to film for the special live television broadcast.
Did Chelsea need to delete videoclips of Marc for the broadcast? (No)

d. The bloggers wondered how many videoclips of Marc you are asking Chelsea to watch for the special live television broadcast.
Did Chelsea need to delete videoclips of Marc for the broadcast? (No)

(32) a. The lawyers wondered how many biographies about himself you are asking Mike to write for the commemorative book.
Was Mike being asked to write posts about himself for the book? (No)

b. The lawyers wondered how many biographies about himself you are asking Mike to verify for the commemorative book.
Was Mike being asked to verify posts about himself for the book? (No)

c. The lawyers wondered how many biographies about Hillary you are asking Mike to write for the commemorative book.
Was Mike being asked to write biographies about Hillary for the book? (Yes)

d. The lawyers wondered how many biographies about Hillary you are asking Mike to verify for the commemorative book.
Was Mike being asked to verify biographies about Hillary for the book? (Yes)

Appendix B

Experiment 2 stimuli

(1) a. The agents wondered how many songs representing her as a songwriter you are asking Heather to compose for the upcoming North American tour album.
Were some songs being composed for a fundraiser? (No)

b. The agents wondered how many songs representing her as a songwriter you are asking Heather to perform for the upcoming North American tour album.
Were some songs being performed for a fundraiser? (No)

c. The agents wondered how many songs representing Heather as a songwriter you are asking her to compose for the upcoming North American tour album.
Were some songs being composed for an album? (Yes)

d. The agents wondered how many songs representing Heather as a songwriter you are asking her to perform for the upcoming North American tour album.
Were some songs being performed for an album? (Yes)

(2) a. The publicists wondered how many posts endorsing him as a model you are asking Roger to create for the upcoming album release.
Were some posts being created for an album release? (Yes)

b. The publicists wondered how many posts endorsing him as a model you are asking Roger to disregard for the upcoming album release.
Were some posts being disregarded for an album release? (Yes)

c. The publicists wondered how many posts endorsing Roger as a model you are asking him to create for the upcoming album release.
Were some posts being created for an awards ceremony? (No)

d. The publicists wondered how many posts endorsing Roger as a model you are asking him to disregard for the upcoming album release.
Were some posts being disregarded for an awards ceremony? (No)

(3) a. The buyers wondered how many drawings representing him as an author you are asking Brad to sketch for the worldwide release of his new book.
Were some drawings being sketched for a magazine? (No)

b. The buyers wondered how many drawings representing him as an author you are asking Brad to dig up for the worldwide release of his new book.
Were some drawings being dug up for a magazine? (No)

c. The buyers wondered how many drawings representing Brad as an author you are asking him to sketch for the worldwide release of his new book.
Were some drawings being sketched for a book? (Yes)

d. The buyers wondered how many drawings representing Brad as an author you are asking him to dig up for the worldwide release of his new book.
Were some drawings being dug up for a book? (Yes)

(4) a. The actors wondered how many exposés revealing her as the star you are asking Jennifer to produce for the latest movie promotion.
Were some exposés being produced for a movie promotion? (Yes)

b. The actors wondered how many exposés revealing her as the star you are asking Jennifer to endorse for the latest movie promotion.
Were some exposés being endorsed for a movie promotion? (Yes)

c. The actors wondered how many exposés revealing Jennifer as the star you are asking her to produce for the latest movie promotion.
Were some exposés being produced for a book release? (No)

d. The actors wondered how many exposés revealing Jennifer as the star you are asking her to endorse for the latest movie promotion.
Were some exposés being endorsed for a book release? (No)

(5) a. The editors wondered how many cartoons depicting her with a diploma you are asking Jessica to sketch for the upcoming edition of the student newspaper.
Was Jessica being asked to sketch cartoons of herself for the newspaper? (Yes or No)

b. The editors wondered how many cartoons depicting her with a diploma you are asking Jessica to approve for the upcoming edition of the student newspaper.
Was Jessica being asked to approve cartoons of herself for the newspaper? (Yes or No)

c. The editors wondered how many cartoons depicting Jessica with a diploma you are asking her to sketch for the upcoming edition of the student newspaper.
Was Jessica being asked to sketch cartoons of herself for the newspaper? (Yes or No)

d. The editors wondered how many cartoons depicting Jessica with a diploma you are asking her to approve for the upcoming edition of the student newspaper.

Was Jessica being asked to approve cartoons of herself for the newspaper? (Yes or No)

(6) a. The organizers wondered how many toy models depicting him with a cape you are asking Charles to build for next weekend's superhero convention.
Was Charles being asked to build toy models of himself for the convention? (Yes or No)

b. The organizers wondered how many toy models depicting him with a cape you are asking Charles to buy at next weekend's superhero convention.
Was Charles being asked to buy toy models of himself at the convention? (Yes or No)

c. The organizers wondered how many toy models depicting Charles with a cape you are asking him to build for next weekend's superhero convention.
Was Charles being asked to build toy models of himself for the convention? (Yes or No)

d. The organizers wondered how many toy models depicting Charles with a cape you are asking him to buy at next weekend's superhero convention.
Was Charles being asked to buy toy models of himself at the convention? (Yes or No)

(7) a. The comedians wondered how many jokes mocking her as a Canadian you are asking Carly to make up for this weekend's sold out stand-up show.
Did Carly need to make up jokes about herself for a stand-up show? (Yes or No)

b. The comedians wondered how many jokes mocking her as a Canadian you are asking Carly to practice for this weekend's sold out stand-up show.
Did Carly need to practice jokes about herself for a stand-up show? (Yes or No)

c. The comedians wondered how many jokes mocking Carly as a Canadian you are asking her to make up for this weekend's sold out stand-up show.
Did Carly need to make up jokes about herself for a stand-up show? (Yes or No)

d. The comedians wondered how many jokes mocking Carly as a Canadian you are asking her to practice for this weekend's sold out stand-up show.
Did Carly need to practice jokes about herself for a stand-up show? (Yes or No)

(8) a. The exhibitors wondered how many caricatures mocking him in a costume you are asking Bob to draw for the upcoming artist showcase.
Was Bob being asked to draw some pictures for the showcase? (Yes or No)

b. The exhibitors wondered how many caricatures mocking him in a costume you are asking Bob to display for the upcoming artist showcase.
Was Bob being asked to display some pictures for the showcase? (Yes or No)

c. The exhibitors wondered how many caricatures mocking Bob in a costume you are asking him to draw for the upcoming artist showcase.
Was Bob being asked to draw some pictures for the showcase? (Yes or No)

d. The exhibitors wondered how many caricatures mocking Bob in a costume you are asking him to display for the upcoming artist showcase.
Was Bob being asked to display some pictures for the showcase? (Yes or No)

(9) a. The directors wondered how many segments identifying her as the admirer you are asking Marisa to draft for the special edition broadcast.
Was Marisa being asked to draft some segments about herself for the broadcast? (Yes or No)

b. The directors wondered how many segments identifying her as the admirer you are asking Marisa to edit for the special edition broadcast.
Was Marisa being asked to edit some segments about herself for the broadcast? (Yes or No)

c. The directors wondered how many segments identifying Marisa as the admirer you are asking her to draft for the special edition broadcast.
Was Marisa being asked to draft some segments about herself for the broadcast? (Yes or No)

d. The directors wondered how many segments identifying Marisa as the admirer you are asking her to edit for the special edition broadcast.
Was Marisa being asked to edit some segments about herself for the broadcast? (Yes or No)

(10) a. The fans wondered how many tweets endorsing him as a host you are asking Jimmy to post for that evening's late night TV episode.
Was Jimmy being asked to post some tweets about himself for the episode? (Yes or No)

b. The fans wondered how many tweets endorsing him as a host you are asking Jimmy to retweet for that evening's late night TV episode.
Was Jimmy being asked to retweet some tweets about himself for the episode? (Yes or No)

c. The fans wondered how many tweets endorsing Jimmy as a host you are asking him to post for that evening's late night TV episode.
Was Jimmy being asked to post some tweets about himself for the episode? (Yes or No)

d. The fans wondered how many tweets endorsing Jimmy as a host you are asking him to retweet for that evening's late night TV episode.
Was Jimmy being asked to retweet some tweets about himself for the episode? (Yes or No)

(11) a. The readers wondered how many articles promoting him as an adventurer you are asking Lance to outline for the special issue of the hiking magazine.
Did Lance need to outline articles about himself for the magazine? (Yes or No)

b. The readers wondered how many articles promoting him as an adventurer you are asking Lance to discredit in the special issue of the hiking magazine.
Did Lance need to discredit articles about himself in the magazine? (Yes or No)

c. The readers wondered how many articles promoting Lance as an adventurer you are asking him to outline for the special issue of the hiking magazine.
Did Lance need to outline articles about himself for the magazine? (Yes or No)

d. The readers wondered how many articles promoting Lance as an adventurer you are asking him to discredit in the special issue of the hiking magazine.
Did Lance need to discredit articles about himself in the magazine? (Yes or No)

(12) a. The administrators wondered how many photographs representing her as an artist you are asking Olivia to shoot for the final class project.
Did Olivia need to shoot some photographs of herself for the project? (Yes or No)

b. The administrators wondered how many photographs representing her as an artist you are asking Olivia to touch up for the final class project.
Did Olivia need to touch up some photographs of herself for the project? (Yes or No)

c. The administrators wondered how many photographs representing Olivia as an artist you are asking her to shoot for the final class project.
Did Olivia need to shoot some photographs of herself for the project? (Yes or No)

d. The administrators wondered how many photographs representing Olivia as an artist you are asking her to touch up for the final class project.
Did Olivia need to touch up some photographs of herself for the project? (Yes or No)

(13) a. The broadcasters wondered how many videos exposing him as the thief you are asking Robert to shoot for the latest promotional advertisement.
Did Robert need to shoot some videos of himself for an advertisement? (Yes or No)

b. The broadcasters wondered how many videos exposing him as the thief you are asking Robert to cut for the latest promotional advertisement.
Did Robert need to cut some videos of himself for an advertisement? (Yes or No)

c. The broadcasters wondered how many videos exposing Robert as the thief you are asking him to shoot for the latest promotional advertisement.
Did Robert need to shoot some videos of himself for an advertisement? (Yes or No)

d. The broadcasters wondered how many videos exposing Robert as the thief you are asking him to cut for the latest promotional advertisement.
Did Robert need to cut some videos of himself for an advertisement? (Yes or No)

(14) a. The roommates wondered how many figurines portraying him with a hat you are asking Steven to make for the mid-semester art project.
Was Steven being asked to make figurines of himself for the project? (Yes or No)

b. The roommates wondered how many figurines portraying him with a hat you are asking Steven

to photograph for the mid-semester art project.
Was Steven being asked to photograph figurines of himself for the project? (Yes or No)

c. The roommates wondered how many figurines portraying Steven with a hat you are asking him to make for the mid-semester art project.
Was Steven being asked to make figurines of himself for the project? (Yes or No)

d. The roommates wondered how many figurines portraying Steven with a hat you are asking him to photograph for the mid-semester art project.
Was Steven being asked to photograph figurines of himself for the project? (Yes or No)

(15) a. The bloggers wondered how many videoclips promoting her as an actress you are asking Chelsea to film for the special live television broadcast.
Did Chelsea need to film some videoclips of herself for the broadcast? (Yes or No)

b. The bloggers wondered how many videoclips promoting her as an actress you are asking Chelsea to watch for the special live television broadcast.
Did Chelsea need to watch some videoclips of herself for the broadcast? (Yes or No)

c. The bloggers wondered how many videoclips promoting Chelsea as an actress you are asking her to film for the special live television broadcast.
Did Chelsea need to film some videoclips of herself for the broadcast? (Yes or No)

d. The bloggers wondered how many videoclips promoting Chelsea as an actress you are asking her to watch for the special live television broadcast.
Did Chelsea need to watch some videoclips of herself for the broadcast? (Yes or No)

(16) a. The lawyers wondered how many biographies recognizing him as a humanitarian you are asking Mike to write for the commemorative book.
Did Mike need to write biographies about himself for a book? (Yes or No)

b. The lawyers wondered how many biographies recognizing him as a humanitarian you are asking Mike to verify for the commemorative book.
Did Mike need to verify biographies about himself for a book? (Yes or No)

c. The lawyers wondered how many biographies recognizing Mike as a humanitarian you are asking him to write for the commemorative book.
Did Mike need to write biographies about himself for a book? (Yes or No)

d. The lawyers wondered how many biographies recognizing Mike as a humanitarian you are asking him to verify for the commemorative book.
Did Mike need to verify biographies about himself for a book? (Yes or No)

(17) a. The producers wondered how many stories describing her as a spokesperson you are asking Larissa to draft for the behind-the-scenes special broadcast interview.

Were some letters being drafted for the interview? (No)

b. The producers wondered how many stories describing her as a spokesperson you are asking Larissa to proofread for the behind-the-scenes special broadcast interview.
Were some letters being proofread for the interview? (No)

c. The producers wondered how many stories describing Larissa as a spokesperson you are asking her to draft for the behind-the-scenes special broadcast interview.
Were some stories being drafted for the interview? (Yes)

d. The producers wondered how many stories describing Larissa as a spokesperson you are asking her to proofread for the behind-the-scenes special broadcast interview.
Were some stories being proofread for the interview? (Yes)

(18) a. The museum wondered how many portraits illustrating him with a mustache you are asking Jasper to paint for next week's grand opening exhibit.
Were some pictures being painted for the exhibit? (Yes)

b. The museum wondered how many portraits illustrating him with a mustache you are asking Jasper to judge at next week's grand opening exhibit.
Were some pictures being judged for the exhibit? (Yes)

c. The museum wondered how many portraits illustrating Jasper with a mustache you are asking him to paint for next week's grand opening exhibit.
Were some figurines being painted for the exhibit? (No)

d. The museum wondered how many portraits illustrating Jasper with a mustache you are asking him to judge at next week's grand opening exhibit.
Were some figurines being judged for the exhibit? (No)

(19) a. The students wondered how many statues honoring him as a hero you are asking Patrick to construct for the end of the year project.
Were some websites being constructed for the project? (No)

b. The students wondered how many statues honoring him as a hero you are asking Patrick to comment on for the end of the year project.
Were some websites being commented on for the project? (No)

c. The students wondered how many statues honoring Patrick as a hero you are asking him to construct for the end of the year project.
Were some sculptures being constructed for the project? (Yes)

d. The students wondered how many statues honoring Patrick as a hero you are asking him to comment on for the end of the year project.
Were some sculptures being commented on for the project? (Yes)

(20) a. The viewers wondered how many rumors exposing her as the reporter you are asking Laura to make up for the upcoming reality TV show.
Was some gossip being made up for the TV show? (Yes)

b. The viewers wondered how many rumors exposing her as the reporter you are asking Laura to discredit for the upcoming reality TV show.
Was some gossip being discredited for the TV show? (Yes)

c. The viewers wondered how many rumors exposing Laura as the reporter you are asking her to make up for the upcoming reality TV show.
Were some jokes being made up for the TV show? (No)

d. The viewers wondered how many rumors exposing Laura as the reporter you are asking her to discredit for the upcoming reality TV show.
Were some jokes being discredited for the TV show? (No)

(21) a. The reporters wondered how many lies discrediting her as a witness you are asking Alexa to invent for tomorrow's live television interview.
Did the lies already exist? (Yes or No)

b. The reporters wondered how many lies discrediting her as a witness you are asking Alexa to deny for tomorrow's live television interview.
Did the lies already exist? (Yes or No)

c. The reporters wondered how many lies discrediting Alexa as a witness you are asking her to invent for tomorrow's live television interview.
Did the lies already exist? (Yes or No)

d. The reporters wondered how many lies discrediting Alexa as a witness you are asking her to deny for tomorrow's live television interview.
Did the lies already exist? (Yes or No)

(22) a. The network wondered how many documentaries praising her as an idol you are asking Betty to commission for the celebration of her ninetieth birthday.
Did the documentaries already exist? (Yes or No)

b. The network wondered how many documentaries praising her as an idol you are asking Betty to watch at the celebration of her ninetieth birthday.
Did the documentaries already exist? (Yes or No)

c. The network wondered how many documentaries praising Betty as an idol you are asking her to commission for the celebration of her ninetieth birthday.
Did the documentaries already exist? (Yes or No)

d. The network wondered how many documentaries praising Betty as an idol you are asking her to watch at the celebration of her ninetieth birthday.

Did the documentaries already exist? (Yes or No)

(23) a. The blogger-community wondered how many rumors revealing him as the spy you are asking Derek to concoct for the social media campaign.
Did the rumors already exist? (Yes or No)

b. The blogger-community wondered how many rumors revealing him as the spy you are asking Derek to deny in the social media campaign.
Did the rumors already exist? (Yes or No)

c. The blogger-community wondered how many rumors revealing Derek as the spy you are asking him to concoct for the social media campaign.
Did the rumors already exist? (Yes or No)

d. The blogger-community wondered how many rumors revealing Derek as the spy you are asking him to deny in the social media campaign.
Did the rumors already exist? (Yes or No)

(24) a. The instructors wondered how many pictures showing her with a paintbrush you are asking Heidi to take for the art class scrapbook.
Did the pictures already exist? (Yes or No)

b. The instructors wondered how many pictures showing her with a paintbrush you are asking Heidi to organize for the art class scrapbook.
Did the pictures already exist? (Yes or No)

c. The instructors wondered how many pictures showing Heidi with a paintbrush you are asking her to take for the art class scrapbook.
Did the pictures already exist? (Yes or No)

d. The instructors wondered how many pictures showing Heidi with a paintbrush you are asking her to organize for the art class scrapbook.
Did the pictures already exist? (Yes or No)

(25) a. The columnists wondered how many images depicting him with a shark you are asking Ryan to create for the latest issue of the scuba diving magazine.
Did the images already exist? (Yes or No)

b. The columnists wondered how many images depicting him with a shark you are asking Ryan to print for the latest issue of the scuba diving magazine.
Did the images already exist? (Yes or No)

c. The columnists wondered how many images depicting Ryan with a shark you are asking him to create for the latest issue of the scuba diving magazine.
Did the images already exist? (Yes or No)

d. The columnists wondered how many images depicting Ryan with a shark you are asking him to print for the latest issue of the scuba diving magazine.
Did the images already exist? (Yes or No)

(26) a. The organizers wondered how many websites distinguishing him as a freelancer you are asking Daniel to develop for next year's art exhibition.
Did the websites already exist? (Yes or No)

b. The organizers wondered how many websites distinguishing him as a freelancer you are asking Daniel to revise for next year's art exhibition.
Did the websites already exist? (Yes or No)

c. The organizers wondered how many websites distinguishing Daniel as a freelancer you are asking him to develop for next year's art exhibition.
Did the websites already exist? (Yes or No)

d. The organizers wondered how many websites distinguishing Daniel as a freelancer you are asking him to revise for next year's art exhibition.
Did the websites already exist? (Yes or No)

(27) a. The journalists wondered how many stories exposing her as the snitch you are asking Ellen to concoct for the upcoming live TV interview.
Did the stories already exist? (Yes or No)

b. The journalists wondered how many stories exposing her as the snitch you are asking Ellen to verify for the upcoming live TV interview.
Did the stories already exist? (Yes or No)

c. The journalists wondered how many stories exposing Ellen as the snitch you are asking her to concoct for the upcoming live TV interview.
Did the stories already exist? (Yes or No)

d. The journalists wondered how many stories exposing Ellen as the snitch you are asking her to verify for the upcoming live TV interview.
Did the stories already exist? (Yes or No)

(28) a. The publishers wondered how many articles honoring her as a politician you are asking Audrey to compose for the autobiographical collection.
Did the articles already exist? (Yes or No)

b. The publishers wondered how many articles honoring her as a politician you are asking Audrey to edit for the autobiographical collection.
Did the articles already exist? (Yes or No)

c. The publishers wondered how many articles honoring Audrey as a politician you are asking her to compose for the autobiographical collection.
Did the articles already exist? (Yes or No)

d. The publishers wondered how many articles honoring Audrey as a politician you are asking her to edit for the autobiographical collection.
Did the articles already exist? (Yes or No)

(29) a. The teachers wondered how many sketches portraying him with an easel you are asking Stuart to draw for the art school application portfolio.
Were some sketches being made for the portfolio? (Yes)

b. The teachers wondered how many sketches portraying him with an easel you are asking Stuart to adjust for the art school application portfolio.
Were some sketches being fixed for the portfolio? (Yes)

c. The teachers wondered how many sketches portraying Stuart with an easel you are asking him to draw for the art school application portfolio.
Were some sketches being destroyed for the portfolio? (No)

d. The teachers wondered how many sketches portraying Stuart with an easel you are asking him to adjust for the art school application portfolio.
Were some sketches being destroyed for the portfolio? (No)

(30) a. The curators wondered how many sculptures portraying her as a princess you are asking Megan to carve for the museum's end of season art show.
Were some sculptures being demolished for an art show? (No)

b. The curators wondered how many sculptures portraying her as a princess you are asking Megan to critique for the museum's end of season art show.
Were some sculptures being demolished for an art show? (No)

c. The curators wondered how many sculptures portraying Megan as a princess you are asking her to carve for the museum's end of season art show.
Were some sculptures being made for an art show? (Yes)

d. The curators wondered how many sculptures portraying Megan as a princess you are asking her to critique for the museum's end of season art show.
Were some sculptures being judged for an art show? (Yes)

(31) a. The hosts wondered how many photographs showcasing her as a volunteer you are asking Christina to take for the online promotional advertisement.
Were some photographs being shot for an advertisement? (Yes)

b. The hosts wondered how many photographs showcasing her as a volunteer you are asking

Christina to copy for the online promotional advertisement.
Were some photographs being duplicated for an advertisement? (Yes)

c. The hosts wondered how many photographs showcasing Christina as a volunteer you are asking her to take for the online promotional advertisement.
Were some photographs being destroyed for an advertisement? (No)

d. The hosts wondered how many photographs showcasing Christina as a volunteer you are asking her to copy for the online promotional advertisement.
Were some photographs being destroyed for an advertisement? (No)

(32) a. The parents wondered how many columns describing him as the lead you are asking Joey to write for the drama club yearbook.
Were some columns being deleted from the yearbook? (No)

b. The parents wondered how many columns describing him as the lead you are asking Joey to shorten for the drama club yearbook.
Were some columns being deleted from the yearbook? (No)

c. The parents wondered how many columns describing Joey as the lead you are asking him to write for the drama club yearbook.
Were some columns being composed for the yearbook? (Yes)

d. The parents wondered how many columns describing Joey as the lead you are asking him to shorten for the drama club yearbook.
Were some columns being condensed for the yearbook? (Yes)

Appendix C

Experiment 3 stimuli

(1) a. The agents wondered how many songs representing her as a songwriter you had asked Taylor Swift to perform for the North American tour.
Were some songs performed for a tour? (Yes)

b. The agents wondered how many songs representing her as a songwriter you had asked Anna Lang to perform for the North American tour.
Were some songs performed for a tour? (Yes)

c. The agents wondered how many songs representing Taylor Swift as a songwriter you had asked her to perform for the North American tour.
Were some songs performed for a tour? (Yes)

d. The agents wondered how many songs representing Anna Lang as a songwriter you had asked her to perform for the North American tour.
Were some songs performed for a tour? (Yes)

(2) a. The publicists wondered how many posts endorsing him as a model you had asked Johnny Depp to promote for the latest movie release.
Were some posts promoted for an album? (No)

b. The publicists wondered how many posts endorsing him as a model you had asked Jason Lane to promote for the latest movie release.
Were some posts promoted for an album? (No)

c. The publicists wondered how many posts endorsing Johnny Depp as a model you had asked him to promote for the latest movie release.
Were some posts promoted for an album? (No)

d. The publicists wondered how many posts endorsing Jason Lane as a model you had asked him to promote for the latest movie release.
Were some posts promoted for an album? (No)

(3) a. The readers wondered how many blog posts representing him as an author you had asked George RR Martin to share for the new book promotion.
Were some blog posts shared for a movie? (No)

b. The readers wondered how many blog posts representing him as an author you had asked Anthony Baldwin to share for the new book promotion.
Were some blog posts shared for a movie? (No)

c. The readers wondered how many blog posts representing George RR Martin as an author you had asked him to share for the new book promotion.
Were some blog posts shared for a movie? (No)

d. The readers wondered how many blog posts representing Anthony Baldwin as an author you had asked him to share for the new book promotion.
Were some blog posts shared for a movie? (No)

(4) a. The journalists wondered how many exposés revealing her as the informant you had asked Barbara Walters to tape for the ABC special broadcast.
Were some exposés taped for a broadcast? (Yes)

b. The journalists wondered how many exposés revealing her as the informant you had asked Stephanie McDonald to tape for the ABC special broadcast.
Were some exposés taped for a broadcast? (Yes)

c. The journalists wondered how many exposés revealing Barbara Walters as the informant you had asked her to tape for the ABC special broadcast.
Were some exposés taped for a broadcast? (Yes)

d. The journalists wondered how many exposés revealing Stephanie McDonald as the informant you had asked her to tape for the ABC special broadcast.
Were some exposés taped for a broadcast? (Yes)

(5) a. The editors wondered how many cartoons depicting her with a diploma you had asked Hillary Clinton to approve for the latest newspaper article.
Was Hillary asked to approve cartoons of herself for the newspaper? (Yes or No)

b. The editors wondered how many cartoons depicting her with a diploma you had asked Jessica Campbell to approve for the latest newspaper article.
Was Jessica asked to approve cartoons of herself for the newspaper? (Yes or No)

c. The editors wondered how many cartoons depicting Hillary Clinton with a diploma you had asked her to approve for the latest newspaper article.
Was Hillary asked to approve cartoons of herself for the newspaper? (Yes or No)

d. The editors wondered how many cartoons depicting Jessica Campbell with a diploma you had asked her to approve for the latest newspaper article.

Was Jessica asked to approve cartoons of herself for the newspaper? (Yes or No)

(6) a. The organizers wondered how many toy models depicting him with a cape you had asked Chris Evans to showcase at the weekend superhero convention.
Was Chris asked to showcase toy models of himself for the convention? (Yes or No)

b. The organizers wondered how many toy models depicting him with a cape you had asked Kyle Coleman to showcase at the weekend superhero convention.
Was Kyle asked to showcase toy models of himself for the convention? (Yes or No)

c. The organizers wondered how many toy models depicting Chris Evans with a cape you had asked him to showcase at the weekend superhero convention.
Was Chris asked to showcase toy models of himself for the convention? (Yes or No)

d. The organizers wondered how many toy models depicting Kyle Coleman with a cape you had asked him to showcase at the weekend superhero convention.
Was Kyle asked to showcase toy models of himself for the convention? (Yes or No)

(7) a. The comedians wondered how many jokes mocking him as a Canadian you had asked Jim Carrey to practice for the late night stand-up show.
Did Jim need to practice jokes about himself for a stand-up show? (Yes or No)

b. The comedians wondered how many jokes mocking him as a Canadian you had asked Rick Johnson to practice for the late night stand-up show.
Did Rick need to practice jokes about himself for a stand-up show? (Yes or No)

c. The comedians wondered how many jokes mocking Jim Carrey as a Canadian you had asked him to practice for the late night stand-up show.
Did Jim need to practice jokes about himself for a stand-up show? (Yes or No)

d. The comedians wondered how many jokes mocking Rick Johnson as a Canadian you had asked him to practice for the late night stand-up show.
Did Rick need to practice jokes about himself for a stand-up show? (Yes or No)

(8) a. The fans wondered how many columns describing her as the lead you had asked Meryl Streep to repost for the social media campaign.
Did Meryl need to repost some columns about herself for the campaign? (Yes or No)

b. The fans wondered how many columns describing her as the lead you had asked Kayla Smith to repost for the social media campaign.
Did Kayla need to repost some columns about herself for the campaign? (Yes or No)

c. The fans wondered how many columns describing Meryl Streep as the lead you had asked her to repost for the social media campaign.
Did Meryl need to repost some columns about herself for the campaign? (Yes or No)

d. The fans wondered how many columns describing Kayla Smith as the lead you had asked her to repost for the social media campaign.
Did Kayla need to repost some columns about herself for the campaign? (Yes or No)

(9) a. The exhibitors wondered how many caricatures mocking him in a costume you had asked Matt Groening to display for the local artist showcase.
Was Matt asked to display some caricatures of himself for the showcase? (Yes or No)

b. The exhibitors wondered how many caricatures mocking him in a costume you had asked Bob Swinton to display for the local artist showcase.
Was Bob asked to display some caricatures of himself for the showcase? (Yes or No)

c. The exhibitors wondered how many caricatures mocking Matt Groening in a costume you had asked him to display for the local artist showcase.
Was Matt asked to display some caricatures of himself for the showcase? (Yes or No)

d. The exhibitors wondered how many caricatures mocking Bob Swinton in a costume you had asked him to display for the local artist showcase.
Was Bob asked to display some caricatures of himself for the showcase? (Yes or No)

(10) a. The directors wondered how many segments identifying her as the admirer you had asked Julia Roberts to support for the special edition broadcast.
Was Julia asked to support some segments about herself for the broadcast? (Yes or No)

b. The directors wondered how many segments identifying her as the admirer you had asked Marisa Walker to support for the special edition broadcast.
Was Marisa asked to support some segments about herself for the broadcast? (Yes or No)

c. The directors wondered how many segments identifying Julia Roberts as the admirer you had asked her to support for the special edition broadcast.
Was Julia asked to support some segments about herself for the broadcast? (Yes or No)

d. The directors wondered how many segments identifying Marisa Walker as the admirer you had asked her to support for the special edition broadcast.
Was Marisa asked to support some segments about herself for the broadcast? (Yes or No)

(11) a. The subscribers wondered how many photographs representing her as an artist you had asked Katy Perry to choose for the national newspaper article.
Did Katy need to choose some photographs of herself for the article? (Yes or No)

b. The subscribers wondered how many photographs representing her as an artist you had asked Karen Taylor to choose for the national newspaper article.
Did Karen need to choose some photographs of herself for the article? (Yes or No)

c. The subscribers wondered how many photographs representing Katy Perry as an artist you had asked her to choose for the national newspaper article.
Did Katy need to choose some photographs of herself for the article? (Yes or No)

d. The subscribers wondered how many photographs representing Karen Taylor as an artist you had asked her to choose for the national newspaper article.
Did Karen need to choose some photographs of herself for the article? (Yes or No)

(12) a. The fans wondered how many tweets endorsing him as a host you had asked Jimmy Kimmel to retweet for the late night TV show.
Did Jimmy need to retweet some tweets about himself for the show? (Yes or No)

b. The fans wondered how many tweets endorsing him as a host you had asked Harry Murphy to retweet for the late night TV show.
Did Harry need to retweet some tweets about himself for the show? (Yes or No)

c. The fans wondered how many tweets endorsing Jimmy Kimmel as a host you had asked him to retweet for the late night TV show.
Did Jimmy need to retweet some tweets about himself for the show? (Yes or No)

d. The fans wondered how many tweets endorsing Harry Murphy as a host you had asked him to retweet for the late night TV show.
Did Harry need to retweet some tweets about himself for the show? (Yes or No)

(13) a. The readers wondered how many articles promoting him as an athlete you had asked Shaquille O'Neal to publicize for the special basketball issue.
Did Shaquille need to publicize articles about himself for the issue? (Yes or No)

b. The readers wondered how many articles promoting him as an athlete you had asked Thomas Murray to publicize for the special basketball issue.
Did Thomas need to publicize articles about himself for the issue? (Yes or No)

c. The readers wondered how many articles promoting Shaquille O'Neal as an athlete you had asked him to publicize for the special basketball issue.
Did Shaquille need to publicize articles about himself for the issue? (Yes or No)

d. The readers wondered how many articles promoting Thomas Murray as an athlete you had asked him to publicize for the special basketball issue.
Did Thomas need to publicize articles about himself for the issue? (Yes or No)

(14) a. The bloggers wondered how many videoclips promoting her as an actress you had asked Angelina Jolie to watch for the special live broadcast.
Did Angelina need to watch some videoclips of herself for the broadcast? (Yes or No)

b. The bloggers wondered how many videoclips promoting her as an actress you had asked

Samantha Gibson to watch for the special live broadcast.
Did Samantha need to watch some videoclips of herself for the broadcast? (Yes or No)

c. The bloggers wondered how many videoclips promoting Angelina Jolie as an actress you had asked her to watch for the special live broadcast.
Did Angelina need to watch some videoclips of herself for the broadcast? (Yes or No)

d. The bloggers wondered how many videoclips promoting Samantha Gibson as an actress you had asked her to watch for the special live broadcast.
Did Samantha need to watch some videoclips of herself for the broadcast? (Yes or No)

(15) a. The viewers wondered how many rumors exposing her as the reporter you had asked Katie Couric to discredit for the new reality show.
Was Katie asked to discredit rumors about herself for the show? (Yes or No)

b. The viewers wondered how many rumors exposing her as the reporter you had asked Laura Stevens to discredit for the new reality show.
Was Laura asked to discredit rumors about herself for the show? (Yes or No)

c. The viewers wondered how many rumors exposing Katie Couric as the reporter you had asked her to discredit for the new reality show.
Was Katie asked to discredit rumors about herself for the show? (Yes or No)

d. The viewers wondered how many rumors exposing Laura Stevens as the reporter you had asked her to discredit for the new reality show.
Was Laura asked to discredit rumors about herself for the show? (Yes or No)

(16) a. The broadcasters wondered how many videos exposing him as the source you had asked Anderson Cooper to cut for the latest news report.
Was Anderson asked to cut videos about himself for the report? (Yes or No)

b. The broadcasters wondered how many videos exposing him as the source you had asked Oliver Snyder to cut for the latest news report.
Was Oliver asked to cut videos about himself for the report? (Yes or No)

c. The broadcasters wondered how many videos exposing Anderson Cooper as the source you had asked him to cut for the latest news report.
Was Anderson asked to cut videos about himself for the report? (Yes or No)

d. The broadcasters wondered how many videos exposing Oliver Snyder as the source you had asked him to cut for the latest news report.
Was Oliver asked to cut videos about himself for the report? (Yes or No)

(17) a. The producers wondered how many stories describing her as a spokesperson you had asked Jennifer Hudson to proofread for the special broadcast interview.

Did Jennifer need to proofread stories about herself for the interview? (Yes or No)

b. The producers wondered how many stories describing her as a spokesperson you had asked Larissa Freeman to proofread for the special broadcast interview.
Did Larissa need to proofread stories about herself for the interview? (Yes or No)

c. The producers wondered how many stories describing Jennifer Hudson as a spokesperson you had asked her to proofread for the special broadcast interview.
Did Jennifer need to proofread stories about herself for the interview? (Yes or No)

d. The producers wondered how many stories describing Larissa Freeman as a spokesperson you had asked her to proofread for the special broadcast interview.
Did Larissa need to proofread stories about herself for the interview? (Yes or No)

(18) a. The fans wondered how many pictures depicting him with a hat you had asked Brad Pitt to provide for the special magazine issue.
Did Brad need to provide pictures of himself for the issue? (Yes or No)

b. The fans wondered how many pictures depicting him with a hat you had asked Craig Lowe to provide for the special magazine issue.
Did Craig need to provide pictures of himself for the issue? (Yes or No)

c. The fans wondered how many pictures depicting Brad Pitt with a hat you had asked him to provide for the special magazine issue.
Did Brad need to provide pictures of himself for the issue? (Yes or No)

d. The fans wondered how many pictures depicting Craig Lowe with a hat you had asked him to provide for the special magazine issue.
Did Craig need to provide pictures of himself for the issue? (Yes or No)

(19) a. The lawyers wondered how many biographies recognizing him as a humanitarian you had asked Sean Penn to verify for the organization's commemorative book.
Was Sean asked to verify biographies about himself for the book? (Yes or No)

b. The lawyers wondered how many biographies recognizing him as a humanitarian you had asked Mike Hill to verify for the organization's commemorative book.
Was Mike asked to verify biographies about himself for the book? (Yes or No)

c. The lawyers wondered how many biographies recognizing Sean Penn as a humanitarian you had asked him to verify for the organization's commemorative book.
Was Sean asked to verify biographies about himself for the book? (Yes or No)

d. The lawyers wondered how many biographies recognizing Mike Hill as a humanitarian you had asked him to verify for the organization's commemorative book.
Was Mike asked to verify biographies about himself for the book? (Yes or No)

(20) a. The network wondered how many documentaries praising her as an idol you had asked Oprah Winfrey to endorse for the new channel launch.
Was Oprah asked to endorse documentaries about herself for the channel launch? (Yes or No)

b. The network wondered how many documentaries praising her as an idol you had asked Betty Wilson to endorse for the new channel launch.
Was Betty asked to endorse documentaries about herself for the channel launch? (Yes or No)

c. The network wondered how many documentaries praising Oprah Winfrey as an idol you had asked her to endorse for the new channel launch.
Was Oprah asked to endorse documentaries about herself for the channel launch? (Yes or No)

d. The network wondered how many documentaries praising Betty Wilson as an idol you had asked her to endorse for the new channel launch.
Was Betty asked to endorse documentaries about herself for the channel launch? (Yes or No)

(21) a. The audience wondered how many songs featuring him as a rapper you had asked Kanye West to record for the latest album release.
Were some songs recorded for an album? (Yes)

b. The audience wondered how many songs featuring him as a rapper you had asked Terence Hines to record for the latest album release.
Were some songs recorded for an album? (Yes)

c. The audience wondered how many songs featuring Kanye West as a rapper you had asked him to record for the latest album release.
Were some songs recorded for an album? (Yes)

d. The audience wondered how many songs featuring Terence Hines as a rapper you had asked him to record for the latest album release.
Were some songs recorded for an album? (Yes)

(22) a. The volunteers wondered how many photographs showcasing her as a philanthropist you had asked Melinda Gates to support for the online recruitment ad.
Were some campaigns supported for an ad? (No)

b. The volunteers wondered how many photographs showcasing her as a philanthropist you had asked Julianne Stone to support for the online recruitment ad.
Were some campaigns supported for an ad? (No)

c. The volunteers wondered how many photographs showcasing Melinda Gates as a philanthropist you had asked her to support for the online recruitment ad.
Were some campaigns supported for an ad? (No)

d. The volunteers wondered how many photographs showcasing Julianne Stone as a philanthropist you had asked her to support for the online recruitment ad.

Were some campaigns supported for an ad? (No)

(23) a. The journalists wondered how many stories exposing her as the investor you had asked Ellen DeGeneres to verify for the recent TV interview.
Were some rumors verified for an interview? (No)

b. The journalists wondered how many stories exposing her as the investor you had asked Carrie Alexander to verify for the recent TV interview.
Were some rumors verified for an interview? (No)

c. The journalists wondered how many stories exposing Ellen DeGeneres as the investor you had asked her to verify for the recent TV interview.
Were some rumors verified for an interview? (No)

d. The journalists wondered how many stories exposing Carrie Alexander as the investor you had asked her to verify for the recent TV interview.
Were some rumors verified for an interview? (No)

(24) a. The voters wondered how many videos promoting him as a candidate you had asked Donald Trump to endorse for the Super Tuesday primaries.
Were some videos endorsed for the primaries? (Yes)

b. The voters wondered how many videos promoting him as a candidate you had asked Patrick Burke to endorse for the Super Tuesday primaries.
Were some videos endorsed for the primaries? (Yes)

c. The voters wondered how many videos promoting Donald Trump as a candidate you had asked him to endorse for the Super Tuesday primaries.
Were some videos endorsed for the primaries? (Yes)

d. The voters wondered how many videos promoting Patrick Burke as a candidate you had asked him to endorse for the Super Tuesday primaries.
Were some videos endorsed for the primaries? (Yes)

(25) a. The publishers wondered how many articles honoring her as a politician you had asked Angela Merkel to edit for the recent autobiographical collection.
Was it the voters who wondered about the articles? (No)

b. The publishers wondered how many articles honoring her as a politician you had asked Katrina Webster to edit for the recent autobiographical collection.
Was it the voters who wondered about the articles? (No)

c. The publishers wondered how many articles honoring Angela Merkel as a politician you had asked her to edit for the recent autobiographical collection.
Was it the voters who wondered about the articles? (No)

d. The publishers wondered how many articles honoring Katrina Webster as a politician you had asked her to edit for the recent autobiographical collection.
Was it the voters who wondered about the articles? (No)

(26) a. The reporters wondered how many lies discrediting him as a witness you had asked Charlie Sheen to deny for the live television interview.
Was it the reporters who wondered about the lies? (Yes)

b. The reporters wondered how many lies discrediting him as a witness you had asked Kevin Holmes to deny for the live television interview.
Was it the reporters who wondered about the lies? (Yes)

c. The reporters wondered how many lies discrediting Charlie Sheen as a witness you had asked him to deny for the live television interview.
Was it the reporters who wondered about the lies? (Yes)

d. The reporters wondered how many lies discrediting Kevin Holmes as a witness you had asked him to deny for the live television interview.
Was it the reporters who wondered about the lies? (Yes)

(27) a. The world leaders wondered how many rumors revealing him as the informant you had asked Vladimir Putin to deny for the community outreach initiative.
Was it the world leaders who wondered about the rumors? (Yes)

b. The world leaders wondered how many rumors revealing him as the informant you had asked Sebastian Weaver to deny for the community outreach initiative.
Was it the world leaders who wondered about the rumors? (Yes)

c. The world leaders wondered how many rumors revealing Vladimir Putin as the informant you had asked him to deny for the community outreach initiative.
Was it the world leaders who wondered about the rumors? (Yes)

d. The world leaders wondered how many rumors revealing Sebastian Weaver as the informant you had asked him to deny for the community outreach initiative.
Was it the world leaders who wondered about the rumors? (Yes)

(28) a. The spectators wondered how many jokes representing her as a comedian you had asked Amy Poehler to tell at the sold out show.
Was it the workers who wondered about the jokes? (No)

b. The spectators wondered how many jokes representing her as a comedian you had asked Megan Simmons to tell at the sold out show.
Was it the workers who wondered about the jokes? (No)

c. The spectators wondered how many jokes representing Amy Poehler as a comedian you had asked her to tell at the sold out show.
Was it the workers who wondered about the jokes? (No)

d. The spectators wondered how many jokes representing Megan Simmons as a comedian you had asked her to tell at the sold out show.
Was it the workers who wondered about the jokes? (No)

(29) a. The columnists wondered how many images depicting him as an Olympian you had asked Michael Phelps to choose for the sports magazine issue.
Were some images deleted for the issue? (No)

b. The columnists wondered how many images depicting him as an Olympian you had asked Ryan Hart to choose for the sports magazine issue.
Were some images deleted for the issue? (No)

c. The columnists wondered how many images depicting Michael Phelps as an Olympian you had asked him to choose for the sports magazine issue.
Were some images deleted for the issue? (No)

d. The columnists wondered how many images depicting Ryan Hart as an Olympian you had asked him to choose for the sports magazine issue.
Were some images deleted for the issue? (No)

(30) a. The players wondered how many videoclips showcasing her as a talent you had asked Serena Williams to share on the social media page.
Were some videoclips shared on social media? (Yes)

b. The players wondered how many videoclips showcasing her as a talent you had asked Allison Matthews to share on the social media page.
Were some videoclips shared on social media? (Yes)

c. The players wondered how many videoclips showcasing Serena Williams as a talent you had asked her to share on the social media page.
Were some videoclips shared on social media? (Yes)

d. The players wondered how many videoclips showcasing Allison Matthews as a talent you had asked her to share on the social media page.
Were some videoclips shared on social media? (Yes)

(31) a. The artists wondered how many pictures showing her with a paintbrush you had asked Yoko Ono to select for the renovated art studio.
Were some pictures selected for the studio? (Yes)

b. The artists wondered how many pictures showing her with a paintbrush you had asked Heidi

Jenkins to select for the renovated art studio.
Were some pictures selected for the studio? (Yes)

c. The artists wondered how many pictures showing Yoko Ono with a paintbrush you had asked her to select for the renovated art studio.
Were some pictures selected for the studio? (Yes)

d. The artists wondered how many pictures showing Heidi Jenkins with a paintbrush you had asked her to select for the renovated art studio.
Were some pictures selected for the studio? (Yes)

(32) a. The organizers wondered how many websites distinguishing him as a writer you had asked Stephen King to approve for the latest book release.
Were some websites removed for the book? (No)

b. The organizers wondered how many websites distinguishing him as a writer you had asked Daniel Moore to approve for the latest book release.
Were some websites removed for the book? (No)

c. The organizers wondered how many websites distinguishing Stephen King as a writer you had asked him to approve for the latest book release.
Were some websites removed for the book? (No)

d. The organizers wondered how many websites distinguishing Daniel Moore as a writer you had asked him to approve for the latest book release.
Were some websites removed for the book? (No)

Appendix D

Experiments 4a-4b stimuli

(1) a. The directors wondered how many segments about herself in Nicole's kitchen you asked Marisa to edit for the live broadcast.
Was Marisa asked to edit some segments for the broadcast? (Yes)

b. The directors wondered how many segments about herself in Andrew's kitchen you asked Marisa to edit for the live broadcast.
Was Marisa asked to edit some segments for the broadcast? (Yes)

c. The directors wondered how many segments about her in Nicole's kitchen you asked Marisa to edit for the live broadcast.
Was Marisa asked to edit some segments for the broadcast? (Yes)

d. The directors wondered how many segments about her in Andrew's kitchen you asked Marisa to edit for the live broadcast.
Was Marisa asked to edit some segments for the broadcast? (Yes)

(2) a. The organizers wondered how many comics about himself on Corey's patio you asked Daniel to repost for the art blog.
Was Daniel asked to repost some comics for an advertisement? (No)

b. The organizers wondered how many comics about himself on Lindsay's patio you asked Daniel to repost for the art blog.
Was Daniel asked to repost some comics for an advertisement? (No)

c. The organizers wondered how many comics about him on Corey's patio you asked Daniel to repost for the art blog.
Was Daniel asked to repost some comics for an advertisement? (No)

d. The organizers wondered how many comics about him on Lindsay's patio you asked Daniel to repost for the art blog.
Was Daniel asked to repost some comics for an advertisement? (No)

(3) a. The reporters wondered how many exposés about herself at Vivian's school you asked Jennifer to endorse for the movie promotion.
Was Jennifer asked to endorse some exposés for a book? (No)

b. The reporters wondered how many exposés about herself at Luke's school you asked Jennifer to endorse for the movie promotion.
Was Jennifer asked to endorse some exposés for a book? (No)

c. The reporters wondered how many exposés about her at Vivian's school you asked Jennifer to endorse for the movie promotion.
Was Jennifer asked to endorse some exposés for a book? (No)

d. The reporters wondered how many exposés about her at Luke's school you asked Jennifer to endorse for the movie promotion.
Was Jennifer asked to endorse some exposés for a book? (No)

(4) a. The employees wondered how many reports about himself at Gavin's workplace you asked Steven to summarize at the weekly meeting.
Was Steven asked to summarize some reports for a meeting? (Yes)

b. The employees wondered how many reports about himself at Vicki's workplace you asked Steven to summarize at the weekly meeting.
Was Steven asked to summarize some reports for a meeting? (Yes)

c. The employees wondered how many reports about him at Gavin's workplace you asked Steven to summarize at the weekly meeting.
Was Steven asked to summarize some reports for a meeting? (Yes)

d. The employees wondered how many reports about him at Vicki's workplace you asked Steven to summarize at the weekly meeting.
Was Steven asked to summarize some reports for a meeting? (Yes)

(5) a. The agents wondered how many songs about herself in Liz's garage you asked Heather to perform for the worldwide tour.
Was Heather asked to perform some songs for a fundraiser? (No)

b. The agents wondered how many songs about herself in Kirk's garage you asked Heather to perform for the worldwide tour.
Was Heather asked to perform some songs for a fundraiser? (No)

c. The agents wondered how many songs about her in Liz's garage you asked Heather to perform for the worldwide tour.
Was Heather asked to perform some songs for a fundraiser? (No)

d. The agents wondered how many songs about her in Kirk's garage you asked Heather to perform for the worldwide tour.

Was Heather asked to perform some songs for a fundraiser? (No)

(6) a. The buyers wondered how many drawings of himself in Jason's store you asked Brad to touch up for the book release.
Was Brad asked to touch up some drawings for the book release? (Yes)

b. The buyers wondered how many drawings of himself in Diane's store you asked Brad to touch up for the book release.
Was Brad asked to touch up some drawings for the book release? (Yes)

c. The buyers wondered how many drawings of him in Jason's store you asked Brad to touch up for the book release.
Was Brad asked to touch up some drawings for the book release? (Yes)

d. The buyers wondered how many drawings of him in Diane's store you asked Brad to touch up for the book release.
Was Brad asked to touch up some drawings for the book release? (Yes)

(7) a. The bloggers wondered how many photographs of herself in Lori's backyard you asked Christina to repost for the online advertisement.
Were the photographs of Christina? (Yes or No)

b. The bloggers wondered how many photographs of herself in Ivan's backyard you asked Christina to repost for the online advertisement.
Were the photographs of Christina? (Yes or No)

c. The bloggers wondered how many photographs of her in Lori's backyard you asked Christina to repost for the online advertisement.
Were the photographs of Christina? (Yes or No)

d. The bloggers wondered how many photographs of her in Ivan's backyard you asked Christina to repost for the online advertisement.
Were the photographs of Christina? (Yes or No)

(8) a. The parents wondered how many columns about himself at Jack's gym you asked Joey to revise for the school yearbook.
Were the columns about Joey? (Yes or No)

b. The parents wondered how many columns about himself at Nancy's gym you asked Joey to revise for the school yearbook.
Were the columns about Joey? (Yes or No)

c. The parents wondered how many columns about him at Jack's gym you asked Joey to revise for the school yearbook.
Were the columns about Joey? (Yes or No)

d. The parents wondered how many columns about him at Nancy's gym you asked Joey to revise for the school yearbook.
Were the columns about Joey? (Yes or No)

(9) a. The agents wondered how many videoclips of herself at Juliet's party you asked Sheila to watch for the television broadcast.
Were the videoclips of Sheila? (Yes or No)

b. The agents wondered how many videoclips of herself at Logan's party you asked Sheila to watch for the television broadcast.
Were the videoclips of Sheila? (Yes or No)

c. The agents wondered how many videoclips of her at Juliet's party you asked Sheila to watch for the television broadcast.
Were the videoclips of Sheila? (Yes or No)

d. The agents wondered how many videoclips of her at Logan's party you asked Sheila to watch for the television broadcast.
Were the videoclips of Sheila? (Yes or No)

(10) a. The lawyers wondered how many blogs about himself on Jacob's website you asked Mike to edit for the grand opening.
Were the blogs about Mike? (Yes or No)

b. The lawyers wondered how many blogs about himself on Evelyn's website you asked Mike to edit for the grand opening.
Were the blogs about Mike? (Yes or No)

c. The lawyers wondered how many blogs about him on Jacob's website you asked Mike to edit for the grand opening.
Were the blogs about Mike? (Yes or No)

d. The lawyers wondered how many blogs about him on Evelyn's website you asked Mike to edit for the grand opening.
Were the blogs about Mike? (Yes or No)

(11) a. The relatives wondered how many pictures of herself in Jessica's house you asked Mary to bring for the photo album.
Were the pictures of Jessica? (Yes or No)

b. The relatives wondered how many pictures of herself in Paul's house you asked Mary to bring for the photo album.
Were the pictures of Paul? (Yes or No)

c. The relatives wondered how many pictures of her in Jessica's house you asked Mary to bring for the photo album.
Were the pictures of Jessica? (Yes or No)

d. The relatives wondered how many pictures of her in Paul's house you asked Mary to bring for the photo album.
Were the pictures of Paul? (Yes or No)

(12) a. The administrators wondered how many reports about himself on Jimmy's desk you asked Charles to bring to the annual conference.
Were the reports about Jimmy? (Yes or No)

b. The administrators wondered how many reports about himself on Clare's desk you asked Charles to bring to the annual conference.
Were the reports about Clare? (Yes or No)

c. The administrators wondered how many reports about him on Jimmy's desk you asked Charles to bring to the annual conference.
Were the reports about Jimmy? (Yes or No)

d. The administrators wondered how many reports about him on Clare's desk you asked Charles to bring to the annual conference.
Were the reports about Clare? (Yes or No)

(13) a. The comedians wondered how many jokes about herself at Holly's bar you asked Carly to practice for the stand-up show.
Were the jokes about Holly? (Yes or No)

b. The comedians wondered how many jokes about herself at Kevin's bar you asked Carly to practice for the stand-up show.
Were the jokes about Kevin? (Yes or No)

c. The comedians wondered how many jokes about her at Holly's bar you asked Carly to practice for the stand-up show.
Were the jokes about Holly? (Yes or No)

d. The comedians wondered how many jokes about her at Kevin's bar you asked Carly to practice for the stand-up show.
Were the jokes about Kevin? (Yes or No)

(14) a. The writers wondered how many articles about himself on Matthew's boat you asked Lance to verify for the travel magazine.
Were the articles about Matthew? (Yes or No)

b. The writers wondered how many articles about himself on Patricia's boat you asked Lance to

verify in the travel magazine.
Were the articles about Patricia? (Yes or No)

c. The writers wondered how many articles about him on Matthew's boat you asked Lance to verify for the travel magazine.
Were the articles about Matthew? (Yes or No)

d. The writers wondered how many articles about him on Patricia's boat you asked Lance to verify in the travel magazine.
Were the articles about Patricia? (Yes or No)

(15) a. The journalists wondered how many stories about herself in Emma's attic you asked Ellen to tell for the TV interview.
Was Ellen asked to tell some jokes for an interview? (No)

b. The journalists wondered how many stories about herself in George's attic you asked Ellen to tell for the TV interview.
Was Ellen asked to tell some jokes for an interview? (No)

c. The journalists wondered how many stories about her in Emma's attic you asked Ellen to tell for the TV interview.
Was Ellen asked to tell some jokes for an interview? (No)

d. The journalists wondered how many stories about her in George's attic you asked Ellen to tell for the TV interview.
Was Ellen asked to tell some jokes for an interview? (No)

(16) a. The relatives wondered how many jokes about himself at Ted's camp you asked Patrick to tell at the family dinner.
Was Patrick asked to tell some jokes at dinner? (Yes)

b. The relatives wondered how many jokes about himself at Sophia's camp you asked Patrick to tell at the family dinner.
Was Patrick asked to tell some jokes at dinner? (Yes)

c. The relatives wondered how many jokes about him at Ted's camp you asked Patrick to tell at the family dinner.
Was Patrick asked to tell some jokes at dinner? (Yes)

d. The relatives wondered how many jokes about him at Sophia's camp you asked Patrick to tell at the family dinner.
Was Patrick asked to tell some jokes at dinner? (Yes)

(17) a. The publishers wondered how many articles about herself at Kylie's event you asked Audrey to correct for the online review.

Was Audrey asked to correct some articles for a review? (Yes)

b. The publishers wondered how many articles about herself at Nathan's event you asked Audrey to correct for the online review.
Was Audrey asked to correct some articles for a review? (Yes)

c. The publishers wondered how many articles about her at Kylie's event you asked Audrey to correct for the online review.
Was Audrey asked to correct some articles for a review? (Yes)

d. The publishers wondered how many articles about her at Nathan's event you asked Audrey to correct for the online review.
Was Audrey asked to correct some articles for a review? (Yes)

(18) a. The producers wondered how many stories about himself at Henry's cottage you asked Larry to proofread for the live interview.
Was Larry asked to proofread some poems for an interview? (No)

b. The producers wondered how many stories about himself at Sarah's cottage you asked Larry to proofread for the live interview.
Was Larry asked to proofread some poems for an interview? (No)

c. The producers wondered how many stories about him at Henry's cottage you asked Larry to proofread for the live interview.
Was Larry asked to proofread some poems for an interview? (No)

d. The producers wondered how many stories about him at Sarah's cottage you asked Larry to proofread for the live interview.
Was Larry asked to proofread some poems for an interview? (No)

(19) a. The network wondered how many documentaries about herself at Allison's spa you asked Betty to watch for the press release.
Was Betty asked to watch some documentaries for the press release? (Yes)

b. The network wondered how many documentaries about herself at Joel's spa you asked Betty to watch for the press release.
Was Betty asked to watch some documentaries for the press release? (Yes)

c. The network wondered how many documentaries about her at Allison's spa you asked Betty to watch for the press release.
Was Betty asked to watch some documentaries for the press release? (Yes)

d. The network wondered how many documentaries about her at Joel's spa you asked Betty to watch for the press release.
Was Betty asked to watch some documentaries for the press release? (Yes)

(20) a. The publicists wondered how many posts about himself in Richard's studio you asked Roger to share for the album release.
Was Roger asked to share some stories for the album release? (No)

b. The publicists wondered how many posts about himself in Leah's studio you asked Roger to share for the album release.
Was Roger asked to share some stories for the album release? (No)

c. The publicists wondered how many posts about him in Richard's studio you asked Roger to share for the album release.
Was Roger asked to share some stories for the album release? (No)

d. The publicists wondered how many posts about him in Leah's studio you asked Roger to share for the album release.
Was Roger asked to share some stories for the album release? (No)

(21) a. The instructors wondered how many pictures of herself in Julia's classroom you asked Heidi to choose for the art scrapbook.
Was it the instructors who wondered about the pictures? (Yes)

b. The instructors wondered how many pictures of herself in Roy's classroom you asked Heidi to choose for the art scrapbook.
Was it the instructors who wondered about the pictures? (Yes)

c. The instructors wondered how many pictures of her in Julia's classroom you asked Heidi to choose for the art scrapbook.
Was it the instructors who wondered about the pictures? (Yes)

d. The instructors wondered how many pictures of her in Roy's classroom you asked Heidi to choose for the art scrapbook.
Was it the instructors who wondered about the pictures? (Yes)

(22) a. The exhibitors wondered how many caricatures of himself at Ethan's museum you asked Bob to show at the artist showcase.
Was it the patrons who wondered about the caricatures? (No)

b. The exhibitors wondered how many caricatures of himself at Kayla's museum you asked Bob to show at the artist showcase.
Was it the patrons who wondered about the caricatures? (No)

c. The exhibitors wondered how many caricatures of him at Ethan's museum you asked Bob to show at the artist showcase.
Was it the patrons who wondered about the caricatures? (No)

d. The exhibitors wondered how many caricatures of him at Kayla's museum you asked Bob to show at the artist showcase.

Was it the patrons who wondered about the caricatures? (No)

(23) a. The administrators wondered how many photographs of herself in Jasmine's office you asked Olivia to touch up for the promotional brochure.
Was it the professors who wondered about the photographs? (No)

b. The administrators wondered how many photographs of herself in Jake's office you asked Olivia to touch up for the promotional brochure.
Was it the professors who wondered about the photographs? (No)

c. The administrators wondered how many photographs of her in Jasmine's office you asked Olivia to touch up for the promotional brochure.
Was it the professors who wondered about the photographs? (No)

d. The administrators wondered how many photographs of her in Jake's office you asked Olivia to touch up for the promotional brochure.
Was it the professors who wondered about the photographs? (No)

(24) a. The reporters wondered how many lies about himself in Sean's report you asked Jeff to deny for the TV interview.
Was it the reporters who wondered about the lies? (Yes)

b. The reporters wondered how many lies about himself in Shelly's report you asked Jeff to deny for the TV interview.
Was it the reporters who wondered about the lies? (Yes)

c. The reporters wondered how many lies about him in Sean's report you asked Jeff to deny for the TV interview.
Was it the reporters who wondered about the lies? (Yes)

d. The reporters wondered how many lies about him in Shelly's report you asked Jeff to deny for the TV interview.
Was it the reporters who wondered about the lies? (Yes)

(25) a. The curators wondered how many portraits of herself in Alicia's parlor you asked Megan to take to the art show.
Was it the tourists who wondered about the portraits? (No)

b. The curators wondered how many portraits of herself in Benjamin's parlor you asked Megan to take to the art show.
Was it the tourists who wondered about the portraits? (No)

c. The curators wondered how many portraits of her in Alicia's parlor you asked Megan to take to the art show.
Was it the tourists who wondered about the portraits? (No)

d. The curators wondered how many portraits of her in Benjamin's parlor you asked Megan to take to the art show.
Was it the tourists who wondered about the portraits? (No)

(26) a. The bloggers wondered how many rumors about himself at Shane's party you asked Derek to deny on social media.
Was it the bloggers who wondered about the rumors? (Yes)

b. The bloggers wondered how many rumors about himself at Hilda's party you asked Derek to deny on social media.
Was it the bloggers who wondered about the rumors? (Yes)

c. The bloggers wondered how many rumors about him at Shane's party you asked Derek to deny on social media.
Was it the bloggers who wondered about the rumors? (Yes)

d. The bloggers wondered how many rumors about him at Hilda's party you asked Derek to deny on social media.
Was it the bloggers who wondered about the rumors? (Yes)

(27) a. The viewers wondered how many rumors about herself at Katie's place you asked Laura to discredit for the reality show.
Did Laura confirm some rumors for the show? (No)

b. The viewers wondered how many rumors about herself at Noah's place you asked Laura to discredit for the reality show.
Did Laura confirm some rumors for the show? (No)

c. The viewers wondered how many rumors about her at Katie's place you asked Laura to discredit for the reality show.
Did Laura confirm some rumors for the show? (No)

d. The viewers wondered how many rumors about her at Noah's place you asked Laura to discredit for the reality show.
Did Laura confirm some rumors for the show? (No)

(28) a. The fans wondered how many tweets about himself at Connor's salon you asked Jimmy to retweet for the evening episode.
Did Jimmy retweet some tweets for the episode? (Yes)

b. The fans wondered how many tweets about himself at Chelsea's salon you asked Jimmy to retweet for the evening episode.
Did Jimmy retweet some tweets for the episode? (Yes)

c. The fans wondered how many tweets about him at Connor's salon you asked Jimmy to retweet for the evening episode.
Did Jimmy retweet some tweets for the episode? (Yes)

d. The fans wondered how many tweets about him at Chelsea's salon you asked Jimmy to retweet for the evening episode.
Did Jimmy retweet some tweets for the episode? (Yes)

(29) a. The columnists wondered how many images of herself at Kristin's winery you asked Lauren to select for the food magazine.
Did Lauren select some images for the magazine? (Yes)

b. The columnists wondered how many images of herself at Caleb's winery you asked Lauren to select for the food magazine.
Did Lauren select some images for the magazine? (Yes)

c. The columnists wondered how many images of her at Kristin's winery you asked Lauren to select for the food magazine.
Did Lauren select some images for the magazine? (Yes)

d. The columnists wondered how many images of her at Caleb's winery you asked Lauren to select for the food magazine.
Did Lauren select some images for the magazine? (Yes)

(30) a. The broadcasters wondered how many videos of himself at Max's ceremony you asked Robert to select for the promotional advertisement.
Did Robert delete some videos for the advertisement? (No)

b. The broadcasters wondered how many videos of himself at Amanda's ceremony you asked Robert to select for the promotional advertisement.
Did Robert delete some videos for the advertisement? (No)

c. The broadcasters wondered how many videos of him at Max's ceremony you asked Robert to select for the promotional advertisement.
Did Robert delete some videos for the advertisement? (No)

d. The broadcasters wondered how many videos of him at Amanda's ceremony you asked Robert to select for the promotional advertisement.
Did Robert delete some videos for the advertisement? (No)

(31) a. The teachers wondered how many sketches of herself in Lucy's office you asked Stephanie to pick for the annual scrapbook.
Did Stephanie pick some sketches for the scrapbook? (Yes)

b. The teachers wondered how many sketches of herself in Tim's office you asked Stephanie to

pick for the annual scrapbook.
Did Stephanie pick some sketches for the scrapbook? (Yes)

c. The teachers wondered how many sketches of her in Lucy's office you asked Stephanie to pick for the annual scrapbook.
Did Stephanie pick some sketches for the scrapbook? (Yes)

d. The teachers wondered how many sketches of her in Tim's office you asked Stephanie to pick for the annual scrapbook.
Did Stephanie pick some sketches for the scrapbook? (Yes)

(32) a. The museum wondered how many portraits of himself in Kyle's foyer you asked Jasper to judge for the new exhibit.
Did Jasper destroy some portraits for the exhibit? (No)

b. The museum wondered how many portraits of himself in Leanne's foyer you asked Jasper to judge for the new exhibit.
Did Jasper destroy some portraits for the exhibit? (No)

c. The museum wondered how many portraits of him in Kyle's foyer you asked Jasper to judge for the new exhibit.
Did Jasper destroy some portraits for the exhibit? (No)

d. The museum wondered how many portraits of him in Leanne's foyer you asked Jasper to judge for the new exhibit.
Did Jasper destroy some portraits for the exhibit? (No)

Appendix E

Experiment 5 stimuli

(1) a. The reporters wondered how many episodes about herself visiting Nicole's restaurant you asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Marisa)

b. The reporters wondered how many episodes about herself visiting Andrew's restaurant you asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

c. The reporters wondered how many episodes about her visiting Nicole's restaurant you asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Marisa)

d. The reporters wondered how many episodes about her visiting Andrew's restaurant you asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

(2) a. The agents wondered how many songs about herself meeting Liz's friend you asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Liz)

b. The agents wondered how many songs about herself meeting Kirk's friend you asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Kirk)

c. The agents wondered how many songs about her meeting Liz's friend you asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Liz)

d. The agents wondered how many songs about her meeting Kirk's friend you asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Kirk)

(3) a. The bloggers wondered how many photographs of herself wearing Lori's hat you asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Christina)

b. The bloggers wondered how many photographs of herself wearing Ivan's hat you asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

c. The bloggers wondered how many photographs of her wearing Lori's hat you asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Christina)

d. The bloggers wondered how many photographs of her wearing Ivan's hat you asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

(4) a. The reporters wondered how many exposés about herself revealing Vivian's motives you asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Vivian)

b. The reporters wondered how many exposés about herself revealing Luke's motives you asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

c. The reporters wondered how many exposés about her revealing Vivian's motives you asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Vivian)

d. The reporters wondered how many exposés about her revealing Luke's motives you asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

(5) a. The superheroes wondered how many comics about himself fighting Mark's nemesis you asked Wade to promote for the movie promotion.
Who were the comics about? (Options: Mark, Wade)

b. The superheroes wondered how many comics about himself fighting Lindsay's nemesis you asked Wade to promote for the movie promotion.
Who were the comics about? (Options: Lindsay, Wade)

c. The superheroes wondered how many comics about him fighting Mark's nemesis you asked Wade to promote for the movie promotion.
Who were the comics about? (Options: Mark, Wade)

d. The superheroes wondered how many comics about him fighting Lindsay's nemesis you asked Wade to promote for the movie promotion.

Who were the comics about? (Options: Lindsay, Wade)

(6) a. The employees wondered how many reports about himself plagiarizing Gavin's work you asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Gavin)

b. The employees wondered how many reports about himself plagiarizing Vicki's work you asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

c. The employees wondered how many reports about him plagiarizing Gavin's work you asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Gavin)

d. The employees wondered how many reports about him plagiarizing Vicki's work you asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

(7) a. The tourists wondered how many blog posts about himself touring Jacob's resort you asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Mike)

b. The tourists wondered how many blog posts about himself touring Evelyn's resort you asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

c. The tourists wondered how many blog posts about him touring Jacob's resort you asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Mike)

d. The tourists wondered how many blog posts about him touring Evelyn's resort you asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

(8) a. The guests wondered how many photos of himself holding Jason's trophy you asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Jason)

b. The guests wondered how many photos of himself holding Diane's trophy you asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

c. The guests wondered how many photos of him holding Jason's trophy you asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Jason)

d. The guests wondered how many photos of him holding Diane's trophy you asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

(9) a. The publishers wondered how many articles about herself supporting Kylie's business you asked Audrey to correct for the online review.
Who were the articles about? (Options: Audrey, Kylie)

b. The publishers wondered how many articles about herself supporting Nathan's business you asked Audrey to correct for the online review.
Who were the articles about? (Options: Audrey, Nathan)

c. The publishers wondered how many articles about her supporting Kylie's business you asked Audrey to correct for the online review.
Who were the articles about? (Options: Audrey, Kylie)

d. The publishers wondered how many articles about her supporting Nathan's business you asked Audrey to correct for the online review.
Who were the articles about? (Options: Audrey, Nathan)

(10) a. The journalists wondered how many stories about herself rehearsing Emma's lines you asked Ellen to tell for the TV interview.
Who were the stories about? (Options: Emma, Ellen)

b. The journalists wondered how many stories about herself rehearsing George's lines you asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

c. The journalists wondered how many stories about her rehearsing Emma's lines you asked Ellen to tell for the TV interview.
Who were the stories about? (Options: Emma, Ellen)

d. The journalists wondered how many stories about her rehearsing George's lines you asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

(11) a. The friends wondered how many videoclips of herself sampling Juliet's snacks you asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Juliet)

b. The friends wondered how many videoclips of herself sampling Logan's snacks you asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

c. The friends wondered how many videoclips of her sampling Juliet's snacks you asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Juliet)

d. The friends wondered how many videoclips of her sampling Logan's snacks you asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

(12) a. The network wondered how many documentaries about herself participating in Allison's class you asked Betty to watch for the press release.
Who were the documentaries about? (Options: Allison, Betty)

b. The network wondered how many documentaries about herself participating in Joel's class you asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

c. The network wondered how many documentaries about her participating in Allison's class you asked Betty to watch for the press release.
Who were the documentaries about? (Options: Allison, Betty)

d. The network wondered how many documentaries about her participating in Joel's class you asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

(13) a. The publicists wondered how many posts about himself attending Richard's show you asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Richard)

b. The publicists wondered how many posts about himself attending Leah's show you asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

c. The publicists wondered how many posts about him attending Richard's show you asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Richard)

d. The publicists wondered how many posts about him attending Leah's show you asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

(14) a. The relatives wondered how many jokes about himself trying on Ted's costume you asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Ted, Patrick)

b. The relatives wondered how many jokes about himself trying on Sophia's costume you asked

Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

c. The relatives wondered how many jokes about him trying on Ted's costume you asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Ted, Patrick)

d. The relatives wondered how many jokes about him trying on Sophia's costume you asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

(15) a. The supervisors wondered how many reports about himself editing Jimmy's articles you asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Jimmy)

b. The supervisors wondered how many reports about himself editing Clare's articles you asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

c. The supervisors wondered how many reports about him editing Jimmy's articles you asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Jimmy)

d. The supervisors wondered how many reports about him editing Clare's articles you asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

(16) a. The stylists wondered how many stories about himself mocking Henry's hair you asked Larry to rehearse for the comedy show.
Who were the stories about? (Options: Henry, Larry)

b. The stylists wondered how many stories about himself mocking Sarah's hair you asked Larry to rehearse for the comedy show.
Who were the stories about? (Options: Sarah, Larry)

c. The stylists wondered how many stories about him mocking Henry's hair you asked Larry to rehearse for the comedy show.
Who were the stories about? (Options: Henry, Larry)

d. The stylists wondered how many stories about him mocking Sarah's hair you asked Larry to rehearse for the comedy show.
Who were the stories about? (Options: Sarah, Larry)

(17) a. The followers wondered how many tweets about herself backing Alicia's defense you asked Megan to retweet for the court date.

Who were the tweets about? (Options: Megan, Alicia)

b. The followers wondered how many tweets about herself backing Benjamin's defense you asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

c. The followers wondered how many tweets about her backing Alicia's defense you asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Alicia)

d. The followers wondered how many tweets about her backing Benjamin's defense you asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

(18) a. The reporters wondered how many lies about himself crashing Sean's car you asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Sean)

b. The reporters wondered how many lies about himself crashing Shelly's car you asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

c. The reporters wondered how many lies about him crashing Sean's car you asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Sean)

d. The reporters wondered how many lies about him crashing Shelly's car you asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

(19) a. The exhibitors wondered how many caricatures of himself kissing Ethan's dog you asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Bob)

b. The exhibitors wondered how many caricatures of himself kissing Kayla's dog you asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

c. The exhibitors wondered how many caricatures of him kissing Ethan's dog you asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Bob)

d. The exhibitors wondered how many caricatures of him kissing Kayla's dog you asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

(20) a. The pet rescuers wondered how many pictures of herself petting Julia's dog you asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Julia, Heidi)

b. The pet rescuers wondered how many pictures of herself petting Roy's dog you asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

c. The pet rescuers wondered how many pictures of her petting Julia's dog you asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Julia, Heidi)

d. The pet rescuers wondered how many pictures of her petting Roy's dog you asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

(21) a. The bloggers wondered how many rumors about himself losing Shane's wallet you asked Derek to deny on the social media page.
Whose wallet was lost? (Options: Derek, Shane)

b. The bloggers wondered how many rumors about himself losing Hilda's wallet you asked Derek to deny on the social media page.
Whose wallet was lost? (Options: Derek, Hilda)

c. The bloggers wondered how many rumors about him losing Shane's wallet you asked Derek to deny on the social media page.
Whose wallet was lost? (Options: Derek, Shane)

d. The bloggers wondered how many rumors about him losing Hilda's wallet you asked Derek to deny on the social media page.
Whose wallet was lost? (Options: Derek, Hilda)

(22) a. The advertisers wondered how many photographs of herself driving Jasmine's car you asked Olivia to touch up for the promotional brochure.
Whose car was being driven? (Options: Olivia, Jasmine)

b. The advertisers wondered how many photographs of herself driving Jake's car you asked Olivia to touch up for the promotional brochure.
Whose car was being driven? (Options: Olivia, Jake)

c. The advertisers wondered how many photographs of her driving Jasmine's car you asked Olivia to touch up for the promotional brochure.
Whose car was being driven? (Options: Olivia, Jasmine)

d. The advertisers wondered how many photographs of her driving Jake's car you asked Olivia to touch up for the promotional brochure.

Whose car was being driven? (Options: Olivia, Jake)

(23) a. The comedians wondered how many jokes about herself mocking Holly's job you asked Carly to practice for the stand-up show.
Whose job was being mocked? (Options: Holly, Carly)

b. The comedians wondered how many jokes about herself mocking Kevin's job you asked Carly to practice for the stand-up show.
Whose job was being mocked? (Options: Kevin, Carly)

c. The comedians wondered how many jokes about her mocking Holly's job you asked Carly to practice for the stand-up show.
Whose job was being mocked? (Options: Holly, Carly)

d. The comedians wondered how many jokes about her mocking Kevin's job you asked Carly to practice for the stand-up show.
Whose job was being mocked? (Options: Kevin, Carly)

(24) a. The writers wondered how many articles about himself mishandling Matthew's case you asked Lance to verify for the newspaper article.
Whose case was mishandled? (Options: Matthew, Lance)

b. The writers wondered how many articles about himself mishandling Patricia's case you asked Lance to verify for the newspaper article.
Whose case was mishandled? (Options: Patricia, Lance)

c. The writers wondered how many articles about him mishandling Matthew's case you asked Lance to verify for the newspaper article.
Whose case was mishandled? (Options: Matthew, Lance)

d. The writers wondered how many articles about him mishandling Patricia's case you asked Lance to verify for the newspaper article.
Whose case was mishandled? (Options: Patricia, Lance)

(25) a. The followers wondered how many posts about herself promoting Lucy's product you asked Stephanie to share for the online advertisement.
Whose product was being promoted? (Options: Lucy, Stephanie)

b. The followers wondered how many posts about herself promoting Tim's product you asked Stephanie to share for the online advertisement.
Whose product was being promoted? (Options: Tim, Stephanie)

c. The followers wondered how many posts about her promoting Lucy's product you asked Stephanie to share for the online advertisement.
Whose product was being promoted? (Options: Lucy, Stephanie)

d. The followers wondered how many posts about her promoting Tim's product you asked Stephanie to share for the online advertisement.
Whose product was being promoted? (Options: Tim, Stephanie)

(26) a. The broadcasters wondered how many videos about himself supporting Max's cause you asked Robert to select for the promotional ad.
Whose cause was being supported? (Options: Max, Robert)

b. The broadcasters wondered how many videos about himself supporting Amanda's cause you asked Robert to select for the promotional ad.
Whose cause was being supported? (Options: Amanda, Robert)

c. The broadcasters wondered how many videos about him supporting Max's cause you asked Robert to select for the promotional ad.
Whose cause was being supported? (Options: Max, Robert)

d. The broadcasters wondered how many videos about him supporting Amanda's cause you asked Robert to select for the promotional ad.
Whose cause was being supported? (Options: Amanda, Robert)

(27) a. The fans wondered how many tweets about himself winning Connor's contest you asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Connor)

b. The fans wondered how many tweets about himself winning Chelsea's contest you asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

c. The fans wondered how many tweets about him winning Connor's contest you asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Connor)

d. The fans wondered how many tweets about him winning Chelsea's contest you asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

(28) a. The viewers wondered how many rumors about herself discrediting Katie's story you asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Katie)

b. The viewers wondered how many rumors about herself discrediting Noah's story you asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

c. The viewers wondered how many rumors about her discrediting Katie's story you asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Katie)

d. The viewers wondered how many rumors about her discrediting Noah's story you asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

(29) a. The reporters wondered how many paragraphs about himself managing Kyle's business you asked Jasper to publish on the social media page.
Who was asked to publish the paragraphs? (Options: Kyle, Jasper)

b. The reporters wondered how many paragraphs about himself managing Leanne's business you asked Jasper to publish on the social media page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

c. The reporters wondered how many paragraphs about him managing Kyle's business you asked Jasper to publish on the social media page.
Who was asked to publish the paragraphs? (Options: Kyle, Jasper)

d. The reporters wondered how many paragraphs about him managing Leanne's business you asked Jasper to publish on the social media page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

(30) a. The columnists wondered how many images of herself hanging Kristin's paintings you asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Kristin, Lauren)

b. The columnists wondered how many images of herself hanging Caleb's paintings you asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

c. The columnists wondered how many images of her hanging Kristin's paintings you asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Kristin, Lauren)

d. The columnists wondered how many images of her hanging Caleb's paintings you asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

(31) a. The yogis wondered how many pictures of herself practicing Jessica's technique you asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Jessica)

b. The yogis wondered how many pictures of herself practicing Paul's technique you asked Mary

to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

c. The yogis wondered how many pictures of her practicing Jessica's technique you asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Jessica)

d. The yogis wondered how many pictures of her practicing Paul's technique you asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

(32) a. The readers wondered how many columns about himself defending Jack's case you asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Joey, Jack)

b. The readers wondered how many columns about himself defending Nancy's case you asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Joey, Nancy)

c. The readers wondered how many columns about him defending Jack's case you asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Joey, Jack)

d. The readers wondered how many columns about him defending Nancy's case you asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Joey, Nancy)

Appendix F

Experiment 6a stimuli

(1) a. The reporters wondered how many episodes about herself visiting the restaurant Nicole had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Marisa)

b. The reporters wondered how many episodes about herself visiting the restaurant Andrew had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

c. The reporters wondered how many episodes about her visiting the restaurant Nicole had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Marisa)

d. The reporters wondered how many episodes about her visiting the restaurant Andrew had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

(2) a. The agents wondered how many songs about herself falling in love Liz had asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Liz)

b. The agents wondered how many songs about herself falling in love Kirk had asked Heather to perform for the come back tour.
Who were the episodes about? (Options: Heather, Kirk)

c. The agents wondered how many songs about her falling in love Liz had asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Liz)

d. The agents wondered how many songs about her falling in love Kirk had asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Kirk)

(3) a. The bloggers wondered how many photographs of herself wearing a hat Lori had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Christina)

b. The bloggers wondered how many photographs of herself wearing a hat Ivan had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

c. The bloggers wondered how many photographs of her wearing a hat Lori had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Christina)

d. The bloggers wondered how many photographs of her wearing a hat Ivan had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

(4) a. The reporters wondered how many exposés about herself revealing the source Vivian had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Vivian)

b. The reporters wondered how many exposés about herself revealing the source Luke had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

c. The reporters wondered how many exposés about her revealing the source Vivian had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Vivian)

d. The reporters wondered how many exposés about her revealing the source Luke had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

(5) a. The readers wondered how many comics about himself fighting the enemy Mark had asked Wade to promote for the movie premiere.
Who were the comics about? (Options: Mark, Wade)

b. The readers wondered how many comics about himself fighting the enemy Lindsay had asked Wade to promote for the movie premiere.
Who were the comics about? (Options: Lindsay, Wade)

c. The readers wondered how many comics about him fighting the enemy Mark had asked Wade to promote for the movie premiere.
Who were the comics about? (Options: Mark, Wade)

d. The readers wondered how many comics about him fighting the enemy Lindsay had asked Wade to promote for the movie premiere.

Who were the comics about? (Options: Lindsay, Wade)

(6) a. The employees wondered how many reports about himself violating the patent Gavin had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Gavin)

b. The employees wondered how many reports about himself violating the patent Vicki had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

c. The employees wondered how many reports about him violating the patent Gavin had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Gavin)

d. The employees wondered how many reports about him violating the patent Vicki had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

(7) a. The tourists wondered how many blog posts about himself touring the resort Jacob had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Mike)

b. The tourists wondered how many blog posts about himself touring the resort Evelyn had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

c. The tourists wondered how many blog posts about him touring the resort Jacob had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Mike)

d. The tourists wondered how many blog posts about him touring the resort Evelyn had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

(8) a. The guests wondered how many photos of himself holding a trophy Jason had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Jason)

b. The guests wondered how many photos of himself holding a trophy Diane had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

c. The guests wondered how many photos of him holding a trophy Jason had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Jason)

d. The guests wondered how many photos of him holding a trophy Diane had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

(9) a. The publishers wondered how many articles about herself supporting a charity Kylie had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Kylie)

b. The publishers wondered how many articles about herself supporting a charity Nathan had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Nathan)

c. The publishers wondered how many articles about her supporting a charity Kylie had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Kylie)

d. The publishers wondered how many articles about her supporting a charity Nathan had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Nathan)

(10) a. The journalists wondered how many stories about herself rehearsing the script Emma had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: Emma, Ellen)

b. The journalists wondered how many stories about herself rehearsing the script George had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

c. The journalists wondered how many stories about her rehearsing the script Emma had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: Emma, Ellen)

d. The journalists wondered how many stories about her rehearsing the script George had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

(11) a. The friends wondered how many videoclips of herself modeling the mascara Juliet had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Juliet)

b. The friends wondered how many videoclips of herself modeling the mascara Logan had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

c. The friends wondered how many videoclips of her modeling the mascara Juliet had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Juliet)

d. The friends wondered how many videoclips of her modeling the mascara Logan had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

(12) a. The network wondered how many documentaries about herself touring the facilities Allison had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Allison, Betty)

b. The network wondered how many documentaries about herself touring the facilities Joel had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

c. The network wondered how many documentaries about her touring the facilities Allison had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Allison, Betty)

d. The network wondered how many documentaries about her touring the facilities Joel had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

(13) a. The publicists wondered how many posts about himself attending the show Richard had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Richard)

b. The publicists wondered how many posts about himself attending the show Leah had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

c. The publicists wondered how many posts about him attending the show Richard had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Richard)

d. The publicists wondered how many posts about him attending the show Leah had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

(14) a. The relatives wondered how many jokes about himself trying on a costume Ted had asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Ted, Patrick)

b. The relatives wondered how many jokes about himself trying on a costume Sophia had asked

Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

c. The relatives wondered how many jokes about him trying on a costume Ted had asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Ted, Patrick)

d. The relatives wondered how many jokes about him trying on a costume Sophia had asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

(15) a. The supervisors wondered how many reports about himself editing the articles Jimmy had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Jimmy)

b. The supervisors wondered how many reports about himself editing the articles Clare had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

c. The supervisors wondered how many reports about him editing the articles Jimmy had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Jimmy)

d. The supervisors wondered how many reports about him editing the articles Clare had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

(16) a. The comedians wondered how many skits about himself wearing a wig Henry had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Henry, Larry)

b. The comedians wondered how many skits about himself wearing a wig Sarah had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Sarah, Larry)

c. The comedians wondered how many skits about him wearing a wig Henry had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Henry, Larry)

d. The comedians wondered how many skits about him wearing a wig Sarah had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Sarah, Larry)

(17) a. The followers wondered how many tweets about herself supporting the defense Alicia had asked Megan to retweet for the court date.

Who were the tweets about? (Options: Megan, Alicia)

b. The followers wondered how many tweets about herself supporting the defense Benjamin had asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

c. The followers wondered how many tweets about her supporting the defense Alicia had asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Alicia)

d. The followers wondered how many tweets about her supporting the defense Benjamin had asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

(18) a. The broadcasters wondered how many lies about himself crashing a Rolls-Royce Sean had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Sean)

b. The broadcasters wondered how many lies about himself crashing a Rolls-Royce Shelly had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

c. The broadcasters wondered how many lies about him crashing a Rolls-Royce Sean had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Sean)

d. The broadcasters wondered how many lies about him crashing a Rolls-Royce Shelly had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

(19) a. The exhibitors wondered how many caricatures of himself kissing the queen Ethan had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Bob)

b. The exhibitors wondered how many caricatures of himself kissing the queen Kayla had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

c. The exhibitors wondered how many caricatures of him kissing the queen Ethan had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Bob)

d. The exhibitors wondered how many caricatures of him kissing the queen Kayla had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

(20) a. The pet rescuers wondered how many pictures of herself petting a dog Julia had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Julia, Heidi)

b. The pet rescuers wondered how many pictures of herself petting a dog Roy had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

c. The pet rescuers wondered how many pictures of her petting a dog Julia had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Julia, Heidi)

d. The pet rescuers wondered how many pictures of her petting a dog Roy had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

(21) a. The bloggers wondered how many rumors about himself losing the bet Shane had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

b. The bloggers wondered how many rumors about himself losing the bet Hilda had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

c. The bloggers wondered how many rumors about him losing the bet Shane had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

d. The bloggers wondered how many rumors about him losing the bet Hilda had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

(22) a. The advertisers wondered how many photographs of herself driving the car Jasmine had asked Olivia to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

b. The advertisers wondered how many photographs of herself driving the car Jake had asked Olivia to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

c. The advertisers wondered how many photographs of her driving the car Jasmine had asked Olivia to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

d. The advertisers wondered how many photographs of her driving the car Jake had asked Olivia to touch up for the promotional brochure.

Who wondered about the photographs? (Options: The followers, The advertisers)

(23) a. The fans wondered how many jokes about herself mocking the audience Holly had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

b. The fans wondered how many jokes about herself mocking the audience Kevin had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

c. The fans wondered how many jokes about her mocking the audience Holly had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

d. The fans wondered how many jokes about her mocking the audience Kevin had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

(24) a. The writers wondered how many articles about himself mishandling the case Matthew had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

b. The writers wondered how many articles about himself mishandling the case Patricia had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

c. The writers wondered how many articles about him mishandling the case Matthew had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

d. The writers wondered how many articles about him mishandling the case Patricia had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

(25) a. The followers wondered how many posts about herself promoting the face cream Lucy had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

b. The followers wondered how many posts about herself promoting the face cream Tim had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

c. The followers wondered how many posts about her promoting the face cream Lucy had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

d. The followers wondered how many posts about her promoting the face cream Tim had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

(26) a. The broadcasters wondered how many videos about himself supporting the relief effort Max had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

b. The broadcasters wondered how many videos about himself supporting the relief effort Amanda had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

c. The broadcasters wondered how many videos about him supporting the relief effort Max had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

d. The broadcasters wondered how many videos about him supporting the relief effort Amanda had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

(27) a. The fans wondered how many tweets about himself winning the contest Connor had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Connor)

b. The fans wondered how many tweets about himself winning the contest Chelsea had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

c. The fans wondered how many tweets about him winning the contest Connor had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Connor)

d. The fans wondered how many tweets about him winning the contest Chelsea had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

(28) a. The viewers wondered how many rumors about herself discrediting the report Katie had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Katie)

b. The viewers wondered how many rumors about herself discrediting the report Noah had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

c. The viewers wondered how many rumors about her discrediting the report Katie had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Katie)

d. The viewers wondered how many rumors about her discrediting the report Noah had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

(29) a. The journalists wondered how many paragraphs about himself managing the business Kyle had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Kyle, Jasper)

b. The journalists wondered how many paragraphs about himself managing the business Leanne had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

c. The journalists wondered how many paragraphs about him managing the business Kyle had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Kyle, Jasper)

d. The journalists wondered how many paragraphs about him managing the business Leanne had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

(30) a. The columnists wondered how many images of herself hanging the paintings Kristin had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Kristin, Lauren)

b. The columnists wondered how many images of herself hanging the paintings Caleb had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

c. The columnists wondered how many images of her hanging the paintings Kristin had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Kristin, Lauren)

d. The columnists wondered how many images of her hanging the paintings Caleb had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

(31) a. The yogis wondered how many pictures of herself practicing the pigeon pose Jessica had asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Jessica)

b. The yogis wondered how many pictures of herself practicing the pigeon pose Paul had asked

Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

c. The yogis wondered how many pictures of her practicing the pigeon pose Jessica had asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Jessica)

d. The yogis wondered how many pictures of her practicing the pigeon pose Paul had asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

(32) a. The readers wondered how many columns about himself defending the death penalty Jack had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Jack, Joey)

b. The readers wondered how many columns about himself defending the death penalty Nancy had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Nancy, Joey)

c. The readers wondered how many columns about him defending the death penalty Jack had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Jack, Joey)

d. The readers wondered how many columns about him defending the death penalty Nancy had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Nancy, Joey)

Appendix G

Experiment 6b stimuli

(1) a. The reporters wondered how many episodes about herself visiting the restaurant Nicole had asked Andrew to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Andrew)

b. The reporters wondered how many episodes about herself visiting the restaurant Andrew had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

c. The reporters wondered how many episodes about her visiting the restaurant Nicole had asked Andrew to edit for the live broadcast.
Who were the episodes about? (Options: Nicole, Andrew)

d. The reporters wondered how many episodes about her visiting the restaurant Andrew had asked Marisa to edit for the live broadcast.
Who were the episodes about? (Options: Andrew, Marisa)

(2) a. The agents wondered how many songs about herself falling in love Liz had asked Kirk to perform for the come back tour.
Who were the songs about? (Options: Kirk, Liz)

b. The agents wondered how many songs about herself falling in love Kirk had asked Heather to perform for the come back tour.
Who were the episodes about? (Options: Heather, Kirk)

c. The agents wondered how many songs about her falling in love Liz had asked Kirk to perform for the come back tour.
Who were the songs about? (Options: Kirk, Liz)

d. The agents wondered how many songs about her falling in love Kirk had asked Heather to perform for the come back tour.
Who were the songs about? (Options: Heather, Kirk)

(3) a. The bloggers wondered how many photographs of herself wearing a hat Lori had asked Ivan to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Ivan)

b. The bloggers wondered how many photographs of herself wearing a hat Ivan had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

c. The bloggers wondered how many photographs of her wearing a hat Lori had asked Ivan to repost for the online advertisement.
Who were the photographs of? (Options: Lori, Ivan)

d. The bloggers wondered how many photographs of her wearing a hat Ivan had asked Christina to repost for the online advertisement.
Who were the photographs of? (Options: Ivan, Christina)

(4) a. The reporters wondered how many exposés about herself revealing the source Vivian had asked Luke to tape for the legal battle.
Who were the exposés about? (Options: Luke, Vivian)

b. The reporters wondered how many exposés about herself revealing the source Luke had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

c. The reporters wondered how many exposés about her revealing the source Vivian had asked Luke to tape for the legal battle.
Who were the exposés about? (Options: Luke, Vivian)

d. The reporters wondered how many exposés about her revealing the source Luke had asked Jennifer to tape for the legal battle.
Who were the exposés about? (Options: Jennifer, Luke)

(5) a. The readers wondered how many comics about himself fighting the enemy Mark had asked Lindsay to promote for the movie premiere.
Who were the comics about? (Options: Mark, Lindsay)

b. The readers wondered how many comics about himself fighting the enemy Lindsay had asked Wade to promote for the movie premiere.
Who were the comics about? (Options: Lindsay, Wade)

c. The readers wondered how many comics about him fighting the enemy Mark had asked Lindsay to promote for the movie premiere.
Who were the comics about? (Options: Mark, Lindsay)

d. The readers wondered how many comics about him fighting the enemy Lindsay had asked Wade to promote for the movie premiere.

Who were the comics about? (Options: Lindsay, Wade)

(6) a. The employees wondered how many reports about himself violating the patent Gavin had asked Vicki to refute at the weekly meeting.
Who were the reports about? (Options: Vicki, Gavin)

b. The employees wondered how many reports about himself violating the patent Vicki had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

c. The employees wondered how many reports about him violating the patent Gavin had asked Vicki to refute at the weekly meeting.
Who were the reports about? (Options: Vicki, Gavin)

d. The employees wondered how many reports about him violating the patent Vicki had asked Steven to refute at the weekly meeting.
Who were the reports about? (Options: Steven, Vicki)

(7) a. The tourists wondered how many blog posts about himself touring the resort Jacob had asked Evelyn to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Evelyn)

b. The tourists wondered how many blog posts about himself touring the resort Evelyn had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

c. The tourists wondered how many blog posts about him touring the resort Jacob had asked Evelyn to edit for the grand opening.
Who were the blog posts about? (Options: Jacob, Evelyn)

d. The tourists wondered how many blog posts about him touring the resort Evelyn had asked Mike to edit for the grand opening.
Who were the blog posts about? (Options: Evelyn, Mike)

(8) a. The guests wondered how many photos of himself holding a trophy Jason had asked Diane to touch up for the party invitation.
Who were the photos of? (Options: Diane, Jason)

b. The guests wondered how many photos of himself holding a trophy Diane had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

c. The guests wondered how many photos of him holding a trophy Jason had asked Diane to touch up for the party invitation.
Who were the photos of? (Options: Diane, Jason)

d. The guests wondered how many photos of him holding a trophy Diane had asked Brad to touch up for the party invitation.
Who were the photos of? (Options: Brad, Diane)

(9) a. The publishers wondered how many articles about herself supporting a charity Kylie had asked Nathan to correct for the online blog.
Who were the articles about? (Options: Nathan, Kylie)

b. The publishers wondered how many articles about herself supporting a charity Nathan had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Nathan)

c. The publishers wondered how many articles about her supporting a charity Kylie had asked Nathan to correct for the online blog.
Who were the articles about? (Options: Nathan, Kylie)

d. The publishers wondered how many articles about her supporting a charity Nathan had asked Audrey to correct for the online blog.
Who were the articles about? (Options: Audrey, Nathan)

(10) a. The journalists wondered how many stories about herself rehearsing the script Emma had asked George to tell for the TV interview.
Who were the stories about? (Options: Emma, George)

b. The journalists wondered how many stories about herself rehearsing the script George had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

c. The journalists wondered how many stories about her rehearsing the script Emma had asked George to tell for the TV interview.
Who were the stories about? (Options: Emma, George)

d. The journalists wondered how many stories about her rehearsing the script George had asked Ellen to tell for the TV interview.
Who were the stories about? (Options: George, Ellen)

(11) a. The friends wondered how many videoclips of herself modeling the mascara Juliet had asked Logan to select for the TV promo.
Who were the videoclips of? (Options: Logan, Juliet)

b. The friends wondered how many videoclips of herself modeling the mascara Logan had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

c. The friends wondered how many videoclips of her modeling the mascara Juliet had asked Logan to select for the TV promo.
Who were the videoclips of? (Options: Logan, Juliet)

d. The friends wondered how many videoclips of her modeling the mascara Logan had asked Sheila to select for the TV promo.
Who were the videoclips of? (Options: Sheila, Logan)

(12) a. The network wondered how many documentaries about herself touring the facilities Allison had asked Joel to watch for the press release.
Who were the documentaries about? (Options: Allison, Joel)

b. The network wondered how many documentaries about herself touring the facilities Joel had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

c. The network wondered how many documentaries about her touring the facilities Allison had asked Joel to watch for the press release.
Who were the documentaries about? (Options: Allison, Joel)

d. The network wondered how many documentaries about her touring the facilities Joel had asked Betty to watch for the press release.
Who were the documentaries about? (Options: Joel, Betty)

(13) a. The publicists wondered how many posts about himself attending the show Richard had asked Leah to share for the album release.
Who were the posts about? (Options: Leah, Richard)

b. The publicists wondered how many posts about himself attending the show Leah had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

c. The publicists wondered how many posts about him attending the show Richard had asked Leah to share for the album release.
Who were the posts about? (Options: Leah, Richard)

d. The publicists wondered how many posts about him attending the show Leah had asked Roger to share for the album release.
Who were the posts about? (Options: Roger, Leah)

(14) a. The relatives wondered how many jokes about himself trying on a costume Ted had asked Sophia to tell at the family dinner.
Who were the jokes about? (Options: Ted, Sophia)

b. The relatives wondered how many jokes about himself trying on a costume Sophia had asked

Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

c. The relatives wondered how many jokes about him trying on a costume Ted had asked Sophia to tell at the family dinner.
Who were the jokes about? (Options: Ted, Sophia)

d. The relatives wondered how many jokes about him trying on a costume Sophia had asked Patrick to tell at the family dinner.
Who were the jokes about? (Options: Sophia, Patrick)

(15) a. The supervisors wondered how many reports about himself editing the articles Jimmy had asked Clare to bring to the performance review.
Who were the reports about? (Options: Clare, Jimmy)

b. The supervisors wondered how many reports about himself editing the articles Clare had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

c. The supervisors wondered how many reports about him editing the articles Jimmy had asked Clare to bring to the performance review.
Who were the reports about? (Options: Clare, Jimmy)

d. The supervisors wondered how many reports about him editing the articles Clare had asked Charles to bring to the performance review.
Who were the reports about? (Options: Charles, Clare)

(16) a. The comedians wondered how many skits about himself wearing a wig Henry had asked Sarah to rehearse for the comedy show.
Who were the skits about? (Options: Henry, Sarah)

b. The comedians wondered how many skits about himself wearing a wig Sarah had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Sarah, Larry)

c. The comedians wondered how many skits about him wearing a wig Henry had asked Sarah to rehearse for the comedy show.
Who were the skits about? (Options: Henry, Sarah)

d. The comedians wondered how many skits about him wearing a wig Sarah had asked Larry to rehearse for the comedy show.
Who were the skits about? (Options: Sarah, Larry)

(17) a. The followers wondered how many tweets about herself supporting the defense Alicia had asked Benjamin to retweet for the court date.

Who were the tweets about? (Options: Benjamin, Alicia)

b. The followers wondered how many tweets about herself supporting the defense Benjamin had asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

c. The followers wondered how many tweets about her supporting the defense Alicia had asked Benjamin to retweet for the court date.
Who were the tweets about? (Options: Benjamin, Alicia)

d. The followers wondered how many tweets about her supporting the defense Benjamin had asked Megan to retweet for the court date.
Who were the tweets about? (Options: Megan, Benjamin)

(18) a. The broadcasters wondered how many lies about himself crashing a Rolls-Royce Sean had asked Shelly to deny for the TV interview.
Who were the lies about? (Options: Shelly, Sean)

b. The broadcasters wondered how many lies about himself crashing a Rolls-Royce Shelly had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

c. The broadcasters wondered how many lies about him crashing a Rolls-Royce Sean had asked Shelly to deny for the TV interview.
Who were the lies about? (Options: Shelly, Sean)

d. The broadcasters wondered how many lies about him crashing a Rolls-Royce Shelly had asked Jeff to deny for the TV interview.
Who were the lies about? (Options: Jeff, Shelly)

(19) a. The exhibitors wondered how many caricatures of himself kissing the queen Ethan had asked Kayla to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Kayla)

b. The exhibitors wondered how many caricatures of himself kissing the queen Kayla had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

c. The exhibitors wondered how many caricatures of him kissing the queen Ethan had asked Kayla to show at the artist showcase.
Who were the caricatures of? (Options: Ethan, Kayla)

d. The exhibitors wondered how many caricatures of him kissing the queen Kayla had asked Bob to show at the artist showcase.
Who were the caricatures of? (Options: Kayla, Bob)

(20) a. The pet rescuers wondered how many pictures of herself petting a dog Julia had asked Roy to choose for the online profiles.
Who were the pictures of? (Options: Julia, Roy)

b. The pet rescuers wondered how many pictures of herself petting a dog Roy had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

c. The pet rescuers wondered how many pictures of her petting a dog Julia had asked Roy to choose for the online profiles.
Who were the pictures of? (Options: Julia, Roy)

d. The pet rescuers wondered how many pictures of her petting a dog Roy had asked Heidi to choose for the online profiles.
Who were the pictures of? (Options: Roy, Heidi)

(21) a. The bloggers wondered how many rumors about himself losing the bet Shane had asked Hilda to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

b. The bloggers wondered how many rumors about himself losing the bet Hilda had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

c. The bloggers wondered how many rumors about him losing the bet Shane had asked Hilda to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

d. The bloggers wondered how many rumors about him losing the bet Hilda had asked Derek to deny on the social media page.
Who wondered about the rumors? (Options: The bloggers, The reporters)

(22) a. The advertisers wondered how many photographs of herself driving the car Jasmine had asked Jake to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

b. The advertisers wondered how many photographs of herself driving the car Jake had asked Olivia to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

c. The advertisers wondered how many photographs of her driving the car Jasmine had asked Jake to touch up for the promotional brochure.
Who wondered about the photographs? (Options: The followers, The advertisers)

d. The advertisers wondered how many photographs of her driving the car Jake had asked Olivia to touch up for the promotional brochure.

Who wondered about the photographs? (Options: The followers, The advertisers)

(23) a. The fans wondered how many jokes about herself mocking the audience Holly had asked Kevin to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

b. The fans wondered how many jokes about herself mocking the audience Kevin had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

c. The fans wondered how many jokes about her mocking the audience Holly had asked Kevin to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

d. The fans wondered how many jokes about her mocking the audience Kevin had asked Carly to practice for the stand-up show.
Who wondered about the jokes? (Options: The fans, The comedians)

(24) a. The writers wondered how many articles about himself mishandling the case Matthew had asked Patricia to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

b. The writers wondered how many articles about himself mishandling the case Patricia had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

c. The writers wondered how many articles about him mishandling the case Matthew had asked Patricia to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

d. The writers wondered how many articles about him mishandling the case Patricia had asked Lance to verify for the newspaper article.
Who wondered about the articles? (Options: The lawyers, The writers)

(25) a. The followers wondered how many posts about herself promoting the face cream Lucy had asked Tim to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

b. The followers wondered how many posts about herself promoting the face cream Tim had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

c. The followers wondered how many posts about her promoting the face cream Lucy had asked Tim to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

d. The followers wondered how many posts about her promoting the face cream Tim had asked Stephanie to share for the online advertisement.
Who wondered about the posts? (Options: The followers, The bloggers)

(26) a. The broadcasters wondered how many videos about himself supporting the relief effort Max had asked Amanda to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

b. The broadcasters wondered how many videos about himself supporting the relief effort Amanda had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

c. The broadcasters wondered how many videos about him supporting the relief effort Max had asked Amanda to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

d. The broadcasters wondered how many videos about him supporting the relief effort Amanda had asked Robert to select for the promotional ad.
Who wondered about the ad? (Options: The publicists, The broadcasters)

(27) a. The fans wondered how many tweets about himself winning the contest Connor had asked Chelsea to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Chelsea, Connor)

b. The fans wondered how many tweets about himself winning the contest Chelsea had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

c. The fans wondered how many tweets about him winning the contest Connor had asked Chelsea to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Chelsea, Connor)

d. The fans wondered how many tweets about him winning the contest Chelsea had asked Jimmy to boast about on the evening talk show.
Who was asked to boast about the tweets? (Options: Jimmy, Chelsea)

(28) a. The viewers wondered how many rumors about herself discrediting the report Katie had asked Noah to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Noah, Katie)

b. The viewers wondered how many rumors about herself discrediting the report Noah had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

c. The viewers wondered how many rumors about her discrediting the report Katie had asked Noah to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Noah, Katie)

d. The viewers wondered how many rumors about her discrediting the report Noah had asked Laura to confirm for the reality show.
Who was asked to confirm the rumors? (Options: Laura, Noah)

(29) a. The journalists wondered how many paragraphs about himself managing the business Kyle had asked Leanne to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Kyle, Leanne)

b. The journalists wondered how many paragraphs about himself managing the business Leanne had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

c. The journalists wondered how many paragraphs about him managing the business Kyle had asked Leanne to publish on the Facebook page.
Who was asked to publish the paragraphs?(Options: Kyle, Leanne)

d. The journalists wondered how many paragraphs about him managing the business Leanne had asked Jasper to publish on the Facebook page.
Who was asked to publish the paragraphs? (Options: Leanne, Jasper)

(30) a. The columnists wondered how many images of herself hanging the paintings Kristin had asked Caleb to select for the art blog.
Who was asked to select the images? (Options: Kristin, Caleb)

b. The columnists wondered how many images of herself hanging the paintings Caleb had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

c. The columnists wondered how many images of her hanging the paintings Kristin had asked Caleb to select for the art blog.
Who was asked to select the images? (Options: Kristin, Caleb)

d. The columnists wondered how many images of her hanging the paintings Caleb had asked Lauren to select for the art blog.
Who was asked to select the images? (Options: Caleb, Lauren)

(31) a. The yogis wondered how many pictures of herself practicing the pigeon pose Jessica had asked Paul to display at the athletic studio.
Who was asked to display the pictures? (Options: Paul, Jessica)

b. The yogis wondered how many pictures of herself practicing the pigeon pose Paul had asked

Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

c. The yogis wondered how many pictures of her practicing the pigeon pose Jessica had asked Paul to display at the athletic studio.
Who was asked to display the pictures? (Options: Paul, Jessica)

d. The yogis wondered how many pictures of her practicing the pigeon pose Paul had asked Mary to display at the athletic studio.
Who was asked to display the pictures? (Options: Mary, Paul)

(32) a. The readers wondered how many columns about himself defending the death penalty Jack had asked Nancy to revise for the local newspaper.
Who was asked to revise the columns? (Options: Jack, Nancy)

b. The readers wondered how many columns about himself defending the death penalty Nancy had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Nancy, Joey)

c. The readers wondered how many columns about him defending the death penalty Jack had asked Nancy to revise for the local newspaper.
Who was asked to revise the columns? (Options: Jack, Nancy)

d. The readers wondered how many columns about him defending the death penalty Nancy had asked Joey to revise for the local newspaper.
Who was asked to revise the columns? (Options: Nancy, Joey)

Appendix H

Experiment 7 stimuli

(1) a. The editors realized that you had asked Jessica to approve five cartoons about herself for yesterday's edition of the newspaper.
Was Jessica asked to approve some cartoons about herself for the newspaper? (Yes)

b. The editors realized that you had asked Jessica to approve five cartoons about her for yesterday's edition of the newspaper.
Was Jessica asked to approve some cartoons about herself for the newspaper? (Yes or No)

c. The editors realized that you had asked Jessica to approve five cartoons about Joseph for yesterday's edition of the newspaper.
Was Jessica asked to approve some cartoons about Joseph for the newspaper? (Yes)

(2) a. The teachers realized that you had asked Stuart to adjust two sketches of himself for the art school application portfolio.
Was Stuart asked to adjust some sketches of himself for the portfolio? (Yes)

b. The teachers realized that you had asked Stuart to adjust two sketches of him for the art school application portfolio.
Was Stuart asked to adjust some sketches of himself for the portfolio? (Yes or No)

c. The teachers realized that you had asked Stuart to adjust two sketches of Lucy for the art school application portfolio.
Was Stuart asked to adjust some sketches of Lucy for the portfolio? (Yes)

(3) a. The comedians realized that you had asked Carly to practice ten jokes about herself for the sold out stand-up show.
Was Carly asked to practice some jokes about herself for a stand-up show? (Yes)

b. The comedians realized that you had asked Carly to practice ten jokes about her for the sold out stand-up show.
Was Carly asked to practice some jokes about herself for a stand-up show? (Yes or No)

c. The comedians realized that you had asked Carly to practice ten jokes about Jared for the sold out stand-up show.
Was Carly asked to practice some jokes about Jared for a stand-up show? (Yes)

(4) a. The museum realized that you had asked Jasper to judge four portraits of himself for last week's grand opening exhibit.
Was Jasper asked to judge some portraits of himself for the exhibit? (Yes)

b. The museum realized that you had asked Jasper to judge four portraits of him for last week's grand opening exhibit.
Was Jasper asked to judge some portraits of himself for the exhibit? (Yes or No)

c. The museum realized that you had asked Jasper to judge four portraits of Celina for last week's grand opening exhibit.
Was Jasper asked to judge some portraits of Celina for the exhibit? (Yes)

(5) a. The hosts considered that you had asked Christina to copy three photographs of herself for the online promotional advertisement.
Was Christina asked to copy some photographs of herself for an advertisement? (Yes)

b. The hosts considered that you had asked Christina to copy three photographs of her for the online promotional advertisement.
Was Christina asked to copy some photographs of herself for an advertisement? (Yes or No)

c. The hosts considered that you had asked Christina to copy three photographs of Kevin for the online promotional advertisement.
Was Christina asked to copy some photographs of Kevin for an advertisement? (Yes)

(6) a. The publicists considered that you had asked Roger to disregard four posts about himself for the evening talk show.
Was Roger asked to disregard some posts about himself for the talk show? (Yes)

b. The publicists considered that you had asked Roger to disregard four posts about him for the evening talk show.
Was Roger asked to disregard some posts about himself for the talk show? (Yes or No)

c. The publicists considered that you had asked Roger to disregard four posts about Stephanie for the evening talk show.
Was Roger asked to disregard some posts about Stephanie for the talk show? (Yes)

(7) a. The reporters realized that you had asked Alexa to deny three lies about herself for yesterday's live television interview.
Was it the police who realized that you had asked Alexa to deny some lies? (No)

b. The reporters realized that you had asked Alexa to deny three lies about her for yesterday's live television interview.
Was it the police who realized that you had asked Alexa to deny some lies? (No)

c. The reporters realized that you had asked Alexa to deny three lies about Sean for yesterday's live television interview.
Was it the police who realized that you had asked Alexa to deny some lies? (No)

(8) a. The network realized that you had asked Betty to watch four documentaries about herself for the prime-time interview special.
Was it the network who realized that you had asked Betty to watch some documentaries? (Yes)

b. The network realized that you had asked Betty to watch four documentaries about her for the prime-time interview special.
Was it the network who realized that you had asked Betty to watch some documentaries? (Yes)

c. The network realized that you had asked Betty to watch four documentaries about Allen for the prime-time interview special.
Was it the network who realized that you had asked Betty to watch some documentaries? (Yes)

(9) a. The blogger-community realized that you had asked Derek to deny three rumors about himself for the social media campaign.
Was it the fans who realized that you had asked Derek to deny some rumors? (No)

b. The blogger-community realized that you had asked Derek to deny three rumors about him for the social media campaign.
Was it the fans who realized that you had asked Derek to deny some rumors? (No)

c. The blogger-community realized that you had asked Derek to deny three rumors about Sarah for the social media campaign.
Was it the fans who realized that you had asked Derek to deny some rumors? (No)

(10) a. The instructors realized that you had asked Heidi to organize three pictures of herself for the art class scrapbook.
Was it the instructors who realized that you had asked Heidi to organize some pictures? (Yes)

b. The instructors realized that you had asked Heidi to organize three pictures of her for the art class scrapbook.
Was it the instructors who realized that you had asked Heidi to organize some pictures? (Yes)

c. The instructors realized that you had asked Heidi to organize three pictures of Nicholas for the art class scrapbook.
Was it the instructors who realized that you had asked Heidi to organize some pictures? (Yes)

(11) a. The columnists considered that you had asked Ryan to print four images of himself for the recent issue of the hiking magazine.
Was it the readers who considered that you had asked Ryan to print some images? (No)

b. The columnists considered that you had asked Ryan to print four images of him for the recent issue of the hiking magazine.
Was it the readers who considered that you had asked Ryan to print some images? (No)

c. The columnists considered that you had asked Ryan to print four images of Diane for the recent issue of the hiking magazine.
Was it the readers who considered that you had asked Ryan to print some images? (No)

(12) a. The broadcasters considered that you had asked Robert to cut two videos of himself for the latest promotional advertisement.
Was it the broadcasters who considered that you had asked Robert to cut some videos? (Yes)

b. The broadcasters considered that you had asked Robert to cut two videos of him for the latest promotional advertisement.
Was it the broadcasters who considered that you had asked Robert to cut some videos? (Yes)

c. The broadcasters considered that you had asked Robert to cut two videos of Amanda for the latest promotional advertisement.
Was it the broadcasters who considered that you had asked Robert to cut some videos? (Yes)

(13) a. The parents realized that you had asked Joey to shorten two entries about himself for the drama club yearbook.
Was Joey asked to shorten some entries for the yearbook? (Yes)

b. The parents realized that you had asked Joey to shorten two entries about him for the drama club yearbook.
Was Joey asked to shorten some entries for the yearbook? (Yes)

c. The parents realized that you had asked Joey to shorten two entries about Susie for the drama club yearbook.
Was Joey asked to shorten some entries for the yearbook? (Yes)

(14) a. The students considered that you had asked Patrick to comment on three statues of himself for the art class project.
Was Patrick asked to comment on some websites for the project? (No)

b. The students considered that you had asked Patrick to comment on three statues of him for the art class project.
Was Patrick asked to comment on some websites for the project? (No)

c. The students considered that you had asked Patrick to comment on three statues of Sophia for

the art class project.
Was Patrick asked to comment on some websites for the project? (No)

(15) a. The viewers realized that you had asked Laura to discredit two rumors about herself for the network's latest reality show.
Was Laura asked to discredit some rumors for the reality show? (Yes)

b. The viewers realized that you had asked Laura to discredit two rumors about her for the network's latest reality show.
Was Laura asked to discredit some rumors for the reality show? (Yes)

c. The viewers realized that you had asked Laura to discredit two rumors about Tom for the network's latest reality show.
Was Laura asked to discredit some rumors for the reality show? (Yes)

(16) a. The producers realized that you had asked Larissa to proofread five stories about herself for the live radio interview.
Was Larissa asked to proofread some letters for an interview? (No)

b. The producers realized that you had asked Larissa to proofread five stories about her for the live radio interview.
Was Larissa asked to proofread some letters for an interview? (No)

c. The producers realized that you had asked Larissa to proofread five stories about Henry for the live radio interview.
Was Larissa asked to proofread some letters for an interview? (No)

(17) a. The lawyers realized that you had asked Mike to verify two biographies about himself for the special commemorative book.
Was Mike asked to verify some biographies for the book? (Yes)

b. The lawyers realized that you had asked Mike to verify two biographies about him for the special commemorative book.
Was Mike asked to verify some biographies for the book? (Yes)

c. The lawyers realized that you had asked Mike to verify two biographies about Hillary for the special commemorative book.
Was Mike asked to verify some biographies for the book? (Yes)

(18) a. The curators considered that you had asked Olivia to touch up four photographs of herself for the museum's new website.
Was Olivia asked to touch up some paintings for the website? (No)

b. The curators considered that you had asked Olivia to touch up four photographs of her for the

museum's new website.
Was Olivia asked to touch up some paintings for the website? (No)

c. The curators considered that you had asked Olivia to touch up four photographs of Chris for the museum's new website.
Was Olivia asked to touch up some paintings for the website? (No)

(19) a. The agents realized that you had asked Heather to perform three songs about herself for the recent North American tour.
Had Heather been asked to perform some songs for a fundraiser? (No)

b. The agents realized that you had asked Heather to perform three songs about her for the recent North American tour.
Had Heather been asked to perform some songs for a fundraiser? (No)

c. The agents realized that you had asked Heather to perform three songs about Bryan for the recent North American tour.
Had Heather been asked to perform some songs for a fundraiser? (No)

(20) a. The buyers realized that you had asked Brad to dig up three drawings of himself for last weekend's book release.
Had Brad been asked to dig up some drawings for a magazine? (No)

b. The buyers realized that you had asked Brad to dig up three drawings of him for last weekend's book release.
Had Brad been asked to dig up some drawings for a magazine? (No)

c. The buyers realized that you had asked Brad to dig up three drawings of Julia for last weekend's book release.
Had Brad been asked to dig up some drawings for a magazine? (No)

(21) a. The journalists realized that you had asked Ellen to verify three stories about herself for the latest live TV interview.
Had Ellen been asked to verify some stories for an article? (No)

b. The journalists realized that you had asked Ellen to verify three stories about her for the latest live TV interview.
Had Ellen been asked to verify some stories for an article? (No)

c. The journalists realized that you had asked Ellen to verify three stories about Bobby for the latest live TV interview.
Had Ellen been asked to verify some stories for an article? (No)

(22) a. The exhibitors realized that you had asked Bob to display three caricatures of himself for the

exhibit's artist showcase.
Had Bob been asked to display some caricatures for a showcase? (Yes)

b. The exhibitors realized that you had asked Bob to display three caricatures of him for the exhibit's artist showcase.
Had Bob been asked to display some caricatures for a showcase? (Yes)

c. The exhibitors realized that you had asked Bob to display three caricatures of Emily for the exhibit's artist showcase.
Had Bob been asked to display some caricatures for a showcase? (Yes)

(23) a. The bloggers considered that you had asked Chelsea to watch three videoclips of herself for the live television broadcast.
Had Chelsea been asked to watch some videoclips for the broadcast? (Yes)

b. The bloggers considered that you had asked Chelsea to watch three videoclips of her for the live television broadcast.
Had Chelsea been asked to watch some videoclips for the broadcast? (Yes)

c. The bloggers considered that you had asked Chelsea to watch three videoclips of Marc for the live television broadcast.
Had Chelsea been asked to watch some videoclips for the broadcast? (Yes)

(24) a. The roommates considered that you had asked Steven to photograph three figurines of himself for the mid-semester art project.
Had Steven been asked to photograph some figurines for a project? (Yes)

b. The roommates considered that you had asked Steven to photograph three figurines of him for the mid-semester art project.
Had Steven been asked to photograph some figurines for a project? (Yes)

c. The roommates considered that you had asked Steven to photograph three figurines of Kimberley for the mid-semester art project.
Had Steven been asked to photograph some figurines for a project? (Yes)

(25) a. The fans realized that you had asked Jimmy to retweet five blurbs about himself for yesterday's late night TV episode.
Was Jimmy asked to retweet some blurbs about Molly for a TV episode? (No)

b. The fans realized that you had asked Jimmy to retweet five blurbs about him for yesterday's late night TV episode.
Was Jimmy asked to retweet some blurbs about Molly for a TV episode? (No)

c. The fans realized that you had asked Jimmy to retweet five blurbs about Molly for yesterday's late night TV episode.

Was Jimmy asked to retweet some blurbs about himself for a TV episode? (No)

(26) a. The directors realized that you had asked Marisa to edit three segments about herself for the special evening broadcast.
Had Marisa been asked to edit some segments about Andrew for the broadcast? (No)

b. The directors realized that you had asked Marisa to edit three segments about her for the special evening broadcast.
Had Marisa been asked to edit segments about Andrew for the broadcast? (No)

c. The directors realized that you had asked Marisa to edit three segments about Andrew for the special evening broadcast.
Had Marisa been asked to edit segments about herself for the broadcast? (No)

(27) a. The organizers realized that you had asked Daniel to revise two websites about himself for the weekend art exhibition.
Had Daniel been asked to revise some website about Sabrina for the exhibition? (No)

b. The organizers realized that you had asked Daniel to revise two websites about him for the weekend art exhibition.
Had Daniel been asked to revise some website about Sabrina for the exhibition? (No)

c. The organizers realized that you had asked Daniel to revise two websites about Sabrina for the weekend art exhibition.
Had Daniel been asked to revise some website about himself for the exhibition? (No)

(28) a. The actors realized that you had asked Jennifer to endorse three exposés about herself for the latest movie promotion.
Was Jennifer asked to endorse some exposés about Liam for a movie promotion? (No)

b. The actors realized that you had asked Jennifer to endorse three exposés about her for the latest movie promotion.
Was Jennifer asked to endorse some exposés about Liam for a movie promotion? (No)

c. The actors realized that you had asked Jennifer to endorse three exposés about Liam for the latest movie promotion.
Was Jennifer asked to endorse some exposés about herself for a movie promotion? (No)

(29) a. The readers considered that you had asked Lance to discredit three articles about himself for the latest edition of the newspaper.
Was Lance asked to discredit some articles about Cindy for the magazine? (No)

b. The readers considered that you had asked Lance to discredit three articles about him for the latest edition of the newspaper.

Was Lance asked to discredit some articles about Cindy for the magazine? (No)

c. The readers considered that you had asked Lance to discredit three articles about Cindy for the latest edition of the newspaper.
Was Lance asked to discredit some articles about himself for the magazine? (No)

(30) a. The publishers considered that you had asked Audrey to edit four articles about herself for the collection of autobiographical stories.
Had Audrey been asked to edit some articles about Nathan for a collection? (No)

b. The publishers considered that you had asked Audrey to edit four articles about her for the collection of autobiographical stories.
Had Audrey been asked to edit some articles about Nathan for a collection? (No)

c. The publishers considered that you had asked Audrey to edit four articles about Nathan for the collection of autobiographical stories.
Had Audrey been asked to edit some articles about herself for a collection? (No)