

The role of structural and discourse-level cues during pronoun resolution

Clare Patterson

October 2013

Supervisor: Dr. Claudia Felser

Doctoral dissertation submitted to the Faculty of Human Sciences at the University of Potsdam in fulfillment of the requirements for the degree Doctor of Philosophy in Allgemeine Sprachwissenschaft

This work is licensed under a Creative Commons License:
Attribution 4.0 International
To view a copy of this license visit
<http://creativecommons.org/licenses/by/4.0/>

1. Gutachterin: PD Dr. Claudia Felser
2. Gutachter: Prof. Dr. Shravan Vasishth
Thesis Defence: Potsdam, 03 April 2014

Published online at the
Institutional Repository of the University of Potsdam:
URL <http://opus.kobv.de/ubp/volltexte/2014/7128/>
URN <urn:nbn:de:kobv:517-opus-71280>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-71280>

Abstract

Pronoun resolution normally takes place without conscious effort or awareness, yet the processes behind it are far from straightforward. A large number of cues and constraints have previously been recognised as playing a role in the identification and integration of potential antecedents, yet there is considerable debate over how these operate within the resolution process. The aim of this thesis is to investigate how the parser handles multiple antecedents in order to understand more about how certain information sources play a role during pronoun resolution. I consider how both structural information and information provided by the prior discourse is used during online processing. This is investigated through several eye tracking during reading experiments that are complemented by a number of offline questionnaire experiments.

I begin by considering how condition B of the Binding Theory (Chomsky 1981; 1986) has been captured in pronoun processing models; some researchers have claimed that processing is faithful to syntactic constraints from the beginning of the search (e.g. Nicol and Swinney 1989), while others have claimed that potential antecedents which are ruled out on structural grounds nonetheless affect processing, because the parser must also pay attention to a potential antecedent's features (e.g. Badecker and Straub 2002). My experimental findings demonstrate that the parser is sensitive to the subtle changes in syntactic configuration which either allow or disallow pronoun reference to a local antecedent, and indicate that the parser is normally faithful to condition B at all stages of processing. Secondly, I test the *Primitives of Binding* hypothesis proposed by Koornneef (2008) based on work by Reuland (2001), which is a modular approach to pronoun resolution in which variable binding (a semantic relationship between pronoun and antecedent) takes place before coreference. I demonstrate that a variable-binding (VB) antecedent is not systematically considered earlier than a coreference (CR) antecedent online. I then go on to explore whether these findings could be attributed to the linear order of the antecedents, and uncover a robust recency preference both online and

offline. I consider what role the factor of recency plays in pronoun resolution and how it can be reconciled with the first-mention advantage (Gernsbacher and Hargreaves 1988; Arnold 2001; Arnold *et al.*, 2007). Finally, I investigate how aspects of the prior discourse affect pronoun resolution. Prior discourse status clearly had an effect on pronoun resolution, but an antecedent's appearance in the previous context was not always facilitative; I propose that this is due to the number of topic switches that a reader must make, leading to a lack of discourse coherence which has a detrimental effect on pronoun resolution.

The sensitivity of the parser to structural cues does not entail that cue types can be easily separated into distinct sequential stages, and I therefore propose that the parser is structurally sensitive but not modular. Aspects of pronoun resolution can be captured within a parallel constraints model of pronoun resolution, however, such a model should be sensitive to the activation of potential antecedents based on discourse factors, and structural cues should be strongly weighted.

Acknowledgements

There are a number of people without whom I would not have been able to finish, or perhaps even start, the writing of this thesis. First and foremost, I would like to thank my supervisor, Claudia Felser, who has been unfailingly supportive, encouraging and generous with her time throughout this project. I have learnt a great deal from her, and she has always been ready to offer advice and insight on any aspect of my work. She has let me research independently, and at the same time has discreetly ensured that I stay on course, which something of a delicate balancing act! I am very grateful for her speed in commenting on draft sections of my thesis, as well as her patience in responding to my many mundane queries. I am also slightly in awe of her ability to spot a typo at twenty paces.

I have been part of two research groups during my doctoral studies, the Psycholinguistics Research Group at Essex and the PRIM group at Potsdam. I have learnt a great deal through presenting my own work and receiving feedback, as well as hearing about and commenting on the work of other members. I am grateful to Harald Clahsen, and again to Claudia, for leading these groups and encouraging debate during meetings.

Ian Cunnings introduced me to the eye-tracking method, and I have learnt an enormous amount from him. He has always been ready to give very detailed and clear advice on a range of technical issues, and is also a brilliant researcher. I have enjoyed working together with him on a number of experiments, some of which are included in this thesis, and I hope that our collaboration continues in the future.

I am extremely grateful for the financial support I have received, first from an ESRC postgraduate studentship from the Language and Linguistics Department at Essex, and then from a PhD studentship at PRIM (funded by an Alexander-von-Humboldt-Professorship to Harald Clahsen).

I am grateful, too, to friends and colleagues from Essex and Potsdam for advice, support, distraction and good company. Here I include a few words to them: To the ever cheerful Kaili, thanks for doing everything a year before me so I could ask you about it! To Elli, Vera and Oliver, I am glad we could make the move from Essex to Potsdam together. Elli and Vera, you also deserve special thanks, along with Helena, Sabrina and other PRIM colleagues, for helping me to settle in to life in Germany and for helping to improve my German (a work in progress). To Michaela, thanks for calmly taking care of so many administrative issues.

I would like to thank my parents, who have always been hugely supportive of everything I do. They have taken an active interest in my progress with this thesis, and they have also supplied enormous quantities of tea and homemade marmalade that have helped to fuel this project!

And a final word to my husband, Matt. Thank you for everything. You have never once doubted the choice I made to return to university, and I would not have taken that step without your encouragement. You have made many sacrifices, large and small, to ensure that I got to this stage, and I am very grateful.

Table of contents

Abstract	<i>iii</i>
Acknowledgements	<i>v</i>
Table of contents	<i>vii</i>
List of figures	<i>xv</i>
List of tables	<i>xvii</i>
Overview of experiments	<i>xxi</i>
1 Introduction	1
1.1 Pronouns and antecedents	1
1.2 Pronoun-antecedent relations	3
1.3 Pronoun processing: some assumptions	6
1.4 Aims and organisation of this thesis	7
1.5 Experimental method	8
1.5.1 Description of the eye tracking method	9
1.5.2 Eye-movement data	9
2 Condition B as a constraint on the processing of pronouns	13
2.1 Introduction	13
2.2 Condition B: theoretical background	14
2.2.1 Labels	14
2.2.2 Condition B defined	15

2.2.3	Condition B: a configurational constraint?	19
2.2.4	An exception to condition B: short-distance pronouns	22
2.3	Condition B in sentence processing	26
2.3.1	Condition B as a constraint on antecedent search	26
2.3.2	Evidence for parallel constraints	35
2.3.3	The processing of SDPs	45
2.3.4	Summary of proposed models and motivation for Experiments 1-3	48
2.4	Experiment 1: the offline interpretation of pronouns in canonical condition B and SDP environments	51
2.4.1	Experiment 1 materials	52
2.4.2	Experiment 1 participants	53
2.4.3	Experiment 1 procedure	54
2.4.4	Experiment 1 results	55
2.5	Experiment 2: processing pronouns in condition B environments	56
2.5.1	Experiment 2 materials	56
2.5.2	Experiment 2 predictions	58
2.5.3	Experiments 2 and 3 method	60
2.5.4	Experiment 2 results	63
2.6	Experiment 3: processing pronouns in SDP environments	73
2.6.1	Experiment 3 materials	73
2.6.2	Experiment 3 predictions	74
2.6.3	Experiment 3 method	75
2.6.4	Experiment 3 results	76
2.7	Discussion	85
2.7.1	Experiment 1: offline interpretations	85
2.7.2	Experiment 2: condition B environments	86
2.7.3	Experiment 3: SDPs	88
2.7.4	Timing	91

2.7.5 What do Experiments 2 and 3 tell us about pronoun processing models? 93

2.8 Summary 97

3 The role of variable binding and coreference in the search for an antecedent 99

3.1 Introduction 99

3.2 Theoretical background: variable binding and coreference 100

3.2.1 Origins of the POB model 100

3.2.2 Condition D 100

3.2.3 What is special about QPs? 104

3.2.4 Variable binding and coreference 105

3.2.5 The POB hypothesis 108

3.3 Evidence bearing on the POB claim 110

3.3.1 Evidence from ellipsis 110

3.3.2 Evidence from overt pronouns 112

3.4 Processing studies with QP antecedents 114

3.5 Experiment 4: investigating a VB-initial search mechanism 120

3.6 Experiment 5: further evidence against a VB-initial search mechanism 124

3.6.1 Experiment 5 materials 125

3.6.2 Experiment 5 predictions 127

3.6.3 Experiment 5 method 130

3.6.4 Experiment 5 results 133

3.7 Discussion 140

3.7.1 Summary of Experiment 5 results 140

3.7.2 Evaluation of the predictions for
Experiment 5 150

3.7.3 Can POB be rescued? 154

3.7.4 QPs as antecedents 155

3.8 Summary 156

4 The role of recency in pronoun resolution 157

4.1 Introduction 157

4.2 Previous literature on first-mention as a cue in
pronoun resolution 158

4.2.1 The structure-building approach 158

4.2.2 The expectancy hypothesis 162

4.2.3 First-mention versus subjecthood: evidence from Finnish 166

4.3 Recency in sentence processing research 171

4.4 Recency as a factor in the processing of pronouns
172

4.5 Experimental evidence for a recency preference in
pronoun resolution 179

4.6 Experiment 6: alternating the order of mention of DP
and QP antecedents 180

4.6.1 Experiment 6 materials 180

4.6.2 Experiment 6 participants 181

4.6.3 Experiment 6 procedure 181

4.6.4 Experiment 6 predictions 182

4.6.5 Experiment 6 results 182

4.7 Experiment 7: the interpretation of pronouns with
VB and CR antecedents 184

4.7.1 Experiment 7 materials 185

4.7.2 Experiment 7 participants and procedure 185

4.7.3	Experiment 7 results	186
4.8	Experiment 8: the effect of order of mention on the interpretation of pronouns with VB and CR antecedents	187
4.8.1	Experiment 8 materials	187
4.8.2	Experiment 8 participants	188
4.8.3	Experiment 8 procedure	188
4.8.4	Experiment 8 predictions	189
4.8.5	Experiment 8 results	189
4.9	Experiment 9: searching for an online recency preference	191
4.9.1	Experiment 9 materials	191
4.9.2	Experiment 9 predictions	194
4.9.3	Experiment 9 method	194
4.9.4	Experiment 9 results	197
4.10	Discussion	202
4.10.1	Summary of findings for the offline experiments	203
4.10.2	Summary of findings for the online results	204
4.10.3	Comparing results from Experiments 4 and 5 to Experiments 6-9	205
4.10.4	A memory-driven recency effect?	209
4.10.5	Recency versus first-mention	211
4.11	Summary and conclusions	214
5	Looking beyond the sentence: the role of prior discourse in pronoun resolution	215
5.1	Introduction	215
5.1.1	Discourse prominence	216
5.2	Topichood in pronoun resolution	218

5.3	Coherence in pronoun resolution	222
5.3.1	The coherence analysis	222
5.3.2	Centering Theory: an outline	224
5.3.3	The testing of CT assumptions in relation to pronoun processing	226
5.4	Experiment 5: CT and the influence of the prior discourse	232
5.4.1	Experiment 5 materials	234
5.4.2	Applying Centering Theory assumptions to the materials for Experiment 5	236
5.4.3	Topichood in the materials for Experiment 5	239
5.4.4	Experiment 5 predictions	240
5.4.5	Experiment 5 results summary	241
5.5	Discussion	249
5.5.1	Assessing antecedent type and the influence of the repeated-name penalty	250
5.5.2	The cost of a topic shift and textual coherence: effects on pronoun resolution	251
5.6	Conclusion	254
6	Conclusions	255
6.1	Summary of experimental findings	255
6.2	A structurally sensitive, non-modular, search mechanism	257
6.3	QPs as antecedents	259
6.3.1	Avoidance of QP antecedents	259
6.3.2	The involvement of discourse in binding relations with QPs	259
6.4	Recency as a cue in pronoun processing	261
6.5	Prior discourse and coherence	261

6.6 Conclusion	262
References	263
Appendix 1	273
Appendix 2	277
Appendix 3	283
Appendix 4	287
Appendix 5	291
Appendix 6	293
Erklärung	297

List of figures

Figure 3.1. Simplified syntactic tree diagram for (3.1), illustrating that the QP c-commands the pronoun. **102**

Figure 3.2. Variable binding: three cross-modular steps. From Koornneef (2008, p. 44). **107**

Figure 3.3. Mean regression-path times (ms) per condition in the precritical region for Experiment 5. **147**

Figure 3.4. Mean regression-path times (ms) per condition in the pronoun region for Experiment 5. **147**

Figure 3.5. Mean total viewing times (ms) in the spillover region, per condition for Experiment 5. **149**

Figure 5.1. Mean regression-path times (ms) per condition in the precritical region for Experiment 5. **246**

Figure 5.2. Mean regression-path times (ms) per condition in the pronoun region for Experiment 5. **247**

List of tables

Table 2.1. Means (ms) for first-fixation durations, first-pass times and regression-path times in each condition in the precritical, critical, spillover, prefinal and final regions in Experiment 2. **65**

Table 2.2. Summary of statistical analyses for first-fixation durations, first-pass times and regression-path times in the precritical, pronoun, spillover, prefinal and final regions of Experiment 2. **66**

Table 2.3. Means (ms) for rereading times and total viewing times in each condition in the precritical, critical, spillover, prefinal and final regions of Experiment 2. **69**

Table 2.4. Summary of statistical analyses for rereading times and total viewing times in the precritical, pronoun, spillover, prefinal and final regions of Experiment 2. **70**

Table 2.5. Means (ms) for first-fixation durations, first-pass times and regression-path times in each condition in the precritical, critical, spillover, prefinal and final regions in Experiment 3. **77**

Table 2.6. Summary of statistical analyses for earlier eye-movement measures in the precritical, pronoun, spillover, prefinal and final regions of Experiment 3. **78**

Table 2.7. Means (ms) for rereading and total viewing times in each condition in the precritical, pronoun, spillover, prefinal and final regions in Experiment 3. **80**

Table 2.8. Summary of statistical analyses for later eye-movement measures in the precritical, critical, spillover, prefinal and final regions of Experiment 3. **81**

Table 2.9. Means (ms) for rereading times and total viewing times, and mean number of regressions-in, for each condition in the local name region in Experiment 3. **84**

Table 2.10. Summary of statistical analyses for rereading times, total viewing times and regressions-in in the local name region of Experiment 3. **84**

Table 3.1. Means in all conditions (in ms) for first fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions for Experiment 5. **135**

Table 3.2. Model outcomes for each region and measure, Experiment 5. **137**

Table 3.3. Means for all conditions (in ms) for first fixation durations, first-pass times, and total viewing times in the QP and DP regions in Experiment 5. **143**

Table 3.4. Outcome of the modelling process for QP and DP regions, Experiment 5. **144**

Table 4.1. Frequency of responses and percentages for Experiment 6. **183**

Table 4.2. Outcome of the Wilcoxon signed rank tests for Experiment 6. **184**

Table 4.3. Frequency of responses and percentages, Experiment 7. **186**

Table 4.4. Mean response scores in each condition for Experiment 8. **190**

Table 4.5. Means in all conditions (in ms) for first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions for Experiment 9. **198**

Table 4.6. Summary of statistical analysis for each region and measure, Experiment 9. **199**

Table 5.1. Means in all conditions (in ms) for first fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions for Experiment 5. **243**

Table 5.2. Model outcomes for each region and measure, Experiment 5. **243**

Overview of the experiments reported in this thesis

Experiment 1

Method: offline questionnaire

Reported in: Chapter 2

Theme: offline antecedent choices in condition B and SDP environments

Example (Condition B):

The boy remembered that Matthew had bought him a new computer game.

Example (SDP):

At the post office, Ernie watched the boy pick up the parcel next to him.

Experiment 2

Method: eye tracking

Reported in: Chapter 2

Theme: pronoun processing in condition B environments

Example:

{Jane/John} remembered that {Mark/Jane/John} had taught him a new song on the guitar.

Experiment 3

Method: eye tracking

Reported in: Chapter 2

Theme: pronoun processing in SDP environments

Example:

{Barry/Megan} saw {Gavin/Megan/Garry} place a gun near him on the ground with great care.

Experiment 4 (conducted by Dr Ian Cummings)

Method: eye tracking

Reported in: Chapter 3

Theme: pronoun processing with VB and CR antecedents

Example:

Every soldier who knew that {James/Helen} was watching was convinced that {he/she} should wave as the parade passed.

Experiment 5

Method: eye tracking

Reported in: Chapter 3; Chapter 5

Theme: pronoun processing with VB and CR antecedents; the effect of prior discourse

Example:

{The queen/The soldiers} felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that {he/she} should wave as the parade passed. The entire town had come out to watch.

Experiment 6

Method: offline questionnaire

Reported in: Chapter 4

Theme: The effect of antecedent order on antecedent choices

Example:

{Every cheerleader/The showgirl} sensed that {the showgirl/every cheerleader} thought that she should follow the team off the field.

Experiment 7

Method: offline questionnaire

Reported in: Chapter 4 (method details in Chapter 2)

Theme: The effect of antecedent order on antecedent choices

Example (QP first):

Every pilot who knew that Peter was working at the control tower today hoped that he would arrive on time.

Example (Name first):

It appeared to Paul that every boxer in the changing room thought that he should fight a lot harder.

Experiment 8

Method: offline questionnaire

Reported in: Chapter 4

Theme: The effect of antecedent order on antecedent choices

Example (QP first):

Every soldier who knew that James was watching was convinced that he should wave as the parade passed.

Example (Name first):

It looked to James that every soldier was completely convinced that he should wave as the parade passed.

Experiment 9

Method: eye tracking

Reported in: Chapter 4

Theme: pronoun processing with VB and CR antecedents, antecedent order reversed

Example:

It looked to {James/Helen} that every soldier was completely convinced that {he/she} should wave as the parade passed.

CHAPTER 1

Introduction

1.1 Pronouns and antecedents

A crucial part of language comprehension is the successful resolution of pronouns. Pronoun resolution normally takes place without conscious effort or awareness, yet the processes behind it are far from straightforward. This is because a pronoun, on its own, carries minimal information and does not have a fixed interpretation; pronouns are understood in relation to other elements of a text or discourse. For example, in (1.1) the pronoun *she* is understood in relation to *Marion Bartoli*:

(1.1) Marion Bartoli became the Wimbledon champion when she was 28 years old.

Pronouns, in having to rely on other expressions for their interpretation, belong to a linguistic class called *anaphors*.¹ Anaphors by their very nature require an *antecedent*, that is, another expression that allows the anaphor to be interpreted. *Marion Bartoli* fulfils this role in (1.1).

The relationship between a pronoun and an antecedent is of course not restricted to single sentences. An antecedent can occur in the same sentence, the previous sentence or even further back in the text or discourse. Some pronouns do not have an overt antecedent in the text or discourse at all. This tends to occur when the pronoun refers to something or someone that can easily be understood from the context. The antecedents for first- and second-person pronouns (*I, you, we*, etc.) in fact are rarely overt. Third-person pronouns more often have antecedents that

¹ Note that I use the terms anaphor and anaphoric here in the general sense of expressions that rely on other expressions for their interpretation, and not in the sense that the term anaphor is used in Binding Theory (Chomsky 1981; 1986). For more on Binding Theory, specifically condition B, see Chapter 2.

are previously mentioned, but it is far from unusual to find examples where an antecedent is not overt. For example, if two people are looking at the last slice of cake, one might say to the other, “Are you going to eat it, or shall I?”. It is easily assumed that *it* refers to the cake, even though the cake has not overtly been mentioned.

While it might seem as though pronouns with overt antecedents should be easier to resolve than those without, in fact the opposite is probably true. It is rarely the case that the antecedent of first- and second-person pronouns is misunderstood, for example. But third-person pronouns in a text can sometimes be harder to resolve because it is not immediately obvious who or what the antecedent is, particularly when there is more than one possibility. If a sentence like (1.1) were to be preceded by information about another female, as shown in (1.2) below, the referent for *she* in the final sentence becomes somewhat harder to determine.

(1.2) Alice was a huge tennis fan and in particular she loved Marion Bartoli. From the age of 16 Alice followed Bartoli’s career, and when she was 22 she even got tickets to Bartoli’s Wimbledon match against Venus Williams. Bartoli became the Wimbledon champion when **she** was 28 years old.

While pronouns are dependent on antecedents for interpretation, they do carry certain types of morphological information. In English, personal pronouns can carry gender information (*she* versus *he*), number information (*she* versus *they*), and information about their grammatical role (*he* is a subject pronoun, whereas *him* is an object pronoun). This information is often referred to as a pronoun’s *features*. A pronoun’s features play a role in the resolution process because they determine some of the features that the antecedent should carry.² But as (1.2) demonstrates, features such as

² Note that (in English) the gender and number features of a pronoun and an antecedent should match, but the grammatical role of the pronoun does not determine the grammatical role of the antecedent.

number and gender do not always narrow down the list of possible antecedents to a single one. Other sources of information also play a role.

It should be noted that, while the pronoun appears to be ambiguous in (1.2) and unambiguous in (1.1), this is not actually correct. Pronouns are always technically ambiguous. (1.1) would not be an incorrect sentence if it turned out that the antecedent for *she* was not *Marion Bartoli*. It might be considered unusual if no other obvious antecedent was available in the context, but it would not be a grammatical error. Despite the inherent ambiguity in all pronouns, in this thesis I occasionally make use of the terms unambiguous and ambiguous to distinguish pronouns that have a single obvious referent from those that do not.

In this thesis I focus on third-person singular pronouns in English. I do not include the pronoun *it* because it has such a wide range of possible referents and is quite distinct from pronouns like *he* and *she*, which are used in English to refer (almost exclusively) to humans. In particular I am interested in pronouns with more than one potential antecedent, as demonstrated in (1.2) above. Before I present the aims of my research, I outline basic concepts relating to pronoun-antecedent relations and the processing of pronouns.

1.2 Pronoun-antecedent relations

The relationship between a pronoun and its antecedent has been approached in several different ways in different branches of linguistics. Syntactic approaches to pronoun-antecedent relations have mainly been concerned with dependencies at the level of the sentence. One significant area of research has been the distribution of pronouns in relation to other anaphors and the effect of this on pronoun-antecedent relations. For example, the pronoun in (1.3) is allowed to refer to *John* but not to *Bill*. If the pronoun is replaced with the reflexive *himself* as in (1.4), the opposite is true: the reflexive refers to *Bill* but not to *John*.

(1.3) John hoped that Bill would buy him a ticket to the opera.

(1.4) John hoped that Bill would buy himself a ticket to the opera.

Furthermore, a slight change to the sentence structure alters the restrictions on reference, as (1.5) demonstrates:

(1.5) John hoped that Bill's aunt would buy him a ticket to the opera.

In (1.5), unlike (1.3), both *Bill* and *John* are potential antecedents for *him*. Such restrictions on pronoun reference are clearly related to syntactic structure, and specifying the nature of these restrictions has been one of the main aspects of syntactic research on pronouns. In particular, these restrictions have been captured in Binding Theory (Chomsky 1981; 1986). Pronouns with antecedents outside the sentence have not received the same degree of attention in syntax because they are not subject to the same type of restrictions.

Semantic approaches have concentrated less on the restrictions on reference at the level of the sentence, and more on characterising the precise nature of the relationship between pronouns and antecedents. Traditionally, it was assumed that pronouns and other anaphors took their meaning directly from the antecedent expression in a sentence or preceding text. However, this direct relationship has been questioned and instead the importance of *coreference* has been emphasised. Coreference is what happens when two expressions refer to the same individual. So in (1.1), the expression *Marion Bartoli* refers to an individual. When the pronoun *she* in (1.1) also refers to the same individual, *she* and the expression *Marion Bartoli* are said to *corefer*.³ This implies that at least some pronoun-antecedent relations take place at a level of representation outside of the syntactic representation (see Burkhardt 2005 for a discussion). This is a very common assumption in

³ There are some cases where coreference is not possible. This can happen, for instance, when the antecedent is not a referring expression. The relationship between a pronoun and its antecedent in these cases is described differently; more details can be found in Chapter 3.

much of the pronoun processing literature, particularly in studies concerned with discourse factors in pronoun resolution.

This level of representation is often called the discourse representation (Kamp 1981; Heim 1982) or the mental model (Johnson-Laird 1983; see also Garnham 2001). It is assumed that during text comprehension, the reader constructs a mental representation of the situation that is being described.⁴ Important components of the discourse representation are events and entities. Events encompass the situations that a text is about, and verbs play an important part in their establishment in the model. Entities are the people or objects that are involved in the events. The situations in the mental representation can be representations of the real world or, equally, an imagined world. They are built up over the course of a text/discourse almost like constructing a more and more elaborately detailed scene in a play, and are not limited to single sentences. These assumptions underpin some semantic and many discourse-related approaches to pronoun resolution.

The different approaches have led to a variety of terminology for describing pronoun-antecedent relations. It is common to say that a pronoun *refers to* an antecedent (also known as a *referent*), even if the relationship is understood as coreference. *Binding* is a notion that is more closely associated with syntactic approaches to pronouns, and can express a very specific type of relation in certain theories. In this thesis I use the terms *antecedent* and *referent* interchangeably, and the relationship between pronoun and antecedent is expressed as linking or referring.⁵ In order to prevent confusion I avoid the terms *binding* and *coreference* except when discussing theories in which these terms have a specific meaning.

⁴ I use the term reader here for convenience, but I refer to the task of language comprehension in any modality.

⁵ I do not use the term linking in the technical sense established by Higginbotham (1983; 1985), but informally, without commitment to a particular theoretical approach.

1.3 Pronoun processing: some assumptions

The way in which pronoun resolution unfolds over time is not fully understood, but research in this area has uncovered certain aspects of the process. Processing begins when, or shortly after, the pronoun is encountered. Encountering a pronoun triggers the search for a suitable antecedent. Like other aspects of sentence processing, pronoun processing is not delayed until the end of a clause or sentence, but takes place incrementally. The search and integration involved in pronoun processing has been described as a two-stage process called bonding and resolution (Garrod and Terras 2000). This was originally proposed for the processing of single antecedents, but the terms have also been used to describe the formation of a candidate set of several antecedents (e.g. Sturt 2003). The bonding stage is triggered when an anaphor is encountered, and potential antecedents are activated. In the resolution phase, the link between the anaphor and the candidate antecedents is evaluated and integrated into the interpretation of the sentence. As such, pronoun processing can be a fairly lengthy process, continuing while subsequent words are read or heard, even though it can begin rapidly upon encountering the pronoun.

It has proved difficult to cleanly separate the effects of this neat division of labour into two separate processes either in terms of their behavioural indexes such as early and late eye-movement measures (see below) or in the information sources that are used at each stage (Pickering *et al.* 2004; Callahan 2008 for ERP data). Nevertheless for eye tracking during reading it has been claimed that bonding processes involve lexical and possibly structural information, and are reflected in earlier measures such as first-fixation durations and first-pass times, while resolution processes involving discourse information are reflected in later measures (Sturt 2003; Foraker and McElree 2007).

There is considerable debate over the timecourse of the anaphoric resolution process with relation to structural information. Some researchers have claimed that processing is faithful to structural restrictions such as those captured in Binding Theory (see above) from the beginning of the search (e.g. Nicol and Swinney 1989). Others have claimed that potential

antecedents which are ruled out on structural grounds nonetheless affect processing, because the parser must also pay attention to a potential antecedent's features such as gender and number (e.g. Badecker and Straub 2002). An additional consideration is that from the perspective of the discourse, pronoun processing seems to involve not just a backward search initiated at the pronoun itself, but a certain amount of prediction based on the building up of discourse expectations (e.g. Rohde, Kehler and Elman 2007; Kehler *et al.* 2008).

Thus, there is a lack of clarity about the precise timecourse of pronoun resolution and how it will emerge in behavioural data. What is clear, though, is that pronoun resolution exploits a number of different sources of information from the syntactic structure, morphological features, and discourse cues.

1.4 Aims and organisation of this thesis

In this thesis I aim to investigate the processing of pronouns with multiple antecedents, in order to understand more about how certain information sources play a role during resolution. I consider how both structural information and information provided by the prior discourse is used in online processing. I report a number of eye tracking during reading experiments, a methodology that is suited to investigating the often lengthy process of pronoun resolution. The online studies are complemented by a number of offline questionnaire studies which check eventual interpretation preferences. I explore the following questions:

- Is the pronoun resolution process faithful to syntactic constraints at all stages, and what are the implications for the claims of certain processing models?
- Is pronoun resolution modular, in that different types of pronoun-antecedent relations are computed sequentially?
- How does an antecedent's position in relation to a pronoun affect resolution?
- How do aspects of the prior discourse affect pronoun resolution?

These questions are explored in four main chapters. Chapter 2 is concerned with the syntactic constraint captured in condition B of the Binding Theory (Chomsky 1981; 1986), and considers how this has been captured in pronoun processing models. Chapter 3 tests the *Primitives of Binding* hypothesis proposed by Koornneef (2008) based on work by Reuland (2001), suggesting a modular approach to pronoun resolution in which variable binding (a semantic relationship between pronoun and antecedent) takes place before coreference. Chapter 4 considers the factor of recency in pronoun resolution and how it has an effect both online and offline. Chapter 5 considers the issue of discourse prominence, and whether prior prominence affects later pronoun resolution. Chapter 6 concludes.

1.5 Experimental method

In this thesis I have chosen to use the eye tracking during reading method for my online experiments. This method has been used to study various aspects of language processing, including pronouns. Rather than giving a snapshot of processing at one or two moments, eye tracking during reading gives a detailed record of processing as it unfolds over time. This makes it particularly suited to the investigation of pronoun resolution which, as noted above, can be a lengthy process. Another advantage of the eye tracking during reading method (henceforth eye tracking) is that the participants are asked only to read texts and respond to comprehension questions.⁶ In comparison to some other methods, this is a relatively natural task, so it encourages more normal processing and is less susceptible to strategic patterns of responses from participants.

⁶ The term eye tracking can also refer to the eye tracking during listening method, also known as visual world eye tracking. In this thesis I use the term *eye tracking* to refer to eye tracking during reading, and *visual world* to refer to eye tracking during listening. A description of the visual world method can be found in Chapter 2, section 2.3.2.

1.5.1 Description of the eye tracking method

In an eye tracking study, participants are seated in front of a computer screen. Short texts (which can be composed of several sentences) or single sentences are presented on the screen one at a time, and participants are asked to read them at their normal reading rate. Normally participants are asked to press a button when they have finished each text/sentence to bring up the next item on the screen. A camera tracks the movements of the eyes as they move over the text. It is common in these experiments to include a number of comprehension questions following some or all of the items, which participants respond to using a response box, mouse or keyboard. This is to ensure that participants are reading for comprehension and not simply moving their eyes over the screen. A full description of the experimental setup and procedure is given in Chapter 2, Experiment 2 (section 2.5.3.2). Rather than repeat the details of the experimental setup and procedure in each chapter, reports of subsequent experiments refer to the description in Chapter 2 throughout, and only details specific to each experiment are reported.

1.5.2 Eye-movement data

The eye movements that occur during reading are not a single smooth sweep through the text from left to right. The movements can instead be characterised by fixations and saccades (see Rayner 1998 for an overview). A fixation is when the eye rests in a particular place. A saccade is a movement of the eye between fixations. The movement is not only forwards (left to right for reading a line of text in English) but can also be backwards as some parts of a text are refixated, or fixated for the first time out of a strict left to right sequence. Not all words in a text receive fixations. Some words, especially (but not exclusively) short function words, are viewed parafoveally. This means that, during a fixation in a particular position, a number of characters can be simultaneously viewed on either the left or the right of the fixation position. This can lead to processing of upcoming material in advance of the next fixation.

The eye-movement data recorded during an experiment is very detailed and a number of different eye-movement measures can be calculated.

Before the measures are calculated for a particular experiment, the texts are divided into a number of *regions of interest* for the purpose of analysis, and then the eye-movement measures are associated with each region. Sometimes whole regions are initially skipped (not fixated during the first pass through a text left to right). Initially skipped regions are generally removed from data analysis. The eye-movement measures reported in this thesis are as follows. *First-fixation duration* is the length of the first fixation in a particular region. *First-pass time* is the sum of the fixation durations in a region before it is exited for the first time. The region can be exited to the right (towards new material) or the left (backwards to inspect previous material). *Regression-path time* is the sum of the fixation durations in a region before it is exited to the right, towards new material. *Rereading time* is the sum of durations in a region after it has been re-entered. *Total viewing time* is the sum of all fixation durations in a region. *Regressions-in* counts the number of regressive fixations into a region of interest. A regressive fixation is a fixation that occurs in a region after subsequent regions have been fixated.

In order to interpret eye-tracking data of the kind outlined above, it is necessary to make several assumptions. The most fundamental assumption, usually referred to as the eye-mind assumption, is that readers fixate the word they are processing (Just and Carpenter 1980). However, studies that have looked at the processing effects of sentence- or discourse-level phenomena have found that effects occur not just on the critical word but also subsequent words, which calls into question the assumption that readers do not move on from a word until processing is complete (see, e.g. Pickering *et al.* 2004 for a discussion). Thus, in eye tracking studies on pronouns it is common to look not just at the pronoun region but also regions beyond the pronoun to assess whether effects take place.

The interpretation of data from the different eye-movement measures also relies on an assumption of the cognitive processes that they reflect. Several sophisticated models of eye-movement control and behavioural processes have been put forward; these models are mainly concerned with lexical

access and the processing cost of individual words, and effects arising from syntactic processing difficulty are less considered (as pointed out by Boston *et al.* 2008). Evidence from Boston *et al.* (2008) also calls into question assumptions about linking cognitive events to specific reading-time measures. As noted above in relation to bonding and resolution, it has been claimed that earlier measures such as first-fixation durations and first-pass times are related to lexical processes (such as accessing a word in the mental lexicon). Integration is supposed to be reflected in more cumulative measures and second-pass times such as regression-path times, rereading and total viewing times. However, it is not clear that this separation is entirely trustworthy. Although I do discuss the issue of timing in relation to my own data in this thesis, I do so with caution in view of the uncertainty surrounding the interpretation of such evidence. What I do rely on is a more straightforward assumption, common to previous studies on pronoun processing as well as other phenomena, that increased processing difficulty is reflected in increased reading times, which may be visible in early or late measures.

CHAPTER 2

Condition B as a constraint on the processing of pronouns

2.1 Introduction

Successful pronoun resolution involves the identification of a suitable antecedent from the surrounding discourse, and it is often the case that more than one potential antecedent is available. However, there are strong restrictions on the selection of antecedents that occur within the same sentence as a pronoun. One of these restrictions is known as condition B of the Binding Theory (Chomsky 1981; 1986). This constraint prevents a pronoun from referring to an antecedent in its local domain. For example, the pronoun *her* in (2.1) cannot refer to *Heather*:

(2.1) Mrs Smith stifled a yawn as Heather showed her some more holiday photos.

Condition B makes certain antecedents inaccessible to the pronoun, which means that the pronoun cannot refer to them. In (2.1), *Heather* is an inaccessible antecedent, as condition B predicts, because *Heather* c-commands the pronoun and appears in the pronoun's local domain (see below for more detail about c-command and the local domain).

While condition B is a good generalisation about the accessibility and inaccessibility of certain antecedents, in certain syntactic structures pronouns appear to be able to circumvent this constraint and refer to a local antecedent. For example, in (2.2) *her* can (optionally) refer to *Heather* even though *Heather* is in the local domain of the pronoun. Pronouns in this type of structure are often referred to as short-distance pronouns (SDPs).

(2.2) Heather reached behind her to retrieve the last set of holiday photos.

One significant challenge for syntactic theory has been to specify principles which disallow *Heather* as an antecedent in (2.1) but not in (2.2).

This chapter considers pronouns that appear in canonical condition B structures, as in (2.1), and SDPs, as in (2.2), in an attempt to understand more about how condition B constrains the accessibility of antecedents during processing. The question has in fact long been a concern in psycholinguistics. A key research issue has been to find out at what point an inaccessible antecedent is excluded from the set of possible antecedents. So far, there is no consensus on this. While some researchers assert that inaccessible antecedents are never considered, giving a primary role to condition B in determining the set of candidate antecedents, others propose that features (e.g. feminine, singular) of both accessible and inaccessible antecedents affect pronoun processing, at least temporarily. I present two eye-tracking experiments testing pronouns in canonical condition B structures and SDPs, and a complementary offline study. My own evidence shows that inaccessible antecedents do not affect processing. However, based on certain theoretical approaches to condition B and consideration of timing issues, I question the details of a processing model in which pragmatic and discourse information is secondary to the application of condition B. I also find that there is a weak tendency for a local antecedent to be favoured during processing, despite such a preference not being apparent in offline judgements.

2.2 Condition B: theoretical background

This section begins by introducing the theoretical background of condition B. The structural nature of the constraint is considered in the light of alternative approaches, in particular highlighting those approaches which incorporate a pragmatic element. Theoretical accounts for SDPs, which appear to be exempt from condition B, are also introduced.

2.2.1 Labels

Condition B has been discussed at great length in the syntax literature and other areas of linguistics since it was first proposed. This has given rise to a

number of different labels for the key concepts. To avoid confusion, I will clarify here the labels that I adopt throughout this chapter. Firstly, I will use the label Binding Theory (BT) to refer to Chomsky's (1981) formulation. This is sufficient to cover the canonical condition B environments and exceptions discussed in this chapter. The reformulation of BT in Chomsky's (1986) work *Knowledge of Language* will be referred to specifically when necessary. I adopt the commonly used term *condition* to refer to the binding conditions, rather than *principle*, but the terms are interchangeable. In keeping with other parts of this thesis, I use the term *reflexive* for a reflexive pronoun: this is called an anaphor in BT. I use the term *pronoun* for a non-reflexive pronoun; this is called a pronominal in BT. I adopt the terms *accessible* and *inaccessible* in describing antecedents that are either allowed or disallowed under the BT conditions. I use the term *governing category* when discussing its BT origins and *binding domain* elsewhere. Finally, this section (2.2) discusses various approaches to defining condition B. For the remainder of the chapter, except where explicitly mentioned in connection with Chomsky's BT definitions, I use the term *condition B* as shorthand for the locality restriction on pronoun binding without assuming a particular theoretical approach.

2.2.2 Condition B defined

BT (Chomsky 1981) is an attempt to capture the distribution of pronouns and reflexives in a systematic way. In doing so, it creates rules for determining whether reference to a particular antecedent is allowed (an accessible antecedent) or disallowed (an inaccessible antecedent). In BT, noun types are divided into three groups, namely reflexives, pronouns and referring expressions (R-expressions), and each group has its own condition. Condition A relates to reflexives, condition B to pronouns and condition C to referring expressions such as names. Condition B is of primary interest in the current chapter, and relatedly condition A because pronouns and reflexives are assumed to be in complementary distribution under BT (see below). A full discussion of condition C is outside the scope of this thesis, but all three conditions are given below for completeness. The conditions of BT are as follows:

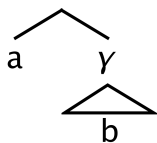
- “(A) An anaphor is bound in its governing category
- (B) A pronominal is free in its governing category
- (C) An R-expression is free.”

(Chomsky 1981, p. 188)

The BT conditions describe the circumstances under which the three noun types can be bound. Binding relates to the link between reflexives or pronouns and their antecedents, but it is more than simply establishing that an antecedent exists. It is a syntactic relationship in which the pronoun or reflexive gets its meaning directly from its antecedent in the same sentence. Binding can only be established when there is a particular syntactic configuration involving the pronoun/reflexive and its antecedent. This configuration is called c-command. The c-command relationship is asymmetrical, in that c-commanding element is higher in the syntactic structure than the c-commanded element. Reuland (2011) describes c-command as follows:

“a c-commands b iff a is a sister to a category γ containing b.

Schematically: [a [γ ...b...]], or



According to this definition and common syntactic assumptions, the subject of a clause will c-command the object, but not vice versa.”

(Reuland 2011, p. 29)

In the diagram above, *a* and γ are sisters because they are both dominated by the same branching node. Anything contained in γ , including γ itself, is

c-commanded by *a*. Likewise, γ c-commands *a*, but the elements within γ do not c-command *a* because they are too low down in the tree.

C-command is a necessary requirement of binding because there can only be a binding relationship if the antecedent c-commands the pronoun/reflexive, and if the two elements are co-indexed, that is, they have the same referent. When there is binding, the c-commanded expression takes its meaning from the c-commanding expression. Under BT, reflexives must be bound, but pronouns are optionally bound and must not be bound inside their governing category. If there is an antecedent in the sentence which c-commands the pronoun and is not within the governing category, binding can take place.

The definition of governing category (later known as the binding domain) has in fact been the subject of much scrutiny over the years, and has changed even within Chomsky's own work (Chomsky 1981; cf Chomsky 1986). The following definition, based on the 1981 work, is sufficient to capture the core cases discussed in this section (but see section 2.2.4 for a discussion of SDPs, where this definition of governing category becomes problematic): the governing category is a noun phrase (NP) or sentence (S) which contains both a governor and the pronoun/reflexive. A governor is an element such as a verb that can assign case. With governing category described in these terms, conditions A and B succinctly describe the distribution of reflexives and pronouns and determine that pronouns and reflexives are in complementary distribution. Consider (2.3):

(2.3a) Ken_i thinks that Boris_j likes himself_{i*/j}.

(2.3b) Ken_i thinks that Boris_j likes him_{i/j*}.

In (2.3) the complementizer *that* heads an S which contains the verb *like* and the reflexive/pronoun, thus constituting a governing category. (2.3a) contains a reflexive which is subject to condition A. *Himself* is bound by

Boris but cannot be bound by *Ken*, as the indexing indicates.⁷ This is because *Boris* is in the governing category of *himself*; condition A stipulates that the reflexive must be bound inside the governing category, so this condition is satisfied when the reflexive is bound by *Boris*. *Ken* is outside of the governing category and is therefore inaccessible. If *himself* were bound by *Ken* it would be in violation of condition A. This gives rise to the correct interpretation of (2.3a) that the object of *Boris*' liking is *Boris* himself, not *Ken* or anybody else. (2.3b) contains the pronoun *him* so condition B applies. This makes *Boris* an inaccessible antecedent, as indicated by the indexing. *Him* can refer to *Ken* in (2.3b), since *Ken* is outside the governing category. It is also possible for *him* to refer to a different antecedent outside the sentence; condition B does not stipulate a particular antecedent. Simply put, a reflexive must have an antecedent in its binding domain and a pronoun must not.

BT is an attractive proposal because it neatly describes the referential properties of reflexives and pronouns in the majority of cases, and does so in purely configurational terms. At the same time it makes the assumption that pronouns and reflexives are in complementary distribution. This is because condition A stipulates binding of the reflexive by the exact antecedent that is not allowed to bind the pronoun in condition B. It should therefore be impossible to find syntactic environments where a reflexive or a pronoun (in the same position) can refer to the same antecedent. But exceptions do exist, and have posed a challenge to the original assumptions of BT. One such exception is pronouns appearing in certain prepositional phrases (PPs), also known as SDPs. Before discussing this particular exception to condition B, however, I will present a brief overview of challenges to the notion that condition B can be framed in purely configurational terms.

⁷ The asterisks indicate a disallowed referential dependency.

2.2.3 Condition B: a configurational constraint?

At the time that condition B was proposed, it was assumed to be a configurational restriction on pronouns; that is, its definition relied on structural notions such as the governing category, and on where the pronoun and potential antecedents are placed in the syntactic structure. However, the notion that ruling out an inaccessible antecedent can be achieved entirely with structural apparatus has been challenged in several different theoretical approaches, particularly by researchers who have claimed that pragmatic principles must also play a role. I summarise three of these approaches below.

The assumption that pragmatic inferences are involved, to a greater or lesser degree, in specifying inaccessible antecedents is not new. This approach has been taken by several researchers (for a discussion, see Huang 1994), most notably Reinhart (1983).⁸ Reinhart proposes that a distinction is needed between *variable binding* and *coreference*, two ways in which a pronoun can be linked to an antecedent. A full description and discussion of variable binding and coreference is given in Chapter 3 (see section 3.2.4). For the purposes of the current chapter, it is sufficient to say that variable binding is a semantic operation that takes place under certain structural restrictions, one of which is called condition B.⁹ Coreference is free from such structural restrictions but is guided by pragmatics, in particular, the notion that the speaker must be as explicit as possible. To see how this operates, consider again example (2.3), repeated here for convenience:

(2.3a) Ken_i thinks that Boris_j likes himself_{i*/j}.

(2.3b) Ken_i thinks that Boris_j likes him_{i/j*}.

⁸ Reinhart's approach is developed in subsequent work, and also forms an important part in Reinhart and Reuland's (1993) influential paper *Reflexivity*. It is also incorporated into Reuland's later work on binding (Reuland 2001; 2011).

⁹ In Reinhart's work condition B is formulated differently from BT, but the effect is still the ruling out of an inaccessible antecedent.

In Reinhart's approach, *him* in (2.3b) can be linked to a potential antecedent via variable binding or via coreference. Both of these routes may link *him* to *Boris*; this is an undesirable outcome which needs to be prevented. The variable-binding route is obstructed because Reinhart's version of condition B applies to variable-binding operations. But the constraint does not apply to the coreference route. What prevents *him* being linked to *Boris* via coreference? This is where the notion of explicitness comes into play. The assumption is that if a speaker wishes to link *him* to *Boris*, they must be as explicit as possible in signalling this. The most explicit way of signalling that *Boris* should be the antecedent is to use a reflexive, as in (2.3a). When a reflexive is not used (as in 2.3b), an assumption can be made that the link to *Boris* is not intended. In this way, coreference is guided by a pragmatic principle, and variable binding is guided by a configurational principle, and both of these block binding of the pronoun by the inaccessible antecedent, just as the original condition B does. Crucially, both configurational and pragmatic principles are involved in ruling out an inaccessible antecedent. While this system may seem far more complicated than the original formulation of BT, the availability of two routes for pronoun interpretation has the advantage of being able to account for a wide variety of different phenomena related to anaphora that cannot be accommodated so elegantly within BT. Importantly, in this approach structural information alone is not enough to rule out an inaccessible antecedent.

A similar proposal has been made by Wexler and colleagues to account for children's processing data in relation to pronouns. A persistent finding in studies with children is that, in contrast to good performance on reflexives, they fail to rule out inaccessible antecedents for pronouns. This is known as the delay of principle B effect (DPBE), and is still very much under debate (e.g. Wexler and Chien 1985; Chien and Wexler 1990; Avrutin and Thornton 1994; Perovic, Modyanova and Wexler 2013 provide evidence in favour of the DPBE, but Conroy *et al.* 2009; Bergmann, Paulus and Fikkert 2012 and others provide an alternative view). The asymmetry between performance on reflexives and pronouns is attributed by Wexler and colleagues to a pragmatic deficit in children. While condition B is assumed

to be intact and operating on variable-binding operations (i.e. following Reinhart's proposal described above), children fail to use pragmatic knowledge to rule out a coreference relation between a pronoun and an inaccessible antecedent. While there is considerable support for such a view, other proposals regarding the DPBE have cited working memory limitations or an inability to consider a speaker's perspective as possible alternative explanations for the data. Nevertheless, Reinhart's distinction between two types of operation for pronoun interpretation, as well as the assumption that ruling out an inaccessible antecedent is more than a purely structural operation, continues to inform mainstream accounts of children's pronoun processing.

A different theoretical account is provided by Huang (1994), who proposes a detailed, neo-Gricean pragmatic account of the distribution of pronouns and reflexives.¹⁰ The ruling out of an inaccessible antecedent for the pronoun is achieved through the interaction of pragmatic principles and a background assumption called the disjoint reference presumption (DRP). The DRP assumes that the arguments of a predicate are supposed to be disjoint, unless marked otherwise. Pragmatic principle I dictates that speakers should give as little information as is necessary, and principle M dictates that speakers should not use a marked expression without reason. During comprehension, the listener expects that the speaker will adhere to these principles, which interact with the DRP. In (2.3b), the verb *like* (the predicate), has two arguments: *Boris*, the subject, and *him*, the object. The DRP simply assumes that the two arguments of *like* are supposed to refer to different entities (i.e. are disjoint). This assumption can be overcome by overt marking. Reflexives are assumed to be marked forms in this approach, so they defeat the DRP assumption, so that reference is not disjoint. For pronouns, which are not marked, the DRP is not defeated, giving rise to disjoint reference; this automatically rules out the inaccessible antecedent. Of course, other pragmatic principles come into play which can

¹⁰ Huang's pragmatic theory of anaphora has been developed in collaboration with Stephen Levinson, and is further expanded in Huang (2000).

account for a wider range of data on anaphora than the core condition B case described here.

This is by no means a comprehensive survey of theoretical approaches to pronouns, which is far beyond the scope of this dissertation. However, it serves to point out that at least several well established approaches to condition B contain not just structural principles but a pragmatic element as well. There is no clear consensus about the precise nature of condition B, but its status as a purely structural principle should be questioned. This issue has in fact received less attention in the adult literature on the application of condition B during processing, while it has been taken more seriously in acquisition research. If condition B is not only a structural constraint but also involves pragmatic reasoning of some kind, this should be reflected in models of adult processing.

Before reviewing processing evidence on condition B, I will first introduce the theoretical background for SDPs in the following section.

2.2.4 An exception to condition B: short-distance pronouns

As mentioned above, there are some well known exceptions to the assumption of the complementary distribution of pronouns and reflexives. Some examples are picture noun phrases, logophoric usage of reflexives, and SDPs. In these exceptional cases, pronouns and reflexives are not in complementary distribution which poses a challenge to BT. SDPs are the focus of this chapter because they provide an interesting contrast to the canonical condition B environment. SDPs are pronouns which appear within certain prepositional phrases (PPs), and which are apparently able to refer to antecedents inside the pronoun's binding domain in violation of condition B. (2.4) is a typical example of an SDP:

(2.4) David_i put the cat beside him_i/himself_i.

In (2.4) it is acceptable for both the pronoun and the reflexive to refer to *David*, showing that pronouns and reflexives are not in complementary

distribution.¹¹ Under the original definition of governing category (an NP or S containing a governor and the reflexive/pronoun) the whole sentence (S) must be the governing category, because it contains a governor (the verb *put*) and the reflexive/pronoun. (There is no smaller NP domain that could be a governing category in (2.4) because there is no NP containing a governor.) *Himself* obeys condition A because it is bound within the governing category. But *him* is in violation of condition B because it too can be bound within the governing category, i.e. the pronoun is not free. Chomsky's revised (1986) definition of the governing category does not fare any better for SDPs. The revised formulation proposes that the governing category should be a complete functioning complex (CFC). The CFC is a broader category than just NP or S, and could, for example, be a PP. One requirement is that a CFC contains a governor, and all the functions associated with that governor must be realised within the CFC. A PP contains a governor (the preposition), and all functions associated with the governor (that is, anything that it assigns case to) are realised within the PP. But there is an additional requirement on CFCs. They must also contain a subject. This requirement is to ensure that CFCs can be compatible with condition A. Considering example (2.4) again, the PP cannot fulfil the requirement to contain a subject. Only the S can do so. This means that under the revised definition of the governing category, the S is still the governing category and condition B is still violated. Effectively, even with the updated definition for the governing category, SDPs still pose a challenge to BT.

Various proposals have been put forward to deal with exceptions such as SDPs. Hestvik (1991), for example, suggested that SDPs could be accounted for within the revised (Chomsky 1986) BT framework if the requirement for a subject in the CFC was removed. This would make the binding domain

¹¹ Note that the acceptability of reflexives in PPs is somewhat dependent on the verb or the particular situation being described, as noted by Reinhart and Reuland (1993) and Büring (2005). I will focus on pronouns, for which acceptability is more robust, and the issue of compatibility with condition B, leaving aside the more complex picture of reflexives inside PPs.

for *him* the PP, rather than the whole S. Since *him* is free within the PP, condition B is not violated. But if the PP is the governing category and not the S, the reflexive would not be bound within the governing category and condition A would therefore be violated. To account for the reflexive, it is assumed that the governing category for a reflexive must contain a binder (to make the reflexive compatible with condition A); since the PP contains no potential binder for *himself*, the binding domain is assumed to be the whole S, because this does contain a potential binder (*David*). So, by removing the subject requirement but specifying BT compatibility, Hestvik (1991) accounts for the acceptability of both *him* and *himself* in sentences such as (2.4) by having a different governing category for the pronoun and the reflexive. The difficulty for such an account is that, by specifying different binding domains for pronouns and reflexives, the ability to predict complementarity of distribution is lost.

Tenny (2004) also puts forward a proposal for SDPs in which the binding domain for pronouns is altered, but in a different way from Hestvik's (1991) structural account. In this proposal, the syntactic structure incorporates certain types of semantic information. This is based on the observation that the acceptability of pronouns and reflexives in PPs is not uniform for a particular structure but varies according to verb class and the perspective of the characters in the event being described, and information about location for certain events. This semantic information is represented in the syntactic structure. A sentience feature takes scope over the V and PP, and defines the binding domain; to be in accordance with condition B, the pronoun must be free inside the domain. The subject (*David* in (2.4)) is outside of the domain defined by the sentience feature, so condition B is not violated when the pronoun binds to *David*. The domain for reflexives is left as it is normally defined, leading again to a different domain for the pronoun and the reflexive. Similar to Tenny's (2004) proposal, Rooryck and Vanden Wyngaerd (2011) also make use of the syntactic representation of semantic information such as perspective and location to define different binding domains for reflexives and pronouns.

Reinhart and Reuland's (1993) account differs from the previous accounts which are concerned with redefining the binding domain. Their framework for the binding of reflexives and pronouns is concerned with co-argumenthood and reflexive marking. As noted above, Reinhart's (1983) formulation of condition B is somewhat different to Chomsky's, emphasising the importance of the properties of the verb and its arguments rather than syntactic domains. For example, a verb such as *wash* (when used transitively) has two arguments, a subject and an object: *Jim* (subject) *washed the floor* (object) / *Jim* (subject) *washed the baby* (object). If the subject and object are to be understood as the same person (as in *Jim washed himself*), Reinhart and Reuland (1993) stipulate that there must be reflexive marking to indicate this. *Himself* carries reflexive marking, so this requirement is satisfied. The absence of reflexive marking (as in *Jim washed him*) indicates that the subject and object must refer to different people, i.e. the pronoun, being without reflexive marking, cannot refer to a coargument of the same verb *wash*. In the case of SDPs such as (2.4), *him* is not an argument of the verb *put*, and as such there is no restriction preventing the pronoun referring to *David* in this case. The licensing of reflexives in such constructions is rather more complex, but as far as pronouns are concerned SDPs do not cause a violation in Reinhart and Reuland's (1993) approach to binding.

These and other accounts demonstrate that SDP environments cannot straightforwardly be accounted for within BT. It is interesting to note that certain proposals for dealing with these exceptions rely to a greater or lesser extent on semantic information (Tenny 2004; Rooryck and Vanden Wyngaerd 2011; to some extent, Reinhart and Reuland 1993), even though this semantic information may be brought into the syntactic structure under some proposals.

From a processing perspective, SDPs provide an interesting contrast to canonical condition B environments. This is because the antecedent search process does not involve ruling out a particular antecedent. Both environments, however, are similar in that they can have two *potential* antecedents within the sentence in subject position. The next section gives

an overview of previous processing studies that have examined condition B and SDPs from a processing perspective.

2.3 Condition B in sentence processing

While theoretical linguistics is concerned with the correct formulation of restrictions on a pronoun's reference, psycholinguistic research has a different concern: the application of such restrictions in real-time processing. Because the processing of a pronoun normally involves the identification of a suitable referent, a key question in processing research is whether the real-time search for a referent is constrained by structural constraints such as condition B. (As noted above, despite the variety of theoretical accounts of condition B that contain at least some element of pragmatic reasoning, condition B in the psycholinguistic literature is normally referred to as a structural or grammatical constraint, which may give rise to the assumption that there is no pragmatic involvement in its application.) So far, psycholinguistic research in this area gives a rather mixed picture, with evidence both for and against a restricted search having been found. This evidence has come from a number of different experimental paradigms, very little, so far, from eye tracking.¹² In this section I will present and discuss previous psycholinguistic research on the application of condition B and on the processing of SDPs in adults.

2.3.1 Condition B as a constraint on antecedent search

One of the most well known studies on condition B is Nicol and Swinney (1989). Their paper claims that the search for antecedents is structurally licensed. The authors present studies on various anaphoric elements including wh-traces, NP-traces, reflexives, pronouns, and PRO as the basis of their hypothesis, but I will concentrate here on the studies relating to reflexives and pronouns.

¹² In fact, Lewis and Chow (2012) report eye-tracking results on the processing of condition B and I make reference to their findings below, but because this is from a presentation and is not yet in published form, I do not present their study in detail.

With respect to reflexives, Nicol and Swinney (1989) summarise a study that was reported in Nicol (1988). The study used the cross-modal priming technique to assess the reactivation of structurally licensed and unlicensed antecedents of a reflexive. In cross-modal priming, participants listen to sentences while looking at a screen; at particular points in the sentence a word is presented on a screen. Participants have to make a decision regarding the word that they see – a common task is to decide whether the word exists in their native language or not. In the experimental conditions, the visually presented word is semantically related to a word that appears in the auditory sentence. The visually presented word in the control conditions is matched on certain features (e.g. length and frequency) to the word in the experimental condition, but is semantically unrelated to the auditory material. The response times to the word in the experimental condition are subtracted from the response times to the word in the control condition. Faster response times in the experimental condition compared to the control condition indicate that a mental representation of earlier material has been activated. Response times at control positions are also measured, to check that the activation of earlier material is related to the point of interest, rather than a general memory trace of all earlier material. An example of the material from Nicol's (1988) study using this technique is given below:

(2.5) The boxer told the skier that the doctor₁ for the team would blame himself₁ [] for the recent injury.

The correct antecedent for *himself* is *doctor*, according to condition A of BT. This is indicated by the indices. *Boxer* and *skier* are outside the binding domain for *himself* and are therefore inaccessible antecedents. The square brackets after *himself* indicate the test point at which the experimental or control word was presented. The results showed significant priming for *doctor* (that is, participants responded faster to words semantically related to *doctor* than words that were unrelated) at the test point, but no significant priming for *boxer* or *skier*, the inaccessible antecedents. This indicates that, on encountering the reflexive, participants reactivated a mental representation of *doctor* but not of the other potential antecedents.

The authors propose that the reason for this pattern of reactivation is that only accessible antecedents are reactivated on encountering the reflexive, rather than all possible antecedents. But another possibility is that *doctor* was still activated during the reading of the following words, and that initial activation (as opposed to reactivation) gave rise to the result. To counter this possibility, pronouns in the same position were also tested:

(2.6) The boxer₁ told the skier₁ that the doctor for the team would blame him₁ [] for the recent injury.

In (2.6), *doctor* is an inaccessible antecedent for the pronoun *him* because it is excluded by condition B, whereas *boxer* and *skier* are accessible antecedents; the complete reversal of the situation for the reflexive. For the pronoun experiment, there was significant priming for *boxer* and *skier*, the accessible antecedents, but not for *doctor*, the inaccessible antecedent. The lack of priming for *doctor* in the pronoun experiment suggests that the priming pattern found for reflexives was not a result of a lingering initial activation of *doctor*. Furthermore, the fact that *doctor* appears not to be reactivated at the pronoun supports the authors' interpretation of the previous results: the reactivation of antecedents for reflexives and pronouns is constrained by conditions A and B of BT, which they term structural licensing.

Nicol and Swinney (1989) go on to describe a similar study which provides slightly more puzzling results:

(2.7) The boxer₁ visited the doctor_k that the swimmer at the competition had [] advised [] him₁ [] to see t_k about the injury.

In this example, *boxer* is the only accessible antecedent for *him* (extra-sentential antecedents are also possible). In the experiment there was no significant priming for any of the antecedents at the first test point after *had*, although it is worth noting that numerically there is (non-significant) priming for both *boxer* and *doctor* at this point, with a 23ms and 32ms difference respectively between responses to control and related words, compared to a 9ms difference for *swimmer*. At the second test point there is

significant priming for *doctor* (56ms). At the third test point after the pronoun, there is significant priming for both *doctor* (42ms) and *boxer* (51ms). So it appears that at the pronoun there is activation of both an accessible antecedent, *boxer*, and an inaccessible one, *doctor*. This stands in contrast to the previous results in which only the accessible antecedents were reactivated. The authors explain that the unexpected activation for *doctor* could be attributed to its reactivation at the second test point; the trace *t*, which in reality appears after *see*, could be assumed to appear after *advised*, triggering reactivation of *doctor* at that point. Since the trace *t* is so close to the pronoun, it is not surprising that *doctor* is still activated at the pronoun, despite not being structurally licensed. Following this same argument, it is also possible that the occurrence of *him* when participants expected *t* (i.e. they expected a syntactic gap) was perceived as a structural violation, which could have disrupted processing.

Despite the somewhat conflicting results of the previous experiment, Nicol and Swinney (1989) conclude overall that binding conditions constrain the reactivation of candidate antecedents, such that reactivation is limited to those antecedents that are accessible. Furthermore, they make the bold claim that structural licensing from the binding conditions takes place *before* pragmatic/discourse factors are considered:

“...the reactivation of prior referents is restricted by grammatical constraints. In the case where such information does not sufficiently constrain the list of potential antecedents to a single one, the pragmatic and other sentence/discourse processing procedures undoubtedly come into play, but, given the present evidence, only at a *later* point in processing.”

(Nicol and Swinney 1989, p. 18)

The evidence presented by Nicol and Swinney (1989) directly addresses the question of how the binding conditions can influence the search for antecedents. The results do show a clear distinction between the activation of accessible and inaccessible antecedents, especially in the case of

reflexives. The evidence for pronouns is less clear, given the activation of an inaccessible antecedent in structures like (2.7). There is no doubt that cross-modal priming data of this sort can be informative about aspects of processing, but it can only provide a snapshot of what is happening at particular points in time. For example, we do not know if other, very rapid processes have already taken place, or if different factors come into play later on during processing. Computing referential dependencies can be a complex process, and evidence from a longer timecourse is therefore required.

Further evidence that inaccessible antecedents are ruled out immediately and therefore do not affect the processing of a pronoun comes from an experiment within Clifton, Kennison and Albrecht's (1997) paper which looks primarily at the processing of the word *her*. *Her* can be used either as a full NP (e.g. *We saw **her** at the shops*) or as a specifier for an NP (e.g. *We saw **her** dog at the shops*). Experiment 2, which was a phrase-by-phrase self-paced reading experiment, tested the unambiguous pronoun *him* and the specifier *his*, to draw a contrast to the ambiguity of *her*. The sentences containing *him* are of particular interest here. In the phrase-by-phrase self-paced reading paradigm, participants are presented with sentences to read on a screen, one phrase at a time. Participants press a button when they have finished reading the phrase to move on to the next one, and as such are in control of the speed of presentation. The time taken to press the button for each phrase is measured. In comparison to the cross-modal priming method, regions at and beyond the pronoun can all be examined, giving a picture of processing over a longer timecourse. The authors presented sentences such as the following (| represents the division of the sentence into phrases for presentation):

(2.8a) The supervisors | paid | him | yesterday | to finish | typing | the manuscript.

(2.8b) The supervisor | paid | him | yesterday | to finish | typing | the manuscript.

The NP *the supervisor(s)* is inaccessible according to BT; binding of *him* by *the supervisor(s)* would be violating condition B. No accessible antecedent was presented. Interestingly, although the authors do not discuss it in their paper, the inaccessible NP is also the first-mentioned antecedent; this is in contrast to Nicol and Swinney (1989), whose first-mentioned antecedent was accessible according to condition B. Only the NP in (2.8b) matched in number with the pronoun. If participants were considering the inaccessible antecedent *the supervisor(s)* during processing, there should have been increased reading times for (2.8a) compared to (2.8b) because of the number mismatch between the NP and the pronoun in (2.8a). But the results showed no difference in reading times between (2.8a) and (2.8b), at or beyond the pronoun region. This suggests that participants were not considering the NP as a potential antecedent for the pronoun, while reading the pronoun or afterwards. As noted, the NP was the first-mentioned entity in the sentence. In the study reported in Nicol and Swinney (1989), one of the accessible antecedents was the first-mentioned antecedent, so it was not clear whether it was targeted because it was accessible or targeted because it was first-mentioned. From the Clifton *et al.* (1997) experiment it seems that there is no default to the first-mentioned antecedent when that antecedent is inaccessible. This adds more weight to the suggestion that something like condition B is in operation, rather than a simple default to a first-mentioned antecedent for pronouns.

Evidence of a very similar sort comes from Clifton, Frazier and Deevy (1999), another self-paced reading study. In this study, the gender of the inaccessible antecedent was manipulated rather than the number, and an accessible antecedent was also presented. Sentences were of the following type:

(2.9a) John thinks that Bill owes him another chance to solve the problem.

(2.9b) John thinks that Betsy owes him another chance to solve the problem.

Bill/Betsy is an inaccessible antecedent for *him*, but *John* is accessible. As per Clifton *et al.* (1997), this study found no difference between the two

conditions; a difference would have been indicative of the inaccessible antecedent's gender affecting processing.

The evidence from these two self-paced reading studies appears to support Nicol and Swinney's (1989) hypothesis that binding conditions are applied before candidate antecedents are evaluated, ensuring that no inaccessible antecedents are considered. This hypothesis is often referred to as the *binding as initial filter* (BAIF) hypothesis. The idea behind BAIF is that grammatical principles take priority over other types of processing. Nicol and Swinney explicitly state that pragmatic and discourse information is relegated to the role of mopping up after the binding conditions, if they fail to identify a unique antecedent. While this is in line with the original formulation of BT, it is rather difficult to reconcile with other approaches to condition B, some of which were outlined in section 2.2.3, which explicitly involve aspects of pragmatic reasoning to rule out an inaccessible antecedent. Nevertheless, the lack of interference from inaccessible antecedents does suggest that matching antecedents on the basis of gender or number features is less important than adherence to condition B. So far, though, the evidence is a little sparse. Nicol and Swinney's clear results from their first experiments are somewhat undermined by the more puzzling results from the second pronoun experiment. The evidence from the two self-paced reading studies points in the direction of the BAIF, but there is certainly room for further evidence from a more time-sensitive method than self-paced reading. Evidence supporting the BAIF from both self-paced reading and eye tracking has been shown by Lewis and Chow (2012), who also found no evidence of interference from inaccessible antecedents. In their final experiment they also manipulated the gender of the accessible antecedent and found that this did affect processing. While it is not possible to assess this study in detail (see footnote 12 above), this study provides support for the BAIF from both self-paced reading and eye-tracking paradigms.

One previous study that has used a more time-sensitive paradigm to assess the timing profile of the binding conditions is Sturt (2003), an influential study on the timing of condition A for reflexives using eye tracking. While

this study concerns reflexives rather than pronouns, I discuss it here because of its relevance in terms of the binding conditions and the processing hypothesis presented. Sturt (2003) manipulated the stereotypical gender match between reflexives and accessible and inaccessible antecedents. In the first experiment the following sentence types were used:

(2.10a) *Accessible match/inaccessible match*

Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked himself with a used syringe needle. There should be an investigation soon.

(2.10b) *Accessible match/inaccessible mismatch*

Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked himself with a used syringe needle. There should be an investigation soon.

(2.10c) *Accessible mismatch/inaccessible match*

Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked herself with a used syringe needle. There should be an investigation soon.

(2.10d) *Accessible mismatch/inaccessible mismatch*

Jonathan was pretty worried at the City Hospital. He remembered that the surgeon had pricked herself with a used syringe needle. There should be an investigation soon.

Sturt (2003) expected that there would be an effect of the stereotypical gender of the accessible antecedent, with disruption when there was a mismatch between the antecedent's stereotypical gender and the gender of the reflexive. Moreover he was particularly interested in finding out whether the gender of the inaccessible antecedent would affect the processing of the reflexive, and if so, at what point. The results showed a clear accessible mismatch effect in the expected direction: participants' reading times increased when, for example, they had to link the female reflexive *herself* to the stereotypically male, accessible antecedent *surgeon*.

This effect was shown in very early measures, first-fixation times and first-pass times on the reflexive itself. There was no reliable effect of the inaccessible antecedent in these early measures. However, in a later measure (rereading times) there was a reliable effect of the inaccessible antecedent: rereading times in the prefinal region, *used syringe* in the above example, were significantly longer when the inaccessible antecedent mismatched in gender with the reflexive. This was the most robust finding regarding the influence of the inaccessible antecedent, but there were also interactions in the precritical region (before the reflexive) and the reflexive regions. A second experiment investigated whether similar effects would be found when the order of the antecedents was reversed, i.e. the inaccessible antecedent was closer to the reflexive than the accessible one. In the second experiment, an effect of the accessible antecedent was found again in first-fixation and first-pass times in the reflexive region. In contrast to experiment 1, however, no effect of the inaccessible mismatch was detected in any of the regions or measures.

Sturt (2003) describes his results in terms of bonding and resolution (Garrod and Terras 2000). The bonding stage is a process in which a link between an anaphoric element and a potential antecedent is made; lexical information is important at this point. In resolution, the links made in the bonding stage are evaluated, with the possible involvement of contextual information. Extending this model specifically to reflexives, Sturt suggests that the bonding stage is triggered when the reflexive is encountered: potential antecedents are activated at this stage. In the resolution phase, the link between the reflexive and the candidate antecedents is evaluated and integrated into the interpretation of the sentence. Sturt suggests that the bonding stage not only involves lexical information but is also constrained by syntactic considerations, so that only antecedents that are accessible in terms of the BT are considered. Inaccessible antecedents, on the other hand, could be considered at the resolution phase. Sturt claimed that binding principles act as an early but defeasible filter during the resolution of reflexives and pronouns, such that inaccessible antecedents could play a role in later processing stages, reflected in later measures such as rereading times.

Sturt's (2003) study is of interest because, firstly, it provides further evidence that the binding conditions play a role early on in processing. This, as far as the earlier stages of processing are concerned, supports the BAIF even if the later stages of processing are assumed to operate differently. Secondly, it provides evidence from a method that is more time-sensitive than self-paced reading; however, it only provides evidence for reflexives, not pronouns. Sturt's proposal of a defeasible filter does differ from the BAIF, though, in suggesting that binding conditions can be defeated at later stages of processing.

So far, I have presented evidence that supports the involvement of binding conditions, and condition B in particular, in the search for antecedents. This evidence is mainly negative evidence, showing that inaccessible antecedents do not affect the processing of pronouns (or, in the case of the reflexives in Sturt's study, do not affect earlier stages of processing). However, several studies show quite the opposite: the inaccessible antecedent having an effect on the processing of pronouns. These studies are considered in the next section.

2.3.2 Evidence for parallel constraints

This section reviews the evidence against the BAIF and similar hypotheses. All the studies below suggest that the search for a suitable antecedent is not constrained by condition B in the way that Nicol and Swinney (1989) propose, but suggest instead that inaccessible antecedents can have an effect on processing and are therefore not immediately excluded from consideration.

Badecker and Straub (2002) is probably the most well known study that presents evidence against the BAIF. The paper describes a series of word-by-word self-paced reading studies. As well as reading the sentences, participants were also presented with a probe word at the end of the sentence, and they had to decide whether or not the probe word appeared in the sentence they had just read. There were comprehension questions at

the end of a quarter of the items presented. In experiment 1, items were presented in the following conditions:¹³

(2.11a) *Multiple match*

John thought that Bill owed him another chance to solve the problem.

(2.11b) *Accessible match*

John thought that Beth owed him another chance to solve the problem.

(2.11c) *Inaccessible match*

Jane thought that Bill owed him another chance to solve the problem.

(2.11d) *No match*

Jane thought that Beth owed him another chance to solve the problem.

The gender match between the pronoun and the two antecedents was manipulated to create the four conditions. *John/Jane* is the accessible antecedent and *Bill/Beth* is the inaccessible antecedent. The crucial finding from this experiment was that reading times in the region following the pronoun were significantly longer in the multiple match condition compared to the accessible match condition, suggesting that when the pronoun was encountered, participants were influenced by the gender of the inaccessible antecedent. In order to check whether this was due to the processing of the names rather than the pronoun (in case it takes longer to establish two similar (gender matching) names in the discourse model) they took conditions (2.11a) and (2.11b) and compared them to sentences in which the pronoun was replaced with a name in a further experiment.

¹³ The observant reader may recall that these are very similar to the items in Clifton *et al.*'s (1999) study, discussed above. In fact, the Clifton *et al.* (1999) study was a partial replication Badecker and Straub's work that was reported earlier in a poster presentation (Badecker and Straub 1994, cited in Badecker and Straub 2002).

They found that there were no significant differences in reading times in the name conditions, but in the pronoun conditions the multiple match condition again produced longer reading times. This effect was found again in a third experiment in which they compared pronouns to reflexives, but this time the effect was further downstream of the pronoun. On the basis of these results, the authors suggested that multiple constraints that are relevant for pronoun processing (including binding constraints) all contribute in parallel, positively or negatively, to an antecedent's activation. Thus, positive activation from one constraint may be cancelled out by inhibition from another. For example, binding condition B will inhibit the activation of inaccessible antecedents. But at the same time, an inaccessible antecedent that has features matching the pronoun, such as *masculine* and *singular*, will receive more activation from morphological constraints than an inaccessible antecedent with (partially) mismatching features. Due to this parallel activation/inhibition, the feature match or mismatch of an inaccessible antecedent will have an influence on processing, in direct contrast to the BAIF hypothesis (which, according to Badecker and Straub 2002, over-emphasises the importance of the binding conditions in comparison to other cues).

This study has been criticised because the probe recognition task and the low number of comprehension questions may not have encouraged participants to make a full interpretation of the pronoun (Phillips, Wagers and Lau 2011).¹⁴ This criticism implies that it was not the processing of the pronoun that drove the effect of the inaccessible antecedent; if so, the only source of difference between the conditions was the gender of the names. Indeed, in experiment 2, the reading times in the two pronoun conditions began to diverge before the pronoun was encountered, suggesting a more general difficulty with processing the multiple matching names. But if matching names rather than pronoun processing was the source of the

¹⁴ Probe studies in general have been criticised because they may encourage participants to create a list of potential probe targets alongside processing the sentence (Gordon, Hendrick and Foster 2000).

disruption, why weren't longer reading times seen for the multiple matching name condition in experiment 2? The results of both experiments together seem to suggest that the effect is driven by the processing of the pronoun. But Phillips *et al.* (2011) also note that the effect has not been replicated by a similar study (Clifton *et al.* 1999, see above). Recall that this was an almost direct replication of Badecker and Straub's (2002) experiment, but with comprehension questions instead of a probe recognition task, and the comprehension questions probed the referent of the pronoun specifically. They failed to find the difference that Badecker and Straub found. It seems plausible to attribute the different results to the different tasks that the participants were asked to do, although it is still in question where the difference in Badecker and Straub's results come from, if not from processing the pronoun. One further point to make about this study is that the effects in experiment 3, despite the materials being the same, occurred later than in the first two experiments. This suggests that, at least to some extent, the effect is modulated by the precise experimental conditions.

Further evidence of interference from a matching inaccessible antecedent comes from Kennison (2003), who carried out a self-paced reading study, with comprehension questions following the sentences rather than a probe task. The following sentence types were used:

(2.12a) Susan | watched | her | yesterday | during | the open rehearsals | of
the | school play.

(2.12b) Carl | watched | her | yesterday | during | the open rehearsals | of
the | school play.

(2.12c) They | watched | her | yesterday | during | the open rehearsals | of
the | school play.

(2.13a) Susan | watched | him | yesterday | during | the open rehearsals | of
the | school play.

(2.13b) Carl | watched | him | yesterday | during | the open rehearsals | of
the | school play.

(2.13c) They | watched | him | yesterday | during | the open rehearsals | of
the | school play.

As per Clifton *et al.* (1997), these sentence types were contrasted with those in which the pronoun was a specifier. I will discuss the results from the sentences involving pronouns as NPs only since these are relevant to the issue at hand. The findings for the first set of sentences (2.12a-c) were as follows: reading times in (2.12b) (*Carl*) were longer than those in (2.12c) (*they*) in regions 4 and 5 (the pronoun was presented in region 3). In region 5, reading times in (2.12a) (*Susan*) were longer than those in (2.12c) (*they*). In region 6, reading times in (2.12a) (*Susan*) were longer than in both other conditions. This seems to suggest that participants' reading times were affected by the number and gender of the inaccessible antecedent.

Participants found it easier to process sentences in which the NP was a plural pronoun. The gender of the NP also affected the reading times, so that a gender matching inaccessible antecedent yielded longer reading times than the gender mismatching inaccessible antecedent in a region downstream from the pronoun. For the sentence types (2.13a-c), where the masculine pronoun was used, the reading times in region 4 were shorter for (2.13c) (*they*) compared to the other two conditions. (2.13b) yielded the longest reading times in this region, but the difference between (2.13b) and (2.13a) did not reach significance; when regions 4 and 5 were pooled, the increased reading times in (2.13b) compared to (2.13a) became significant. The pattern of results is similar for the masculine and feminine pronouns; firstly there was a clear reading time advantage when the inaccessible antecedent was a plural pronoun. Secondly, longer reading times for the gender matching inaccessible antecedents are apparent in both sentence types but emerge later (region 6) for the feminine pronoun. The later appearance of the effect in sentence type 12 could be due to the experiment containing an equal number of sentences where *her* appears as a specifier rather than a full NP (since this was one of the experimental manipulations). *Her* is ambiguous in this respect while *him/his* is not. Participants could have been processing the disambiguating region, which

directly followed the pronoun, before attempting to link the pronoun to an antecedent in (2.12), while in (2.13) there was no ambiguity so the search for an antecedent could begin sooner, leading to earlier effects.

Despite the similarity of experiment type and materials, Kennison's (2003) experiment yielded different results from those in the Clifton *et al.* (1997) study presented in the previous section. While Clifton *et al.* found no influence of a number matching inaccessible antecedent, Kennison claimed that participants were affected by the number and gender of the inaccessible antecedent. Since Kennison's sentences did not contain any other potential antecedents, it is possible that participants attempted to link the pronoun to the inaccessible antecedent as a last resort, finding no other alternative in regions 4-6. But if that is the case, why was this not seen in the Clifton *et al.* study, whose sentences also contained no accessible antecedent? It is interesting that in Kennison's study, number stood out as having more robust effects than gender; conditions with the number mismatching inaccessible antecedent consistently yielded the shortest reading times. However, it is difficult to interpret this effect because there is no gender neutral number matching inaccessible antecedent condition for comparison; the comparison with the name conditions is unsatisfactory because, firstly, it is comparing a pronominal antecedent (*they*) to a proper name antecedent, and secondly, because the gender match/mismatch appears to have its own influence so it is not a good baseline for comparison. For this reason it is better to concentrate on the comparison between the gender match/mismatch conditions. These effects are weaker and later. In (2.12) they appear in region 6, and in (2.13) they appear only when reading times from region 4 and 5 are pooled. A late effect of gender match, with the matching gender antecedent yielding longer reading times, suggests that participants were considering the inaccessible antecedent at a late stage. As mentioned above, it could be that participants attempted to link the pronoun to the inaccessible antecedent as a last resort, finding no other alternative in regions 4-6. Indeed, a later experiment presented in the same paper shows that when a plausible antecedent is presented in the preceding discourse, the effect of the inaccessible antecedent goes away. This strongly suggests that the effect is driven by a lack of accessible

antecedent in the discourse. Nevertheless, it does work against the BAIF hypothesis, which predicts no effect of an inaccessible antecedent at any point.

Further, more time-sensitive evidence of interference from an inaccessible antecedent comes from two studies by Clackson and colleagues (Clackson, Felser and Clahsen 2011; Clackson, Heyer and Clahsen 2012), who compared the processing of reflexives and pronouns in visual world experiments. In the visual world paradigm, participants listen to sentences or short stories while looking at a visual display. The visual display shows pictures related to the auditory material, and eye movements on the display are recorded. Clackson *et al.*'s (2011) study tested both adults' and children's processing of pronouns and reflexives; I discuss the findings for the adults here. In the experimental items, the gender of the accessible and inaccessible antecedents was manipulated to create a double match and single match for the reflexive and the pronoun. Example stories are shown below.

(2.14a/b) Double match

Peter was waiting outside the corner shop. He watched as Mr. Jones bought a huge box of popcorn for {himself/him} over the counter.

(2.14c/d) Single match

Susan was waiting outside the corner shop. She watched as Mr. Jones bought a huge box of popcorn for {himself/her} over the counter.

The screen displayed four pictures: a separate picture of each character in the story (Peter and Mr Jones in (2.14a/b); Susan and Mr Jones in (2.14c/d)), the object in the story (e.g. *popcorn*) and an unrelated distractor. The online measure was the proportion of looks to each antecedent (coded as accessible or inaccessible) over a two second time window following the onset of the reflexive or pronoun in the story. Comprehension questions followed each item, but the questions did not directly probe the referent of

the pronoun. In both the reflexive and pronoun conditions, participants looked significantly more at the accessible than the inaccessible antecedent, reflecting their offline interpretations. In the reflexive conditions, participants were slower to look at the accessible antecedent in the double match conditions, but this was not statistically reliable. They found no reliable difference between proportion of looks to the accessible antecedent in the single and double match conditions for the reflexives. Conversely, in the conditions containing pronouns, there was a significant interaction between antecedent (accessible/inaccessible) and condition (single/double match), with a bigger difference in looks to the accessible versus the inaccessible picture in the single match condition. This was a fairly stable pattern over the 2s time window.

Before discussing these results I will briefly describe the results from a follow up study (Clackson *et al.* 2012). The materials and set-up were very similar to those in the 2011 study, but in this study the offline task was not to answer comprehension questions. Instead, participants were shown an object at the start of each trial (for example, in relation to (2.14) the object would be *popcorn*), then listened to the story, and at the end of the trial, answered "*Who is it for?*", a question which directly probed the referent of the reflexive or pronoun. The visual display showed a distractor character, not mentioned in the story and mismatching in gender with the pronoun or reflexive, instead of a distractor object. Participants were less accurate at answering the question correctly in the double match conditions containing a pronoun, compared to the single match conditions (86% versus 96%). Accuracy scores for the reflexives did not show such a difference (double match 95%; single match 96%). For both reflexives and pronouns, the single match conditions were responded to faster than the double match conditions. The online measure of gaze direction showed that participants looked away from the picture of the inaccessible antecedent more quickly in the single match conditions compared with the double match conditions; this was true for both reflexives and pronouns.

Clackson's studies clearly show that participants were distracted by a gender-matching, structurally inaccessible antecedent. This was true for

pronouns across both experiments, but only true for reflexives in one experiment. With regard to the timing profile, the effect of the inaccessible antecedent appeared quite quickly after hearing the pronoun/reflexive and tended to remain stable throughout the time period observed. The first question to be answered then is why did reflexives elicit different behavioural responses across the two studies which used the same materials? It is plausible to suggest that the task in the second experiment, which explicitly probed the referent of a pronoun/reflexive, encouraged participants to process the referents differently, or put more focus on the characters in the story which could have led to the elevated effects of the inaccessible antecedent in the second study. On the other hand, the accessible antecedent appeared as a name in the introductory sentence and as a subject pronoun in the critical sentence. The subject pronoun signals an entity that is highly salient in the discourse (e.g. Givón 1983). Since pronouns are sensitive to discourse salience, this gives an advantage to the accessible antecedent over the inaccessible antecedent, but in spite of this the inaccessible antecedent has an effect. Nevertheless, the offline results show that participants were less sure of the eventual referent of a pronoun in the double match conditions, suggesting that they were less able to process these sentences properly. Despite having pictures on the screen to look at, the task is demanding in terms of memory because the stimulus is auditory. Indeed the combination of a demanding memory task, and the presentation of the pictures of the inaccessible antecedent, may combine to give rise to more robust interference effects than are seen in studies using other paradigms.

The three sets of studies presented in this section demonstrate that inaccessible antecedents can have an effect on the processing of a pronoun, thus posing a challenge to the BAIF hypothesis. The evidence fits into the hypothesis put forward by Badecker and Straub (2002), that multiple constraints in parallel affect the search for an antecedent. This is explained by Badecker and Straub in terms of activation: multiple cues contribute in parallel, positively or negatively, to an antecedent's activation. For an inaccessible antecedent, a binding constraint might contribute negatively to activation in comparison to an accessible antecedent while gender or

number feature match with the pronoun may contribute positively, leading to interference from a matching but inaccessible antecedent during processing.

This parallel constraints approach is very similar to the view put forward by Patil, Vasishth and Lewis (2011a; 2011b; 2012), who also advocate the importance of feature-matching and predict interference during retrievals during reflexive processing. This is within the framework of direct-access retrieval models (e.g. Lewis and Vasishth 2005; Lewis, Vasishth and Van Dyke 2006; Foraker and McElree 2007; Van Dyke 2007). These models make use of up to date research on memory and have been advanced also through the use of computational modelling. It has been proposed within this framework that long-distance dependencies, for example reflexive resolution, subject-verb agreement and filler-gap dependencies (among others), are supported by a single memory retrieval mechanism. Retrieval is achieved via cue matching; items are stored in memory with certain bundles of features (e.g. *singular, masculine*). When retrieval is triggered, for example on encountering a reflexive that has the features *singular* and *masculine*, the mechanism will attempt to match the cues from the trigger to the features of items stored in memory. The position of the item in memory is not important, because the cues give direct access to that item, so the number of intervening stored (non-matching) items should not affect the time it takes to carry out the retrieval. On the other hand, the activation of stored items does decay over time when they are not reactivated. A retrieval can be affected by interference, which occurs if the cues used for the retrieval partially match several items in memory. One issue of debate in this area has been the extent to which the same mechanism applies to various different dependencies. For example, if the retrieval mechanism is subject to interference effects during subject-verb agreement processing, as has been shown in various studies (e.g. Clifton *et al.* 1999; Wagers, Lau and Phillips 2009), and the same retrieval mechanism is used in reflexive processing, there should also be retrieval interference effects in reflexive processing (Patil *et al.*, 2011a; 2011b; 2012), and, by extension, all other dependencies that rely on the same kind of retrieval. This assumption has been challenged by Dillon (2011; Dillon *et al.*, 2013), who suggests that

retrieval for reflexives is guided mainly by structural cues, which does not lead to interference. It is not yet clear how structural cues such as c-command could be implemented within a direct access model; some suggestions are put forward in Alcocer and Phillips (submitted). But even assuming that structural cues of this kind can be implemented, whether or not certain cue types are more important in one type of dependency than in another is still a question that needs to be resolved.

2.3.3 The processing of SDPs

As described in section 2.2.4, SDPs represent an exception to condition B as formulated in BT. They might be accounted for by redefining the binding domain or by incorporating semantic information into the syntactic representation. The processing of SDPs, though, has rarely been investigated. The one study (that I am aware of) which investigated the processing of SDPs is Sekerina, Stromswold and Hestvik (2004), discussed below.

Sekerina *et al.*'s (2004) study on SDPs compared adults' and children's processing using the visual world method. An offline questionnaire was also carried out to test the interpretation of the type of structure used in the experiment. They were interested in the difference between processing pronouns and reflexives inside PPs, and this factor was manipulated in the questionnaire materials:

(2.15) In these pictures, you see a boy, a man and a box. The boy has placed the box on the ground. Which picture shows that the boy has placed the box behind {himself/him}?

The sentences were presented with two pictures; for the example given above, there was one picture in which the box is behind the boy, and one in which the box is behind the man. *The boy* is the sentence-internal referent, and is in focus in the discourse according to a number of factors – for example it is mentioned more times, it is the subject of the second sentence, it is mentioned first in the discourse. *The man* is the sentence-external referent. The task was to read the text and decide whether the sentence

matched one picture or the other, or both. I focus here on the results from the adult participants, and not from the children. In the reflexive conditions, participants chose the sentence internal referent 95% of the time, and in the pronoun conditions they chose the sentence-internal referent 79% of the time, indicating a preference for the sentence internal referent, and some referential uncertainty in the case of the pronouns compared to the reflexives. These offline results could, however, have been affected by the fact that there were no fillers in the questionnaire, so participants may have noticed that both pronouns and reflexives were being represented and attempted to distinguish their responses to each condition by choosing the sentence-internal referent less in the pronoun condition. Nevertheless, this offline task demonstrates that participants are happy to accept and may even prefer reference to an antecedent within the (BT defined) binding domain for pronouns.

The online task was very similar to the offline task, but eye movements were also recorded while the participants looked at the pictures. The 24 experimental items were mixed with 30 fillers. The middle sentence (see (2.15) above) was left out of the online task, so participants heard the introduction of the characters and objects, and were then asked the question "*Which picture shows...*". As well as eye movements, responses were recorded using a push button box, and the time taken to respond was analysed. Adults' responses to the questions matched those in the offline task. Response times in the pronoun condition were significantly slower than in the reflexive condition. The result from the eye-tracking data was that the proportion of fixations on the sentence-internal picture was significantly lower in the pronoun condition than the reflexive condition; this was from the onset of the pronoun/reflexive for three seconds. The difference remained stable for the whole period of time. These findings seem to suggest that, as far as pronouns in SDP structures are concerned, people have a general tendency to interpret them with a sentence-internal antecedent as opposed to a sentence external one, but that they experience some ambiguity during processing in comparison to reflexives.

There are a number of problems with this study. The number of adult participants was very low; only 16 ended up in the final analysis of the online part, and a large amount of data had to be discarded. More importantly, the low number of fillers in the online task (and total lack of fillers in the offline questionnaire) might have alerted the participants to the phenomenon that was being investigated, giving rise to response strategies. In the eye-tracking part of the experiment, participants had to make a decision on pictures that were ambiguous; an additional problem with the questions was that if participants always chose the sentence-internal picture, they would be correct (because it is the correct answer for the reflexives and is a possibility for the pronouns), so the picture stimuli were not balanced. Given the salience of the sentence-internal antecedent, it is not surprising that there was a preference for this referent as opposed to the sentence external one. Nevertheless, despite these difficulties, the results from this study do show that for pronouns in SDP structures, participants may experience some ambiguity and this is reflected in both the offline and online results. Participants seem to prefer the sentence-internal referent, which confirms that condition B does not exclude the antecedent within the binding domain in these cases.

While Sekerina *et al.* (2004) is the only online study of SDPs, there have been a number of studies (e.g. Runner, Sussman and Tanenhaus 2003; Runner, Sussman and Tanenhaus 2006; Kaiser *et al.* 2009) on pronouns and reflexives in picture noun phrases (PNPs), which are another exception to condition B. A full review of these studies is outside the scope of this chapter, but I give here a brief example of one experiment which demonstrates that offline choices and online processing are not always guided by BT. (2.16), from Kaiser *et al.* (2009) gives an example of a pronoun and a reflexive in a PNP:

(2.16) Peter {told/heard from} Andrew about the picture of
 {him/himself} on the wall.

Similar to SDPs, pronouns and reflexives inside PNPs appear not to be in complementary distribution, in violation of BT. BT predicts that the

reflexive must be bound by the subject, *Peter*, and that the pronoun cannot be bound by *Peter*. In one of the visual world and offline choice experiments reported in Kaiser *et al.* (2009), both the appearance of a reflexive or pronoun, and the type of main verb (the manipulations are shown in parentheses) were manipulated as shown in (2.16). The participants' choices about the pictures for reflexives showed a strong preference for the subject, *Peter*. For the pronouns, they demonstrated a preference for the object, *Andrew* (in accordance with BT) but this was not as strong a preference as the reflexive preference for subjects. The pronoun choices were also significantly affected by the verb manipulation: participants tended to choose the subject, *Peter*, more often when the verb made the subject the perceiver of the information (*heard from* in (2.16)). Eye-movement data reflected these choices, with the verbal manipulation affecting the eye movements in the pronoun condition more than in the reflexive condition. Participants began to respond to the semantic information within approximately 200ms of the onset of the pronoun.

Together with the results of Sekerina *et al.*'s (2004) study, it is clear that offline antecedent choices for pronouns can be incompatible with BT, and furthermore, the online processing of pronouns appears to be guided by semantic factors from an early stage.

In the following section I will summarise the main approaches to implementing condition B in processing models that I have presented.

2.3.4 Summary of proposed models and motivation for Experiments 1-3

I will briefly summarise the various processing models presented above regarding the online application of condition B. I will also consider the timing of the effects from some previous studies, and then motivate Experiments (1-3) which are reported in the following sections.¹⁵

¹⁵ The three experiments are also reported in Patterson, Trompelt and Felser (in prep.).

2.3.4.1 Models

Nicol and Swinney (1989) propose the BAIF account, in which binding conditions act from the beginning of the antecedent search to prevent inaccessible antecedents from ever being considered. Pragmatic, discourse and morphological information are brought into play after the binding constraints. Importantly, the BAIF account predicts no interference from inaccessible antecedents at any point. This model is supported by cross-modal priming evidence (Nicol 1988; Nicol and Swinney 1989) although one experiment does not entirely fit with the predictions. Self-paced reading studies (Clifton *et al.* 1997; Clifton *et al.* 1999) have also found no interference from inaccessible antecedents, and there is also evidence from Lewis and Chow (2012) from eye tracking.

Sturt's (2003) study on reflexive processing also supports an initial stage in which binding constraints are applied, like the BAIF, but also suggests that binding conditions can be overridden at a later integration stage, unlike the BAIF. Evidence for this is from reflexives only, not pronouns.

Badecker and Straub (2002) propose a parallel constraints model in which both binding constraints and cues such as gender and number affect the activation of antecedents simultaneously, leading to interference effects from inaccessible antecedents based on feature match or mismatch. This is supported by their self-paced reading/probe recognition experiments, to some extent by the findings from Kennison (2003) (but note that the effect is very late and can be attributed to a lack of an accessible antecedent), and by Clackson *et al.*'s (2011; 2012) visual world eye-tracking experiments. The importance of gender and/or number features in retrieval is also emphasised by Patil *et al.* (2011a; 2011b; 2012) whose model predicts interference effects from inaccessible antecedents, although it is not clear how far the assumptions for reflexives can be applied to pronouns. I will refer to this approach as the *feature matching hypothesis*.

2.3.4.2 Timing

It is not always possible to make comparisons about timing from studies that have employed such a range of experimental techniques. However,

since I am using a time-sensitive method in my online experiments, I will consider the timing of both interference effects and effects from accessible antecedents. Overall, the timing of the effects seen in the above studies for pronouns is usually somewhat delayed. The timing profile of interference effects was very late in Kennison's (2003) study, occurring two or three words after the pronoun, and Badecker and Straub's (2002) effects were in the region following the pronoun, not on the pronoun itself. Clackson's studies (Clackson *et al.* 2011; 2012) used a more time-sensitive paradigm, but failed to find a profile that altered over time as Sturt (2003) did for reflexives; rather, they found a stable interference effect that began approximately 200ms after the onset of the pronoun and remained constant during the 2000ms duration that was examined in the analysis.¹⁶ The evidence regarding timing mainly refers to the timing of interference effects, rather than effects induced by a mismatching accessible antecedent. Badecker and Straub's study does however provide evidence regarding a mismatching accessible antecedent – as with the interference effect, this effect is seen in the region following the pronoun rather than on the pronoun itself.

2.3.4.3 The current experiments

The range of evidence and different models presented above demonstrates that it is far from clear whether or not inaccessible antecedents have an effect on the processing of pronouns. This issue deserves further exploration, and further eye-tracking evidence would be welcome. It should also be possible to gain further insight into the timecourse of processing pronouns with two sentence-internal antecedents using this method. Three experiments (one offline, two online) are described below which aim to further explore the application of condition B during processing and the consideration of multiple antecedents, using structures that are typically subject to condition B (Experiment 2) and structures

¹⁶ 200ms is approximately the amount of time taken to programme an eye movement.

containing SDPs (Experiment 3).¹⁷ The following questions will be explored:

- Does the inaccessible antecedent affect pronoun processing at any point?
- Does the order/timing of considering the two antecedents differ according to whether or not condition B applies?

In the condition B environments the accessible antecedent is always the first-mentioned antecedent. If, as expected, sensitivity to the accessible antecedent in the condition B environments is observed, it might then be attributable to a preference for a first-mentioned antecedent (Gernsbacher and Hargreaves 1988; Gordon, Grosz and Gilliom 1993; Arnold, Brown-Schmidt and Trueswell 2007; see also Chapter 4 for a discussion of a first-mention preference). An additional motivation for Experiment 3 is therefore to check for such a preference in the SDP environments.

As well as looking at evidence from the two online experiments, it is important to consider how pronouns in these two structure types are interpreted offline. This is especially important for the SDPs because they are theoretically ambiguous.

2.4 Experiment 1: the offline interpretation of pronouns in canonical condition B and SDP environments

Experiment 1 was a questionnaire task conducted via the internet. As stated above, the purpose of Experiment 1 was to check the offline interpretations of pronouns in the two syntactic environments (canonical condition B and SDP) that were tested in the online Experiments 2 and 3. This is particularly important because Experiments 2 and 3 did not include

¹⁷ Note that the experimental items for Experiments 2 and 3 were combined and run in one experimental session.

an offline pronoun interpretation component in order to avoid drawing attention to the phenomena under investigation.

2.4.1 Experiment 1 materials

Materials for Experiment 1 were composed of two experimental sentence types, all single sentences, and fillers. The full set of experimental materials is provided in Appendix 1. There were ten canonical condition B sentences, each of which contained an object pronoun and two preceding potential antecedents, both matching the pronoun in gender. Half the condition B items contained a masculine pronoun and half a feminine pronoun. The antecedent nearest the pronoun was inaccessible according to condition B. Antecedents were a mixture of proper names and nouns with definitional gender. The verbs in the complementizer phrase (beginning at *that*) are all interpretable when followed by a reflexive, to avoid any biases arising from the verbs themselves. An example condition B item is given below in (2.17):

(2.17) The boy remembered that Matthew had bought him a new computer game.

There were ten SDP sentences, each of which contained an object pronoun and two preceding potential antecedents, both matching the pronoun in gender. Half the items contained a masculine pronoun and half a feminine pronoun. The pronoun was inside a PP, such that neither antecedent was ruled out on structural grounds. An effort was made to ensure that neither antecedent was ruled out on grounds of plausibility. Each sentence contained a PP to set the scene (see examples) which appeared either at the start or at the end of the sentence (half each). Antecedents were a mixture of proper names and nouns with definitional gender. An example of a PP-first SDP item and a PP-last SDP item is given below:

(2.18a) *PP-first*

At the post office, Ernie watched the boy pick up the parcel next to him.

(2.18b) *PP-last*

Harry heard William pull the curtain around him in the quiet hospital ward.

There were 22 filler sentences. These consisted of ten sentences containing quantifier phrases as antecedents (reported as Experiment 7 in Chapter 4, section 4.7, and listed in Appendix 1) and 12 other sentences containing two antecedents and one pronoun/reflexive in various configurations: four sentences contained reflexives; two sentences followed the condition B structure, but had antecedents of different genders; one sentence contained backwards anaphora; five sentences contained subject pronouns. Nine sentences were unambiguous because they contained pronouns that matched in gender with only one antecedent.¹⁸ These unambiguous filler sentences were used in order to check that participants were engaged with the task. The responses are discussed below in the results section.

2.4.2 Experiment 1 participants

Participants were recruited via email and word of mouth to people who were known to be native speakers of English. The questionnaire was also advertised on an English-language forum on the internet. Because the experiment was carried out remotely and not directly under the control of an experimenter, a larger number of participants was recruited. There were 83 participants (33 male), mean age 40 years (age range 19-72 years). All participants reported that they were native speakers of English. No participants reported having dyslexia or any language disorders. 87% chose British English as their variety of English; this matches the profile of the participants in the online eye-tracking study which was carried out in Essex, UK. The educational profile is also similar to the participants in the

¹⁸ While these sentences were not completely unambiguous, because pronouns can refer to antecedents outside the immediate discourse, they were unambiguous with regard to the task in the questionnaire. As described in the following section, participants were given three options for pronoun reference; antecedent 1, antecedent 2 or either. For these unambiguous items only one answer was correct, because choosing the other two options would result in a gender violation.

online experiments: 88% of participants in Experiment 1 were university educated.

2.4.3 Experiment 1 procedure

The questionnaire was administered via the internet using SurveyGizmo (surveygizmo.com). The first page of the questionnaire was a welcome page, telling participants what to expect in terms of their personal data, the estimated time the survey would take and a reminder that only native English speakers should take part. The second page collected participant data (summarised above), and summarised the data and consent issues. The third page contained the instructions and two examples.

Participants were asked to read each sentence and decide who the pronoun probably referred to. The word *probably* was included in the instructions to account for the fact that a pronoun can refer to an entity outside of the discourse. Because the sentences were presented in isolation, without any surrounding context giving rise to other potential antecedents, it is unlikely that participants would interpret the pronoun as referring to an extra-discourse antecedent. The use of *probably* takes account of the fact that another interpretation is possible, although unlikely. After each sentence the same question appeared: “Who does [pronoun] refer to?”. They were given three choices – *antecedent 1*, *antecedent 2* or *either*. An example is given below.

(2.19)

1. The boy remembered that Matthew had bought him a new computer game.

Who does 'him' refer to?

The boy

Matthew

Either

The order of the two antecedents was varied throughout the questionnaire. The *either* option always appeared at the bottom.

The items were pseudo-randomised so that no two items of the same type appeared next to each other, and so that the sentence types were distributed evenly throughout the questionnaire.

2.4.4 Experiment 1 results

Nine unambiguous items from the set of fillers (as described in the materials section) were checked to ensure that participants were giving sensible answers to the questionnaire. Most participants (88%) gave the expected answers for all nine items. Ten participants gave fewer than nine expected answers. Of these, eight participants gave eight expected answers (score: 89%) and two gave seven expected answers (score: 78%). Given the high percentages of expected answers, all 83 participants were left in the analysis for the experimental items. The overall percentage of expected answers was 98%. *Either* was chosen at a rate of 0.5%, NAs were 0.5%, and 1% of answers were for the non-matching antecedent.

2.4.4.1 Condition B environments

Responses were coded as *either*, *accessible antecedent* or *inaccessible antecedent*. One item (A-08) was removed from the analysis because it contained the pronoun *her* followed by a bare noun phrase; this makes *her* ambiguous, as it can either be an object pronoun or a possessive pronoun. This left nine experimental items for analysis. The overall percentage of *accessible antecedent* responses was 98% (*either* responses, 1%; *inaccessible antecedent* responses, 1%). A Wilcoxon signed rank test confirmed that the participant and item percentages of *accessible antecedent* responses differed significantly from chance-level (33%): $V_1=3486$, $p_1<0.01$; $V_2=45$, $p_2<0.01$.

2.4.4.2 SDP environments

Responses were coded as *either*, *local antecedent* or *non-local antecedent*. The non-local antecedent was the subject of the matrix clause. Overall percentages were as follows: *either*, 60%; *local antecedent*, 18%; *non-local antecedent*, 22%. A Wilcoxon signed rank test confirmed that the percentages for *either* were significantly higher than chance level (33%): $V_1=3092$, $p_1<0.01$; $V_2=54$, $p_2<0.01$. This confirms that participants found the SDP structures ambiguous, and the majority of the time they chose the

either option. However, participants also responded *local antecedent* or *non-local antecedent* 40% of the time. A Wilcoxon signed rank test (paired) compared the percentages for *local antecedent* against the percentages for *non-local antecedent*. There was no significant difference between them: $V_1=773.5, p_1=0.15; V_2=23.5, p_2=0.72$). When participants did not give the *either* response, there was no clear preference for the *local* or the *non-local antecedent*; these options were chosen in equal amounts.

2.4.4.3 Summary

To summarize, Experiment 1 shows that for pronouns in condition B environments, participants chose the accessible antecedent almost all of the time. There was much less certainty about pronouns in the SDP structures; in general, they were found to be ambiguous as indicated by the largest number of responses for the *either* option, however, the *local antecedent* and *non-local antecedent* options were chosen (in equal amounts) at a rate of 40%, reflecting uncertainty about the interpretation of these pronouns but not showing a preference for local or non-local antecedents.

The next two sections describe Experiments 2 and 3, the online experiments. The experimental items for Experiments 2 and 3 were combined and run in one experimental session, but they are analysed and reported separately.

2.5 Experiment 2: processing pronouns in condition B environments

Experiment 2 was an eye-tracking study investigating the processing of pronouns in condition B environments.

2.5.1 Experiment 2 materials

24 experimental items were constructed. They were composed of three sentences, a lead-in sentence, a critical sentence that contained the pronoun and two potential antecedents, and a wrap-up sentence. The verbs in the complementizer phrase (beginning at *that*) were all interpretable when followed by a reflexive, to avoid any biases arising from the verbs

themselves. In the critical sentence, the two antecedents appeared before the pronoun; they were both names. The names were matched in letter and syllable length, and were typically masculine or typically feminine names. Names were counter-balanced across the items to avoid any frequency effects. The first name (henceforth the non-local antecedent) was always the first word of the critical sentence and was the subject of the main verb. It was always an accessible antecedent by virtue of being outside the local domain and therefore was exempt from being ruled out by condition B. The second name (henceforth the local antecedent) was always the subject of the verb in the complementizer phrase, introduced by *that*. The local antecedent was always an inaccessible antecedent because it was in the local domain of the pronoun and c-commanded the pronoun, which means it is ruled out as an antecedent for the pronoun according to condition B. Half the pronouns were masculine and half were feminine. They were always object pronouns. The gender match between the two names and the pronoun was manipulated to create three conditions. Examples are given below. A full list of the experimental items is given in Appendix 2.

(2.20a) *Double match condition*

Band practice was beginning to get rather dull.

John remembered that Mark had taught him a new song on the guitar.

That really lifted everyone's spirits!

(2.20b) *Local mismatch condition*

Band practice was beginning to get rather dull.

John remembered that Jane had taught him a new song on the guitar.

That really lifted everyone's spirits!

(2.20c) *Non-local mismatch condition*

Band practice was beginning to get rather dull.

Jane remembered that John had taught him a new song on the guitar.

That really lifted everyone's spirits!

A fourth condition, in which both the local and the non-local antecedent mismatched in gender with the pronoun, was not included. This was because the items in the non-local mismatch condition (2.20c) were expected to create disruption because they did not present a gender-matching antecedent. Including a fourth no-match condition would have doubled the number of disruptive items that participants saw, which could have drawn participants' attention to the experimental items, thereby affecting normal processing.

2.5.2 Experiment 2 predictions

BAIF hypothesis

If condition B is applied before any other type of information is considered, the local antecedent will be ruled out in all conditions, leaving only the non-local antecedent for consideration. This should result in increased reading times when the non-local antecedent mismatches in gender with the pronoun (2.20c) compared to the other conditions due to increased processing difficulty. There should be no interference from the inaccessible antecedent. This means that there should be no difference between (2.20a) and (2.20b) because the gender match between the pronoun and the local antecedent should not have any effect, having already been ruled out of consideration by condition B.

Defeasible filter hypothesis

If Sturt's (2003) findings for reflexives can be directly transferred to pronoun processing there should be an effect of non-local mismatch, with longer reading times in (2.20c), as above, followed by an effect of local match (differences in reading time between (2.20a) and (2.20b)). These effects would be due to (i) an earlier bonding stage in which binding condition B rules out the local antecedent, leaving only the non-local antecedent for consideration which causes processing difficulty in the non-local mismatch condition; and (ii) a later integration stage in which binding conditions can be violated so that the local antecedent can be considered. Consideration of the local antecedent may lead either to longer reading times when the local antecedent mismatches ($2.20b > 2.20a$) or to longer

reading times when the local antecedent matches (2.20a > 2.20b) because of competition between the two antecedents.

Feature matching hypothesis

If feature matching plays as important a role, or indeed a more important role, than binding condition B in the initial retrieval of antecedents, then the gender match of the local antecedent should affect processing. This will lead either to longer reading times when the local antecedent mismatches (2.20b > 2.20a), or to longer reading times when the local antecedent matches (2.20a > 2.20b) because of retrieval interference. Note that in Badecker and Straub's (2002) study, the double match condition had longer reading times following the pronoun than the accessible match condition (2.20a > 2.20b). Under this model there should also be a mismatch effect of the non-local antecedent (longer reading times in (2.20c) compared to (2.20b)) when it is identified as the only accessible antecedent.

Predictions regarding the timing of effects

Predictions regarding the timing of all effects very difficult to make in all three prediction sets, partly because previous studies have used different methodologies. Activation effects of the accessible antecedent in Nicol and Swinney (1989) occurred at the pronoun region, but the reaction times measured were not reading times of the pronoun itself but to responses to a related task, so it is unclear how immediate this effect really is. While the application of condition B is rapid in the BAIF, leading to the immediate exclusion of the inaccessible antecedent, the rapid identification of the *correct* antecedent cannot be assumed. The defeasible filter hypothesis does predict a sequence for the effects, with effects for the accessible antecedent occurring before effects for the inaccessible antecedent. For the feature matching hypothesis, interference is normally assumed to be an early effect in retrieval models, but Badecker and Straub's (2002) findings were mainly in regions downstream of the pronoun, which is clearly not the earliest measurable time window in a self-paced reading study. For this reason I will not make specific predictions regarding the timing of effects. I will, however, consider the issue of timing as part of the discussion.

2.5.3 Experiments 2 and 3 method

The materials were arranged in a Latin-square design, resulting in three lists. The order of items in the lists were reversed to create a further three lists, to minimise learning or tiredness effects on particular items. In addition to the 24 experimental items for Experiment 2 and 18 experimental items for Experiment 3, there were 28 fillers and 16 pseudo-fillers. The fillers contained a variety of sentence structures, different to both types of experimental items but of approximately equivalent length. The 16 pseudo-fillers were of two types: eight items contained reflexives, and eight items contained object pronouns where the inaccessible antecedent was in first-mention position. The pseudo-fillers were included so that participants did not habituate to the type of anaphoric element or the position of the inaccessible antecedent in the experimental items, thus minimizing the risk of participants developing a strategy. Following 59 of the trials (approximately two thirds) there were yes/no comprehension questions. This was to ensure that the participants were reading and understanding the trials that were presented. Comprehension questions did not explicitly probe the referent of the pronoun except in two questions, in order not to draw attention to the purpose of the experiment or encourage strategic reading behaviour.

2.5.3.1 Experiments 2 and 3 participants

Participants were 36 native speakers of English (11 males) who were recruited from the University of Essex and the surrounding community. Participants' mean age was 26 years (range 18-54 years). All confirmed English was their first language, and three participants reported an additional first language. 32 of the participants (89%) were educated to university level. All participants were paid £6 for their participation. Some participants wore contact lenses or glasses during the experiment, and all reported that they could read from the screen successfully.

2.5.3.2 Experiments 2 and 3 procedure

The data was recorded using the Eyelink 1000 system (SR Research, Canada). Participants were seated in front of a PC monitor at a distance of 70cm, with their chin on a chinrest and their forehead resting against a bar.

The camera which tracked their eye movements was placed on the desk below the monitor. Before the experiment began, the recording of participants' eye movements was calibrated on a nine-point grid. The text on the screen was presented in Courier New font, size 18, and displayed across up to three lines of text on screen. Text was displayed in black on a white background. Viewing was binocular, but only data from the right eye was recorded, with the pupil being tracked at a 500Hz sampling rate. There was a drift correction procedure before each trial in which a target appeared in the top left of the screen (above the position where the text would start in the next trial), and participants had to press a button while fixating on the target. After set up, participants were presented with a screen containing instructions, followed by practice trials, two of which were followed by practice comprehension questions. Responses to the comprehension questions and to the drift correction screen were made using a USB gamepad connected to the host PC. Participants were instructed to sit still for the duration of the experiment. Following two thirds of the trials in the experiment there were comprehension questions. This was to ensure that the participants were reading and understanding the trials that were presented. Comprehension questions did not explicitly probe the referent of the pronoun, in order not to draw attention to the purpose of the experiment or encourage strategic reading behaviour. After the eye-tracking part of the experiment was complete, participants took a working memory test.¹⁹ The experiment session took around 50 minutes in total.

2.5.3.3 Data preparation

Individual fixations shorter than 80ms and within one degree of visual angle of a neighbouring fixation were merged with the neighbouring fixation. Other fixations shorter than 80ms, and longer than 1200ms, were

¹⁹ A working memory test was included in the experimental session because previous studies have revealed an effect of working memory capacity on the processing of reflexives (e.g. Cunnings and Felser 2013). An initial analysis of the current data using ANOVAs revealed that working memory score did not have an effect on the results of the current experiment, and the working memory score was not included in the mixed models analysis.

removed. Within a trial, regions that were initially skipped were removed from the analysis. Three individual trials were removed from the analysis due to excessive track loss, representing 0.2% of the data. The eye-movement measures examined in Experiments 2 and 3 were first-fixation times, first-pass times, regression-path times, rereading times and total-viewing times (eye-movement measures are described in Chapter 1, section 1.5.2).

The experimental items were divided into regions of interest for the purpose of the analysis. Eye-movement data from the following five regions was analysed for Experiment 2: precritical, pronoun, spillover, prefinal and final. (2.21) shows how the text in the critical sentence was divided into regions.

(2.21)

Band practice was beginning to get rather dull.

John remembered that Mark had | taught | him | a new | song on | the guitar. |

That really lifted everyone's spirits!

The **precritical region** contained the word before the pronoun. It always contained a single word which was the past participle of the verb in the complementizer phrase (*taught* in (2.21)). First-fixation durations, first-pass times and regression-path times in the precritical region are measured in this region before the pronoun is encountered. The earlier measures in this region are therefore inspected to ensure that any effects seen in the pronoun region (or later) begin during or after the pronoun is encountered. Rereading times in the precritical region reflect processing after the pronoun is encountered. Total viewing times in this region, being an omnibus measure composed of first-pass times and rereading times, could reflect processing before or after processing of the pronoun. The **pronoun region** contained only the pronoun (*him* in (2.21)). The **spillover region** contained the two words following the pronoun. The **prefinal region** contained the two words following the spillover region, and the **final region** contained the two words following the prefinal region. The final region was always the end of the sentence.

2.5.4 Experiment 2 results

2.5.4.1 Comprehension questions

The overall correct response rate to the comprehension questions was 88%, range 72-96%.

2.5.4.2 Eye-movement data

The skipping rates in the analysed regions were as follows: precritical, 13%; pronoun, 48%; spillover, 13%; prefinal, 11%; final, 8%. Before the statistical analysis the dataset was transformed using a logarithmic transformation ($\log_e(x+1)$).²⁰ Untransformed and transformed data was plotted against a normal distribution line and visually inspected. Where the transformed data better fitted the normal distribution, it was used in the analysis in place of the untransformed data. This was the case for all regions and measures except the rereading times (all regions). Untransformed data was used in the analysis of the rereading times. Absent rereading times contributed a value of zero to the rereading times analysis, as per Sturt (2003).

The data was analysed using linear mixed-effects modelling, using the lme4 package in R (Bates, Maechler and Bolker 2012). For the fixed part of the model, the three-level factor *condition* was included. The baseline for *condition* was the local mismatch condition (2.20b). An additional model was created with the factor *condition* re-levelled so that the baseline was the double match condition (2.20a); this was to obtain a model estimate of the comparison between conditions a and c. For the random part of the model, the full random effects structure (by-participant and by-item random intercepts and slopes) was preferred (Barr *et al.* 2013). However, in cases of non-convergence, the random slopes were removed from the model one by one, beginning with the by-item random slope, until a converging model was found (as suggested in Barr *et al.* 2013). *P*-values determining

²⁰ $\log_e(x+1)$: 1 was added to each point in the dataset before transformation so that no datapoint would equal zero (zero cannot be log-transformed).

significance of the t -value were derived from the upper bound of the t -statistic (Baayen 2008).

The presentation of the means and statistical analyses of the eye-movement data is divided into two sections, *earlier measures* and *later measures*. First-fixation durations, first-pass times and regression-path times are presented under *earlier measures*, and rereading times and total viewing times are presented under *later measures*. The classification of regression-path times as *earlier* or *later* is rather arbitrary, and its inclusion under *earlier measures* here is merely for convenience of presentation. Similarly, total viewing times is not a late measure but a cumulative one, but it has been placed in the *late* category simply for ease of presentation.

Earlier measures

Results for the first-fixation durations, first-pass times and regression-path times are presented first. Table 2.1 below shows the means in the earlier measures for the precritical, pronoun, spillover, prefinal and final regions. The outcome of the statistical analysis for these regions and measures is presented in Table 2.2. The results from the earlier measures in each region are subsequently summarised.

Table 2.1. Means (in milliseconds) for first-fixation durations, first-pass times and regression-path times in each condition in the precritical, critical, spillover, prefinal and final regions in Experiment 2. Standard deviations are given in italics.

	Condition					
	Double match		Local mismatch		Non-local mismatch	
Precritical region						
First-fixation duration	221	<i>(78)</i>	221	<i>(73)</i>	218	<i>(74)</i>
First-pass time	261	<i>(119)</i>	253	<i>(106)</i>	264	<i>(119)</i>
Regression-path time	356	<i>(258)</i>	341	<i>(301)</i>	340	<i>(339)</i>
Pronoun region						
First-fixation duration	219	<i>(72)</i>	220	<i>(65)</i>	221	<i>(76)</i>
First-pass time	227	<i>(97)</i>	235	<i>(83)</i>	234	<i>(109)</i>
Regression-path time	307	<i>(307)</i>	277	<i>(180)</i>	317	<i>(324)</i>
Spillover region						
First-fixation duration	208	<i>(76)</i>	213	<i>(72)</i>	205	<i>(59)</i>
First-pass time	259	<i>(163)</i>	262	<i>(164)</i>	264	<i>(153)</i>
Regression-path time	366	<i>(393)</i>	336	<i>(315)</i>	434	<i>(465)</i>
Prefinal region						
First-fixation duration	210	<i>(78)</i>	205	<i>(71)</i>	205	<i>(77)</i>
First-pass time	294	<i>(162)</i>	295	<i>(170)</i>	286	<i>(159)</i>
Regression-path time	378	<i>(271)</i>	422	<i>(390)</i>	505	<i>(657)</i>
Final region						
First-fixation duration	214	<i>(84)</i>	224	<i>(105)</i>	212	<i>(91)</i>
First-pass time	308	<i>(200)</i>	319	<i>(196)</i>	303	<i>(196)</i>
Regression-path time	915	<i>(1368)</i>	837	<i>(1044)</i>	1196	<i>(1670)</i>

Table 2.2. Summary of statistical analyses for first-fixation durations, first-pass times and regression-path times in the precritical, pronoun, spillover, prefinal and final regions of Experiment 2. The estimate for each contrast and its corresponding standard error and t-value is given. SE=standard error. (*) = $p \leq .10$, * = $p < .05$.

	First-fixation durations		First-pass times		Regression-path times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region						
<i>double match (a) vs local mismatch (b)</i>	<0.01 (0.03)	0.12	-0.02 (0.03)	-0.69	-0.06 (0.05)	-1.08
<i>double match (a) vs non-local mismatch (c)</i>	-0.01 (0.03)	-0.49	-0.03 (0.03)	-0.89	-0.08 (0.05)	-1.55
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.02 (0.03)	-0.61	-0.01 (0.03)	-0.20	-0.02 (0.05)	-0.38
Pronoun region						
<i>double match (a) vs local mismatch (b)</i>	0.01 (0.03)	0.19	0.04 (0.04)	1.03	0.01 (0.06)	0.15
<i>double match (a) vs non-local mismatch (c)</i>	-0.01 (0.03)	-0.42	-0.01 (0.04)	-0.23	0.02 (0.07)	0.35
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.02 (0.03)	-0.59	-0.05 (0.04)	-1.19	0.02 (0.07)	0.24
Spillover region						
<i>double match (a) vs local mismatch (b)</i>	0.03 (0.03)	0.95	0.03 (0.05)	0.54	-0.02 (0.07)	-0.37
<i>double match (a) vs non-local mismatch (c)</i>	< -0.01 (0.03)	-0.09	0.03 (0.05)	0.57	0.14 (0.07)	1.92(*)
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.03 (0.03)	-1.19	-0.01 (0.04)	-0.15	0.16 (0.07)	2.32*
Prefinal region						
<i>double match (a) vs local mismatch (b)</i>	-0.02 (0.03)	-0.77	< -0.01 (0.05)	-0.08	0.02 (0.06)	0.38
<i>double match (a) vs non-local mismatch (c)</i>	-0.03 (0.03)	-1.11	-0.03 (0.04)	-0.80	0.08 (0.07)	1.27
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.01 (0.03)	-0.30	-0.03 (0.05)	-0.63	0.06 (0.07)	0.84
Final region						
<i>double match (a) vs local mismatch (b)</i>	0.02 (0.03)	0.70	0.03 (0.04)	0.72	-0.02 (0.07)	-0.24
<i>double match (a) vs non-local mismatch (c)</i>	-0.02 (0.03)	-0.59	-0.03 (0.04)	-0.67	0.17 (0.07)	2.33*
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.04 (0.03)	-1.30	-0.06 (0.04)	-1.41	0.18 (0.07)	2.57*

Precritical region

Differences between conditions in earlier measures of the precritical region are not expected, because the material in the three conditions is identical up to this point except for the two names (which are matched for letter and syllable length). First-fixation, first-pass times and regression-path times in the precritical region are very similar in all three conditions. There is no effect of condition in any of the earlier measures.

Pronoun region

First-fixation durations and first-pass times are similar across all conditions in this region. Differences between the conditions emerge in the regression-path times which are longest in the non-local mismatch condition and shortest in the local mismatch condition. However, the high standard deviations for this measure indicate that the actual values are dispersed widely around the means and the statistical analysis reveals that there is no significant effect of condition at this point.

Spillover region

First-fixation durations and first-pass times are similar across all conditions in this region. Some differences emerge in the regression-path times. The longest regression-path times are in the non-local mismatch condition and the shortest are in the local mismatch condition. The statistical analysis reveals a marginally significant difference between the double match and non-local mismatch conditions (2.20a-2.20c), and a significant difference between the local mismatch and non-local mismatch conditions (2.20b-2.20c).

Prefinal region

In this region again we see very little difference between the conditions in the first fixation durations or the first-pass times. Similar to the pronoun and spillover regions, the non-local mismatch condition has the longest regression-path times. However, there was no significant effect of condition in any of the measures in this region.

Final region

In the final region the two earlier measures show small differences, with longer first-fixation durations and first-pass times in the local mismatch condition. This is not confirmed in the statistical analysis, which showed no significant effects in these two measures. There is a clear main effect of condition in the regression-path times, with the longest regression-path times in the non-local mismatch condition (2.20c) and the shortest in the local mismatch condition (2.20b). This is confirmed statistically, with significant differences between both the double match and non-local mismatch conditions (2.20a–2.20c) and the local mismatch and non-local mismatch conditions (2.20b–2.20c).

Later measures

Rereading and total viewing times for the precritical, critical, spillover, prefinal and final regions are presented in this section. Table 2.3 shows the means for the later measures in all regions. The outcome of the statistical analysis is presented in Table 2.4. The results for each region are subsequently summarised.

Table 2.3. Means (in milliseconds) for rereading times and total viewing times in each condition in the precritical, critical, spillover, prefinal and final regions of Experiment 2. Standard deviations are given in italics.

	Condition					
	Double match		Local mismatch		Non-local mismatch	
Precritical region						
Rereading times	160	<i>(231)</i>	151	<i>(239)</i>	267	<i>(371)</i>
Total viewing times	421	<i>(252)</i>	404	<i>(265)</i>	521	<i>(377)</i>
Pronoun region						
Rereading times	102	<i>(179)</i>	71	<i>(141)</i>	199	<i>(348)</i>
Total viewing times	329	<i>(215)</i>	305	<i>(177)</i>	433	<i>(356)</i>
Spillover region						
Rereading times	152	<i>(236)</i>	151	<i>(240)</i>	257	<i>(350)</i>
Total viewing times	411	<i>(274)</i>	412	<i>(286)</i>	522	<i>(384)</i>
Prefinal region						
Rereading times	172	<i>(253)</i>	159	<i>(221)</i>	214	<i>(312)</i>
Total viewing times	466	<i>(297)</i>	454	<i>(278)</i>	500	<i>(348)</i>
Final region						
Rereading times	126	<i>(255)</i>	105	<i>(208)</i>	120	<i>(226)</i>
Total viewing times	434	<i>(342)</i>	424	<i>(296)</i>	424	<i>(311)</i>

Table 2.4. Summary of statistical analyses for rereading times and total viewing times in the precritical, pronoun, spillover, prefinal and final regions of Experiment 2. The estimate for each contrast and its corresponding standard error and t-value is given. SE=standard error. (*) = $p \leq .10$, * = $p < .05$, ** = $p < .01$.

	Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region				
<i>double match (a) vs local mismatch (b)</i>	-10.45 (22.93)	-0.46	-0.06 (0.05)	-1.15
<i>double match (a) vs non-local mismatch (c)</i>	106.67 (32.69)	3.26**	0.18 (0.07)	2.67*
<i>local mismatch (b) vs non-local mismatch (c)</i>	116.22 (35.21)	3.30**	0.24 (0.07)	3.39**
Pronoun region				
<i>double match (a) vs local mismatch (b)</i>	-30.24 (24.58)	-1.23	-0.02 (0.06)	-0.34
<i>double match (a) vs non-local mismatch (c)</i>	83.01 (33.77)	2.46*	0.18 (0.07)	2.67*
<i>local mismatch (b) vs non-local mismatch (c)</i>	113.25 (34.96)	3.24**	0.20 (0.07)	2.76*
Spillover region				
<i>double match (a) vs local mismatch (b)</i>	-1.12 (23.39)	-0.05	<0.01 (0.06)	-0.05
<i>double match (a) vs non-local mismatch (c)</i>	102.45 (33.78)	3.03**	0.21 (0.06)	3.32**
<i>local mismatch (b) vs non-local mismatch (c)</i>	103.57 (32.35)	3.20**	0.21 (0.06)	3.43**
Prefinal region				
<i>double match (a) vs local mismatch (b)</i>	-11.42 (20.90)	-0.55	-0.02 (0.05)	-0.38
<i>double match (a) vs non-local mismatch (c)</i>	41.20 (27.19)	1.52	0.04 (0.05)	0.78
<i>local mismatch (b) vs non-local mismatch (c)</i>	53.15 (26.64)	2.00(*)	0.06 (0.05)	1.12
Final region				
<i>double match (a) vs local mismatch (b)</i>	-24.80 (21.41)	-1.16	-0.01 (0.04)	-0.33
<i>double match (a) vs non-local mismatch (c)</i>	-8.15 (19.35)	-0.42	-0.02 (0.04)	-0.48
<i>local mismatch (b) vs non-local mismatch (c)</i>	16.99 (18.69)	0.91	-0.01 (0.04)	-0.13

Precritical region

In the precritical region the rereading and total viewing times in the non-local mismatch condition (2.20c) were much higher in comparison to the double match and the local mismatch conditions. This was confirmed statistically, with rereading and total viewing times being significantly longer in the non-local mismatch condition compared to the other two conditions. There is also a slight tendency (not statistically significant) for the local mismatch condition to elicit the shortest total viewing times, although this total could have been affected by a similar tendency in the first-pass times of this region.

Pronoun region

Similar to the the precritical region, there was a small difference between the double match (2.20a) and the local mismatch (2.20b) conditions in both measures of the pronoun region, but much longer rereading and total viewing times in the non-local mismatch condition (2.20c). This was confirmed statistically, with rereading and total viewing times being significantly longer in the non-local mismatch condition compared to the other two conditions. The small advantage for the local mismatch condition in both measures was not statistically significant.

Spillover region

In the spillover region rereading times and total viewing times are again significantly longer in the non-local mismatch condition (2.20c) compared to either the double match condition or the local mismatch condition. There is no difference between the double match (2.20a) and the local mismatch (2.20b) conditions.

Prefinal region

Rereading times and total viewing times are again longest in the non-local mismatch condition (2.20c) in the prefinal region. However, the differences do not reach statistical significance in this region, although the contrast between the local mismatch (2.20b) and the non-local mismatch (2.20c) conditions is marginally significant. The slight advantage for the local

mismatch condition is also seen in the prefinal region but again it is not significant.

Final region

Rereading times are shortest in the local mismatch condition (2.20b) and similar in the other two conditions. Total viewing times are slightly longer for the double match condition (2.20a) than for the other two conditions. However none of these small differences reach statistical significance.

2.5.4.3 Summary of results from Experiment 2

There were no differences between the conditions in the very earliest measures (first fixation times and first-pass times) in any of the regions analysed, and on the whole there were no numerical trends indicating differences between the conditions. Differences between the conditions do emerge in the regression-path times. In the spillover region there was a marginally significant difference between the double match (2.20a) and the non-local match (2.20c) conditions, and a significant difference between the local mismatch (2.20b) and the non-local mismatch condition (2.20c). In the final region, regression-path times were significantly longer in the non-local mismatch condition (2.20c) than both the double match condition (2.20a) and the local mismatch condition (2.20b).

Turning now to the later measures, rereading and total viewing times, the pattern of results is similar across the precritical, pronoun and spillover regions. In all of these regions the non-local mismatch condition (2.20c) had longer rereading and total viewing times than the other two conditions. There were no significant differences between the double match (2.20a) and the local mismatch (2.20b) conditions. The non-local mismatch condition had longer rereading times than the local mismatch condition in the prefinal region, but this difference was only marginally significant.

These results will be discussed in section 2.7 after the results from Experiment 3 are presented.

2.6 Experiment 3: processing pronouns in SDP environments

Experiment 3 was an eye-tracking study investigating the processing of pronouns in SDP environments, as motivated above in section 2.3.4.

2.6.1 Experiment 3 materials

18 experimental items were constructed for Experiment 3. They were composed of three sentences, a lead-in sentence, a critical sentence that contained the pronoun and two potential antecedents, and a wrap-up sentence. In the critical sentence, the two antecedents appeared before the pronoun and they were both names. The names were matched in letter and syllable length, and were typically masculine or typically feminine names. Names were counter-balanced across the items to avoid any frequency effects. The first name (henceforth the non-local antecedent) was always the first word of the critical sentence and was the subject of the first verb. The first verb, always a verb of perception, was followed by an infinitival complement, the subject of which was the second name (henceforth the local antecedent). The complement contained an inanimate object and a preposition that was followed by a pronoun. The structure was such that either antecedent could be a legitimate antecedent for the pronoun (a typical SDP environment). The scenarios were carefully constructed so that neither name was a more plausible antecedent for the pronoun. Half the pronouns were masculine and half were feminine. The gender match between the two names and the pronoun was manipulated to create three conditions. Examples are given below. A complete list of experimental items can be found in Appendix 2.

(2.22a) Double match condition

Suddenly the lights went on and there were police everywhere.
Barry saw Gavin place a gun near him on the ground with great care.
The robbery was definitely over now.

(2.22b) *Local mismatch condition*

Suddenly the lights went on and there were police everywhere.
Barry saw Megan place a gun near him on the ground with great care.
The robbery was definitely over now.

(2.22c) *Non-local mismatch condition*

Suddenly the lights went on and there were police everywhere.
Megan saw Barry place a gun near him on the ground with great care.
The robbery was definitely over now.

As per Experiment 2, a fourth condition, in which both antecedents mismatched the pronoun, was not included to avoid presenting a large number of items with no matching antecedents.

2.6.2 Experiment 3 predictions

First-mention preference

If the first-mentioned antecedent is perceived as being more prominent than subsequently mentioned antecedents, there should be longer reading times in the non-local mismatch condition (2.22c), when the non-local (first-mentioned) antecedent mismatches in gender with the pronoun, compared to the other two conditions.

Feature matching hypothesis

Because both antecedents are accessible for the pronoun, and obvious plausibility biases have been avoided, the SDP environments may offer a better opportunity than condition B environments for feature-match effects to manifest because other potential cues have been kept to a minimum. If feature matching plays an important role in the initial retrieval of antecedents, then the gender match of the local antecedent should affect processing. This will lead either to longer reading times when the local antecedent mismatches (2.22b > 2.22a), or to longer reading times when the local antecedent matches (2.22a > 2.22b) because of retrieval interference.

2.6.3 Experiment 3 method

See section 2.5.3 for a description of the arrangement of materials into lists.

2.6.3.1 Participants

See section 2.5.3.

2.6.3.2 Procedure

See section 2.5.3.

2.6.3.3 Data preparation

The eye-movement measures examined in Experiment 3 are first-fixation times, first-pass times, regression-path times, rereading times, total-viewing times and regressions-in. The deletion and merging of short and long fixations and the track loss percentage is reported in section 2.5.3.3.

The experimental items were divided into regions of interest for the purpose of the analysis. Eye-movement data from the following five regions was analysed for Experiment 3: precritical, pronoun, spillover, prefinal and final. An additional region, the local name region, was examined in a post-hoc analysis. (2.23) shows how the text in the critical sentence was divided into regions.

(2.23)

Suddenly the lights went on and there were police everywhere.

*Barry saw | Gavin | place a gun | near | him | on the | ground with |
great care. |*

The robbery was definitely over now.

The **local name** region contained only the second-mentioned name (see section 2.6.4.4 for the post-hoc analysis). The **precritical region** contained the word before the pronoun. It contained one or two words which formed the preposition (*near* in (2.23)). As per Experiment 2, first-fixation durations, first-pass times and regression-path times in the precritical region are measured before the pronoun is encountered. The earlier measures in this region are therefore inspected to ensure that any effects

seen in the pronoun region (or later) begin during or after the pronoun is encountered. Rereading times in the precritical region reflect processing after the pronoun is encountered. Total viewing times in this region, being an omnibus measure composed of first-pass times and rereading times, could reflect processing before or after processing of the pronoun. The **pronoun region** contained only the pronoun (*him* in (2.23)). The **spillover region** contained the two words following the pronoun. The **prefinal region** contained the two words following the spillover region, and the **final region** contained the two words following the prefinal region. The final region was always the end of the sentence.

2.6.4 Experiment 3 results

2.6.4.1 Comprehension questions

As per Experiment 2 (see section 2.5.4.1).

2.6.4.2 Eye-movement data

The skipping rates in the analysed regions were as follows: precritical, 15%; pronoun, 45%; spillover, 20%; prefinal, 9%; final, 22%. Data log transformations and composition of the statistical models is described in section 2.5.4.2.

As per Experiment 2, the presentation of the means and statistical analyses of the eye-movement data is divided into two sections, *earlier measures* and *later measures*.

Earlier measures

Table 2.5 shows the means in the earlier measures for the precritical, pronoun, spillover, prefinal and final regions. The outcome of the statistical analysis for these regions and measures is presented in Table 2.6. The results from the earlier measures in each region are subsequently summarised.

Table 2.5. Means (in milliseconds) for first-fixation durations, first-pass times and regression-path times in each condition in the precritical, critical, spillover, prefinal and final regions in Experiment 3. Standard deviations are given in italics.

	Condition					
	Double match (a)		Local mismatch (b)		Non-local mismatch (c)	
Precritical region						
First-fixation durations	222	<i>(84)</i>	215	<i>(80)</i>	219	<i>(96)</i>
First-pass times	262	<i>(142)</i>	246	<i>(132)</i>	247	<i>(132)</i>
Regression-path times	334	<i>(326)</i>	353	<i>(523)</i>	351	<i>(383)</i>
Pronoun region						
First-fixation durations	217	<i>(99)</i>	222	<i>(71)</i>	208	<i>(71)</i>
First-pass times	223	<i>(102)</i>	232	<i>(89)</i>	223	<i>(101)</i>
Regression-path times	268	<i>(212)</i>	299	<i>(227)</i>	273	<i>(168)</i>
Spillover region						
First-fixation durations	208	<i>(72)</i>	200	<i>(63)</i>	213	<i>(72)</i>
First-pass times	259	<i>(215)</i>	239	<i>(117)</i>	253	<i>(136)</i>
Regression-path times	326	<i>(317)</i>	344	<i>(292)</i>	336	<i>(353)</i>
Prefinal region						
First-fixation durations	195	<i>(65)</i>	200	<i>(79)</i>	203	<i>(80)</i>
First-pass times	291	<i>(172)</i>	287	<i>(169)</i>	294	<i>(192)</i>
Regression-path times	417	<i>(417)</i>	454	<i>(602)</i>	475	<i>(536)</i>
Final region						
First-fixation durations	200	<i>(95)</i>	211	<i>(88)</i>	206	<i>(102)</i>
First-pass times	239	<i>(135)</i>	255	<i>(156)</i>	244	<i>(139)</i>
Regression-path times	512	<i>(619)</i>	633	<i>(1068)</i>	595	<i>(831)</i>

Table 2.6. Summary of statistical analyses for earlier eye-movement measures in the precritical, pronoun, spillover, prefinal and final regions of Experiment 3. The estimate for each contrast and its corresponding standard error and t-value is given. SE=standard error.

	First-fixation durations		First-pass times		Regression-path times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region						
<i>double match (a) vs local mismatch (b)</i>	-0.04 (0.03)	-1.30	-0.07 (0.04)	-1.49	-0.05 (0.06)	-0.86
<i>double match (a) vs non-local mismatch (c)</i>	-0.03 (0.03)	-0.78	-0.06 (0.05)	-1.33	-0.01 (0.07)	-0.16
<i>local mismatch (b) vs non-local mismatch (c)</i>	0.01 (0.03)	0.40	<0.01 (0.04)	0.06	0.04 (0.06)	0.65
Pronoun region						
<i>double match (a) vs local mismatch (b)</i>	0.06 (0.04)	1.49	0.07 (0.05)	1.38	0.09 (0.06)	1.32
<i>double match (a) vs non-local mismatch (c)</i>	<0.01 (0.04)	<0.01	0.02 (0.05)	0.39	0.06 (0.07)	0.82
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.06 (0.04)	-1.50	-0.04 (0.05)	-0.94	-0.03 (0.07)	-0.39
Spillover region						
<i>double match (a) vs local mismatch (b)</i>	-0.03 (0.03)	-0.79	-0.02 (0.05)	-0.42	0.08 (0.08)	0.95
<i>double match (a) vs non-local mismatch (c)</i>	0.02 (0.03)	0.82	0.02 (0.04)	0.39	0.01 (0.06)	0.23
<i>local mismatch (b) vs non-local mismatch (c)</i>	0.05 (0.03)	1.58	0.04 (0.05)	0.82	-0.06 (0.07)	-0.86
Prefinal region						
<i>double match (a) vs local mismatch (b)</i>	0.01 (0.03)	0.44	-0.03 (0.05)	-0.60	0.01 (0.07)	0.16
<i>double match (a) vs non-local mismatch (c)</i>	0.04 (0.03)	1.13	-0.01 (0.05)	-0.21	0.03 (0.07)	0.42
<i>local mismatch (b) vs non-local mismatch (c)</i>	0.02 (0.03)	0.65	0.02 (0.05)	0.38	0.02 (0.07)	0.25
Final region						
<i>double match (a) vs local mismatch (b)</i>	0.07 (0.05)	1.50	0.05 (0.06)	0.80	0.09 (0.09)	0.97
<i>double match (a) vs non-local mismatch (c)</i>	0.01 (0.05)	0.30	0.01 (0.05)	0.23	0.06 (0.10)	0.61
<i>local mismatch (b) vs non-local mismatch (c)</i>	-0.06 (0.04)	-1.25	-0.03 (0.05)	-0.74	-0.03 (0.10)	-0.30

Precritical region

As expected in the precritical region, the earlier measures are very similar across the three conditions, although there does appear to be a slight advantage for the double match condition in the regression-path times. The statistical analysis reveals that none of the differences in the earlier measures of the precritical region are significant.

Pronoun region

In the pronoun region the differences between the conditions appear to be small, but there is a trend in all three measures for the local mismatch condition to have the longest reading times. Differences were not statistically significant, although the higher *t*-values for the double match compared to the local mismatch conditions appear to support this trend.

Spillover region

There are only small differences between conditions in the spillover region, and the differences are not statistically significant. The regression-path times follow the same pattern as the overall trend in the pronoun region (longer regression-path times in the local mismatch condition), but the first-fixation durations and first-pass times show the opposite pattern, with an advantage in the local-mismatch condition.

Prefinal region

The first fixation duration and the first-pass reading times are very similar in all three conditions in this region. For the regression-path times the double match condition appears to have a small advantage, but there were no significant effects in this region.

Final region

Similar to the pronoun region, in the final region there is a trend for the local mismatch condition to elicit the longest reading times, in particular in regression-path times. There were no statistically significant differences in this region.

Later measures

Results for the later measures (rereading times and total-viewing times) are presented below. Table 2.7 shows the means in the later measures for the precritical, pronoun, spillover, prefinal and final regions. The outcome of the statistical analysis for these regions and measures is presented in Table 2.8. The results for each region are subsequently summarised.

Table 2.7. Means (in milliseconds) for rereading and total viewing times in each condition in the precritical, pronoun, spillover, prefinal and final regions in Experiment 3. Standard deviations are given in italics.

	Condition					
	Double match (a)		Local mismatch (b)		Non-local mismatch (c)	
Precritical region						
Rereading times	112	<i>(183)</i>	154	<i>(229)</i>	123	<i>(203)</i>
Total viewing times	374	<i>(237)</i>	399	<i>(263)</i>	370	<i>(236)</i>
Pronoun region						
Rereading times	78	<i>(169)</i>	74	<i>(125)</i>	73	<i>(133)</i>
Total viewing times	302	<i>(197)</i>	306	<i>(155)</i>	296	<i>(180)</i>
Spillover region						
Rereading times	127	<i>(214)</i>	123	<i>(214)</i>	136	<i>(259)</i>
Total viewing times	386	<i>(298)</i>	362	<i>(232)</i>	389	<i>(300)</i>
Prefinal region						
Rereading times	134	<i>(199)</i>	171	<i>(261)</i>	144	<i>(247)</i>
Total viewing times	425	<i>(253)</i>	458	<i>(299)</i>	438	<i>(320)</i>
Final region						
Rereading times	49	<i>(116)</i>	76	<i>(159)</i>	51	<i>(125)</i>
Total viewing times	288	<i>(167)</i>	331	<i>(233)</i>	294	<i>(202)</i>

Table 2.8. Summary of statistical analyses for later eye-movement measures in the precritical, critical, spillover, prefinal and final regions of Experiment 3. The estimate for each contrast and its corresponding standard error and t-value is given. SE=standard error.

	Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region				
<i>double match (a) vs local mismatch (b)</i>	42.42 (29.61)	1.43	0.06 (0.08)	0.71
<i>double match (a) vs non-local mismatch (c)</i>	7.90 (20.36)	0.39	-0.01 (0.07)	-0.13
<i>local mismatch (b) vs non-local mismatch (c)</i>	-34.52 (28.85)	-1.20	-0.07 (0.07)	-0.89
Pronoun region				
<i>double match (a) vs local mismatch (b)</i>	-6.31 (25.33)	-0.25	0.04 (0.08)	0.57
<i>double match (a) vs non-local mismatch (c)</i>	-6.28 (24.48)	-0.27	0.01 (0.07)	0.09
<i>local mismatch (b) vs non-local mismatch (c)</i>	0.54 (18.08)	0.03	-0.03 (0.07)	-0.44
Spillover region				
<i>double match (a) vs local mismatch (b)</i>	-0.44 (24.09)	-0.02	<0.01 (0.06)	0.04
<i>double match (a) vs non-local mismatch (c)</i>	13.04 (31.06)	0.42	0.01 (0.06)	0.11
<i>local mismatch (b) vs non-local mismatch (c)</i>	13.02 (29.00)	0.45	<0.01 (0.07)	0.06
Prefinal region				
<i>double match (a) vs local mismatch (b)</i>	36.78 (25.02)	1.47	0.05 (0.06)	0.88
<i>double match (a) vs non-local mismatch (c)</i>	10.73 (23.17)	0.46	-0.02 (0.06)	-0.36
<i>local mismatch (b) vs non-local mismatch (c)</i>	-26.68 (22.98)	-1.16	-0.07 (0.06)	-1.22
Final region				
<i>double match (a) vs local mismatch (b)</i>	24.91 (16.22)	1.54	0.08 (0.07)	1.18
<i>double match (a) vs non-local mismatch (c)</i>	4.61 (14.09)	0.33	<0.01 (0.06)	-0.07
<i>local mismatch (b) vs non-local mismatch (c)</i>	-20.29 (17.21)	-1.18	-0.09 (0.06)	-1.46

Precritical region

In this region there is a trend for longer rereading and total viewing times in the local mismatch condition in comparison to the other two conditions, but these differences were not significant.

Pronoun region

In the pronoun region rereading and total viewing times were very similar between conditions, and there were no significant differences here.

Spillover region

There are only small differences between conditions in the spillover region, and the differences are not statistically significant.

Prefinal region

There is a trend for longer rereading and total viewing times in the local mismatch condition in the prefinal region. There were no statistically significant differences although the higher *t*-value for the comparison between the double match and the local mismatch conditions does support the trend.

Final region

In the final region the same trend for longer rereading and total viewing times in the local mismatch condition is apparent, although the differences are not statistically significant.

2.6.4.3 Summary of results from Experiment 3

There was no significant effect of condition in any measure in any of the regions of interest studied. However, there is a trend that is seen in both earlier and later measures. This trend is a consistent disadvantage for the local mismatch condition compared to the other two conditions, and can be seen in the prefinal region (rereading and total viewing times), pronoun region (first-fixation durations, first-pass times and regression-path times), spillover region (regression-path times), prefinal (rereading and total viewing times) and the final region (first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times).

In order to further explore the trend, the local name region was also inspected in an additional analysis. If the trend is meaningful it suggests that participants experience more difficulty when the local name mismatches with the pronoun compared to the other two conditions. As such, evidence of processing difficulty may be visible in the local name region itself. Earlier measures in this region are not of interest because the pronoun had not been encountered at that point. But rereading and total viewing times are composed of (or, in the case of total viewing times, include) second-pass fixations by which time later regions had been encountered. An additional measure, regressions-in, is also included. This measure reflects how many regressive eye-movements were made into a particular region. The results of the additional analysis are reported below.²¹

2.6.4.4 Additional analysis, Experiment 3

Results for the local name region are presented below. Table 2.9 shows the mean rereading times, total viewing times and mean number of regressions-in in the local name region. The outcome of the statistical analysis for this region is presented in Table 2.10. The results are subsequently summarised.

²¹ For the first name region, there was a lot of missing data, probably due to the name being the first word on a line leading to landing inaccuracies, and probable parafoveal processing from the next word. For this reason it was decided not to pursue analysis on this region.

Table 2.9. Means (in milliseconds) for rereading times and total viewing times, and mean number of regressions-in, for each condition in the local name region in Experiment 3.

Standard deviations are given in italics.

	Condition					
	Double match		Local mismatch		Non-local mismatch	
Rereading times	255	<i>(281)</i>	299	<i>(340)</i>	253	<i>(325)</i>
Total viewing times	464	<i>(305)</i>	524	<i>(342)</i>	475	<i>(336)</i>
Regressions-in	0.42	<i>(0.59)</i>	0.65	<i>(0.98)</i>	0.42	<i>(0.64)</i>

Table 2.10. Summary of statistical analyses for rereading times, total viewing times and regressions-in in the local name region of Experiment 3. The estimate for each contrast and its corresponding standard error and t-value is given. SE=standard error. (*) = $p \leq .10$, * = $p < .05$.

	Rereading times		Total viewing times		Regressions-in	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
<i>double match (a) vs local mismatch (b)</i>	39.39 <i>(36.91)</i>	1.07	0.11 <i>(0.07)</i>	1.73(*)	0.20 <i>(0.10)</i>	2.03*
<i>double match (a) vs non-local mismatch (c)</i>	-5.30 <i>(30.47)</i>	-0.17	0.01 <i>(0.06)</i>	0.15	0.01 <i>(0.07)</i>	0.01
<i>local mismatch (b) vs non-local mismatch (c)</i>	-44.69 <i>(33.70)</i>	-1.33	-0.10 <i>(0.06)</i>	-1.64(*)	-0.20 <i>(0.09)</i>	-2.23*

The rereading and total viewing times for the local name region both clearly reflect the trend that was seen in the other regions, namely, a disadvantage for the local mismatch condition. This trend is reflected statistically in the total viewing times in which there are marginally significant differences between the local mismatch condition and the other two conditions. There was also a higher mean number of regressive eye movements into the local name region in the local mismatch condition

compared to the other two conditions. This was confirmed with statistically significant differences between the local mismatch condition and the other two conditions.

2.7 Discussion

The purpose of Experiments 1-3 was to explore the role of condition B in the online resolution of pronouns, to provide further eye-tracking evidence regarding interference from inaccessible antecedents and consideration of accessible antecedents. This was investigated using canonical condition B environments and also SDP environments. In this section, I will summarise the evidence from Experiments 1-3, and then discuss how it fits in with previous findings and hypotheses about structural constraints in online processing.

2.7.1 Experiment 1: offline interpretations

Firstly, Experiment 1, the offline questionnaire, clearly showed differences in the offline interpretation of pronouns in condition B environments versus SDP environments. For canonical condition B environments, participants interpreted the pronoun as referring to the accessible antecedent in almost all cases (98% of responses). This is a clear indication that the interpretation of pronouns in this syntactic environment is in line with condition B, which states that a pronoun cannot be bound by a c-commanding antecedent within its binding domain. In contrast, for the SDP environments participants' responses were more varied. The majority of responses (60%) were *either*, showing that participants found the pronoun ambiguous. This is in line with theoretical assumptions that pronouns in SDP environments allow reference to an antecedent that is within the pronoun's binding domain (as originally defined in BT). However, when participants did not choose *either* as a response, they picked the local antecedent or the non-local antecedent at approximately the same rate (18% versus 22%). Overall the offline results for pronouns in SDP environments demonstrate some uncertainty about their interpretation. This is perhaps worthy of further investigation on its own, but what is important for the current study is, firstly, that the ambiguity of

pronouns in SDP environments is recognised and is in sharp contrast to the interpretations in the condition B environments; and, secondly, that there is no offline preference for either the local antecedent or the non-local antecedent.

2.7.2 Experiment 2: condition B environments

Experiment 2 investigated the online processing of pronouns in condition B environments. A reminder of the conditions (critical sentence only) is given below in (2.24):

(2.24a) Double match condition

John remembered that Mark had taught him a new song on the guitar.

(2.24b) Local mismatch condition

John remembered that Jane had taught him a new song on the guitar.

(2.24c) Non-local mismatch condition

Jane remembered that John had taught him a new song on the guitar.

To summarise the findings, first fixation times and first-pass times did not reveal differences between the conditions in any region. All differences emerged in regression-path or rereading times, reflected also in total viewing times. Rereading and total viewing times in the precritical and pronoun regions were significantly longer in the non-local mismatch condition (2.24c) than both the double match (2.24a) and the local mismatch (2.24b) conditions. In the spillover region there was a marginally significant difference in regression-path times between the double match and the non-local match conditions (2.24a–2.24c), and a significant difference between the local mismatch and the non-local mismatch conditions (2.24b–2.24c). Rereading and total viewing times in this region again were significantly longer in the non-local mismatch condition compared to the other two conditions. In the prefinal region there was a marginal difference in the same direction between the non-local mismatch condition and the local

mismatch condition in rereading times. In the final region, regression-path times were significantly longer in the non-local mismatch condition (2.24c) than both the double match condition (2.24a) and the local mismatch condition (2.24b). There were no significant differences between the double match and the local mismatch conditions.

Overall these results show a very consistent picture. Firstly, participants showed more processing difficulty, reflected in longer reading times, in the non-local mismatch condition compared to the other two conditions. In the non-local mismatch condition the accessible antecedent mismatches in gender with the pronoun, whereas in the other two conditions the gender of the accessible antecedent matches the gender of the pronoun. The mismatch in the non-local mismatch condition caused difficulty because no gender-matching accessible antecedent is presented. The fact that this difficulty is observed over several regions is not surprising because there is no other potential antecedent for the pronoun in the discourse (except for one that is ruled out by condition B), so the participants cannot easily recover from the processing difficulty. Secondly, at no point is there a significant difference between the double match condition and the local mismatch condition. Although there is a small trend in this direction in rereading times in the pronoun region, the difference does not get close to significance, and is also not seen in other regions/measures. The double match *vs.* local mismatch comparison is the test of interference from an inaccessible (local) antecedent. There is no evidence of such interference in Experiment 2, i.e. we found no evidence that the gender match between the inaccessible antecedent and the pronoun affects processing.

The lack of an interference effect, coupled with disruption when the accessible antecedent mismatches in gender with the pronoun, fits in with the predictions of the BAIF hypothesis (and also follows the offline results from Experiment 1). Under BAIF, binding condition B excludes the inaccessible antecedent from the set of potential antecedents so that it does not affect processing at any point, predicting gender match/mismatch effects only for the accessible antecedent. The results of Experiment 2 are not consistent with the other two sets of predictions, namely the defeasible

filter or the feature matching hypothesis. Both of these hypotheses predicted interference from the inaccessible antecedent, which would manifest as a difference between the double match and the local mismatch conditions. The hypotheses differed in their predictions about the timing of the interference. Nevertheless, the results of Experiment 2 are incompatible with both hypotheses because no interference was detected before, during or after the processing difficulty caused by the mismatching accessible antecedent.

2.7.3 Experiment 3: SDPs

While in Experiment 2 there was one accessible and one inaccessible antecedent, this was not the case in Experiment 3. In Experiment 3, both antecedents were accessible; this is supported by the results of Experiment 1. Having two accessible antecedents allowed for further testing of the feature match hypothesis. It also tested a possible first-mention preference; while the results of Experiment 2 indicate online adherence to condition B, they could also have come from a preference for the first-mentioned antecedent, which has often been observed in pronoun processing (e.g. Arnold *et al.* 2007). The conditions for Experiment 3 are repeated as (2.25) below.

(2.25a) *Double match condition*

Barry saw Gavin place a gun near him on the ground with great care.

(2.25b) *Local mismatch condition*

Barry saw Megan place a gun near him on the ground with great care.

(2.25c) *Non-local mismatch condition*

Megan saw Barry place a gun near him on the ground with great care.

In Experiment 3 there was no significant effect of condition in any measure in any of the regions of interest. However, there was a persistent trend visible in both early and late measures, for a consistent disadvantage for

the local mismatch condition (2.25b) compared to the other two conditions. The trend was visible in later measures in the the prefinal region, earlier measures in the pronoun region, regression-path times in the spillover region, later measures in the prefinal region and all measures in the final region. A post-hoc analysis of rereading and total viewing times and regressions-in to the local name region revealed a marginal disadvantage for the local mismatch condition in total viewing times, and significantly more regressions-in to the local name region in the local mismatch condition. Both these confirm the trend that was observed in the main analysis, which is processing difficulty with the local mismatch condition. The results here should be interpreted with caution because the trend did not reach significance or marginal significance in the main analysis in any region measure, only in the post-hoc analysis. In addition, the differences observed in total viewing times in the local name region were not reflected in the rereading times, where there was a trend only. This may mean that the differences in the total viewing times arose from earlier measures, which reflect processing before the pronoun is encountered. Another consideration is that while the regressions-in to the local name region in the local mismatch condition indicate difficulty with the mismatching local name, the opposite pattern may well have occurred in the non-local name region in the double match and the non-local mismatch conditions when it mismatched with the pronoun, but this region was not analysed (see footnote 21). So the effect in the regressions-in measure in the local name region are not as good an indicator of general processing difficulty as the later regions.

Having pointed out the difficulties with the data analysis in Experiment 3, it is still possible to observe a very persistent trend which does not fit in with the predictions of a first-mention preference or of a feature matching hypothesis. The first-mention hypothesis predicts longer reading times in the non-local mismatch condition (2.25c) compared to the other two conditions. This is clearly not seen in Experiment 3, nor is there even a trend in this direction. This contrasts with the findings of Experiment 2 in which the non-local mismatch condition caused considerable processing difficulty. The contrast with regard to the non-local mismatch condition

between Experiments 2 and 3 demonstrates that the results of Experiment 2 were unlikely to have been driven by a first-mention preference.

It was hypothesised that in the SDP environments the parser may be more sensitive to the gender features of potential antecedents, which could have led to two outcomes. Either retrieval interference/competition when two antecedents had features matching the pronoun, or an advantage for the double match condition because it had more feature-matching antecedents. Because both antecedents are accessible for the pronoun in SDP environments and obvious plausibility biases were avoided in the creation of the materials, it was expected that the SDP environments would offer a better opportunity than condition B environments for feature-match effects to manifest, because other potential cues had been kept to a minimum. But no feature-match effects were observed in Experiment 3. At first this outcome may seem surprising. After all, the gender match of the antecedent with the pronoun must be taken into account at some point during processing. Given that neither antecedent was excluded on structural grounds, evaluating the antecedents' gender match with the pronoun appears to be a sensible way for the parser to proceed. It should be noted, though, that each condition in Experiment 3 presents *at least one* gender-matching antecedent. It is possible that participants were happy to proceed with processing as long as there was an accessible, gender matching antecedent in the discourse. It is also possible that, on finding no plausibility biases in one direction or another at the point of the pronoun, participants continued reading beyond the pronoun region to assess whether later information would help to determine reference. In either case, it is clear that participants neither gained a reading time advantage nor were slowed down by the presence of two matching accessible antecedents in comparison to one matching antecedent, which is incompatible with a feature matching hypothesis.

This outcome could also indicate that in the double match conditions, participants were satisfied with the ambiguity and did not engage in further processing to determine which of the two antecedents was preferred. The fact that only a small number of the comprehension

questions asked for the referent of a pronoun may have given rise to this (see also Greene, McKoon and Ratcliff (1992), who claim that an antecedent is not identified unless the experimental task explicitly requires it).

While the materials and method of Experiment 3 differ from Sekerina *et al.*'s (2004) experiment on SDPs, both experiments demonstrate that the ambiguity of SDPs is visible online. One interesting contrast is that, while Sekerina *et al.*'s results show, if anything, a tendency for pronouns to be resolved to the non-local antecedent (although the non-local antecedent in Experiment 3 was sentence internal while in Sekerina *et al.*'s experiment the non-local antecedent was outside of the critical sentence), in Experiment 3 there was a trend in the opposite direction, towards the local antecedent. Recall that Sekerina *et al.*'s experiment contained reflexives, and involved a judgement task. Because the reflexives could only be resolved to the local antecedent in Sekerina *et al.*'s experiment, the tendency to resolve pronouns to the non-local antecedent could have come from differentiating between pronouns and reflexives. In Experiment 3, however, reflexives were not systematically tested, so there was no need for participants to differentiate in the same way.

The local mismatch trend in Experiment 3 was not reflected in the offline interpretations in Experiment 1. Offline, there is no preference for either antecedent, and on the whole the pronouns are interpreted as ambiguous. This could indicate that there some kind of processing advantage for the local antecedent that is overruled when participants are given more time to reflect on their responses. A weak trend on its own is difficult to interpret, however. The issue of locality/recency is explored later in this thesis in Chapter 4.

2.7.4 Timing

For Experiment 3 we cannot properly assess the timing of any effects because the local mismatch effect was visible only as a trend. For Experiment 2, predictions regarding the timing of the effects were difficult to make, as noted in section 2.5.2.4, partly because previous studies have used a variety of different methodologies. Given that no interference effects

were detected, the timing of an interference effect obviously cannot be discussed. The timing of the processing difficulty arising from a mismatching accessible antecedent can be considered, even if it is on rather a speculative basis given the absence of precise predictions about the timing of this effect. Considering previous studies, the timing of the accessible mismatch effect is similar in timing to the one in Badecker and Straub's (2002) study, where there was a reading-time advantage for a matching accessible antecedent detected in the spillover region (one region after the pronoun), in comparison to both the no match condition and the inaccessible match condition.²² In contrast, the inaccessible mismatch effect for the reflexive in Sturt's (2003) experiments 1 and 2 was detected from first-fixation times onwards, making it an early effect. The timing of the effect in Experiment 2 in the current study appears closer to the timing in Badecker and Straub's study rather than Sturt's, being found in regression-path times in the spillover and final regions and only in rereading/total times in the earlier regions. It would be difficult to classify this as an early effect.

A later effect for pronouns in comparison to reflexives fits in well with two related considerations: first, pronouns are sensitive to a range of cues or information types which can help to determine the correct antecedent, so considering all these information sources may require more time. Second, the nature of condition B, unlike condition A for reflexives, involves excluding rather than identifying an antecedent, which could lead to the accessible antecedent being more rapidly identified for reflexives as opposed to pronouns. The label 'late effect' should, however, be applied with caution. It is generally difficult to establish that a late effect really occurs late, or whether it is a behavioural residue of an effect that occurred much earlier (Pickering *et al.* 2004). In this sense, early effects are unambiguously early, but late effects could be due to late processes or residues of earlier processes. An additional consideration is that the high

²² In the comparison with the no match condition, the reading times start to diverge in the pronoun region itself but the difference there is not significant.

skipping rates, particularly in the pronoun region led to a large amount of data being excluded from the analysis. This may have prevented effects from being detected in the pronoun region. Nevertheless, it did not prevent robust effects in the rereading and total viewing times in the pronoun region from being detected, which means that if there were effects in earlier measures they would have to be much smaller than the effects in the later measures. I tentatively propose, taking into consideration the difficulty of labelling effects, that the identification of the accessible antecedent was not immediate. This is important in the assessment of the BAIF model below.

2.7.5 What do Experiments 2 and 3 tell us about pronoun processing models?

Both Experiments 2 and 3 showed no effects that could be characterised as retrieval interference, where double match conditions have significantly different reading times compared to single match conditions when (or after) the pronoun is encountered. This is not compatible with Badecker and Straub's (2002) hypothesis that binding constraints and cues such as gender and number affect the activation of antecedents simultaneously. I therefore find no evidence directly supporting the parallel constraints model or feature-based retrieval of antecedents. There is also no evidence for Sturt's (2003) defeasible filter model, in which the initial set of antecedents are only those which are compatible with BT, which can be defeated later as inaccessible antecedents are considered.

The current experiments found no evidence against the proposal that the parser is insensitive to structural cues. I suggest that not only is the parser able to remain faithful to condition B at all stages of processing, it is able to use the subtle structural cues that distinguish canonical condition B environments from SDP environments, leading to different patterns of results in Experiments 2 and 3. A structurally sensitive parser is most compatible with the BAIF (Nicol and Swinney 1989), under which only BT compatible antecedents are considered from the beginning of processing the pronoun. The lack of interference effects found in Experiments 2 and 3 fit in with that assumption. The timing of the effect found could also be interpreted in support of the two-stage process that the BAIF suggests. For

example, if there is an initial stage at which only binding conditions are relevant, the lack of any early effects in the experiments demonstrates that gender information is not relevant at this stage and only the syntactic configuration is considered, which is the same in all three conditions. Having excluded the inaccessible antecedent at the initial stage, the parser then engages other types of information at the second stage in order to search for and evaluate the remaining candidate, contributing to the later effects of the accessible antecedent. However, before adopting such an approach, there are several questions over the details of the BAIF model that need to be considered. While the results from the current experiments cannot be used to evaluate the specific points that I question below, it is important to reflect on the details of any model that my results appear to support.

Firstly, the BAIF model assumes two stages for processing a pronoun: the implementation of grammatical constraints, namely condition B in the current experiments, then the consideration of pragmatic and discourse information. Pragmatic and discourse information is specifically relegated to the role of mopping up after the binding conditions, if they fail to identify a unique antecedent. However, from the short survey of theoretical approaches to condition B that were outlined in section 2.2.3 it is not at all clear that condition B involves only structural information. Pragmatic information may well be involved in the exclusion of inaccessible antecedents. It is rather difficult to maintain, then, that pragmatic information is used to exclude inaccessible antecedents but is at the same time held back until after inaccessible antecedents are excluded. This point in particular highlights the need to carefully consider the theoretical assumptions behind processing models.

A second problem with the BAIF is that it assumes that condition B can usually be sufficient not only to exclude inaccessible antecedents but also to identify the correct antecedent. They state [emphasis added]:

“...the reactivation of prior referents is restricted by grammatical constraints. *In the case where such information does not*

sufficiently constrain the list of potential antecedents to a single one, the pragmatic and other sentence/discourse processing procedures undoubtedly come into play, but, given the present evidence, only at a later point in processing.”

(Nicol and Swinney 1989, p. 18)

While this does not make any specific predictions regarding timing, it is suggestive of a rapid process in which the correct antecedent should be identified as soon as binding conditions take effect. This is difficult to reconcile with the timing of the results of Experiment 2, where the accessible antecedent's mismatch effect was not visible until later measures. This approach may work for binding condition A for reflexives; an antecedent for a reflexive can be identified via structural cues, and an earlier stage at which binding condition A applies fits in with Sturt's (2003) findings and also other studies which show rapid effects for reflexives (e.g. Felser and Cunnings 2012). However, the assumption that binding condition B, too, can uniquely identify an antecedent in some cases implies that this process should take place at an early stage and give rise to earlier effects of the accessible antecedent than were seen in Experiment 2. Given that condition B involves ruling out an antecedent rather than identifying an antecedent, assuming that binding conditions A and B operate in exactly the same way is problematic. This may, however, be an over-interpretation of Nicol and Swinney's original assumptions about the rapidity of such processes.

Finally, I think that the largest problem facing the BAIF as it is stated by Nicol and Swinney (1989) is the suggestion that discourse processing is withheld until a *later* stage of processing. There is, in fact, no direct evidence about this point. In order to uphold this proposal, it is necessary to manipulate both the discourse context and the accessibility of antecedents and evaluate the resulting processing patterns. To my knowledge, this has not yet been done and may prove a fruitful area for further research.

I do not wish to dispense with the BAIF hypothesis, but I point out these problems in the hope that it can be refined to account for theoretical assumptions and to point a way to future developments. The main advantage of the BAIF is the importance it places on the faithfulness of the parser to condition B in all stages of pronoun processing, which is consistent with the evidence I have presented above and also previous evidence from other reading studies as outlined in section 2.3.1. But faithfulness to condition B does not necessarily entail an earlier stage of processing that only makes use of structural cues. It might also be possible, for example, to incorporate the faithfulness to condition B into a parallel constraints model like the one put forward by Badecker and Straub (2002). This could be achieved if a higher weighting could be given to a structural cues (however they are implemented) throughout the retrieval and evaluation process so that inaccessible antecedents are not considered. Cue weighting is suggested by Kaiser and Trueswell (2008) in the form-specific multiple constraints model (see also Brown-Schmidt, Byron and Tanenhaus 2005), and could be adopted here to account for the parser's faithfulness to condition B and sensitivity to the syntactic environment.

A parallel constraints model with a higher weighting for such cues may also allow for some previous results showing interference effects to be incorporated. It is possible that what divides the evidence for and against the feature match hypothesis is how taxed the memory resources are during a given experiment. In particular, Badecker and Straub (2002) and Clackson *et al.* (2011) both involve tasks in which are arguably more demanding than reading (in particular, more demanding than an eye-tracking study where the whole experimental item is on the screen). If additional memory demands lead to disruption in cue weights, it may be more difficult or require more time to implement a structural cue, giving rise to feature-match effects. If this is the case, then it suggests that feature match *is* involved in retrieval, but that under less taxing experimental situations participants are able to modulate this by rapidly recruiting condition B to eliminate matching but inaccessible antecedents. Such an approach could account for a wider set of previous results than the BAIF

alone, without compromising on the importance of the structural cue which can be in operation from an early stage.

The results of the current experiments can rule out a parallel constraints model in which the feature match of an inaccessible antecedent always has a consequence for processing. But they cannot distinguish between a two-stage model such as the BAIF and a parallel constraints model in which structural cues are strongly weighted. This remains to be explored, and could be tested by the systematic manipulation of task type or working memory load, for example. However, the BAIF in its present form accounts for a narrower set of results, and it remains to be seen whether discourse cues are held back in comparison with structural cues in the way that Nicol and Swinney (1989) originally suggested.

2.8 Summary

This chapter has considered how condition B affects the processing of pronouns by rendering certain antecedents inaccessible to the pronoun. The processing of SDPs, a well known exception to condition B, was also investigated. I outlined the theoretical background to condition B and SDPs, showing that condition B may involve pragmatic considerations as well as configurational ones. One offline and two online experiments were reported in which the processing of pronouns in canonical condition B environments and SDP environments was investigated. I found that there was no evidence for a feature-based retrieval in either of the two environments. In the condition B environments participants were sensitive to the gender of the accessible (non-local) antecedent but not the inaccessible (local) antecedent. In the SDP environments there was a tendency for participants to be sensitive to the gender of the local antecedent. While the results are most compatible with the BAIF hypothesis, I question certain details of the hypothesis and suggest that alternative ways of capturing the results are possible. Importantly, though, the faithfulness of the parser to condition B at all stages of processing should be upheld. In addition, the differing profiles of antecedent selection in condition B environments and SDP environments demonstrates that the

parser engages in detailed syntactic processing, and that has an influence on the availability of antecedents during pronoun resolution.

CHAPTER 3

The role of variable binding and coreference in the search for an antecedent

3.1 Introduction

In the previous chapter it was shown that the parser is sensitive to the subtle changes in syntactic configuration which either allow or disallow pronoun reference to a local antecedent. Moreover, the evidence in Chapter 2 (along with evidence from previous studies) indicates that the parser is normally faithful to condition B at all stages of processing. This suggests that the initial retrieval processes involved in pronoun resolution are not driven solely by matching on the basis of morphological features. Rather, the syntactic information is used from the beginning, and is usually a strong enough constraint to prevent the consideration of a feature-matching but inaccessible antecedent. This is compatible with models of pronoun processing which assume an important role for structural information.

In the current chapter a different structural constraint is considered, which allows us to assess a hypothesis based on a more complex model of pronoun processing, namely, *Primitives of Binding* (henceforth POB). This model can be described as a syntax-first model, and thus would align itself with models such as BAIF which was discussed in the previous chapter. In the current chapter a specific claim of the POB model is evaluated; one which relates to a structural constraint on pronoun resolution. I will first outline what I refer to as the POB model, and give the necessary theoretical background. I then evaluate the existing evidence for the POB model before reporting two experiments which further investigate one of the central claims associated with this model.

3.2 Theoretical background: variable binding and coreference

3.2.1 Origins of the POB model

Primitives of Binding is the title of a paper by Reuland (2001) in which a syntactic approach to anaphora is outlined. It is built on previous work, in particular Reinhart and Reuland (1993), and it sits broadly within the Minimalist Program (Chomsky 1995). The 2001 paper does *not* present a processing model; it is a syntactic framework that captures the distribution of anaphoric elements and attempts to integrate them with proposals about general properties of language and the processing system. Reuland has continued to develop this approach to anaphora since 2001, most comprehensively in *Anaphora and Language Design* (Reuland 2011). In the meantime, an explicit *processing* model based on this framework was developed and presented in Koornneef's (2008) thesis (supervised by Reuland). The processing claims in Koornneef's thesis have been incorporated into Reuland's (2011) work. In this chapter I am evaluating a specific processing claim regarding pronouns that is made in Koornneef's processing model (as detailed in the 2008 thesis) which is based on Reuland's theoretical work. The processing model is what I refer to as the POB model throughout this chapter, and I refer to the specific claim being evaluated as the POB hypothesis.

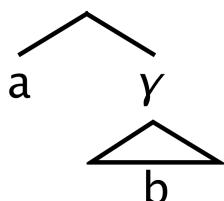
3.2.2 Condition D

I turn now to the structural constraint that is important for the POB hypothesis being evaluated in the current chapter. Aside from the BT binding conditions (see Chapter 2, section 2.2.2), there is another way in which pronouns are restricted in their referencing potential that can be captured in configurational terms. For a singular pronoun to refer to a quantified phrase (QP) such as *every soldier*, the pronoun needs to be c-

commanded by the QP.^{23,24} C-command was introduced in the previous chapter (see Chapter 2, section 2.2.2). It describes an asymmetrical relationship between elements in a syntactic structure. Reuland's (2011) description of c-command was given in Chapter 2 and is repeated here for convenience:

"a c-commands b iff a is a sister to a category γ containing b.

Schematically: [a [γ ...b...]], or



According to this definition and common syntactic assumptions, the subject of a clause will c-command the object, but not vice versa."

(Reuland 2011, p. 29)

C-command is not simply about dominance (being higher in the tree), although it is often the case in the sentences discussed in this chapter that

23 In English, the plural pronoun "they" can sometimes be used to refer to QPs. In these cases the QP is taken to denote a whole group, and the c-command condition is not strictly adhered to. The discussion in this thesis is restricted to cases where a singular pronoun is used to refer to a QP.

24 There are some exceptions to this stipulation, notably so-called donkey anaphora (e.g. Evans 1977; Elbourne 2005) as well as telescoping (Roberts 1989). I will not be concerned with these exceptions here because I am dealing with core cases which follow the c-command rule, although telescoping is mentioned in section 3.4 in relation to Carminati, Frazier and Rayner's (2002) study.

the c-commanding element is higher in the tree; rather, it is a particular configuration relating to sisterhood. The notion of c-command is very important in specifying the conditions under which a singular pronoun may or may not refer to a QP. Consider the following example.

(3.1) Every politician wished that he told the truth.

In this example, the QP c-commands the pronoun. This is illustrated in a simplified tree diagram (Figure 3.1). The QP is the sister of the VP, and the pronoun is contained within the VP. The QP c-commands the pronoun, and as such *he* can be bound by *every politician*. (Equally, the pronoun *he* could refer to another person not mentioned in the sentence. Binding is not obligatory under these conditions.)

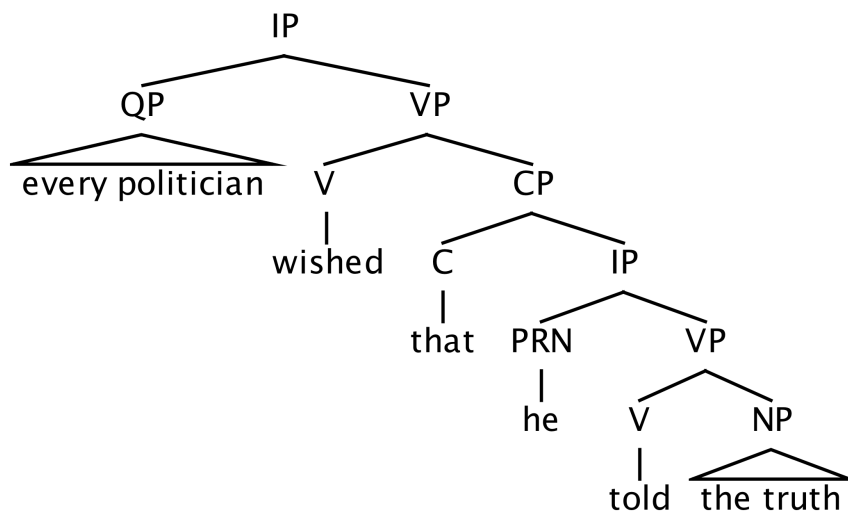


Figure 3.1. Simplified syntactic tree diagram for (3.1), illustrating that the QP c-commands the pronoun.

Compare this to (3.2).

(3.2) Every politician saw the newspaper. He was on the front page.

In this example, the QP cannot c-command the pronoun because the two elements are in different sentences; c-command can only apply in the domain of the sentence. Linking the pronoun *he* to the QP in (3.2) is therefore problematic; despite the fact that politicians are often found on the front page of the newspaper and there is no other salient referent given in the context, the link between the pronoun and the QP nevertheless breaks down. The c-command requirement on QPs is widely recognised and has even been nicknamed *Condition D* (Sportiche 2013) following Chomsky's BT conditions. Sportiche summarises the c-command requirement succinctly as follows:

"A pronoun P can be bound by a quantified expression QE only if
QE c-commands P."

(Sportiche 2013, p. 196)

I will adopt the term condition D throughout this chapter. Note that this restriction does not apply to non-quantified phrases such as names. If we replace the QP in examples (3.1) and (3.2) with a non-quantified determiner phrase (DP) *the politician* or a name such as *Dave*, the requirement that the pronoun be c-commanded by the antecedent is no longer relevant. This is illustrated in examples (3.3) and (3.4).

(3.3a) The politician wished that he told the truth.

(3.3b) Dave wished that he told the truth.

(3.4a) The politician saw the newspaper. He was on the front page.

(3.4b) Dave saw the newspaper. He was on the front page.

In (3.4) the pronoun can refer to *the politician* or to *Dave* without any problem, even though the antecedent does not c-command the pronoun.

3.2.3 What is special about QPs?

The difference between (3.4) and (3.2) demonstrates that there is something special about linking a pronoun to a QP. The configurational requirements for doing so are set out in condition D but that does not explain why such a difference exists. Indeed, it is not only in reference to pronouns that QPs behave differently from non-quantified phrases. QPs have long been of interest in semantics, in fact, they form the basis for a large number of semantic theories. On the surface, a QP such as *every politician* seems to behave in the same way as a non-quantified DP or a name in terms of the syntactic positions it can occupy. But QPs have some particular quirks that set them apart from the other categories, and that can be attributed to their variable reference. For example, consider (3.5) and (3.6) (adapted from Partee 2013). (3.5b) expresses roughly the same meaning as (3.5a). But (3.6b) definitely does not express the same meaning as (3.6a).

(3.5a) Dave wants to win.

(3.5b) Dave wants Dave to win.

(3.6a) Every politician wants to win.

(3.6b) Every politician wants every politician to win.

Because a QP such as *every politician* does not refer to one individual, the task of evaluating the truth conditions of a sentence containing *every politician* is more complicated than evaluating a sentence containing a name or a DP. In effect, a sentence such as (3.6a) expresses a summary of propositions about individual politicians. *Every (x)* is normally interpreted as a universal quantifier. Individual variables are needed for the interpretation of quantified propositions. A variable does not have fixed reference, but can take on given values, so that in a simple sentence such as ‘*Every politician lies*’, the quantifier *every* is paired with a variable, usually called *x*, to give rise to the following (rough) interpretation: ‘for every thing called *x*, if *x* is a politician then it lies’. The quantifier is said to bind the

variable (x); this is known as variable binding. This can be represented as follows:

$$\forall x(\text{POLITICIAN}(x) \rightarrow \text{LIE}(x))$$

The symbol \forall represents the universal quantifier, which binds (a copy of) the variable x . There are a variety of proposals for the precise way in which quantified expressions are represented at the level of semantics, but what is important for the current discussion is the notion of variable binding and the fact that a QP such as *every politician* does not project an individual into the discourse representation. This stands in contrast to the properties of DPs and names. The distinction here has consequences for the way in which pronouns are linked to QPs and DPs, as we will see below.

3.2.4 Variable binding and coreference

Pronouns, too, can act as variables. So if we are to extend our example above to include a pronoun as in (3.7a), we come out with the (simplified) representation shown in (3.7b).

(3.7a) Every politician knows that he lies.

(3.7b) $\forall x(\text{POLITICIAN}(x) \rightarrow (x \text{ knows that } x \text{ lies}))$

In (3.7b), the pronoun acts as a variable (x) and is bound by the universal quantifier. This type of analysis, where pronoun-antecedent relations are conceived of as variable binding in environments such as (3.7a), is the one adopted in POB. Note that variable binding of this kind can only take place when the QP c-commands the pronoun. In order to account for pronouns that are not c-commanded by their antecedents, another operation called coreference is required.

A non-quantified expression such as a DP or a name is able to refer, that is, to point to an individual in the discourse representation. For example, if I were to write a newspaper article which mentions the name Angela Merkel

it is assumed that readers would construct a mental representation of Angela Merkel as part of the ongoing discourse processing.²⁵ A pronoun, too, is able to point to an individual in the discourse representation. For example if, in the next sentence of my newspaper article, I write *she*, readers will probably assume that I am referring to Angela Merkel. When two expressions (*Angela Merkel* and *she*) both point to the same individual in the discourse representation, they are said to corefer. Unlike variable binding, coreference does not require a particular syntactic configuration such as c-command. Having both variable binding and coreference available in the parser can then account for pronouns that are linked to either referring or non-referring expressions. This functionality is referred to in POB as the dual-route architecture.²⁶

Variable binding and coreference form part of a larger system that deals with anaphora in POB. Three distinct modules are postulated, which are arranged in a hierarchy. At the top of the hierarchy sits the syntax module in which reflexives are processed via A-binding (we are not concerned with this process here). The semantic module comes next. This is where variable binding takes place, which can be thought of as involving the semantic representation. At the bottom of the POB hierarchy is the discourse module, where coreference is processed. The hierarchy itself is one of assumed economy. Processes that involve fewer cross-modular steps are assumed to be the most computationally efficient. Figure 3.2 illustrates the model performing a variable-binding operation.

²⁵ At the time of writing, Angela Merkel is the Chancellor of Germany.

²⁶ There are some researchers (e.g. Heim 1998; Kayne 2002) who challenge the dual-route architecture. The debate is ongoing, but the distinction between variable binding and coreference is not outside of the mainstream in syntax.

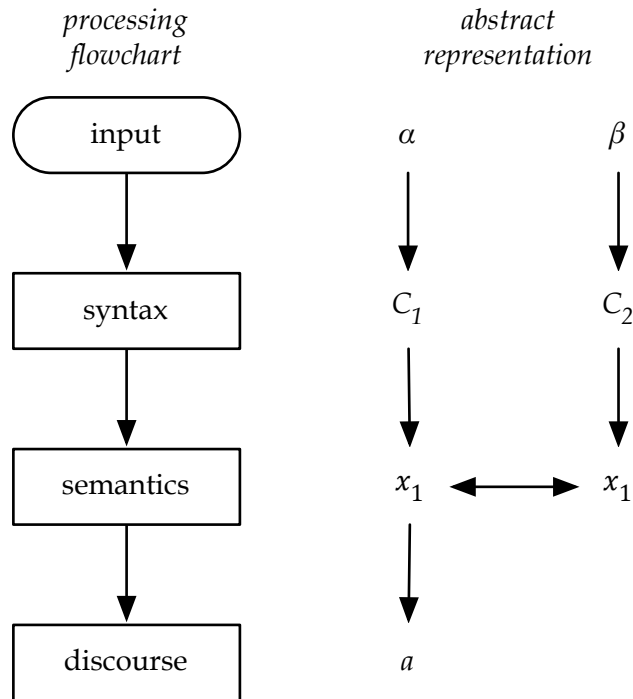


Figure 3.2. Variable binding: three cross-modular steps. From Koornneef (2008, p. 44).

Figure 3.2 shows how variable binding is processed in the POB model. At the top are two different expressions, α and β , which are first processed in the syntax module. Since they do not form an A-chain in the syntax module, the two objects separately move to the semantics module, counting as two cross-modular steps. Within the semantics module, variable binding takes place. Variable binding connects the two expressions (shown in Figure 3.2 by the horizontal arrow), so the output from the semantics module is only one object. This object is moved to the discourse module to be integrated into the discourse representation. Movement of one object from the semantic to the discourse module counts as one cross-modular step, making three cross-modular steps in total for the variable-binding operation. Coreference is more costly by comparison. During coreference processing, the input begins in the same way as before, but this time the output of the semantics module is two separate objects, since variable binding has not identified that they are the same. This counts as two cross-modular steps (the transfer of two separate objects from the

semantic module to the discourse module). In the discourse module, the objects are identified as pointing to the same discourse representation. A total of four cross-modular steps (two from the output of the syntax module and two from the output of the semantics module) is therefore necessary for coreference processing in the POB model. It is assumed that the parser will always opt for a more computationally efficient operation if it is available; A-chain formation (for reflexives) will take place if the conditions are met, variable binding will take place in preference over coreference if the conditions are met. In this way variable binding and coreference are captured within the POB model as operations that take place in different modules, and, importantly, variable binding is assumed to be more economical and therefore preferred by the parser over coreference. This notion of economy is not unique to POB or Reuland's work; it is also seen in other syntactic approaches to anaphora such as those based on Optimality Theory (e.g. Hendriks and Spender 2006). It also features in Burkhardt's (2005) work which is discussed below.

3.2.5 The POB hypothesis

The POB model is mainly concerned with economy. Economy is important because it is how the model captures the distributional properties of various anaphoric elements (this is not directly of concern in the current chapter). Most important is the claim that the parser favours the most economical process in resolving anaphora. A-chain formation is more economical than variable binding, which is in turn more economical than coreference. There is also a more specific claim that relates to the process of searching for antecedents. That is, upon encountering a pronoun, the parser initially searches for and evaluates variable-binding antecedents before any coreference antecedents are considered. (A variable-binding (VB) antecedent is one that c-commands the pronoun as summarised in condition D above. A coreference (CR) antecedent is one that can only be

linked to the pronoun via coreference.)²⁷ This processing claim from POB about antecedent search is the central claim evaluated in the experiments described below. It is important to note that this is a claim about timecourse; it proposes an initial preference, which can be overcome at a later point in processing. Overcoming the initial preference necessitates reanalysis and is assumed to be costly, but possible. The implication of this claim is that, in a sentence containing a pronoun and two potential antecedents, the parser will search first for c-commanding antecedents that are potential variable binders. Secondly, non-c-commanding antecedents are considered. The claim is not that pronoun processing ends if a successful VB antecedent is found. Rather, variable binding is an initial preference, after which other processes can take place. This POB claim has fascinating implications for the parser's use of grammatical constraints during processing and certainly warrants further investigation. Normally, researchers have observed a pattern of pronoun interpretation offline that is subject to a constraint. Condition B is a good example. The question for such constraints is whether and when the constraint applies during processing and whether it is violable. The POB claim, on the other hand, proposes that a constraint operates during processing but its effects are not always visible in offline interpretations.

The implications of the POB claim for the parser have to be separated from the implications of a condition D constraint on its own. Condition D, like condition B, is visible at the level of offline interpretations. Of course there are exceptions to the constraint, which means that there are structures in which a pronoun can be linked to a QP antecedent that does not c-command it. But in canonical cases involving QPs, condition D applies. The

²⁷ While CR antecedents may also c-command the pronoun, leading to some debate about the distinction between coreference and variable binding in such a configuration (see, e.g. Büring 2005 for a discussion of this issue), the CR antecedents discussed in this thesis do not c-command the pronoun and so cannot be considered as potential variable binders (although an alternative analysis is briefly considered in Chapter 4, section 4.10.3).

questions relating to processing would be whether and when c-command plays a role online, and whether the c-command constraint is temporarily violated while the parser considers non-c-commanding QP antecedents that are ruled out in eventual interpretation.²⁸ It is important to distinguish these questions, which are worthy of investigation in themselves, from the questions arising from the POB claim above. Investigating the POB claim involves testing whether or not VB antecedents are considered *before* CR antecedents during processing.

3.3 Evidence bearing on the POB claim

3.3.1 Evidence from ellipsis

The first set of evidence which indirectly supports the POB claim of priority for variable binding comes from the processing of ellipsis structures. According to POB, the parser computes relations at the level of semantics (VB) before computing discourse relations (CR). Koornneef (2008) proposes that this configuration of the parser drives the interpretation preferences in ellipsis structures containing pronouns. Consider (3.8) (from Frazier and Clifton 2006):

(3.8) Fred kicked a cockroach away from his sleeping bag and Henry did too.

The elided VP (*did too*) is normally assumed to stand in for a copy of the VP from the previous clause, in this case *kicked a cockroach away from his sleeping bag*. There are two possible interpretations of this example. One interpretation (known as the sloppy reading) is that Henry kicked a cockroach away from his own sleeping bag. Under this interpretation, *his* in the elided VP refers to Henry, with a binding relation between *Henry* and *his*. The other interpretation (known as the strict reading) is that Henry

²⁸ This issue is currently being investigated See Cummings, Patterson and Felser (2013).

kicked a cockroach away from Fred's sleeping bag. Under this reading, the elided *his* refers to Fred. The relation between the elided *his* and *Fred* must be coreference, since *Fred* does not c-command *his*. So the sloppy reading is associated with a variable-binding relation and the strict reading is associated with coreference. Koornneef reviews evidence from a number of offline studies that show a preference for the sloppy interpretation in this type of structure. This is not always reflected in online studies, in which both strict and sloppy interpretations appear to be activated (e.g. Hestvik 1995). But Koornneef argues that studies using cross-modal priming may not capture the timecourse of processing that is so crucial to the POB hypothesis. Koornneef then presents evidence from an online study that he conducted (in Dutch) which demonstrates a sloppy interpretation preference via longer second-pass reading times in VP ellipsis structures that have a strict (i.e. dispreferred) interpretation bias (although the first-pass reading times go in the opposite direction: a finding that Koornneef attributes to a confound in the length of the materials). This preference for interpretations that involve variable-binding relations over those involving coreference is interpreted as evidence for the POB claim. The parser first identifies a VB antecedent in all cases, but when the discourse biases the interpretation towards a strict reading the parser reanalyses this variable binding and opts instead for coreference. The reanalysis process requires extra processing, which causes the longer second-pass reading times in the strict-bias contexts.

While the evidence appears to support the POB hypothesis that variable binding is computed before coreference, there are other explanations for the results. For example, one could assume that in environments with a sloppy bias only the sloppy reading is computed, and in strict bias environments only the strict reading is computed. Since the strict reading involves a longer-distance dependency (referring to an antecedent from the previous clause instead of the same clause) it could take longer to compute this reading (Clark and Sengul 1979). This would result in longer reading times in the strict reading environments and may also generate more

backwards looks. Such a story would be compatible with the results found by Koornneef. So, the results from the ellipsis structures, while they appear to be in the right direction for the POB hypothesis, could also support other explanations in which an initial preference for variable binding plays no part. Additionally, we should be cautious in taking a set of assumptions about pronouns within ellipsis structures and applying them directly to environments containing overt pronouns. This is because in ellipsis structures it is not clear how much material is actually reconstructed during processing, which makes it difficult to pinpoint precisely which elements within the VP are causing the effects.

3.3.2 Evidence from overt pronouns

Koornneef (2008; also reported in Koornneef, Wijnen and Reuland 2006) has also investigated the initial preference for VB antecedents in a structure with an overt pronoun in an eye-tracking study. The experimental items contained a pronoun which was preceded by two potential antecedents (see (3.9) below: the original items are in Dutch, and the examples below are the English translations provided by Koornneef). One of the antecedents for the pronoun was a QP such as *every worker*. This antecedent c-commanded the pronoun and as such could be linked to the pronoun via variable binding. The second potential antecedent was a name such as *Paul* which did not c-command the pronoun but could be linked to the pronoun via coreference. As such, the pronoun could be linked either to a VB antecedent (QP) or a CR antecedent (name). Using pretested biasing contexts, two conditions for these items were created: one in which the discourse biased towards the QP as an antecedent for the pronoun, and one in which the discourse biased towards the CR antecedent.

(3.9a) VB antecedent bias

A working day in the factories alongside the North Sea Canal is always very tough. Especially today a lot of workers could barely cope. *Every worker who just like Paul was running out of energy,*

thought it was very nice that he could go home early this afternoon.

Hopefully a hot shower would ease the pain.

(3.9b) *CR antecedent bias*

A working day in the factories alongside the North Sea Canal is always very tough. Especially today the old worker Paul could barely cope. *Every worker who noticed that Paul was running out of energy, thought it was very nice that he could go home early this afternoon.* Hopefully a hot shower would ease the pain.

The POB hypothesis assumes that in both conditions the parser should search for and evaluate the VB antecedent first. In the VB antecedent bias condition, the discourse information would then also point to the same antecedent, but in the CR antecedent bias condition the discourse information would conflict with the initial VB antecedent and trigger a reanalysis so that the CR antecedent would be interpreted as the antecedent for the pronoun. This leads to a prediction for longer reading times in the CR condition compared to the VB condition. Koornneef's results confirmed his prediction; longer second-pass reading times were observed in the CR antecedent bias condition.

Koornneef's experiment for overt pronouns in fact has a similar problem to the experiment with ellipsis. That is, the design allows for an assessment of which condition took longer to process as a whole, but it does not offer insight into the supposed sequential consideration of the antecedents.

There is no way of knowing, from the current design, which antecedent is being considered at which point. We have to assume that overall, readers interpreted the QP as the antecedent in the VB-bias condition and the name in the CR antecedent, but there is no way to get direct evidence of which antecedent is being considered at a given moment. Additionally, the discourse biases are achieved via different syntactic structures, which may well have played an independent role in elevating or reducing the reading time in one condition over another. So while the results again point in the

right direction, they do not offer strong evidence in favour of a VB-initial processing strategy.

The evidence so far regarding the POB hypothesis is rather limited. On one hand, POB provides a nice explanation for a sloppy-reading preference in VP ellipsis, but other explanations for the experimental results are also possible. There may be evidence for the POB hypothesis from Koornneef's experiment with overt pronouns in Dutch, but the experimental design does not allow for a detailed examination of the timecourse of processing. Clearly, the issue requires further investigation. In order to continue the investigation of the POB hypothesis, Experiments 4 and 5 were carried out; these are reported below. As with Koornneef's study, Experiments 4 and 5 use QPs as potential antecedents. Before reporting these experiments, therefore, I will present a brief overview of previous processing studies that have investigated QP antecedents.

3.4 Processing studies with QP antecedents

As discussed in section 3.2.3 above, QPs have certain properties that make them distinct from names and DPs. Despite their importance in syntax and semantics, however, QPs have received little attention in the pronoun processing literature. Psycholinguistic studies concerned with other aspects of QPs, that is, the range and scope of meanings expressed by quantifiers rather than their status as antecedents, have focused on issues of scale and in particular on the issue of contrast sets in expressions such as *few in few people came to the party* (e.g. Moxey and Sanford 1993). Quantifier scope ambiguities have also been widely considered. A review of such work is far outside the scope of this thesis. I will limit the discussion to an overview of processing studies that have looked at QP as antecedents for pronouns.

The main study in this area is Carminati, Frazier and Rayner (2002) who investigated the processing of variable binding. The aim of their study was to test the strength of condition D; the authors noted that while condition D

successfully characterises the binding relationship between QPs and antecedents in many cases, there exist a number of exceptions. The exceptions either exhibit variable binding without c-command, or a QP linked to a pronoun via coreference. The first experiment tested the processing consequences for linking a pronoun to a QP as opposed to a DP, and whether this was affected by c-command. In an eye-tracking study, participants read sentences of the following type:

(3.10a) *Quantifier-conjoined*

Every British soldier aimed and then he killed an enemy soldier.

(3.10b) *Quantifier-complement*

Every British soldier thought that he killed an enemy soldier.

(3.10c) *NP-conjoined*

The old British soldier aimed and then he killed an enemy soldier.

(3.10d) *NP-complement*

The old British soldier thought that he killed an enemy soldier.

In the conjoined conditions (3.10a and 3.10c), the QP/DP does not c-command the pronoun *he*, whereas in the complement conditions (3.10b and 3.10d), the QP/NP does c-command *he*. If condition D holds, (3.10a) will be problematic, since there is no c-command relationship hence no legitimate way to link the pronoun to the QP. For the DP in (3.10c) the lack of c-command is not a problem because coreference is available. This should lead to a processing advantage for (3.10c) over (3.10a). No such discrepancy should be detectable for (3.10b) versus (3.10d), because the antecedents c-command the pronoun in that structure, making variable binding available. Note that the structure of the sentences differs between the conjoined and the complement conditions, which means that they are not directly comparable in an analysis. For this reason, the precise prediction for strict adherence to condition D is as follows: if c-command is necessary for linking a pronoun to a QP, there should be a larger difference

in the reading time measures between (3.10a) and (3.10c) than between (3.10b) and (3.10d). A similar prediction could also be derived from the POB model.

The results from the first experiment showed, first of all, that sentences containing QPs took longer to read than sentences containing NPs. This was measured using a time-to-complete measure from the pronoun region onwards, which means that after participants entered the pronoun region it took longer for them to finish reading the sentences containing the QPs than those containing the NPs. Increased regressions from the end of the sentence region appear to have landed in the pronoun region or earlier, leading to the QP sentences having higher rereading times in the pronoun and previous regions. The results also showed a difference in the time-to-complete measure between the sentences containing complement clauses and those that contained conjoined clauses. This difference is likely to be attributable to the increased rereading times in the region after the pronoun. The authors suggested that the difference here could be due to the material in the conjoined clauses being more predictable than the material in the complement clauses. It is also possible that structural complexity plays a role as well because the complement clause is embedded inside the matrix clause while the conjoined clause is a simple clause following the first simple clause. There was no interaction between antecedent type and clause type.

The results are interesting for two reasons. First of all, they are not expected under a pure condition D constraint. The lack of c-command in (3.10a) did not cause a discernable slow-down in reading times, which indicates that linking a pronoun to a QP could take place in the absence of a c-command relationship. In fact, the second experiment reported in the paper supports this finding, showing that QPs can be linked to pronouns via a process called telescoping. Telescoping is one of the known exceptions to the condition D constraint. When the QP is involved in describing an event that is made up of a set of predictable actions, it is not necessary for the QP to c-

command pronouns that are used to describe subsequent actions. The example given in Carminati *et al.* (2002) is as follows:

(3.11) Each degree candidate walked to the stage. He took his diploma from the dean and walked to his seat.

In (3.11), while the QP does not c-command *he* in the second sentence, obtaining a VB-like reading is not difficult because the event described in the second sentence is highly predictable from the description of the first event. Telescoping appears to break down when the subsequent events are less predictable. Returning to Carminati *et al.*'s first experiment, it seems that the actions described in the second clause (e.g. killing an enemy soldier) are predictable from the context (a soldier aiming). This is also true for the other materials in their first experiment, many of which include the word "then" in the conjoined condition, strongly indicating an action/event following on from the first event. However, in the absence of offline judgements about the materials, it is difficult to assess whether the pronoun was indeed interpreted as referring to the QP at all. Carminati *et al.* suggested that their findings are in line with suggestions made by Bosch (1983). Bosch asserts that pronouns can be linked to QPs anaphorically or syntactically. The syntactic route is very much like Reuland's variable binding at the level of semantics (despite the difference in labelling) and relies on a structural relationship (possibly c-command). But in the absence of c-command, a pronoun can be linked to an assumed referent arising from the presence of the QP. Such a link would then not be reliant on a syntactic configuration.

Secondly, and perhaps more importantly for the materials in Experiments 4 and 5 reported below (as well as Koornneef's 2008 experiment) the QPs proved to be rather more time-consuming to process as antecedents than the DPs. One possible reason for this is the complexity of the QP compared to the DP. The notion that QPs are more complex to process than DPs is fairly uncontroversial, and is captured in Burkhardt's (2005) syntax-

discourse model. Burkhardt proposes a cost-based hierarchy based on the representational complexity of an antecedent. The higher the representational complexity, the higher the processing cost. Burkhardt distinguishes between non-referential (lowest cost), referential (medium cost) and referential quantified (highest cost) antecedents. Unlike Reuland (2001), Büring (2005) and others, Burkhardt considers a phrase such as *every soldier* as referential and quantified. Because *soldier* is referential, the whole QP is assumed to be referential and is therefore represented at the level of discourse, in much the same way as Pesetsky (1987) assumes that wh-phrases such as *which car* are discourse-linked. Referential antecedents are DPs such as *the soldier*, and non-referential antecedents are 'light' quantifiers such as *everyone* which have no referential content. The experimental evidence that Burkhardt presents unfortunately does not cover the distinction between referential quantified antecedents (*every soldier*) and referential antecedents (*the soldier*), focusing instead on drawing a distinction between quantified (*everyone*) and referential antecedents (*the soldier*) under her classification. So, while the findings of Carminati *et al.* (2002) fit into this hierarchy, there is no direct supporting evidence from Burkhardt for a difference between DPs and QPs. Partial evidence of a difference comes from a self-paced reading study by Warren (2003), who manipulated the NPs in matrix subject position and embedded subject position of a sentence such as (3.12).

(3.12a) *QP in matrix subject position*

Every editor who the journalist genuinely liked applauded the paper's unbiased coverage of the murder.

(3.12b) *QP in embedded subject position*

The journalist who every editor genuinely liked applauded the paper's unbiased coverage of the murder.

In matrix position, a QP such as *every editor* (3.12a) elicited longer reading times than a DP such as *the journalist* in matrix position (3.12b). (No such

cost was observed at the embedded subject position.) Nevertheless, the comparison between the DP and the QP is on two different nouns (*editor* and *journalist*), which could have independently affected the results. Also, these results do not bear on the issue of processing a QP as an antecedent but on initial processing times when the QP is first encountered. As such, they should be treated with caution when making assumptions about QPs as antecedents.

Finally, two conditions from one of Koornneef's (2008) experiments, experiment 4, actually provides possible evidence for QPs being costly antecedents (cf Koornneef 2008, experiment 5, for different findings). Participants read sentences such as the following (translations from the Dutch materials are taken from Koornneef 2008):

(3.13a) *VB-only condition*

A working day in the factories alongside the North Sea Canal is always very tough. Especially today a lot of workers could barely cope. *Every worker who was running out of energy, thought it was very nice that he could go home early this afternoon.* Hopefully a hot shower would ease the pain.

(3.13b) *CR-only condition*

A working day in the factories alongside the North Sea Canal is always very tough. Especially today the old worker Paul could barely cope. *Paul was running out of energy. It was very nice that he could go home early this afternoon.* Hopefully a hot shower would ease the pain.

Rereading times at the pronoun region were higher in the VB-only condition compared to the CR-only condition. This could indicate that it is more difficult to process a QP compared to a name, although, as in Koornneef's previous experiment, the two structures are different making the comparison somewhat inexact; the pronoun part of the CR-only

condition contains two sentences whereas the VB-only condition contains one more complex sentence.

In sum, the evidence regarding QPs as antecedents is rather incomplete. Carminati *et al.*'s (2002) results are the clearest, showing a processing slowdown when an antecedent is a QP compared to a DP. But this finding is not echoed in Koornneef's (2008) results, although Koornneef did find a slowdown for a QP antecedent compared to a name antecedent (which may or may not be attributable to the antecedents themselves). Koornneef argues that even though in isolation it could be more costly to link a pronoun to a QP than another type of antecedent, this does not affect the serial ordering in the POB model.

The studies presented here give a rather inconclusive overall picture of the role of variable binding and coreference in pronoun processing. While there is indirect evidence from studies on ellipsis that there is a variable-binding preference, direct evidence for its operation in pronoun processing is limited. Additionally, QPs may in fact be more costly as antecedents than their non-quantified counterparts, although again the firm evidence for this is limited to one study (Carminati *et al.* 2002). What is clear is that the POB model does offer an exciting opportunity to explore whether underlying constraints operate in a serial manner during processing. The POB claim may or may not be undermined by the fact that on their own, QPs appear to be costly as antecedents. The following two sections describe two experiments which further investigated the POB claim that the parser initially searches for a VB-antecedent before considering a CR antecedent.

3.5 Experiment 4: investigating a VB-initial search mechanism

Experiment 4, described in this section, directly bears on the *POB* claim of a VB-initial processing strategy, and is reported in Cunnings, Patterson and Felser (accepted). The paper reports two online experiments and one offline

experiment on the topic of variable binding and coreference. The first experiment directly addresses the POB hypothesis, while the second (online) and the third (offline) experiments follow up on a related issue of the linear ordering of antecedents. The second and third experiments from the paper are reported in Chapter 4 of this thesis. Because of the importance of the paper's first experiment to the issues discussed in the current chapter, as well as its relevance to the experiments in the following chapters, I summarise the experiment below, and refer to it as Experiment 4.²⁹ A fuller description, including details of the analysis and results, can be found in the paper itself.

Experiment 4 was designed to further investigate the POB hypothesis of a VB-initial search mechanism, whilst addressing some of the design problems in the Koornneef (2008) study outlined above. It was an eye-tracking study in English, with 27 participants (who were all native speakers of English). Similar to Koornneef's experiment, experimental materials contained a pronoun and two potential antecedents, one a QP that c-commanded the pronoun and the other a name that did not c-command the pronoun. These constituted the VB antecedent and the CR antecedent respectively. Rather than manipulating the discourse context, the sentences were placed within neutral discourse contexts that did not mention any other potential antecedents. The (stereotypical) gender of the antecedents was manipulated in a match-mismatch paradigm to create four conditions, as illustrated below in (3.14). The examples below show where the line breaks appeared in the experimental materials. The QP and the name appeared on the same line, and the pronoun appeared in the middle of the next line. The full set of experimental materials is given in Appendix 3.

²⁹ Data collection and analysis for Experiment 4 was carried out by Ian Cummings, and for this reason it is not treated as one of my own experiments which are reported in detail in this thesis. The experiment has been numbered in sequence for ease of reference because it is discussed in the current and subsequent chapters.

(3.14a) *QP Match, Name Match*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(3.14b) *QP Match, Name Mismatch*

The squadron paraded through town. Every soldier who knew that Helen was watching was convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(3.14c) *QP Mismatch, Name Match*

The squadron paraded through town. Every soldier who knew that Helen was watching was convinced that she should wave as the parade passed. The entire town was extremely proud that day.

(3.14d) *QP Mismatch, Name Mismatch*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that she should wave as the parade passed. The entire town was extremely proud that day.

The stereotypical role nouns were selected from a pretested list of role nouns.³⁰ It has previously been demonstrated that stereotypical role nouns of this type can give rise to mismatch effects during anaphoric resolution (e.g. Kennison and Trofe 2003; Sturt 2003; Kreiner, Sturt and Garrod 2008; Cunnings and Felser 2013).

The POB hypothesis of VB-initial processing claims that, on encountering the pronoun, the parser initially searches for VB antecedents, in this case, *every soldier*. *James/Helen* cannot be considered as a VB antecedent because

³⁰ The stereotypical role nouns were pretested by Ian Cunnings. Participants (who did not take part in Experiments 4 or 5) were asked to rate a set of role names on a 7 point scale, with “typically male” at one end of the scale and “typically female” at the other end. Only those role names that were highly rated as typically male or female were included in the materials.

these names do not c-command the pronoun, meaning that they can only be linked to the pronoun via coreference. In two of the conditions (3.14c and 3.14d) the stereotypical gender of the QP mismatches that of the pronoun, so this mismatch should cause reading disruption when the QP is considered. This should lead to longer reading times/more regressions in the QP mismatch conditions compared to the QP match conditions. Any effects of the gender match between the name and the pronoun should occur afterwards, according to the POB hypothesis.

Eye-movement data from the pronoun region (*that he / that she*), the spillover region which was the following two words (*should wave*), the prefinal region which was the two words following the spillover region (*as the*) and the final region (*parade passed*) was analysed. There was a main effect of name in the following regions and measures: rereading and total viewing times in the pronoun region; regression-path times, rereading times and total viewing times in the spillover region. There was a marginal effect of name in the regression-path times in the prefinal and final regions. In all regions and measures the name mismatch condition had longer reading times than the name match condition. There were no effects of the QP in any region or measure.

The results from Experiment 4 point in exactly the opposite direction from the POB hypothesis. Higher reading times for the mismatching name compared to the matching name, and no effect of the QP's gender, indicate that readers were only considering the name as an antecedent for the pronoun and not the QP at any point. This is incompatible with the POB hypothesis. There are one or two issues to consider, however, in relation to Experiment 4. Firstly, the experimental materials are confounded in that the QP was always further away from the pronoun than the name, which could have given the name an advantage over the QP. In the POB hypothesis this should not matter; the parser should be searching for a c-commanding antecedent and should therefore find only the QP in this initial phase, regardless of the position; nevertheless, the influence of

relative ordering of the antecedents should be considered. A second consideration is that perhaps the effect of the QP occurred so early that it was not detectable. This is unlikely for two reasons. Firstly, the effects of the name antecedent in Experiment 4 first emerged in the pronoun region in first-pass times. This is an early measure, so if the QP effect was quicker than that it could be rather difficult to detect it with behavioural measures at all. Secondly, the disadvantage for the CR-bias condition in Koornneef's experiment was detected at second-pass times. If this disadvantage reflects a reanalysis process (undoing the VB relation and computing a CR relation instead) as Koornneef suggests, then it is not very compatible with an extremely quick initial VB link that can be undone before it can be detected at first-pass times.

Experiment 4 produced clear results, based on a match-mismatch design, that are incompatible with the POB hypothesis. Experiment 5 presents further evidence on this issue, with a slightly altered experimental design. Experiment 5 is reported in full below, and the differences between Experiments 4 and 5 are detailed in the materials section.

3.6 Experiment 5: further evidence against a VB-initial search mechanism

Experiment 5 in fact investigated two sets of questions.³¹ The first was whether there is initial consideration of the VB antecedent during pronoun processing, as predicted under the POB model, or indeed whether there is a lack of interest in the QP antecedent during processing, as found in Experiment 4. The second set of questions relates to how processing is affected by prior mention in the discourse. The full results and analysis of the experiment are described below. However, the discussion section limits the discussion to address the first set of questions, which are relevant to the

³¹ This experiment has been reported in Patterson (2012), with a slightly reduced number of participants. A reanalysis of the data is presented in Patterson (2013).

current chapter. Discussion of issues relating to the prior context are explored in Chapter 5.

3.6.1 Experiment 5 materials

The materials were adapted from the materials in Experiment 4. They were comprised of 24 short scenarios, of three sentences each. The first sentence was the context sentence, the second was the critical sentence and the third was the wrap-up sentence (see (3.15) below). The context sentence mentions one of the two potential antecedents: this was systematically varied. The critical sentence contained the pronoun and both potential antecedents. The first potential antecedent in the critical sentence was a QP (e.g. *every soldier*), and the second was a DP (e.g. *the queen*). All the nouns in the QPs were stereotypical role names, as per Experiment 4. All the DPs were nouns with definitional gender. The overall set of QPs did not differ in length (letters) or frequency from the set of DPs, although it was not possible to match length and frequency in each individual item. To avoid biases, the texts were composed so that either antecedent could plausibly be the referent of the pronoun, but the gender cues determined the correct antecedent. (3.15) shows the four conditions for Experiment 5. The QP and the name always appeared on the same line, and the pronoun appeared in the middle of the next line on the screen. The full set of experimental materials is given in Appendix 4.

(3.15a) *DP antecedent, context match*

The queen felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was
absolutely convinced that she should wave as the parade passed.
The entire town had come out to watch.

(3.15b) *DP antecedent, context mismatch*

The soldiers felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was

absolutely convinced that she should wave as the parade passed.
The entire town had come out to watch.

(3.15c) *QP antecedent, context match*

The soldiers felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was
absolutely convinced that he should wave as the parade passed.
The entire town had come out to watch.

(3.15d) *QP antecedent, context mismatch*

The queen felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was
absolutely convinced that he should wave as the parade passed.
The entire town had come out to watch.

Two factors were systematically varied to create the four experimental conditions shown above. These factors were *antecedent* and *context*. *Antecedent* refers to whether the pronoun matches in gender with the QP antecedent or the DP antecedent. *Context* refers to whether the context sentence matches or mismatches with the pronoun.

The design of this experiment differs from Experiment 4 in three ways. Firstly, Experiment 4 did not involve any manipulation in the context sentence. The contexts in Experiment 4 were neutral in that they did not introduce either referent. Instead, they set the scene for the critical sentence by introducing the event or situation. The context manipulation used in the current experiment is discussed fully in Chapter 5. Secondly, Experiment 4 used proper names as the non-quantified antecedents (as per Koornneef 2008), whereas DPs such as *the queen* are used in Experiment 5. Proper names are thought to have particularly stable discourse representations (Garrod and Sanford 1990), and might therefore be more attractive as antecedents than QPs, giving an unfair advantage to the name antecedents. The use of DPs in the current experiment avoids this problem. Finally, Experiment 4 manipulated the gender match between the pronoun and the

antecedents in a typical match–mismatch paradigm. The current experiment presents a single-match paradigm in which two conditions are QP match and two conditions are DP match. This means that no potentially confusing conditions (double match and no match) are presented, and also allows for the interaction between the *context* and *antecedent* factors. An additional advantage of this design is that processing effects are measured when the parser is forced to resolve to the QP. The Experiment 4 results showed a clear preference for the name antecedent and no effect of the QP antecedent, which may suggest that the QP was not being considered at all. But it could be that the match–mismatch paradigm did not give a dispreferred (QP) antecedent enough opportunity to be considered. If the name is really the preferred antecedent for independent reasons, the parser will be satisfied when the name matches and may not engage in a further search for an antecedent. This is the case in Experiment 4 name match and double match conditions. Only when the name mismatches, then, is there a chance that the QP can be considered as a backup option. If that is the case, the potential for the QP to have an effect on the reading times is down to the difference between just two conditions, namely the difference between the QP match condition and the no match condition. This restriction may lead to the difference not being detected because of a lack of power; effectively, the dispreferred antecedent has less of a chance of affecting the reading times than the preferred antecedent. The design of Experiment 5 avoids this problem.

3.6.2 Experiment 5 predictions

In this section I will deal in more detail with the predictions relating to the *antecedent* factor. The predictions relating to the *context* factor are presented in Chapter 5.

3.6.2.1 Prediction set (i): POB

In the materials there are two conditions (3.15a and 3.15b) where only the DP matches in (stereotypical) gender with the pronoun, and two conditions (3.15c and 3.15d) where only the QP matches in (stereotypical) gender with

the pronoun. Under POB assumptions, procedures at the level of semantics take place before procedures at the discourse level. During processing, it is hypothesised that when a pronoun is encountered, the parser first seeks out a c-commanding antecedent. In all conditions of Experiment 5 there is a c-commanding QP which can bind the pronoun, but in the DP match conditions (3.15a and 3.15b) the QP mismatches in stereotypical gender with the pronoun. This mismatch in the QP's gender should lead to processing disruption, which will manifest as longer reading times in the DP conditions compared to the QP conditions, at and beyond the pronoun region. In Koornneef's (2008) experiment, disruption of this kind was actually only detected in second-pass measures, so disruption in the rereading times in the pronoun region and later regions might be expected. While predictions about a specific reading time measure cannot be made, a general ordering can be imposed on the effects. If the discourse information is taken into account after the initial search for a c-commanding antecedent, then any effects of the DP antecedent's gender (and the context manipulation) should be in evidence after initial effects of the QP antecedent's gender. As such, an initial slowdown in the QP antecedent conditions might be followed by a later slowdown in the DP antecedent conditions.

3.6.2.2 Prediction set (ii): anti-QP

Experiment 4 showed effects of the name antecedent and no effect of the QP. This could suggest that the QP is not considered at any point as a potential antecedent, and that on encountering the pronoun the parser will attempt to link the pronoun only to the DP. On this basis, then, when the DP mismatches in gender with the pronoun (3.15c and 3.15d) there will be disruption, compared to when the DP matches (3.15a and 3.15b). Following the results of Experiment 4, these effects should be evident in first-pass times in the pronoun region and later. A retrieval mechanism that makes use of number or gender features (see Chapter 2, section 2.3.2) may also make this prediction. This is because it is not clear how a QP's semantic number would be encoded in memory; a QP might, for example, be

encoded as a set with a plural number feature, making it less likely to be retrieved than a singular DP when a singular pronoun is encountered.

3.6.2.3 Prediction set (iii): QP as backup

So far, the predictions are for a simple effect of antecedent match, with the effect predicted to be in opposite directions depending on whether the POB or an anti-QP hypothesis is followed. However, the anti-QP hypothesis might be too strong. If the DP is preferred but the QP is considered as a backup option, the disruption in (3.15c) and (3.15d) may be short-lived and limited to earlier regions or measures, and later disruption will not be expected because the QP provides a matching antecedent in these conditions.

3.6.2.4 Prediction set (iv): interactions with context

The context manipulation might have a uniform effect on the DP and the QP antecedent alike, in which case a simple effect of context would be expected. This is discussed fully in Chapter 5. In the predictions below, I will consider how prediction sets (i) to (iii) above would be affected if context interacts with antecedent type. I will not discuss the direction of the context effect here, only its potential impact on the antecedent factor.

Under POB assumptions, the search for a c-commanding operator to bind the pronoun should be exempt from considerations of the discourse. Thus there should be no difference between (3.15c) and (3.15d), when the search at the level of semantics finds a satisfactory (matching) operator to bind the pronoun. The POB mechanisms do assume, however, that after the initial search, other procedures can take place at the level of discourse. This includes linking the pronoun to the DP via coreference. Thus, there could be an effect of context in the DP match conditions (3.15a and 3.15b). This effect would manifest as an interaction between antecedent and context, where there would be no difference between the two QP match conditions (3.15c) and (3.15d) but a difference between (3.15a) and (3.15b).

For the anti-QP prediction set, similar to the POB an interaction between antecedent and context is predicted, but for different reasons. The anti-QP predicts that the QP is not considered as an antecedent at all. If this is the case, effects of the context match or mismatch would not be predicted in the QP match conditions. But the prior context may affect the processing of the preferred (DP) antecedent, leading to a difference between (3.15a) and (3.15b).

For the QP as backup prediction set, effects of the context can be expected on both antecedents, but the effects may emerge at different times. If the DP antecedent is preferred and is therefore considered first or faster, the context effect should appear earlier for the DP match conditions compared to the QP match conditions.

3.6.3 Experiment 5 method

The materials were arranged in a Latin-square design, resulting in four lists. The order of items in the lists were reversed to create a further four lists, to minimise learning or tiredness effects on particular items. In addition to the 24 experimental items, there were 48 fillers, of which 12 were pseudo-fillers. The pseudo-fillers contained a pronoun and two potential antecedents, but were of a different structure to the experimental items. The fillers were of a variety of different syntactic structures. Half of the fillers were sentences without pronouns but of a comparable length to the experimental items (they contained names and occupations, like the experimental items); there were also four items containing reflexives, and 14 containing pronouns of a different type to those in the experimental items.

3.6.3.1 Experiment 5 participants

35 native speakers of English, who were recruited from the University of Essex and the surrounding community, were tested. The data from 32 participants was used in the analysis; three participants were removed due to track loss. The mean age of the 32 participants was 29, range 18-60; the

group contained 13 males. All participants were paid £7 for their participation. All participants confirmed that they had not been diagnosed with any language difficulty/disorder. Some participants wore contact lenses or glasses during the experiment, and all reported that they could read from the screen successfully.

3.6.3.2 Experiment 5 procedure

In this experiment the eye movement data was recorded using the head-mounted Eyelink II system (by SR Research, Canada). Participants were seated in front of a PC monitor at a distance of 80cm. They wore a headset carrying two eye cameras and a head camera, with a rubber cap underneath the headset to prevent it from slipping during the experiment. The head camera compensated for small head movements that might be made during recording via the tracking of LEDs at the four corners of the display monitor. The position of the eye cameras was adjusted for each participant. Other aspects of the procedure are as described in Chapter 2, section 2.5.3.2.

Following two thirds of the trials there were comprehension questions. This was to ensure that the participants were reading and understanding the trials that were presented. Comprehension questions did not explicitly probe the referent of the pronoun, in order not to draw attention to the purpose of the experiment or encourage strategic reading behaviour. After the eye-tracking part of the experiment was complete, participants took a working memory test and finally they did a questionnaire experiment (Experiment 6 reported in Chapter 4).³² The experiment session took around one hour in total.

³² A working memory test was included in the experimental session because previous studies have revealed an effect of working memory capacity on the processing of reflexives (e.g. Cunnings and Felser 2013). An initial analysis of the current data using ANOVAs revealed that working memory score did not have an effect on the results of the current experiment, and the working memory score was not included in the mixed models analysis.

3.6.3.3 Data preparation

Individual fixations shorter than 80ms and within one degree of visual angle of a neighbouring fixation were merged with the neighbouring fixation. Other fixations shorter than 80ms, and longer than 1200ms, were removed. Within a trial, regions that were initially skipped were removed from the analysis. The eye-movement measures examined in the current experiment are first-fixation times, first-pass times, regression-path times, rereading times and total viewing times. In the current experiment, data from three participants was removed due to excessive track loss, leaving 32 participants in the analysis. In addition, three individual trials were removed from the analysis due to excessive track loss, representing 0.5% of the data. The critical sentence for each item was divided into the following regions of interest for the analysis:

(3.16)

| Every soldier | *who knew that* | the queen | *was watching intently*
was absolutely | convinced | that she | should wave | as the |
parade passed. |

- **[Every soldier]**: QP region, containing the QP (post-hoc analysis only)
- **[the queen]**: DP region, containing the DP (post-hoc analysis only)
- **[convinced]**: precritical region, containing one word preceding the pronoun; this was always the past participle of a verb whose subject was the QP
- **[that (s)he]**: pronoun region, containing *that* and the pronoun
- **[should wave]**: spillover region, containing two words following the pronoun
- **[as the]**: prefinal region, containing two words following the spillover region
- **[parade passed]**: final region, containing two words following the prefinal region, always the last two words of the sentence.

3.6.4 Experiment 5 results

3.6.4.1 Comprehension questions

Three questions were removed from the analysis of the question scores because they were ambiguous and received low response accuracy rates. For the remaining questions, the overall correct response rate was 84%, range 71-94%.

3.6.4.2 Eye-movement data

The skipping rates in the analysed regions were as follows: precritical, 22%; pronoun, 21%; spillover, 8%; prefinal, 15%; final, 11%. The means and standard deviations for all measures in all analysed regions are shown in Table 3.1. Before the statistical analysis the dataset was transformed using a logarithmic transformation ($\log_e(x+1)$).³³ Untransformed and transformed data was plotted against a normal distribution line and visually inspected. Where the transformed data better fitted the normal distribution, it was used in the analysis in place of the untransformed data. This was the case for all regions and measures except the rereading times (all regions). Untransformed data was used in the analysis of the rereading times.

The data was analysed using linear mixed-effects modelling, using the lme4 package in R (Bates, Maechler and Bolker 2012). For the fixed part of the model, the two-level factors *antecedent* (DP/QP) and *context* (match/mismatch) were included, and the *antecedent* \times *context* interaction. In all models the baseline for *antecedent* was DP and the baseline for *context* was match. The inclusion of each fixed factor in the model was tested by comparing a model containing the factor of interest to a model without the factor of interest, using a Likelihood ratio test. The first test compared a model containing only *antecedent* and *context*, without the interaction, to a model with the two factors plus an interaction. If the additional interaction

³³ $\log_e(x+1)$: 1 was added to each point in the dataset before transformation so that no data point would equal zero (zero cannot be log-transformed).

failed to improve the earlier model (determined by $p < 0.10$ on the Likelihood ratio test) *antecedent* and *context* were each tested for inclusion in the same way. If neither factor on its own improved the model, the empty model was considered the best model (i.e. the fixed factors did not have an effect). For the random part of the model, the full random effects structure was preferred (Barr *et al.* 2013). However, in cases of non-convergence, effects were removed from the model one by one, starting with the interaction, until a converging model was found (as suggested in Barr *et al.* 2013). Likelihood ratio tests for the fixed factors were always conducted using models that had the same random effects structure, to avoid additional differences in degrees of freedom between the models being compared. Where the interaction between the two fixed factors was found to be significant, an additional model was created with the factor *antecedent* re-levelled (so that the baseline was QP); this was to obtain the coefficient for the effect of context at this level. Model estimates from the winning model in each region/measure are shown in Table 3.2. *P*-values determining significance of the *t*-value were derived from the upper bound of the *t*-statistic (Baayen 2008). The positive or negative value of the estimate and corresponding *t*-value indicates the direction of the effect: because the factors are not centred, a significant *t*-value for the factor *antecedent* reveals a difference between the QP and DP conditions when context=match; a significant *t*-value for the factor *context* reveals a difference between the match and the mismatch conditions when antecedent=DP; a significant *t*-value for *context* in the re-levelled model reveals a difference between match and mismatch conditions when antecedent=QP. A significant interaction itself reveals a difference between the two mismatch conditions (i.e. DP antecedent, context mismatch versus QP antecedent, context mismatch).

Table 3.1. Means in all conditions in milliseconds (with standard deviations shown in brackets) for first fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions from precritical to final for Experiment 5.

	First fixation mean		First-pass mean		Regression-path mean		Rereading mean		Total mean	
Precritical region										
DP, context match	211	(68)	244	(124)	312	(253)	126	(251)	370	(298)
DP, context mismatch	233	(87)	262	(126)	359	(424)	113	(226)	375	(254)
QP, context match	222	(73)	240	(90)	379	(424)	124	(259)	364	(273)
QP, context mismatch	210	(62)	231	(96)	330	(315)	128	(260)	359	(273)
Pronoun region										
DP, context match	199	(53)	258	(150)	358	(378)	163	(212)	421	(248)
DP, context mismatch	212	(80)	265	(148)	329	(277)	153	(279)	418	(311)
QP, context match	206	(60)	273	(124)	404	(430)	203	(362)	476	(388)
QP, context mismatch	213	(75)	250	(118)	320	(316)	188	(296)	438	(327)
Spillover region										
DP, context match	204	(80)	292	(188)	430	(459)	243	(336)	536	(380)
DP, context mismatch	195	(58)	273	(172)	340	(270)	184	(300)	457	(324)
QP, context match	198	(71)	291	(184)	447	(535)	214	(350)	505	(434)
QP, context mismatch	204	(68)	288	(157)	436	(471)	232	(317)	520	(347)
Prefinal region										
DP, context match	211	(61)	281	(144)	438	(371)	182	(289)	464	(314)
DP, context mismatch	211	(64)	273	(138)	492	(721)	173	(278)	446	(306)
QP, context match	206	(60)	285	(152)	417	(458)	140	(227)	425	(259)
QP, context mismatch	217	(72)	311	(161)	448	(340)	205	(335)	516	(377)
Final region										
DP, context match	224	(79)	312	(180)	991	(1573)	129	(220)	441	(271)
DP, context mismatch	202	(67)	310	(177)	769	(1019)	126	(280)	436	(345)
QP, context match	203	(73)	297	(165)	825	(1331)	140	(256)	437	(314)
QP, context mismatch	213	(75)	312	(199)	812	(986)	134	(237)	446	(329)

Table 3.2

	First-fixation durations		First-pass times		Regression-path times		Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region										
Antecedent (match)	0.05 (0.04)	1.27	0.02 (0.04)	0.55	0.14 (0.08)	1.72(*)	-	-	-	-
Context	0.10 (0.04)	2.41*	0.10 (0.05)	1.81(*)	0.12 (0.08)	1.52	-	-	-	-
Antecedent x Context	-0.14 (0.05)	-3.03**	-0.13 (0.06)	-2.43*	-0.20 (0.08)	-2.39*	-	-	-	-
Context (antecedent=QP)	-0.04 (0.04)	-0.98	-0.04 (0.05)	-0.66	-0.09 (0.08)	-1.13	-	-	-	-
Pronoun region										
Antecedent (match)	-	-	-	-	-	-	-	-	-	-
Context	-	-	-	-	-0.10 (0.05)	-2.17*	-	-	-	-
Antecedent x Context	-	-	-	-	-	-	-	-	-	-
Spillover region										
Antecedent (match)	-	-	-	-	-0.01 (0.07)	-0.07	-35.26 (34.60)	-1.02	-0.12 (0.07)	-1.70(*)
Context	-	-	-	-	-0.14 (0.07)	-1.86(*)	-64.28 (32.04)	-2.01*	-0.14 (0.06)	-2.29*
Antecedent x Context	-	-	-	-	0.17 (0.08)	1.95(*)	78.57 (43.94)	1.79(*)	0.24 (0.08)	2.87**
Context (antecedent=QP)	-	-	-	-	0.03 (0.07)	0.42	14.28 (31.58)	0.45	0.09 (0.06)	1.42
Prefinal region										
Antecedent (match)	-	-	0.01 (0.05)	0.24	-	-	-	-	-0.05 (0.06)	-0.86
Context	-	-	<-0.01 (0.05)	-0.10	-	-	-	-	-0.04 (0.06)	-0.61
Antecedent x Context	-	-	0.11 (0.06)	1.79(*)	-	-	-	-	0.19 (0.10)	1.79(*)
Context (antecedent=QP)	-	-	0.11 (0.05)	2.24*	-	-	-	-	0.15 (0.07)	2.22*
Final region										
Antecedent (match)	-0.10 (0.03)	-2.82**	-	-	-	-	-	-	-	-
Context	-0.10 (0.03)	-2.83**	-	-	-	-	-	-	-	-
Antecedent x Context	0.14 (0.05)	2.90**	-	-	-	-	-	-	-	-
Context (antecedent=QP)	0.04 (0.03)	1.27	-	-	-	-	-	-	-	-

Table 3.2. Model outcomes for each region and measure, Experiment 5. The estimate for the adjustment for each factor in the model is given in the estimate column, and the standard error of the estimate is shown in brackets. The t-value for each estimate is given with the significance level indicated: () = $p < .10$, * = $p < .05$, ** = $p < .001$. The baseline for antecedent was DP and the baseline for context was match. Coefficient estimates and t-values are not given (indicated by '-') when the outcome of the modelling process was that the empty model was the best model.*

3.6.4.3 Eye-movement data description

In the precritical region, first-fixation durations, first-pass times and regression-path times show a significant interaction between antecedent and context. The mismatch condition has shorter first-fixation durations, first-pass times and regression-path times when the antecedent is a QP compared to when it is a DP. In first-fixation durations there is also a significant main effect of context. This shows that for the DP antecedent, the mismatching context has significantly higher first-fixation durations than the matching context. The difference between match and mismatch in the QP antecedent conditions was non-significant. The main effect of antecedent was also non-significant. A very similar pattern is seen in the first-pass times of the precritical region. There is a significant interaction between antecedent and context, and a marginally significant effect of context. The main effect of antecedent is non-significant as is the effect of context in the QP antecedent conditions. In regression-path times (see Figure 3.3), in addition to the significant interaction, there is a marginally significant effect of antecedent showing that the match condition has marginally longer regression-path times in the QP compared to the DP antecedent condition. The effect of context is not significant in the DP conditions or the QP antecedent conditions. For the rereading times and the total viewing times in this region, no effects or interactions significantly improved model fit.

In the pronoun region, the analysis showed no significant effects in the early measures (first-fixation durations and first-pass times). For the

regression-path times, the best model was one in which only context was included as a fixed factor. There was a significant main effect of context without any interaction; this shows that overall the match conditions had longer regression-path times than the mismatch conditions. This pattern of results can be seen as a numerical trend in the rereading times in the pronoun region but the analysis revealed no significant effects here or in the total viewing times.

In the spillover region, the analysis showed no significant effects in the early measures (first-fixation durations and first-pass times), although numerically the first-pass times follow the same pattern as the later measures. The total viewing times show a significant interaction of antecedent and context. This is reflected in both regression-path times and rereading times as a marginal interaction.¹² This shows that the QP antecedent, context mismatch (QPcmm) condition has longer total viewing times (and marginally longer regression and rereading times) than the DP antecedent, context mismatch (DPcmm) condition. In regression-path times there is a marginally significant main effect of context, showing that the DP antecedent, context match (DPcm) condition had marginally longer regression-path times than the Dcmm condition. There was no effect of context in the QP antecedent conditions and no main effect of antecedent. A very similar pattern is seen in the rereading times. In the presence of a marginally significant interaction between antecedent and context, there is a significant main effect of context, showing longer rereading times in the DPcm condition compared to the DPcmm condition. The main effect of antecedent is non-significant as is the effect of context in the QP antecedent conditions. In total viewing times, in addition to the significant interaction there is a marginally significant effect of antecedent showing that there are marginally longer total viewing times in the DPcm condition compared to

¹² Recall that the data for the rereading times was not log-transformed before analysis, unlike the data from the other measures. This is why the coefficient appears much larger for the rereading times in comparison to other measures.

the QPcm condition. There is a significant effect of context in the DP antecedent conditions with longer total viewing times for DPcm condition compared to the DPcmm condition. There is no effect of context in the QP antecedent conditions.

In the prefinal region, the total viewing times showed a marginally significant interaction between antecedent and context; this was also seen in the first-pass times. This shows marginally longer total viewing and first-pass times in the QPcmm condition compared to the DPcmm condition. There is no main effect of context in either measure when DP is the baseline but there is a main effect of context when QP is the baseline, with longer first-pass and total-viewing times in the QPcmm condition compared to the QPcm condition. There is also no effect of antecedent in either measure showing that there is no difference between the QPcm and the DPcm conditions. There were no significant effects or interactions in the first-fixation durations, regression-path times or rereading times, although the numerical trends are mainly in the same direction as the effects (little difference between match and mismatch for the DP antecedent conditions, but a numerically bigger difference between them in QP antecedent conditions, where mismatch produces higher reading times).

In the final region, it is only the first-fixation durations that show any effects. Here, there is a significant interaction between antecedent and context, as well as main effects of antecedent and context in the DP antecedent conditions, but not in the QP antecedent conditions. The interaction shows that the QPcmm condition has longer first-fixation durations compared to the DPcmm condition. The effect of antecedent shows that first-fixation durations are longer for the DPcm condition than for the QPcm condition. The effect of context shows that the DPcm condition has significantly longer first-fixation durations than the DPcmm condition. In the QP antecedent conditions, there is no effect of context.

3.7 Discussion

The effects of the context manipulation in Experiment 5 are commented upon here when relevant, but the main discussion of the context issues is in Chapter 5. Because of the complex pattern of results, I will begin by summarising and interpreting the pattern of results in Experiment 5 across the different regions. Afterwards, I will discuss the predictions of the various hypotheses outlined above, what can be learned from Experiments 4 and 5 and what the implications are for the POB model.

3.7.1 Summary of Experiment 5 results

Precritical region

The precritical region is analysed in case any effect of the experimental manipulation is visible before the pronoun is encountered. Any effect seen in first-pass measures that begins before the pronoun region is unlikely to be attributed to pronoun processing.¹³ In Experiment 5 there are some effects at the precritical region, but the effects are in a different direction from those at the pronoun region and beyond. This means that the effects seen at the pronoun region and beyond are likely to be a result of processing the pronoun. While orthogonal to the pronoun-related hypotheses central to this chapter, the effects in precritical region are discussed in a bit more detail here because they contribute to the understanding of the overall pattern of results for Experiment 5. The precritical region contained the verb for which the QP was the subject (in all four conditions). For the DP antecedent conditions, early measures were longer for the mismatching condition than for the matching condition, and no such difference was seen in the QP antecedent conditions; in fact there was a trend in the opposite direction. The only difference between the conditions up to and including the precritical region is that the context

¹³ Note that in Experiment 5, the pronoun region contains the word *that* before the pronoun (*he/she*), to capture any parafoveal processing of the pronoun. This means that effects in the precritical region are unlikely to be attributable to parafoveal processing of the pronoun.

sentences are different; the DPcm and the QPcmm conditions (3.15a and 3.15d) contain *the queen* in the context sentence and the DPcmm and the QPcm conditions (3.15b and 3.15c) begin with *the soldiers*. It is not the context manipulation itself (match or mismatch with the pronoun) that is having an effect here because the pronoun has not yet been encountered. Instead, the way the context is set up (introducing *the queen* or *the soldiers*) has an effect on processing the verb. Conditions beginning with *the queen* were easier to process at the precritical region than those beginning with *the soldiers*. The four conditions are shown again here for reference as (3.17).

(3.17a) *DP antecedent, context match*

The queen felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed.
The entire town had come out to watch.

(3.17b) *DP antecedent, context mismatch*

The soldiers felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed.
The entire town had come out to watch.

(3.17c) *QP antecedent, context match*

The soldiers felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed.
The entire town had come out to watch.

(3.17d) *QP antecedent, context mismatch*

The queen felt really quite uneasy about the squadron parade.
Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed.
The entire town had come out to watch.

The question is why, when processing a verb for which the QP is always the subject, does prior mention of *the soldiers* induce a cost compared to prior mention of *the queen*? Processing of a verb is often assumed to involve the reactivation of a verb's arguments, including the subject. If the processing of the subject, the QP, was affected by the context sentence, this may in turn have had an effect on the processing of the verb. To this end, a post-hoc analysis of early reading times at the QP region (*every soldier*) was carried out.¹⁴ Tables 3.3 and 3.4 show the means and model outcomes for three measures in the QP region and also in the DP region (*the queen*).

¹⁴ Data treatment and modelling were the same as for the main analysis. The data was recoded to create a two-level factor called contNP, which indicated whether the context sentence began with *the soldiers* or *the queen*. The baseline was *soldiers*.

Table 3.3. Means for all conditions in milliseconds (with standard deviations shown in brackets) for first fixation durations, first-pass times, and total viewing times in the QP and DP regions in Experiment 5. For ease of reference, contexts containing the soldiers in the example conditions are marked (S) and contexts containing the queen in the example conditions are marked (Q).

	First-fixation durations		First-pass times		Total viewing times	
QP region						
DP, context match (Q)	191	(82)	496	(220)	692	(381)
DP, context mismatch (S)	193	(69)	248	(203)	658	(426)
QP, context match (S)	200	(83)	437	(183)	635	(358)
QP, context mismatch (Q)	186	(77)	483	(232)	696	(390)
DP region						
DP, context match (Q)	202	(67)	408	(207)	662	(401)
DP, context mismatch (S)	210	(69)	478	(272)	792	(502)
QP, context match (S)	209	(62)	492	(342)	821	(581)
QP, context mismatch (Q)	214	(73)	417	(230)	691	(482)

Table 3.4. Outcome of the modelling process for QP and DP regions, Experiment 5. The estimate for the adjustment for the contNP factor in the model is given in the estimate column, and the standard error of the estimate is shown in brackets. The t-value is given with the significance level indicated: * = $p < .05$, ** = $p < .001$. The baseline for contNP was soldiers. Coefficient estimates and t-values are not given (indicated by '-') when the outcome of the modelling process was that the empty model was the best model. SE=standard error.

	First-fixation durations		First-pass times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
QP region	-		0.12 (0.03)	4.78**	0.09 (0.04)	2.37*
DP region	-		-0.13 (0.03)	-3.92**	-0.16 (0.05)	-2.87*

The post-hoc analysis shows that first-pass times and total viewing times of the QP were affected by the context sentence, with shorter reading times when the context sentence began with *the soldiers*. This suggests that the QP was initially easier to process when a set of soldiers had already been introduced by the NP *the soldiers*. Analysis of the DP region shows that when the DP was introduced in the context sentence (*the queen*), the DP was easier to process than when *the soldiers* were introduced in the context sentence. Both these findings suggest that encountering an NP for a second time, even if its form is slightly different (as in the case for *the soldiers...every soldier*), reduces the time needed to process the second occurrence of the NP.

Linking the analysis of the QP region to the processing of the verb in the precritical region, the findings seem rather counter-intuitive. As far as the precritical region is concerned, in conditions with the plural NP (*the*

soldiers) beginning the context sentence, processing the verb (which always had a QP subject) required more time than in conditions beginning *the queen*; in the QP region, conditions beginning with *the soldiers* required less time than conditions beginning *the queen*. While I can only speculate about the reason for this contrast between the QP region and the precritical region, it is reminiscent of findings in the memory literature where it has been argued that more processing effort during encoding can lead to reduced effort in later reaccess, because the extra effort involved in encoding increases salience (see, for example, the discussion in Van Gompel and Majid 2004). I speculate that the extra effort in encoding *every soldier* when only *the queen* has previously been introduced could have led to reduced processing times when *every soldier* was reactivated at the verb in the precritical region.

Pronoun region

At first glance it seems that the pattern of results from the precritical region is not reflected in the pronoun region. In the pronoun region there was a main effect of context in the regression-path times; readers took longer to leave the pronoun region to the right (to read further material) in the context match conditions compared to the context mismatch conditions. This suggests an elevated processing cost for the matching contexts. This is the opposite of what was predicted for the context manipulation (see Chapter 5 for a discussion); the context match conditions were expected to boost the prominence of the antecedent, making it easier to process. However, a closer look at the precritical and pronoun regions together shows that, while the pattern in the two DP antecedent conditions reverses, the pattern for the QP antecedent conditions remains unchanged from the precritical region to the pronoun region. If the precritical region is taken into account, the main effect of context in the pronoun region could be driven by a change in pattern for the DP antecedent conditions and no change in pattern for the QP antecedent conditions. These effects can be seen clearly in the following figures which show the regression-path times in the precritical region (Figure 3.3) and the pronoun region (Figure 3.4).

The result of this change in the DP antecedent conditions is that the DP and QP antecedent conditions pattern together in the pronoun region, leading to an effect of context with higher regression-path times for the context match conditions compared to context mismatch conditions.

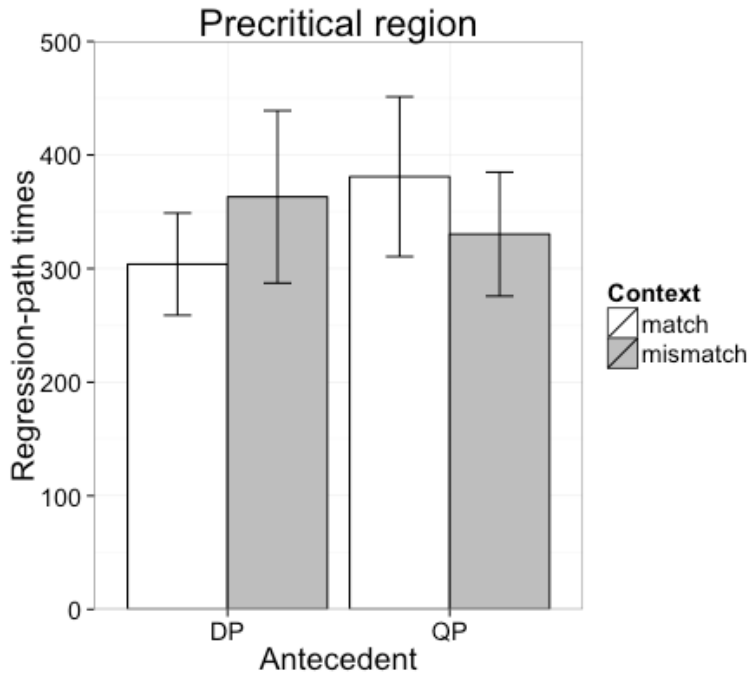


Figure 3.3. Mean regression-path times (in milliseconds) per condition in the precritical region for Experiment 5. The error bars are 95% confidence intervals.

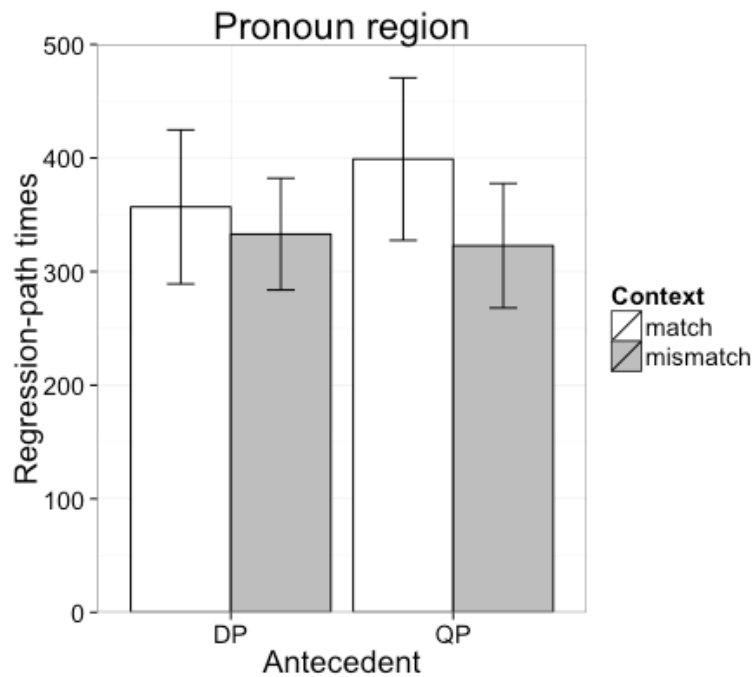


Figure 3.4. Mean regression-path times (in milliseconds) per condition in the pronoun region for Experiment 5. The error bars are 95% confidence intervals.

So, the effects in the pronoun region should not be viewed in isolation. The QP antecedent conditions show the same pattern in the precritical and the pronoun region, so the difference between match and mismatch conditions seen here should not be interpreted as an effect of processing the pronoun. Looking at just the DP antecedent conditions, there is a clear change between the precritical and the pronoun regions. The pattern in the pronoun region for the DP conditions can be interpreted as an effect of encountering the pronoun.¹⁵

¹⁵ Another possibility, though, is that the effects are a continuation of the verb processing from the precritical region. In support of my assertion that the pattern of results relating to the two DP conditions arise from processing the pronoun rather than the preceding verb, I point out that the pattern for the DPs in the pronoun region continues to the spillover region and the prefinal region, and the same is in fact true of the QP pattern that begins in the prefinal region (see below); once the pattern changes from the one seen in the precritical region, it continues into regions that are quite far away from the verb.

Spillover region

In the spillover region, the rereading and total viewing times are significantly longer in the DPcm condition than the DPcmm condition, as shown by the effect of context. This is also seen in the regression-path times but the effect is marginally significant. The total viewing times are longer in the QPcmm condition than the DPcmm condition, and this is also marginally significant in the regression-path and rereading times. Finally, there are marginally longer total viewing times in the DPcm condition compared to the QP QPcm condition.

Comparing this to the pronoun region, the pattern that emerged there for the DP antecedent conditions is seen more clearly in the spillover region. The cm conditions are more problematic than the cmm conditions for the DP antecedent. There are also differences emerging between the DP antecedent and the QP antecedent conditions, visible in the total viewing times; the QPcmm condition has longer total viewing times compared to the DPcmm condition, but the DPcm condition has marginally longer total viewing times than the QPcm condition. This shows that the DP antecedent and the QP antecedent are responding differently to the context manipulation. However, while there is a significant difference between the cm and cmm conditions for the DP antecedent, the difference between the cm and cmm conditions for the QP antecedent is not yet significant. Figure 3.5 shows the mean total viewing times for the spillover region in each condition.

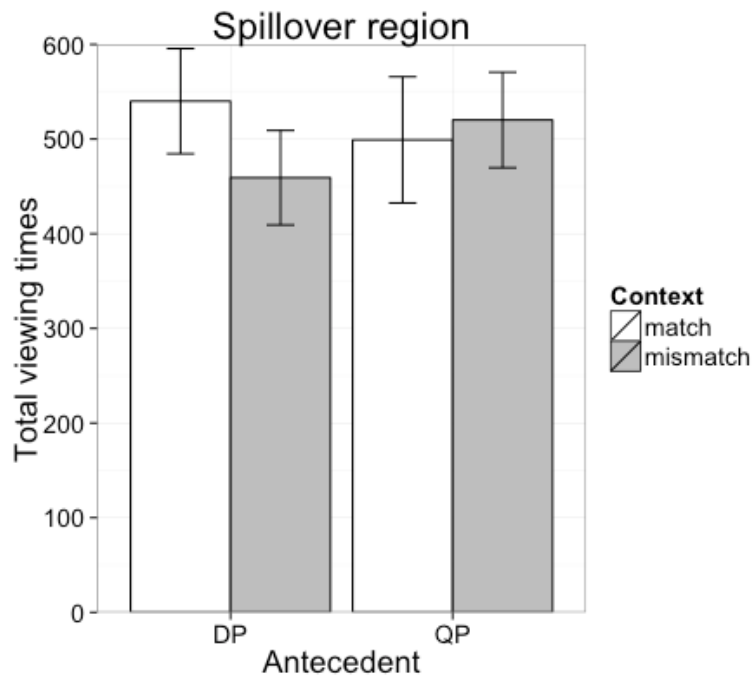


Figure 3.5. Mean total viewing times (in milliseconds) in the spillover region, per condition for Experiment 5. The error bars are 95% confidence intervals.

Prefinal region

In the prefinal region there are effects in the first-pass times and the total viewing times. In both measures the QPcmm condition is significantly longer than the QPcm condition, and also the QPcmm condition is marginally longer than the DPcmm condition. The QP antecedent in this region is affected differently by the context manipulation than the DP was in the region before. Here, there is an advantage for the matching context rather than a cost. The DP in this region does not show effects of the manipulation of context.

Final region

There are significant effects in the first-fixation durations in the final region. First-fixation durations are significantly longer in the DPcm condition compared to the DPcmm condition, and also compared to the QPcm condition. The QPcmm condition has significantly longer first-fixation durations than the DPcmm condition. This is very similar to the pattern of results in the spillover region, however, it is difficult to interpret

these results for two reasons. Firstly, the final region is far away from the pronoun and, being at the end of the sentence, is subject to general sentence wrap-up effects which may not be anything to do with the pronoun. Secondly, the effect is found here in first-fixation durations only, so it is a very fleeting pattern. For this reason the effects in this region will not be discussed any further.

Overall pattern

The overall pattern of results is that the context and antecedent factors interact throughout. Beginning in the pronoun region, the context match is more costly in the DP antecedent conditions than the context mismatch, and beginning as a trend in the spillover region and reaching significance in the prefinal region, the context match is *less* costly in the QP antecedent conditions than the context mismatch. Thus, the effect of context emerges in a later region for the QP antecedent than the DP antecedent, and the effect of context goes in opposite direction for the QP antecedent and the DP antecedent. However, no simple main effects (that were not in the presence of an interaction) of the factor antecedent were found.

3.7.2 Evaluation of the predictions for Experiment 5

Even though the pattern of results from Experiment 5 is fairly complicated to interpret, the prediction derived from the POB hypothesis for Experiment 5 is straightforward: there should be processing difficulty at least initially for the DP antecedent conditions compared to the QP antecedent conditions, leading to overall longer reading times in the DP antecedent conditions compared to the QP antecedent conditions; this would have been seen in a main effect of antecedent. This is because POB assumes that a c-commanding VB antecedent is sought out before any CR antecedents can be considered, causing disruption when the QP mismatches in stereotypical gender with the pronoun. The results of Experiment 5 show that at no point were there shorter reading times for the QP antecedent conditions compared to the DP antecedent conditions. In

fact, at the earliest region in which the gender match between the pronoun and the QP could have been assessed (the pronoun region), there was a main effect of context. As discussed above, the context effect in the pronoun can be attributed to the context manipulation affecting the DP antecedent conditions but not the QP antecedent conditions. This suggests that the DP was being considered as an antecedent, and that this process was affected by the context manipulation. But there was no evidence that the QP was considered at or before that point. Even if one assumes that the QP antecedent was being considered at that point but that the context manipulation was having no effect, there was certainly no disruption for the DP antecedent compared to the QP antecedent conditions. Thus, Experiment 5 finds no support for the POB hypothesis that VB antecedents are sought and evaluated before CR antecedents. This is also in line with the findings of Experiment 4.

An anti-QP hypothesis was formed on the basis of the results from the Experiment 4, which appeared to suggest that the QP was not considered at any point during processing as a potential antecedent. If this had been the case in Experiment 5, there should have been long-lasting reading-time disruption in the QP antecedent conditions compared to the DP antecedent conditions. Results from Experiment 5 however do not support the assertion that the QP is completely ignored during processing. At no point were there consistently shorter reading times for the DP antecedent conditions compared to the QP antecedent conditions, and certainly not the long-lasting effect that the anti-QP hypothesis predicts. It seems that this hypothesis was an over-interpretation of the results from Experiment 4. Indeed, to preview the results reported in Chapter 4, there is evidence that a QP is considered and is even preferred if it is closer to the pronoun. As suggested in section 3.6.1, it might be that the design of Experiment 4 gave less opportunity for effects of a dispreferred but still viable antecedent to be picked up. Experiment 5 redressed this balance. If the QP was dispreferred but still considered, there should have been some initial disruption in the QP antecedent conditions because of the mismatch of the preferred

antecedent, but this would have been short-lived because of the availability of a matching QP. This prediction was not completely borne out either. But this hypothesis does seem to fit the pattern of results very well if the effects of context are taken into account. What is clear from the pattern of results over all regions is that the context manipulation had an effect throughout and that the context effect went in different directions for the DP antecedent conditions than for the QP antecedent conditions. While the context match condition was more costly for the DP antecedent, the context mismatch condition was more costly for the QP antecedent. It appears that the context started to take effect for the DP condition in the pronoun region, reaching significance in the spillover region, whereas for the QP antecedent it emerged in the spillover region and became significant in early measures of the prefinal region. While this pattern is interesting in itself (and is discussed in Chapter 5), the interaction of antecedent and context can also be seen as an index for an antecedent being considered. As such, the DP antecedent appears to have been processed more quickly than the QP antecedent, but the fact that both antecedents are affected by the context manipulation indicates that the QP was considered when the DP did not match.

The later consideration for the QP antecedent fits in with an assumption that the QP is either dispreferred as an antecedent or is more costly to process compared to a DP. This fits in well with the results from Experiment 4. In that experiment, readers clearly preferred to link the pronoun to the name rather than the QP. The match-mismatch design of Experiment 4 brought this preference out, but the results of Experiment 5 demonstrate that the QP is not completely ignored. Another possible reason for the difference in results between Experiments 4 and 5 is that names were used as CR antecedents in Experiment 4 and DPs in Experiment 5. As observed above, names do appear to have a particularly stable discourse representation in comparison to other types of NP. This may affect how easy it is to retrieve, or it may make it more difficult to consider alternatives. It is not possible to evaluate this in the current

experiments, but it should be noted that, despite swapping names for DPs in Experiment 5, there was still no initial consideration of the QP antecedent.

The QP may, however, take longer to retrieve as an antecedent than a DP or a name, as demonstrated by Carminati *et al.* (2002) and supported by results of Experiment 5. At the moment I can only speculate on the possible reasons for this. One reason could be the properties of the QP itself: there could be a general dislike of linking pronouns to QPs because of their complexity, so the parser tries to avoid binding if possible but will do so if no better options are available. This implies that antecedent complexity/internal structure is one of the factors which is taken into account when searching for an antecedent. However, it is difficult to see how this is helpful to the parser during sentence interpretation. If the parser seeks simpler antecedents over more complex ones, it will presumably end up with undesired interpretations a lot of the time because there is no evidence to suggest that pronouns tend to refer to simpler antecedents. A different scenario is one in which the internal complexity of an antecedent means that it takes longer to recover it from memory and form a dependency. This is in line with Burkhardt's (2005) syntax-discourse model. In Burkhardt's model an antecedent has a file card in memory, which in the case of a QP such as *every slider* is formed of a set and a restrictor; when it is accessed, information about both the set and the restrictor have to be accessed, using more processing resources. Viewed in this way, the results of Experiment 5 come about because accessing the QP's file card takes longer, resulting in later effects of the context. This may also be the case if the feature match between a singular pronoun and a QP, being sub-optimal, makes retrieval more difficult. But it is difficult to also explain the results of Experiment 4 with these assumptions; if the QP can be accessed but takes longer, why weren't there later effects of the QP in that experiment? A final possibility to consider is that it is not the antecedents themselves but their position in the sentence that privileges the DP (or the name in Experiment 4) over the QP. Being closer to the pronoun

might bring about processing advantages for the nearest antecedent, either because of a memory advantage or for discourse/information structural reasons. The position of the antecedent is explored in Chapter 4. Evidence from Experiments in Chapter 4 help to clarify the results both of Experiment 5 and Experiment 4.

3.7.3 Can POB be rescued?

In the light of the current results from Experiment 5 and those from Experiment 4 the evidence against the POB hypothesis is beginning to stack up, at least as far as English is concerned. The question is whether any part of the POB hypothesis can be rescued or maintained in the light of the current evidence. In fact, a closer examination of the POB model reveals that the hypothesis about the search process, while fully in keeping with the principle of economy, does not necessarily follow from the nature of the model itself. The POB model, being based on Reuland's (2001) syntactic approach, is very detailed in how it produces grammatically acceptable output and rules out ungrammatical structures. This is achieved through the different modules and the economy principle which dictates the relationships between the modules. The core of the model can function quite satisfactorily to evaluate one pronoun-antecedent link at a time, determining whether the link is acceptable and specifying the level at which the link is assumed to take place. But generating only syntactically well-formed pronoun-antecedent links is more akin a production task than a comprehension task. During comprehension of a pronoun there is a search process that is triggered on encountering an anaphoric element, and the parser may need to evaluate more than one potential antecedent, sequentially or simultaneously. There is nothing intrinsic to the core of the POB model which specifies that more economical antecedents need to be picked up and evaluated before more costly antecedents; this claim is an *addition* to the model. In effect, POB is not really a search model. The POB claim investigated in this chapter also completely ignores the role of discourse prominence. For example, most researchers agree that a pronoun is preferentially linked to the most prominent antecedent. If the most

prominent antecedent does not c-command the pronoun, but another antecedent in the discourse does, the POB hypothesis of VB-initial processing dictates that the VB antecedent be evaluated first. Then any CR antecedents would be evaluated, and in the discourse module the discourse prominence of the CR-antecedent would be taken into account, triggering a re-analysis process that has to undo the VB link between the pronoun and the less prominent VB antecedent. This appears to be a rather inefficient use of the available cues to pronoun interpretation.

This is not to say that POB has nothing at all to say about processing. For example, POB could perhaps be characterised as the process through which pronoun-antecedent relations are evaluated after the search process has been completed. In this way, the intrinsic processing costs of particular operations can be assumed and processing hypotheses could be derived based on the intrinsic costs associated with anaphor-antecedent relations that take place in particular modules. If, on the other hand, it is to be taken seriously as a search model, further detail is needed on how it evaluates multiple antecedents, and it would need to be adjusted in light of the current findings. It may also need to revise its position on the use of discourse information, since this is presumably available at the point that an antecedent is picked up from the mental representation. In the current formulation, this information is suppressed until the final module in the POB model is reached.

Despite these problems, the POB hypothesis does attempt to take seriously the role of variable binding and coreference during pronoun interpretation, an issue that has been largely ignored in the processing literature up until this point. Thus, it is worthy of investigation so that better hypotheses can be drawn about how these two operations may affect processing.

3.7.4 QPs as antecedents

Related to the limited attention on processing variable binding and coreference, there is little previous processing research on QPs as

antecedents. Carminati *et al.*'s (2002) results suggest that QPs are more costly as antecedents when directly compared to DPs, and Koornneef's (2008) results suggested that QPs are more costly than names (but recall that the evidence from that experiment was confounded with regard to syntactic structures, and that a different experiment found no difference between QPs and DPs in the same structure). Clearly, further research needs to be carried out to get a fuller picture of the processing of QP antecedents. The results from Experiment 4 seemed to suggest either that QPs were not considered at all or they were strongly dispreferred, but results from Experiment 5 provide evidence that the QPs can be considered as antecedents, albeit with delay in comparison to the DP antecedents. A remaining question is whether this preference for the DP comes about because of properties of the antecedents themselves, or because of their position in the sentence. This issue is explored in Chapter 4.

3.8 Summary

This chapter set out to evaluate the POB hypothesis that VB antecedents are considered before CR antecedents during antecedent search. Two eye-tracking Experiments (4 and 5) were reported which evaluated this claim. There was no evidence from either experiment that VB antecedents are considered first. Rather, Experiment 4 demonstrated that a CR antecedent (a name) was considered, but that the VB antecedent (a QP) seemed not to be considered at all. Experiment 5 showed that both VB and CR antecedents were affected by a context manipulation, demonstrating indirectly that they were both considered as antecedents, and the effect of the context emerged in an earlier region for the CR antecedent compared to the VB antecedent. (The context effects are fully discussed in Chapter 5.) The search for antecedents is not affected by the type of operation, VB or CR, used to link the pronoun with an antecedent. This suggests that the POB model may be better suited to the evaluation of antecedents rather than as a search mechanism.

CHAPTER 4

The role of recency in pronoun resolution

4.1 Introduction

In the previous chapter, it was shown that a variable-binding (VB) antecedent was not systematically considered earlier than a coreference (CR) antecedent, nor was a VB antecedent given any kind of preference over a CR antecedent. Instead, it was the CR antecedent that had priority: only the CR antecedent was considered during processing in Experiment 4 and the CR antecedent was considered before the VB antecedent in Experiment 5. As discussed, the advantage for the CR antecedent could have been due to its proximity to the pronoun, since in all conditions it was closer to the pronoun than the VB antecedent. Alternatively, properties of the antecedents themselves (names and DPs versus QPs) or the resolution route (CR versus VB) may have given an advantage to the CR antecedent. In either case, the POB hypothesis cannot account for these results because it predicts that in any configuration a VB antecedent should be considered before a CR antecedent. Nevertheless, the online processing preference for the most recent antecedent should be considered, additionally in the light of findings from Experiment 3 in Chapter 2.

The results from Experiments 3, 4 and 5 seem to suggest that an antecedent's locality or recency in relation to the pronoun plays a role in pronoun processing. On the surface this may seem a rather obvious assumption. On the basis that there are limited memory resources available for sentence processing (which leads to increased difficulty in integrating syntactic dependencies at increasing distances, e.g. Gibson 1998), it seems logical to assume that an antecedent that is more recent is more felicitous, from the point of view of processing efficiency, than one that is further away, other things being equal. However, in much of the literature on pronoun processing the importance of the first-mentioned antecedent has been emphasised (e.g. Gernsbacher and Hargreaves 1988; Gordon, Grosz

and Gilliom 1993; Arnold, Brown-Schmidt and Trueswell 2007). And while both first-mention and recency appear to be acknowledged as pronoun resolution cues in the literature, there have been very few serious attempts to reconcile the two accounts.

In this chapter I will present the results from one eye-tracking experiment and three offline questionnaire studies that follow up Experiment 4 from Chapter 3, exploring recency as a factor in the processing of pronouns. I will demonstrate the existence of a recency preference during processing, and consider possible underlying causes for such a preference. Before presenting the experimental results, I will first review the previous literature on both first-mention and recency in pronoun resolution. Although first-mention is not directly tested in the experiments I present in this chapter, I begin by reviewing the literature on this topic in order to consider how to reconcile the seemingly opposing cues of first-mention and recency, especially given the emphasis on the first-mention advantage in much pronoun processing literature.

4.2 Previous literature on first-mention as a cue in pronoun resolution

4.2.1 The structure-building approach

First-mention is frequently cited as an important cue in the processing of pronouns, especially within literature that is concerned with discourse aspects of pronoun resolution. The claim that first-mention plays an important role in pronoun processing can be traced back to Gernsbacher's work (Gernsbacher and Hargreaves 1988; Gernsbacher 1989) although the idea of first-mentioned entities having a prominent status in memory can be traced back much further. Gernsbacher's (1989) paper presents a series of probe recognition experiments that explored the memory mechanisms of suppression and enhancement as they are involved in referential access. While not the primary focus of the experiments in that paper, the factor of first-mention showed up persistently throughout the experiments and Gernsbacher briefly discusses the idea of a first-mention advantage. This

idea is discussed in more detail, and tested more systematically, in Gernsbacher and Hargreaves (1988).³⁸ The reason for the first-mention advantage, they propose, is that first-mentioned information is particularly prominent in memory. This is what they called the structure-building approach. In this approach, the first-mentioned entity forms the foundation of a mental discourse structure. In other words, the reader/listener assigns more importance to the first-mentioned entity because they believe it will be more central to the discourse. This importance is reflected in easier enhancement or reduced suppression of this entity, so it is easier to retrieve when it is referred to again further into the discourse. While Gernsbacher's work has been repeatedly cited in the pronoun processing literature, in actual fact the evidence relating specifically to pronouns is not substantial. I briefly summarise the findings below.

The probe recognition experiments in Gernsbacher (1989) presented two-sentence sequences in word-by-word fashion. The first sentence contained two names, and the second sentence contained an anaphor that referred to one of the names. Participants were required to read the sentences and respond to probes at various test points. Probes were either the first NP (NP1) or the second NP (NP2). Several factors were systematically varied: anaphor type (name versus pronoun), probe name (antecedent versus non-antecedent), and test point (before versus after the anaphor). Examples of the materials for experiment 1 are given below in (4.1) (the square brackets [] indicate possible test points):

(4.1a) *Antecedent is NP1*

Bill handed John some tickets to a concert but [] he/Bill [] took the tickets back immediately. (Probe: Bill / John)

³⁸ While the 1988 paper was obviously the first of the two papers to be published, referencing of the 1989 paper in the 1988 one suggests that the 1989 paper was prepared first, and that the experiments in the 1988 paper were follow-up studies.

(4.1b) *Antecedent is NP2*

Ann predicted that Pam would lose the track race but [] she/Pam
[] came in first very easily. (Probe: Ann / Pam)

Reaction times to the probes were measured. The difference between reaction times before and after the anaphor was taken as indication of enhancement or suppression of an NP. For example, in NP2-type sentences (4.1b), when the anaphor was *Pam*, reaction times to the probe *Pam* were faster after the anaphor than before, but reaction times to the probe *Ann* were slower after the anaphor than before. This was taken as an indication that the anaphor *Pam* both enhances the antecedent in memory and suppresses the non-antecedent (*Ann*).

There was, of course, an effect of first-mention. This was shown by an interaction between antecedent position (NP1 versus NP2) and probe name (antecedent versus non-antecedent), so that when the antecedent is NP1 (4.1a) there were faster overall responses to the first-mentioned antecedent probe (*Bill*), and when the antecedent is NP2 (4.1b) there were faster overall responses to the first-mentioned non-antecedent probe (*Ann*). This effect was seen throughout the series of experiments, despite changing the test point (both before and after the anaphor), manipulating gender match, and introducing new entities in the second sentence. However, the effects of suppression and enhancement were mainly triggered by the name anaphors rather than the pronouns. It was only moving the second test point to the end of the sentence that brought about the same suppression and enhancement effects that were shown by the name anaphors. The pronouns must have contributed to some extent to the first-mention effect, otherwise there would have been a three way interaction (antecedent position x probe name x anaphor type). So it is not the case that the first-mention advantage is only applicable to name anaphors and not pronouns in these experiments. Rather, the difficulty of getting enhancement and suppression effects from the pronouns suggests that the method may not be entirely suitable for investigating the processing of pronouns specifically. Added to this, the two sentence types (NP1 and NP2) have very different structures, and the NPs take different thematic roles in each

sentence type. But in five out of six experiments, they were analysed together. The difference between responses to first- and second-mentioned entities could have been due to properties of the sentence structure, or indeed different semantic or pragmatic inferences. The two sentence types are not really comparable, but are treated as though they are. Furthermore, probe-recognition tasks have been criticised because the task itself could encourage participants to mentally replay the sentence, which would give the first-mentioned entity an advantage. Albeit there is no direct way of assessing whether this is the case, it is worth bearing in mind that the task, combined with the word-by-word presentation of the stimuli, does not encourage natural processing, despite some comprehension questions.

Gernsbacher and Hargreaves (1988) report a series of experiments that provide more evidence for the structure-building approach. In these studies, the factor of antecedent position is varied within items rather than between, which is a better design. The antecedents appear in a range of syntactic roles, which are carefully controlled. However, none of the experiments in the paper addresses pronouns, rather, name antecedents are used. As was clear from the Gernsbacher (1989) paper, the effects of name antecedents and pronouns are not necessarily the same.³⁹ Overall, the evidence from these studies regarding the advantage of first-mention for pronoun processing is rather limited.

These criticisms are not intended to discredit the findings of the papers, which demonstrate a remarkably persistent first-mention advantage. It is rather to exercise caution in transferring the findings directly into models of pronoun processing, and to express surprise at how the findings have been assumed to apply directly to pronouns in literature up to the present day. The effects are more robust when it comes to name anaphors than

³⁹ For example, the repeated-name penalty (Gordon *et al.* 1993) (see Chapter 5) shows that the use of names or pronouns to refer to a prominent entity has a large effect on processing.

pronouns, and the method and materials may well not be suited to the investigation of pronoun processing.

4.2.2 The expectancy hypothesis

The advantage of first-mention in English has been further investigated by Arnold (Arnold *et al.* 2000; Arnold 2001; Arnold *et al.* 2007), out of which the expectancy hypothesis has been proposed. This hypothesis is described more fully below, but the basic assumption is that listeners expect certain discourse entities to be referred to again in a text; these entities will be pronominalised on subsequent mentions. First-mentioned entities are expected to be rementioned and are therefore likely to be the referents of pronouns. While Arnold's work has focussed to some extent on children's use of reference, there has been some work on pronoun processing in adults. I will present one study (Arnold *et al.* 2000) in which the effect of first-mention is directly tested online with adult participants. The study tested the use of gender and accessibility in online pronoun resolution. Referent accessibility here refers to discourse prominence as mediated by first-mention. Recency is mentioned in the paper as another factor that could contribute to accessibility (see also Arnold 1998), but it is not systematically tested here, nor is it discussed how first-mention and recency could interact or co-exist. Two visual world experiments⁴⁰ were carried out which manipulated the pronoun's match with two possible antecedents by using gender information and accessibility (first-mention). The participants listened to short stories whilst looking at pictures depicting the scene while their eye movements were monitored. The materials were as follows:

(4.2)

Donald is bringing some mail to {Mickey/Minnie} while a violent storm is beginning.

⁴⁰ For a description of the visual world experimental paradigm see Chapter 2 section 2.3.2.

{He's/She's} carrying an umbrella, and it looks like they're both going to need it.

The content within parentheses was systematically varied, so that four conditions were created: *different/first mention*, *different/second mention*, *same/first mention*, *same/second mention*. The different/same distinction refers to whether the two antecedents were of a different or the same gender (*Donald + Minnie* vs *Donald + Mickey*). The first/second mention distinction refers to whether the pronoun refers to the first character or the second character. The task was to decide whether the picture matched the sentence that the participants heard. In each experimental condition the picture matched the sentence. So in (4.2), the accompanying picture showed the referent of the pronoun (*Donald/Mickey/Minnie*, depending on the condition) holding the umbrella so it is possible to work out the referent of the pronoun even in the same-gender conditions. The results revealed that looks converged on the target referent quickly (from 400ms after the onset of the pronoun) in the first three conditions. In the fourth condition (*same/second mention*), there was no difference in the amount of looks to the target (*Mickey*) compared to the competitor (*Donald*), although in the picture it was *Mickey* who was holding the umbrella. There was no difference in the end-of-trial response accuracy scores between the conditions.

In the second experiment, the accessibility of the first-mentioned antecedent was increased by adding an interim sentence in which *Donald* is pronominalised with *he*:

(4.3)

Donald is bringing some mail to {Mickey/Minnie} while a violent storm is beginning.

He's sauntering down the hill while a violent storm is beginning.

{He's/She's} carrying an umbrella, and it looks like they're both going to need it.

This extra manipulation decreased the end-of-trial response accuracy scores in the *same/second mention* condition. The pattern of looks in the

online measures was similar to experiment 1, with the first three conditions patterning together, showing more looks to the target, and the fourth condition again showing no difference in looks to the target and the competitor (and in one segment, more looks to the competitor). The onset of these effects occurred one segment earlier than in experiment 1, from 200ms after the onset of the pronoun.

These results clearly show that participants expected the pronoun to refer to the first-mentioned antecedent, but also that such a preference could be overridden by gender information. So, both gender and accessibility show early effects on online pronoun processing. In these experiments the factor of accessibility, whilst defined primarily in terms of first-mention, is in fact confounded with subjecthood because only the first-mentioned antecedent (*Donald* in the examples) is a subject, while *Mickey/Minnie* is an indirect object. Indeed, it may not be first-mention or even subjecthood driving these effects, rather a parallelism effect, that is, a match between the grammatical role of the pronoun (subject) and the antecedent (Stevenson, Nelson and Stenning 1995). It is interesting to note that the effects were observable earlier in experiment 2 than in experiment 1. The difference between the materials in the two experiments is the higher accessibility of *Donald* in experiment 2, as signalled by the pronominalisation in the second sentence. The more rapid effects in experiment 2 therefore indicate that accessibility of the antecedents can affect the timecourse of pronoun resolution, with higher accessibility leading to more rapid responses. The authors conclude that the evidence here points to rapid use of both gender and accessibility information, and that neither is temporally delayed (they are both rapidly deployed), nor is one type of information used before another.

The assumption in Arnold *et al.*'s (2000) study is that first-mention contributes to accessibility, and accessibility is crucial to pronoun resolution. Accessibility forms an important part of the expectancy hypothesis (Arnold *et al.* 2007; Arnold 2001). The expectancy hypothesis is an expression of a constraint-based model of reference resolution. According to the hypothesis, a pronoun is likely to be linked to the most

accessible antecedent in the discourse. Accessibility is determined by the listener's expectations of which entities are going to be mentioned again, based on discourse cues in the text. The more likely that an entity will be mentioned again, the higher its accessibility within the discourse model, hence the higher likelihood of such an entity being referred to with a pronoun. The listener's expectations are determined by a number of linguistic and non-linguistic factors. Arnold follows Gernsbacher's (1989) view that first-mention is an important contributor to accessibility because first-mentioned entities are likely to be central to the discourse. Factors other than first-mention can also contribute to an entity's accessibility: subjecthood, entities in focus constructions, and entities in a parallel grammatical role to the pronoun.

The expectancy hypothesis goes some way in describing why a particular entity might be highly accessible, or highly activated in the discourse model. It does not make any suggestions as to how the various factors that contribute to accessibility interact with one another, or under what circumstances particular factors play a more or less important role. For example, the expectancy hypothesis emphasises the importance of first-mention, but does not explore how it may conflict with recency.

First-mention is also important in other discourse models, for example Centering Theory (Grosz, Joshi and Weinstein 1995). This is a framework for tracking prominent entities as a discourse unfolds, and is discussed in some detail in Chapter 5. This framework contains some rough metrics to assess the notion of prominence, but it too sidesteps the issue of how different factors can interact with each other.

The evidence regarding first-mention in English is incomplete because first-mentioned entities are very often grammatical subjects. This means that any first-mention preference could in fact be preference for resolving pronouns to the subject, instead of (or as well as) a preference for resolving them to the first-mentioned entity. Given the difficulty of untangling

subjecthood and first-mention in English, it is logical to assess the role of first-mention in other languages which do not display this confound.^{41,42} If the first-mention advantage is entirely attributable to a preference for subjects, for example, it would go some way to resolve the apparent opposition between a first-mention preference and a recency preference.

4.2.3 First-mention versus subjecthood: evidence from Finnish

In this section I discuss two papers that investigate first-mention and subjecthood as cues during pronoun processing in Finnish, a language in which a first-mention and subject preference can more easily be distinguished.

The first paper is Järvikivi *et al.* (2005), which tested the resolution of the Finnish gender-neutral third person pronoun *hän* (*he/she*). The paper reports a visual-world experiment in which the order of two antecedents was manipulated so that the first sentence was either in subject-verb-object (SVO) order or object-verb-subject (OVS) order. SVO is the unmarked structure in Finnish, but OVS is used commonly, usually to present the object as given information and the subject as new information.

Participants listened to scenarios such (4.4) below, while they looked at pictures of the two people and the location mentioned in the stories.

Translations are taken from the paper.

(4.4a) SVO

Tony Blair kätteli George Bushia valkoisessa talossa. Hän halusi keskustella Irakin tilanteesta.

⁴¹ It is, of course, possible to have a non-subject in first-mention position in an English sentence, but the markedness of such structures makes them rather unsuitable for comparison purposes.

⁴² The need to assess experimental findings in languages other than English is, of course, not limited to this phenomenon.

Tony Blair (subject) shook hands with George Bush (object) in the White House. He wanted to discuss the situation in Iraq.

(4.4b) OVS

George Bushia kätteli Tony Blair valkoisessa talossa. Hän halusi keskustella Irakin tilanteesta.

George Bush (object) shook hands with Tony Blair (subject) in the White House. He wanted to discuss the situation in Iraq.

The suffix *-ia* on *George Bush* indicates partitive singular, marking it as the object of the verb. The pronoun *hän* could refer to either of the antecedents. Firstly, the percentage of first visits and the number of visits to the subject or object picture were analysed. For percentage of first visits, there was a main effect of grammatical role, with more first visits to the subject picture than the object picture. There was also an effect of order of mention, with more first visits to the first-mentioned antecedent than the second-mentioned. There was no interaction between the effects. In the OVS structures, there was no difference between the percentage of visits to the subject and the object.

Secondly, a time-course analysis, using the fixations on a particular picture, was carried out. In this analysis there was a main effect of grammatical role from the second time-window (210ms after pronoun offset) onwards. There was a main effect of order of mention from the third time window (420ms after pronoun offset) onwards. There was no interaction between these effects. These main effects were in the same direction as those found in percentage of first visits, namely more looks to the subject picture and more looks to the first-mentioned picture. No offline measure for the interpretation of the sentences was reported.

The authors drew several conclusions from their study. First, that the lack of an interaction between the grammatical role and the order of mention meant that both factors had an independent effect on the antecedent preferences. Second, that their results undermined a straightforward first-mention account, which would have predicted a first-mention preference in both SVO and OVS structures, with no effect of grammatical role. Third,

that a straightforward subject-preference account is also not supported, because the order of mention did have an effect on the results. They state, finally, that pronoun resolution “is determined by a delicate interplay of several factors” (p. 264).

There is more than one way to look at the results of this study. First of all, it is not clear what is meant by the statement that “grammatical role and order of mention...had independent effects on resolution of the ambiguous pronouns” (p. 264) based on a lack of interaction between the effects, while at the end of the paper they propose “a delicate interplay of several factors” (p. 264). Imagine, for example, that they had found an interaction between the effects in the percentage of first visits. In the way that they presented the results, an interaction could have suggested, for example, a pure subject preference (e.g. more fixations on the subject when it is presented first compared to second, fewer fixations on the object but no difference between first and second position). Such an interaction would not show that the factors of subjecthood and order of mention were interacting, rather it would suggest a primary role for subjecthood during pronoun resolution. This is just one example, but it demonstrates that the way they have thought about interactions between the factors is confusing, perhaps partly because of the way that they present the results. They present the conditions as grammatical role (subject, object) and order of mention (first, second). They could also have coded the order of mention SVO and OVS. Presenting the results this way around would have shown, in fact, something that looks a bit like an interaction. I will give an example of the results under the SVO and OVS coding, with the percentage of first visits measure. Under this coding, in the SOV structure there were more first visits to the subject than the object (35.8% versus 14.9%). In the OVS structure there was no difference between first visits to the subject and object (24.8% versus 24.5%). This could be interpreted as a subject preference, clearly seen in the SVO structure, which interacts with a first-mention preference, so that when the subject is not presented first, the subject preference disappears because it is competing with a first-mention preference. Note that the first-mention preference does not emerge in the OVS structure; if it did, we could dispense with the subject preference idea

altogether, and just attribute the results purely to a first-mention preference.

The study lacks an offline measure for the final interpretation of pronouns. While making it unclear whether the online measurements represent just processing, or if they also reflect people's final interpretation of the sentence, the study does show that relying on one heuristic as a preferred pronoun resolution strategy is not effective. There is not a clear first-mention preference on its own, nor is there a subject preference on its own. Rather, both these factors do have some effect during pronoun resolution so that when a subject is presented in first position, both the factors combine to promote this antecedent. It may also reveal a timing difference; note that the effect of grammatical role emerged at an earlier time window than the effect of order-of-mention.

However, these findings from Finnish may be complicated by the fact that Finnish has an additional anaphoric form that can be used to refer to human entities, that was not used in the Järvikivi *et al.* (2005) study. The demonstrative *tämä* in Finnish is commonly used to refer to a less salient entity in the discourse, whereas *hän* tends to refer to the most salient entity. The distinction between *hän* and *tämä* was explored in a series of visual world studies in Kaiser and Trueswell (2008), following on from offline results reported in Kaiser (2003; 2005). Kaiser and Trueswell explore the influence of word order (SVO versus OVS) and grammatical role (subject versus object) in the online resolution of *hän* and *tämä*. In an offline story continuation pretest, participants completed stories in four conditions: SVO *hän*, OVS *hän*, SVO *tämä*, OVS *tämä*. In the first two conditions, participants mainly interpreted *hän* as referring to the subject, whether it appeared before the verb (SVO *hän*) or after (OVS *hän*). In the SVO *tämä* condition participants strongly preferred to interpret *tämä* as referring to the object, but in the OVS *tämä* condition there was a weaker preference for the subject. Offline, *hän* appears not to be influenced by word order, only grammatical role, whereas *tämä* is influenced more by the word order. The results from the online study more or less reflect the offline preferences. In the online experiment, participants listened to stories that contained an

observer character and two further story characters which were depicted on the visual display and were used in the SVO/OVS set up. The final sentence of the story began with *hän* or *tämä*, and participants' eye movements towards the depicted characters were analysed from 800 ms before the onset of the anaphor. In the first 400 ms after the onset of the anaphor, there are more looks to the first-mentioned character in the SVO conditions than in the OVS conditions. Between 400 and 800 ms after onset, the same word order effect is visible and there is also an interaction between word order and anaphor type. This pattern persists in later time segments. In the SVO *hän* condition participants look more to the subject (the first-mentioned antecedent) and in the OVS *hän* condition participants look more to the object. The processing of *hän*, then, appears not to be influenced by the word order but only by the grammatical role of the antecedent. In the SVO *tämä* condition there was a brief period when looks to the subject increased, which was unexpected. But apart from this blip, both SVO *tämä* and OVS *tämä* conditions patterned together in having a lower proportion of looks to the first-mentioned antecedent, showing that it is not influenced by grammatical role, in contrast to *hän*.

These results contrast somewhat with the results of Järvikivi *et al.* (2005), who found that *hän* was influenced by both grammatical role and order of mention to some degree. Kaiser and Trueswell (2008) attribute this difference to the lack of contexts in Järvikivi *et al.*'s study which may have caused particular difficulty with the marked OVS structures. Kaiser and Trueswell do not want to claim that a single factor (grammatical role) influences the processing of *hän*. Rather, they want to demonstrate that *hän* and *tämä* are differently sensitive to the cues that might promote one antecedent over another in the discourse. They relate their findings to a form-specific multiple constraints model, in which pronoun resolution cues are weighted differently for different anaphoric forms.

The evidence from both these studies together shows that for Finnish, pronouns are sensitive to subjecthood and are less sensitive to first-mention, even though they do display some degree of first-mention advantage in Järvikivi *et al.*'s (2005) study. The sensitivity to certain cues

may be influenced by the availability of anaphor types in a particular language. For English, the relatively fixed word order and the lack of a third-person demonstrative equivalent of *tämä* may mean that the first-mention advantage plays a stronger role in English than in Finnish.

Overall, what we know about first mention as a factor in pronoun resolution can be summarised as follows: first mention has long been acknowledged as playing a role in pronoun processing and is linked to the idea of prominence or accessibility in discourse (e.g. Arnold *et al.* 2000; Grosz *et al.* 1995). First-mention often overlaps with subjecthood, especially in English, but appears not to be completely confounded with it. Evidence from Finnish shows that first mention plays a role in addition to subjecthood (Järvikivi *et al.* 2005) but it is unclear whether the availability of anaphoric forms influences cue sensitivity. First-mention can be overridden, without apparent problems, by gender information (Arnold *et al.* 2000).

While first-mention is often cited as a factor in pronoun processing research, recency has received comparatively less attention. But recency *is* often discussed in relation to other aspects of sentence processing. Before reviewing previous research on recency in pronoun processing, then, I include a very brief section on how recency is viewed in sentence processing research.

4.3 Recency in sentence processing research

Recency (also referred to as locality) is a pervasive factor in sentence processing research as a whole. For example, recency has been put forward to explain individual observed phenomena in sentence processing, such as clause attachment preferences (Frazier and Fodor 1978), ambiguity resolution preferences (Sturt, Scheepers and Pickering 2002), and grammatical illusions in subject-verb agreement (Tabor, Galantucci and Richardson 2004; cf. Phillips, Wagers and Lau 2011). It is also an important component within more general theories of sentence processing such as Gibson's DLT (Gibson 1998; 2000), and to some degree within direct access

retrieval models (e.g. Lewis and Vasishth 2005, Van Dyke and Lewis 2003, but see Lewis, Vasishth and Van Dyke 2006 for a discussion of the relationship between decay, locality and interference). Bartek *et al.* (2011) provide evidence that locality affects not just parsing during more complex syntactic operations such as A-bar movement, but also during simple subject-verb relations. Recency or locality in sentence processing is mainly viewed as memory effect: representations are assumed to decay over time, with the consequence that retrieval of more decayed memory items is more difficult than retrieval of less decayed (more recent) items. Since recency effects appear in many types of structure during processing and “may indeed be a ubiquitous feature of human sentence comprehension” (Bartek *et al.* 2011, p. 1195), it seems logical to assume that recency effects should also be in evidence during pronoun processing.⁴³ However, the pronoun processing evidence is limited. Studies that have investigated recency in pronoun processing are reviewed in the following section.

4.4 Recency as a factor in the processing of pronouns

As mentioned above, recency is a somewhat neglected theme in terms of pronoun processing, despite its importance in other aspects of sentence processing. It is often mentioned in the pronoun processing literature as a cue, but there has been little research that systematically studies it. This could be because first mention is also assumed to play an important role and has received comparatively more attention in the literature, at least in English. On the surface, it seems easier to find examples where first-mention is an important factor than examples where recency is important.

An early and frequently cited study of the phenomenon is Clark and Sengul (1979). They gave participants texts composed of four sentences; in the fourth (final) sentence, a pronoun referred to an antecedent in a

⁴³ Despite the apparent ubiquity of locality effects, Lewis *et al.* (2006) point out that they do not always arise in contexts where they might be expected.

previous sentence.⁴⁴ They systematically varied the sentence in which the antecedent appeared (first, second or third sentence) and measured how long it took for participants to press a button indicating that they understood the text. Participants were reliably faster to press the button when the antecedent appeared in sentence 3 (the most recent sentence) than when it appeared in sentence 1 or 2 (there was a numerical but unreliable difference between these two conditions). They linked this to a discontinuity model in which information in the current and previous sentence is privileged in memory, compared to sentences further back. They then carried out two further experiments, this time using only noun phrase anaphors and not pronouns, to determine whether the crucial boundary for the privileged information is a sentence boundary or a clause boundary. They manipulated clause type (subordinate or main) and linear position of the clause and found that, regardless of the clause type, there were significantly faster responses when antecedents were in the last clause of the previous sentence. They do acknowledge that it was not possible to keep the grammatical roles constant across the experiments, and it is difficult to assess without the materials or even examples of the materials, which have not been provided.

It is difficult to know from this study whether any first-mention effects would normally be expected, both because the materials are not available to assess and because first-mention is not clearly defined. We know from this study that overall comprehension time is affected by the distance of the matching antecedent from the pronoun, but the experimental paradigm does not tie the effect precisely to the processing of the pronoun, because the response is measured after the end of the sentence.

Another early study on distance is Ehrlich and Rayner (1983) who manipulated the distance of the pronoun from its antecedent in an eye-

⁴⁴ Noun phrase anaphors were also used; pronouns and NP anaphors were systematically varied in the experiment but no differences in response latencies between them were found.

tracking experiment. The distances were classified in absolute terms as *near*, *intermediate* and *far* (see (4.5) below) with the description of these conditions described in terms of the amount of intervening material. In the intermediate condition there was one character name intervening between the pronoun and its antecedent, and in the far condition there were two intervening character names.

(4.5a) Near condition

A group of people who shared an interest in photography had recently started writing a newsletter of their activities. In fact, in one room Mark was mailing a copy of the paper to **Susan**. **She** was very involved in photography and spent every weekend taking pictures.

(4.5b) Intermediate condition

A group of people who shared an interest in photography had recently started writing a newsletter of their activities. In fact, in one room **Mark** was mailing a copy of the paper to Susan. **He** was very involved in photography and spent every weekend taking pictures.

(4.5c) Far condition

A group of people who shared an interest in photography had recently started writing a newsletter of their activities. **Mark** wrote most of the copy but the other members did a lot of the work as well. In fact, in one room Cathy was mailing a copy of the paper to Susan. **He** was very involved in photography and spent every weekend taking pictures.

Reading times in the region following the pronoun were significantly slower in the far condition. The antecedents in the near and intermediate conditions were always in the previous sentence, which contained only one clause. But in the far condition, not only was the correct antecedent two sentences back from the pronoun, there was also an intervening subordinate clause in the sentence containing the antecedent, further increasing the distance.

This study ties the distance effect more precisely to the pronoun, given that the measures where differences were found were taken in the region following the pronoun. The results of both these early studies suggest that increasing the distance between the antecedent and the pronoun incurs a processing cost. We cannot be sure if this difficulty is in part attributable to intervening potential but non-matching antecedents, or whether it is solely due to the amount of intervening material. Thus the evidence for recency so far differs from evidence for a first-mention preference. Results from these two studies suggest that, whatever else is driving the resolution of the pronoun, the further away the antecedent is the more processing effort will be involved in establishing a dependency. This does not necessarily entail that the parser will automatically search for a nearby antecedent.

Further evidence for a distance effect, this time in ERPs, comes from Streb, Hennighausen and Rösler (2004) who investigated the ERP signatures of different types of anaphors. Using materials structured in a very similar way to those in Clark and Sengul (1979), they systematically varied the distance between pronouns and their antecedents in German. They also tested other types of anaphor but I will focus on the results from the pronouns here. A target sentence, which contained a pronoun, was preceded by three context sentences. The correct antecedent for the pronoun appeared in context sentence one, two or three, and could be identified by matching gender features. (4.6) shows the structure of the materials, translated into English (translations are taken from the paper).

(4.6a) Far distance condition

Context 1: **Lisa** strolls across a bazaar.

Context 2: Peter sells gems to tourists.

Context 3: The gems are cut excellently.

Target: Then **she** will buy a diamond from the trader.

(4.6b) Medium distance condition

Context 1: The weather is beautiful today.

Context 2: **Gerhart** is an experienced mountaineer.

Context 3: Anna wants to go on a walking tour.

Target: Then **he** shows the ascent to the tourist.

(4.6c) *Near condition*

Context 1: Beate has a small “boardy home” for animals.

Context 2: Everywhere in the house are animals.

Context 3: **Tom** is an old cat.

Target: Today **it** scratched the door of the woman.

They ran two experiments with these materials. The first was a reaction time study, in which participants were asked a comprehension question about the preceding text while their response latencies were recorded. The second was an ERP experiment, in which ERPs at the onset of the subject of the target sentence (the pronoun) were recorded. ERPs at 100ms before the critical word were used as a baseline. They found that for the reaction time study, responses to the comprehension question in the *near* condition were significantly faster than the other two conditions. The *medium distance* and *far distance* conditions did not significantly differ from one another. For the ERPs, however, the three conditions all differed from each other significantly in the 360-480ms time window over central, parietal and temporal electrodes, with the *near* condition evoking the smallest negativity, the *medium distance* condition evoking the largest negativity, and the *far distance* condition lying in between the two. Both sets of results suggest that the processing of the pronoun in the *near* condition is the easiest of the three. However, while the response latencies to the comprehension questions suggest that the number of intervening sentences (one or two) between the pronoun and the antecedent does not make a difference, the ERP data shows otherwise. Between the *medium distance* and the *far distance* conditions, the *medium distance* one is more difficult to process. This is not straightforwardly expected on a purely recency-based assumption. The authors point out that first-mention may play a role here, given that the antecedent in the *far distance* condition is the first-mentioned in the text, whereas the antecedent in the *medium distance* condition is not first-mentioned. Recall that this experiment was carried out in German, which is a language that has gendered nouns. A consequence of this is that pronouns (that have been translated here as *he* and *she*) can refer to all

types of nouns, not just animate ones. In (4.6b), the *medium distance* condition, 'the weather' is the first-mentioned entity. In English, this may not appear to be a plausible antecedent for *he* or *she* on grounds of animacy, but in German 'the weather' can only be ruled out on grounds of gender mismatch, making it a potential, if incorrect, antecedent for the pronoun. If the authors are right about the first-mention effect, a non-matching but first-mentioned potential antecedent might make the *medium distance* condition particularly difficult, even though the antecedent can be rapidly ruled out on grounds of plausibility and gender. Any distance effects seen in the experiment might have been particularly affected by the word-by-word presentation of the stimuli. It should also be mentioned that the description of the study design does not match up with the statistical analysis. The authors describe 300 semantically different scenarios that they wrote for the experiment, in which the pronoun/name anaphor was one, two or three sentences away. This, and the examples of the materials they give, suggest a between-groups design for the factor *distance*. But the ANOVA that they carried out in the statistical analysis was repeated-measures for the factor *distance*. Whether this is a case of unclear description or a problem with the statistical analysis is not certain. However, the results for this study should perhaps be regarded with caution. Nevertheless, if sound, their results imply an interesting interaction then between the factors of recency and first-mention. Clearly, the nearest condition is the easiest to process, evoking the smallest negativity. But contrary to expectations, the *medium distance* condition evoked the greatest processing difficulty. This contrasts with Ehrlich and Rayner's (1983) findings in which only the *far* condition evoked difficulty.

It is difficult to draw comparisons between all three studies because of the different experimental paradigms and structure of the materials. More importantly, in all three studies there is a confound between distance (in terms of number of words or phrase structure) and intervening NPs, which means that it is not possible to unravel the individual effects these factors might have. If Streb *et al.* (2004) are correct that the antecedent in the *far distance* condition is more prominent because of first-mention, it would provide some insight into how first-mention and recency could interact.

The evidence on recency so far simply suggests that the processing effort involved in resolving a pronoun increases as the distance between the pronoun and the antecedent increases. This may be due to distance (number of words, or number of phrases between pronoun and antecedent) or may be due to the number of intervening antecedents. The effect of this distance may be tempered, as Streb *et al.*'s (2004) ERP results show, by first-mention when the previous sentence/clause does not contain a suitable antecedent. As such the evidence from the pronoun studies in this section points to recency effects in pronoun processing being a by-product of memory operations. This is also how it is viewed (mainly) in sentence-processing models. Seen in this way, recency is very different from a cue such as first-mention. The parser does not have to be sensitive to recency during processing, because recency effects arise automatically as memory resources are exploited.

However, the results from Experiment 4 reported in Chapter 3 suggest something different. In that experiment, two potential antecedents were presented, and there was evidence that only the most recent antecedent was considered during processing. If memory operations make it more costly to retrieve a distant antecedent, a delayed effect of the distant antecedent would be expected. Instead, it seemed that only the recent antecedent was considered; no delayed effect of the distant antecedent was found. In Experiment 5, conversely, there were visible effects of the more distant antecedent, but these effects began in a later region than those of the recent antecedent. It was proposed that the different experimental designs brought about the difference in results for Experiments 4 and 5. The match-mismatch design (Experiment 4) is particularly suited to bringing out a preference for one antecedent over another, while the design of Experiment 5 allowed effects of a less preferred antecedent to also be visible (see Chapter 3, section 3.6.1). Taken together, the results of Experiments 4 and 5 suggest a processing preference for the most recent antecedent, which can nonetheless be overcome by gender cues from a more distant antecedent (similar to the gender effects overriding the more accessible antecedent in Arnold *et al.*'s (2000) study discussed above). These results are suggestive of a *preference* for the recent antecedent, rather than a simple distance=cost

metric. It seems more likely, then that the effects in Experiments 4 and 5 were not entirely driven by memory limitations but had an additional source.

4.5 Experimental evidence for a recency preference in pronoun resolution

In Experiment 4 in Chapter 3, there was a recency effect such that the most distant antecedent from the pronoun was not considered during processing. In Experiment 5 there was also some evidence that participants considered the more recent antecedent earlier than the more distant one. These are puzzling results in the light of the first-mention advantage, and a memory-based explanation is somewhat unsatisfactory. One possible reason for the results in Experiments 4 and 5 is that the type of syntactic operation between the pronoun and the antecedent played a role. Recall that the distant antecedent was always a QP and the recent antecedent was a name (Experiment 4) or a DP (Experiment 5). The syntactic structure was such that the QP could only be linked to the pronoun via variable binding, whilst the DP/name could only be linked via coreference. There could be a general preference for CR antecedents in English, which promotes the name/DP as an antecedent. Another possibility is that it is not the syntactic operation that is a factor, rather, there is a dispreference for QPs as antecedents for independent reasons (for example an uncertainty with regard to number agreement with pronouns). Two facts need to be established: whether the order of mention of the antecedents plays a role, and secondly, if it does play a role, whether order of mention affects only processing online, or also offline resolution preferences. In the following sections, three offline experiments and one online experiment are described which test these possibilities by adjusting the order of the antecedents. Not only should these experiments shed further light on the outcome of Experiments 4 and 5, they should also contribute to our understanding of first-mention and recency in the processing of pronouns.

4.6 Experiment 6: alternating the order of mention of DP and QP antecedents

Experiment 6 was an offline questionnaire. The purpose of the questionnaire was to investigate offline preferences for DP and QP antecedents, and to determine whether these preferences were affected by order of mention.

4.6.1 Experiment 6 materials

There were 12 experimental items in two conditions. All items contained one DP, one QP and a pronoun (either *he* or *she*). The QPs were taken from the list of gender stereotyped nouns that were used in Experiments 4 and 5 and the DPs were all nouns of definitional gender. Both the DP and the QP matched the pronoun in gender. Half the items contained feminine pronouns and half contained masculine pronouns. The order of the QP and the DP in the sentence was systematically varied creating two conditions, one in which the DP was closest to the pronoun, shown in (4.7a), and one in which the QP was closest to the pronoun, shown in (4.7b).

(4.7a) *DP nearest*

Every cheerleader sensed that the showgirl thought that she should follow the team off the field.

(4.7b) *QP nearest*

The showgirl sensed that every cheerleader thought that she should follow the team off the field.

Both antecedents c-commanded the pronoun and they were both grammatical subjects. Therefore, both antecedents were accessible to the pronoun grammatically, and were potential binders for the pronoun. An effort was made to avoid any plausibility bias for the antecedents in a sentence and the antecedent positions. The complete list of experimental items is given in Appendix 5.

There were 20 fillers in all, 15 unambiguous fillers and five ambiguous fillers.⁴⁵ Both masculine and feminine pronouns and antecedents were presented in the fillers. The large number of unambiguous fillers was to counterbalance the experimental items, which were all ambiguous. The 15 unambiguous sentences were disambiguated by gender (one antecedent of each gender was presented in these sentences). Eight of the pronouns presented in the fillers were reflexive pronouns. The antecedents were either names, gender-stereotyped NPs or less gender-stereotyped NPs (e.g. *the owner*).

4.6.2 Experiment 6 participants

28 participants (12 males) completed the questionnaire for Experiment 6. All 28 had taken part in Experiment 5 (described in Chapter 3). Their mean age was 30 years, range 18-60. All participants were paid £7 for their participation in the experimental session, which included both Experiments 5 and 6. All participants confirmed that they had not been diagnosed with any language difficulty/disorder.

4.6.3 Experiment 6 procedure

All participants were given a paper questionnaire after having completed the eye-tracking (Experiment 5) part of the session. The questionnaire contained 32 questions, 12 of which were experimental items and 20 of which were fillers (see above). Each question took the form of a single sentence containing a pronoun and two potential antecedents. The sentence was always followed by a question asking who the pronoun refers to, for example *Who does 'he' refer to?*. On the next line both antecedents were presented with a box and participants were asked to indicate, by ticking the box, the most likely referent of the pronoun. The task was forced choice, in that no *either* or *don't know* response options were provided. Participants

⁴⁵ As noted in Chapter 1, any pronoun that is not reflexive is technically ambiguous because it could refer to someone outside the current sentence/discourse. Here I am using the term unambiguous because participants were presented with a choice of the two antecedents who were introduced in the sentence, and they were not given an option to choose a referent outside the sentence.

were asked to answer quickly based on their intuitive preference, and not to go back and change answers to earlier questions, and they were told that some sentences would be ambiguous. The items were presented in a Latin-square design creating two questionnaires, so that each questionnaire contained 6 experimental items from each of the two conditions. The order of items and fillers was pseudo-randomised so that the experimental items were evenly distributed throughout the questionnaire and so that two experimental items would not be presented one after another. The order of questions was reversed to make a further two versions of the questionnaires, to minimise possible task-related effects.

4.6.4 Experiment 6 predictions

If the online results from Experiments 4 and 5 were driven by a general unwillingness to link a pronoun to a QP antecedent, the DP antecedent should be chosen more often than the QP antecedent, regardless of the order of the antecedents (i.e. in both conditions).

If there is a general preference to link the pronoun to the most recent antecedent when both antecedents appear in the same sentence as the pronoun, then the DP should be chosen more often in the DP nearest condition (4.7a) and the QP should be chosen more often in the QP nearest condition (4.7b).

4.6.5 Experiment 6 results

The frequency of responses for each condition are presented in Table 4.1. Percentages are calculated per condition.

Table 4.1 Frequency of responses and percentages (calculated per condition) for Experiment 6.

	DP responses	QP responses	Total
(a) DP nearest	121	41	162
	75%	25%	100%
(b) QP nearest	78	84	162
	48%	52%	100%
Total	199	125	

Overall, there are more DP responses than QP responses. It appears that responses are guided by the proximity of the antecedent to the pronoun, so that when the DP is nearest the pronoun it is the favoured response, and when the QP is nearest the pronoun the QP is the favoured response. However, this trend is much more marked in the DP nearest condition (4.7a).

The response rates were analysed using Wilcoxon signed rank tests by items and participants for two dependent variables.⁴⁶ The first dependent variable analysed (set 1) was the proportion of DP responses per condition averaged over items and participants. The second dependent variable analysed (set 2) was the proportion of responses per condition that were the nearest antecedent to the pronoun (i.e. a DP response in the DP nearest condition and a QP response in the QP nearest condition). The results of the tests are given in Table 4.2.

⁴⁶ When the data is collapsed over items or subjects, there are not enough observations to determine whether or not the dependent variable is normally distributed. For this reason the more conservative Wilcoxon signed rank test has been used as opposed to a t-test because it does not assume a normal distribution.

Table 4.2 Outcome of the Wilcoxon signed rank tests by items and participants for Experiment 6. Set 1 dependent variable = proportion of DP responses. Set 2 dependent variable = proportion of nearest NP responses.

	Item analysis		Participant analysis	
	W	<i>p</i> -value	W	<i>p</i> -value
Set 1	70.5	0.01	249	<0.01
Set 2	66.5	0.03	212.5	<0.01

The set 1 results show that the proportion of DP responses in (4.7a) is significantly different from the proportion of DP responses in (4.7b), in both the item and the participant analysis. This shows that the manipulation of the antecedent order affects the antecedent choice. The set 2 results show that the proportion of nearest responses (when the antecedent nearest the pronoun was chosen) is significantly different across conditions in both the item and participant analysis. The DP is considerably more popular than the QP in the DP nearest condition (4.7a), but the QP is only slightly more popular than the DP even when it is the closest antecedent to the pronoun (4.7b). This demonstrates that responses were affected by the type of antecedent (DP or QP). Responses guided purely by a recency preference, for example, should have had roughly the same proportion of nearest responses across both conditions. Combining the set 1 and set 2 results, it appears that the factors of recency and NP type (DP versus QP) both have an effect on choosing a referent in an offline task. The results are discussed in more detail in section 4.10.

4.7 Experiment 7: the interpretation of pronouns with VB and CR antecedents

Experiment 7 was an offline questionnaire task. The items for Experiment 7 acted as fillers for Experiment 1 which was reported in Chapter 2 (see section 2.4). The purpose of Experiment 7 was to test the interpretation of

pronouns that had both a potential VB antecedent and a potential CR antecedent.

4.7.1 Experiment 7 materials

There were 10 experimental items in Experiment 7. All items contained a pronoun (*he* or *she*) and two potential antecedents, a VB antecedent which was a QP, and a CR antecedent which was a name. The QP antecedent c-commanded the pronoun, and the CR antecedent did not c-command the pronoun. The QPs were taken from the list of gender stereotyped nouns that were used in Experiments 4 and 5. Both antecedents matched the pronoun in gender. Five of the items presented the VB antecedent before the CR antecedent, and five of the items presented the CR antecedent before the VB antecedent. However, the order was not systematically varied within one experimental item; instead, two separate sets of items were created. An example of each type is given below, and the full list of items is given in Appendix 1.

(4.8) *QP first*

Every pilot who knew that Peter was working at the control tower today hoped that he would arrive on time.

(4.9) *Name first*

It appeared to Paul that every boxer in the changing room thought that he should fight a lot harder.

For a description of the other experimental items and fillers in Experiment 7, the reader is referred to Chapter 2, section 2.4.

4.7.2 Experiment 7 participants and procedure

Full details of the participants and the procedure can be found in Chapter 2, section 2.4. The questionnaire was completed by 83 native English speakers via the internet. Participants were asked to read each sentence and decide who the pronoun probably referred to. The word *probably* was included in the instructions to account for the fact that a pronoun can refer to an entity outside of the discourse. After each sentence the same question

appeared: *Who does [pronoun] refer to?*. Participants were given three choices: *antecedent 1, antecedent 2 or either*. An example is given below.

(4.10)

Every pilot who knew that Peter was working at the control tower today hoped that he would arrive on time.

Who does 'he' refer to?

Every pilot

Peter

Either

The order of the two antecedents in the response section was varied throughout the questionnaire. The *either* option always appeared last.

4.7.3 Experiment 7 results

The frequency of responses and percentage for each item type are presented in Table 4.3 below.

Table 4.3. Frequency of responses and percentages for each item type, Experiment 7

	QP responses	Name responses	Either responses
QP first	97 23%	201 48%	115 28%
Name first	96 23%	116 28%	203 49%
Total	193 23%	317 38%	318 38%

Overall the QP antecedent is chosen less than the name antecedent. The pronoun was interpreted as being ambiguous at around the same rate that the name antecedent was chosen. However, there was some variation between the two item types. While the QP antecedent was chosen at the same rate whether it appeared first or last, the name antecedent was chosen more often when it was the most recent antecedent (QP first items)

compared to when it appeared first (name first items). There were more *either* responses in the name first items than in the QP first conditions.

Because the items were not systematically varied to create two conditions, the responses were not compared between the two item types. Wilcoxon signed rank tests were used to check three of the response rates against chance level (33.33%). In the QP first items, the name was chosen at above chance level (significant only by participants, not by items: $V_1=2671$, $p_1<0.01$; $V_2=12$, $p_2=0.14$). In the name first items *either* was chosen at above chance level ($V_1=2396$, $p_1<0.01$; $V_2=15$, $p_2=0.03$). In total, the QP antecedent was chosen at a rate that was marginally below chance ($V_1=795$, $p_1<0.01$; $V_2=13$, $p_2=0.08$).

4.8 Experiment 8: the effect of order of mention on the interpretation of pronouns with VB and CR antecedents

Since Experiment 7 only used a very small number of items and without systematic variation of the items to create two comparable conditions, Experiment 8 was carried out to test the effect of order of mention on VB and CR antecedents.⁴⁷

4.8.1 Experiment 8 materials

There were 24 experimental items in two conditions, QP first and name first. The QP first items were the double match conditions from Experiment 4 (see Chapter 3, section 3.5; for a full list of the materials for Experiment 4 see Appendix 3) and the name first items were the double match conditions from Experiment 9 (reported in the current chapter, section 4.9; for a full list of Experiment 9 materials see Appendix 6). The neutral context

⁴⁷ This experiment was run in collaboration with Ian Cummings. The materials were designed by Ian Cummings, being an adaptation of the materials for Experiment 4 (described in Chapter 3, section 3.5). Data collection was carried out by both Ian Cummings and me, and I carried out the data analysis. This experiment is also reported in full in Cummings, Patterson and Felser (accepted).

sentence was retained from Experiments 4 and 9, but the wrap-up sentence was removed. The structures were the same as those in Experiment 7: all items contained a pronoun (*he* or *she*) and two potential antecedents, a VB antecedent which was a QP, and a CR antecedent which was a name. The QP antecedent c-commanded the pronoun, and the CR antecedent did not c-command the pronoun. Both antecedents matched the pronoun in gender.

(4.11a) *QP first*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that he should wave as the parade passed.

(4.11b) *Name first*

The squadron paraded through town. It looked to James that every soldier was completely convinced that he should wave as the parade passed.

In addition to the 24 experimental items there were 48 fillers items in the questionnaire. Unambiguous and ambiguous fillers were included.

4.8.2 Experiment 8 participants

24 participants (4 males) took part in Experiment 8. All were native speakers of English, and their mean age was 24 years. Participants were not remunerated for their participation. None of the participants for Experiment 8 had taken part in Experiments 4, 5 or 9.

4.8.3 Experiment 8 procedure

The 24 experimental items and 48 fillers were combined into a questionnaire, and the order of items was pseudo-randomised so that the experimental items were evenly distributed and no two experimental items would appear one after another. The items were presented in a Latin-square design creating two questionnaires, so that each questionnaire contained 12 experimental items from each of the two conditions. The pronoun was underlined in each item. Participants were asked to indicate

who the pronoun was most likely to refer to, and each item was followed by a five-point response scale: *equally likely* was in the mid-point of the scale, and each side had a point for *fairly likely* and *very likely*, with the name presented at one end of the scale and the QP presented at the other end. Examples were provided in the instructions, to demonstrate how to use the scale. The questionnaire was administered via email.

4.8.4 Experiment 8 predictions

If the online results from Experiments 4 and 5 were driven by a general unwillingness to link a pronoun to a QP antecedent, the name antecedent should receive more *fairly likely* and *very likely* responses than the QP antecedent, regardless of the order of the antecedents (i.e. in both conditions).

If there is a general preference to link the pronoun to the most recent antecedent when both antecedents appear in the same sentence as the pronoun, then the name should receive more *fairly likely* and *very likely* responses in condition a, shown in (4.11a), and the QP should receive more *fairly likely* and *very likely* responses in condition b, shown in (4.11b).

4.8.5 Experiment 8 results

The response scale was converted to scores with higher scores for QP responses, as follows:

- Name very likely: 1 point
- Name fairly likely: 2 points
- Equally likely: 3 points
- QP fairly likely: 4 points
- QP very likely: 5 points

Table 4.4 gives the mean response scores and standard deviations per condition.

Table 4.4. Mean response scores with standard deviations (shown in italics) in each condition for Experiment 8.

	Score	
(a) QP first	2.64	<i>(1.46)</i>
(b) Name first	3.56	<i>(1.44)</i>

The mean score for (4.11a), where the QP appears before the name, is below three, indicating a preference for the name antecedent in this condition. In the name first condition (4.11b), where the QP is nearest the pronoun, the mean score is above three indicating a preference for the QP. A paired Wilcoxon test was carried out to determine, firstly, that the scores in the two conditions were significantly different from each other. The test confirmed that they were different: $V_1=11, p_1<0.01, V_2=29.5, p_2<0.01$. Both scores were then compared to the value 3, to test whether they were significantly different from the mid-point of the scale. The score for condition a ($V_1=48.5, p_1<0.01, V_2=67.5, p_2=0.06$) and for condition b ($V_1=249.5, p_1<0.01, V_2=243, p_2=0.01$) confirmed this (the test for condition a was marginally significant by items). This confirms the pattern observed above that there was a preference for the name antecedent in the QP first condition and a preference for the QP in the name first condition, i.e. the local antecedent was preferred in each condition.

The results from the three offline experiments are discussed below in section 4.10.

4.9 Experiment 9: searching for an online recency preference

Experiment 9 is a follow-up experiment to Experiment 4, described in Chapter 3.⁴⁸ In Experiment 4 it was found that participants only considered the name antecedent and not the QP antecedent during processing, but in Experiment 4 the two antecedents were always presented with the name antecedent as the most recent antecedent. In Experiment 9 the order of the antecedents was reversed so that the QP antecedent is always the most recent antecedent. This is to demonstrate whether or not the position of the antecedents would play a role during processing.

4.9.1 Experiment 9 materials

The experimental materials were 24 texts that were adaptations of materials in Experiment 4. They each contained a neutral lead-in sentence, a critical sentence and a wrap-up sentence. The lead-in and wrap-up sentences did not contain potential antecedents for the pronoun. The critical sentence contained a singular subject pronoun (*he* or *she*), and two potential antecedents. The first antecedent was always a name and the second antecedent was always a QP containing a noun with stereotypical gender. The name antecedent was not the first word of the critical sentence, rather it was introduced by the phrase “It looked/seemed/appeared to xx...”. Half the pronouns were male and half female. The gender match between the pronoun and both antecedents was manipulated to create four conditions: double match, no match, name only match, QP only match. An example of a text in the four conditions is given below.

⁴⁸ This experiment was run in collaboration with Ian Cummings. The materials were designed by Ian Cummings, being an adaptation of the materials for Experiment 4 discussed in Chapter 3, and reported in Cummings *et al.* (accepted). The experiment was run by me, and the original data analysis was shared between Ian Cummings and me. The data analysis presented below is my own work.

(4.12a) *Double match*

The squadron paraded through town.
It looked to James that every soldier
was completely convinced that he should wave as the parade passed.
The entire town was extremely proud that day.

(4.12b) *QP match*

The squadron paraded through town.
It looked to Helen that every soldier
was completely convinced that he should wave as the parade passed.
The entire town was extremely proud that day.

(4.12c) *Name match*

The squadron paraded through town.
It looked to Helen that every soldier
was completely convinced that she should wave as the parade passed.
The entire town was extremely proud that day.

(4.12d) *No match*

The squadron paraded through town.
It looked to James that every soldier
was completely convinced that she should wave as the parade passed.
The entire town was extremely proud that day.

The examples above show where the line breaks appeared in the experimental materials. The name and the QP appeared on the same line, and the pronoun appeared in the middle of the next line. The full set of experimental materials can be found in Appendix 6.

The materials were designed to be as close as possible to those of Experiment 4, while reversing the order of the antecedents. However, it was not possible to simply swap the antecedent order and leave the rest of the structure of the sentence unaffected. This is because in Experiment 4 the QP was followed by a relative clause, and the name was contained within the relative clause. There were no commas around the relative clause following the QP, which indicates that it can be assumed to be a restrictive relative clause, which usually has the function of narrowing down the set

of possible referents. In (4.13) below (taken from Experiment 4) the restrictive relative clause is in parenthesis:

(4.13) Every soldier [who saw that James was watching]...

The QP *every soldier* suggests a set of soldiers, and the information in the relative clause functions as a restriction on that set, implying that, out of a set of soldiers, only some of them saw that James was watching, and it is those soldiers that we are talking about. Reversing the order of the antecedents gives us a name *James* that is followed by a restrictive relative clause. But *James* already refers only to one person, so it is impossible to restrict the set any further. The second problem with swapping the antecedents directly is that the relationship to the pronoun would change. Putting *James* as the subject of the matrix clause would mean that it c-commanded the pronoun, and thus could be linked to the pronoun via variable binding. Having the QP within the relative clause would mean that it was not in a c-commanding position with respect to the pronoun, so variable binding would be ruled out. As discussed in Chapter 3, a QP such as *every soldier* cannot refer directly to an individual, so coreference is impossible. In that case, the QP becomes inaccessible to the pronoun. In order to overcome these two problems, and to maintain the VB and CR relationships between the antecedents and the pronoun, a slightly different structure was used. The name was placed into a cleft at the start of the sentence, and the QP was contained in a main clause following *that*.

Fillers

There were 48 fillers and 12 pseudo-fillers (pseudo-fillers are discussed below). There were four fillers that had a similar structure to the experimental items but did not contain a pronoun, so they differed from the structure of the experimental items in a subtle way, to encourage careful reading of the materials. Other fillers contained names, role names (e.g. *ballet dancer*) and QPs in a range of syntactic structures, with a range of lengths, to distract from the experimental items.

Pseudo fillers

Twelve items were included that contained either a pronoun or a reflexive. These had a different syntactic structure from the experimental items. It is important for the fillers to contain, firstly, pronouns in other syntactic structures, so that participants do not get used to seeing pronouns used always in the same structure, which could lead to a reduced effort in processing or boredom which could affect the outcomes of the experiments. Presenting a number of items that share a structure can act as a distractor to ensure that participants do not become over-familiar with the experimental structures.

4.9.2 Experiment 9 predictions

If the preference for a recent antecedent in Experiment 4 was driven mainly by the position of the antecedents relative to the pronoun, then the same preference for a recent antecedent should also be found in Experiment 9. This will be visible in main effects of QP match, with shorter reading times in the double match and QP match conditions compared to the name match and no match conditions.

If the preference for a recent antecedent in Experiment 4 was driven mainly by a preference for the name antecedent over the QP antecedent, or a preference for the CR antecedent over the VB antecedent, there should be a main effect of name match, with shorter reading times in (4.12a) and (4.12c) (double match and name match) compared to (4.12b) and (4.12d) (QP match and no match).

4.9.3 Experiment 9 method

The materials were arranged in a Latin-square design, resulting in four lists. The experiment was divided into four blocks, with a break between each block. Forward and reverse orders within each block were constructed and the ordering of each block was different for each participant. No two experimental items were presented consecutively.

4.9.3.1 Experiment 9 participants

Participants were 31 adults (9 males) recruited from the University of Essex community. The age range was 18-46, with a mean age of 24. All participants confirmed that they had not been diagnosed with any language difficulty/disorder. Some participants wore contact lenses or glasses during the experiment, and all reported that they could read from the screen successfully. Participants who took part in this experiment had not taken part in Experiments 4 or 5. All participants were paid £6 for their participation.

4.9.3.2 Experiment 9 procedure

In this experiment the eye-movement data was recorded using the head-mounted Eyelink II system (by SR Research, Canada). See Chapter 3, section 3.6.3 for a description of the Eyelink II system. All other aspects of the procedure are as described in Chapter 2, section 2.5.3.2.

Following two thirds of the trials there were comprehension questions. Comprehension questions did not explicitly probe the referent of the pronoun, in order not to draw attention to the purpose of the experiment or encourage strategic reading behaviour. The experiment session took around 45 minutes in total.

4.9.3.3 Experiment 9 data preparation

Individual fixations shorter than 80ms and within one degree of visual angle of a neighbouring fixation were merged with the neighbouring fixation. Other fixations shorter than 80ms, and longer than 800ms, were removed. Within a trial, regions that were initially skipped were removed from the analysis. The eye-movement measures examined in the current experiment are first-fixation times, first-pass times, regression-path times, rereading times and total viewing times. Initially skipped regions were also removed. Track loss accounted for 1.6% of the data.

The critical sentence for each item was divided into regions of interest for the analysis as follows:

(4.14)

It looked to James that every soldier | was completely convinced |
that he | should wave | as the | parade passed. |

- **[was completely convinced]**: precritical region, the verb preceding the pronoun whose subject was always the QP
- **[that (s)he]**: pronoun region, containing the word *that* and the pronoun
- **[should wave]**: spillover region, containing two words following the pronoun
- **[as the]**: prefinal region, containing two words following the spillover region
- **[parade passed]**: final region, containing two words following the prefinal region, always the last two words of the sentence.

4.9.3.4 Data analysis

Before the statistical analysis the dataset was transformed using a logarithmic transformation ($\log_e(x+1)$). Untransformed and transformed data was plotted against a normal distribution line and visually inspected. Where the transformed data better fitted the normal distribution, it was used in the analysis in place of the untransformed data. This was the case for all regions and measures except the rereading times (all regions). Untransformed data was used in the analysis of the rereading times. The data was analysed with linear mixed effects models using the lmer function in the lme4 package in R (Bates, Maechler and Bolker 2012).

For the fixed part of the model, two effects, *QP* and *name*, were included, as well as a *QP* \times *name* interaction. Each fixed factor was a categorical variable with two levels: match, when the antecedent in question matched with the pronoun in gender, and mismatch, when the antecedent in question mismatched with the pronoun in gender. For the random part of the model maximal random effects were included (following Barr *et al.* 2013). The full random structure was random intercepts for subjects and items, and random slopes for subjects and items by each fixed effect and the

interaction. In cases of non-convergence, one term was removed from the model and it was re-run, with this process repeated until a converging model was found, as suggested by Barr *et al.* (2013). The first term removed was always the random slope for the interaction. *P*-values determining significance of the *t*-value were derived from the upper bound of the *t*-statistic (Baayen 2008).

4.9.4 Experiment 9 results

4.9.4.1 Comprehension questions

The overall response rate to the comprehension questions was 92%. All participants individually scored over 70%.

4.9.4.2 Eye-movement data

The skipping rates in the analysed regions were as follows: precritical, 0.5%; pronoun, 21%; spillover, 7%; prefinal, 12%; final, 7%. The means and standard deviations for all measures in all analysed regions are shown in Table 4.5.

Table 4.5. Means in all conditions in milliseconds (with standard deviations shown in brackets) for first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions from precritical to final for Experiment 9.

	First-fixation mean	First-pass mean	Regression- path mean	Rereading mean	Total mean
Precritical region					
Double match	199 (67)	847 (465)	959 (861)	286 (483)	1133 (705)
QP match	207 (86)	906 (512)	1013 (820)	344 (592)	1249 (817)
Name match	207 (83)	896 (547)	1001 (637)	444 (793)	1340 (1012)
No match	216 (96)	870 (573)	941 (620)	466 (855)	1336 (1046)
Pronoun region					
Double match	218 (78)	286 (161)	336 (235)	154 (525)	440 (306)
QP match	215 (76)	301 (175)	408 (387)	195 (417)	496 (456)
Name match	228 (84)	307 (181)	438 (464)	228 (329)	535 (396)
No match	229 (85)	337 (213)	488 (609)	279 (473)	616 (545)
Spillover region					
Double match	205 (70)	315 (199)	455 (515)	209 (347)	524 (375)
QP match	212 (72)	314 (170)	467 (548)	221 (325)	534 (372)
Name match	213 (80)	309 (191)	491 (733)	237 (316)	546 (368)
No match	207 (72)	275 (148)	603 (934)	327 (457)	602 (493)
Prefinal region					
Double match	219 (80)	299 (156)	404 (491)	191 (271)	490 (318)
QP match	222 (70)	309 (164)	493 (543)	145 (218)	454 (259)
Name match	215 (76)	306 (184)	440 (424)	184 (280)	490 (340)
No match	218 (74)	296 (175)	524 (963)	178 (253)	474 (300)
Final region					
Double match	217 (79)	345 (197)	757 (790)	105 (229)	450 (286)
QP match	217 (84)	336 (209)	649 (607)	137 (327)	473 (365)
Name match	218 (89)	356 (253)	945 (1078)	170 (299)	527 (408)
No match	226 (90)	353 (243)	1104 (1637)	140 (275)	493 (348)

In the precritical region, the first-fixation durations are very similar across the four conditions, but the double match condition (4.12a) has the shortest and the no match condition (4.12d) the longest mean duration. First-pass times and regression-path times in this region are long compared to the other regions, which is likely to be due to the length of the region. The rereading and total viewing times in this region appear to be longer for the name match and no match conditions (4.12c and 4.12d) compared to the double match and QP match conditions (4.12a and 4.12b).

In the pronoun region, in all measures the reading times are longer in the name match and no match conditions (4.12c and 4.12d) compared to the double match and QP match conditions (4.12a and 4.12b). In addition, the no match condition has the longest reading times in nearly all measures. The same pattern is seen in the regression-path times, rereading times and total viewing times in the spillover region. The first-pass times, however, go in the opposite direction with shorter first-pass times in the name match and no match conditions (4.12c and 4.12d) compared to the double match and QP match conditions (4.12a and 4.12b). Viewed in conjunction with the opposite pattern in the regression-path times, this is likely to indicate that participants were quicker to exit the spillover region to the left (reviewing earlier material) when the QP mismatched the pronoun.

In the prefinal region the first-fixation durations and the first-pass times are very similar across all conditions. Regression-path times are longer in the QP match and no match conditions (4.12b and 4.12d) compared to the double match and name match conditions (4.12a and 4.12c) but rereading times are shortest in the QP match condition. Total viewing times are longer in the double match and name match conditions compared to the QP match and no match conditions (4.12a and 4.12c vs. 4.12b and 4.12d). In the regression-path times in the final region the name match and no match conditions are longer compared to the double match and QP match conditions (4.12c and 4.12d vs. 4.12a and 4.12b), and a similar pattern is visible in the rereading times. Total viewing times in the name match and no match conditions are longer compared to the double match and QP match conditions (4.12c and 4.12d vs. 4.12a and 4.12b).

The outcome of the statistical analysis is shown in Table 4.6.

Table 4.6. Summary of statistical analysis for each region and measure, Experiment 9. The estimate for the adjustment for each factor in the model is given in the 'estimate' column, and the standard error of the estimate is shown in brackets. The t-value for each estimate is given with the significance level indicated: () = $p < .10$, * = $p < .05$, ** = $p < .001$. SE=standard error.*

Table 4.6

	First-fixation durations		First-pass times		Regression-path times		Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region										
QP	0.03 (0.03)	0.75	0.04 (0.05)	0.80	0.08 (0.06)	1.37	153.06 (77.65)	1.97*	0.14 (0.05)	2.97**
Name	0.03 (0.04)	0.69	0.07 (0.04)	1.49	0.09 (0.05)	1.74(*)	55.98 (75.52)	0.74	0.11 (0.05)	2.00*
QP x name	<0.01 (0.04)	0.06	-0.11 (0.07)	-1.55	-0.15 (0.07)	-2.27*	-34.13 (94.61)	-0.36	-0.14 (0.08)	-1.81(*)
Pronoun region										
QP	0.04 (0.04)	1.13	0.05 (0.05)	1.01	0.16 (0.08)	1.89(*)	68.92 (40.77)	1.69(*)	0.18 (0.06)	3.01**
Name	-0.01 (0.03)	-0.24	0.04 (0.05)	0.87	0.11 (0.07)	1.42	34.90 (47.11)	0.74	0.08 (0.06)	1.20
QP x name	0.01 (0.05)	0.29	0.03 (0.07)	0.43	-0.04 (0.12)	-0.31	-3.67 (60.81)	-0.06	-0.05 (0.09)	-0.60
Spillover region										
QP	0.04 (0.03)	1.20	<0.01 (0.05)	0.04	-0.01 (0.06)	-0.12	24.67 (45.49)	0.54	0.03 (0.07)	0.48
Name	0.03 (0.03)	1.01	0.02 (0.05)	0.33	-0.01 (0.06)	-0.09	7.33 (42.69)	0.17	0.01 (0.06)	0.21
QP x name	-0.07 (0.05)	-1.36	-0.11 (0.07)	-1.50	0.06 (0.10)	0.62	84.45 (70.27)	1.20	0.05 (0.10)	0.57
Prefinal region										
QP	-0.01 (0.04)	-0.33	-0.01 (0.06)	-0.11	0.07 (0.08)	0.97	8.12 (25.68)	0.32	<0.01 (0.07)	0.04
Name	0.02 (0.03)	0.72	0.04 (0.05)	0.68	0.16 (0.07)	2.23*	-38.80 (30.33)	-1.28	-0.02 (0.07)	-0.36
QP x name	-0.01 (0.05)	-0.29	-0.05 (0.09)	-0.62	-0.12 (0.10)	-1.19	27.58 (35.56)	0.78	0.01 (0.09)	0.08
Final region										
QP	-0.02 (0.04)	-0.55	-0.04 (0.07)	-0.57	0.13 (0.08)	1.63	62.61 (30.05)	2.08*	0.07 (0.06)	1.16
Name	-0.02 (0.04)	-0.49	-0.06 (0.06)	-1.04	-0.10 (0.09)	-1.16	27.91 (33.37)	0.84	-0.01 (0.06)	-0.16
QP x name	0.06 (0.05)	1.22	0.10 (0.09)	1.19	0.13 (0.11)	1.19	-56.91 (51.16)	-1.11	<-0.01 (0.08)	-0.03

First-fixation durations and first-pass times show no significant effects in the precritical region. This is expected because the pronoun has not yet been encountered. The regression-path times show a marginal effect for name match and a significant interaction between QP and name. This is unexpected. Although regression-path times are not a very early measure, they represent summed durations before the region has been exited to the right, i.e. before the pronoun has been encountered. It is difficult to ascribe these effects to parafoveal processing of the pronoun, because the pronoun region includes the word *that*, making it less likely that the pronoun has been viewed from the precritical region. The effect in the regression-path times could reflect gender-related effects between the two antecedents (as opposed to pronoun-antecedent effects). Regression-path times in the double match and the no match conditions (4.12a and 4.12d) appear to be shorter than the regression-path times in the single match conditions (4.12b and 4.12c) where the two antecedents have different gender features. It is possible that the establishment of a second discourse entity might incur less of a processing cost when the gender features match those of a previous discourse entity. This is purely speculative, although it should be noted that Badecker and Straub (2002) found a trend in the same direction in their experiment 2, where there were slightly longer reading times in single match name condition compared to the multiple match name condition.

The rereading and total viewing times in the precritical region show a rather different pattern. This is not surprising because the rereading times reflect processing after the pronoun has been encountered. Firstly, there is a significant main effect of QP in rereading times and total viewing times, which are shorter when the QP matches the pronoun (4.12a and 4.12b) compared to when it mismatches (4.12c and 4.12d). Secondly, there is a main effect of name in the total viewing times; given that this is an aggregate measure, the effect is probably due to the (marginal) effect of name in the regression-path times as opposed to rereading times, where there is no such effect. Similarly, there is a marginally significant interaction between QP and name in the total viewing times reflecting the interaction in the regression-path times discussed above. While the effect of name and the interaction in this region is likely to be due to processing

before the pronoun is encountered, the effect of QP in this region can be attributed to processing after the pronoun is encountered.

In the pronoun region there are marginal main effects of QP in regression-path times, rereading times and a main effect in total viewing times. There is no effect of name in this region and no interaction. This confirms the pattern that was visible from the means, namely longer regression-path, rereading and total viewing times in the name match and no match conditions (4.12c and 4.12d) compared to the double match and QP match conditions (4.12a and 4.12b). This trend was also visible in the earlier measures but this was not confirmed statistically. In the spillover region the trends that were visible in the means were not confirmed statistically. There were no main effects of QP or name in the spillover region and no interaction. In the prefinal region there is a main effect of name match in the regression-path times, with longer regression-path times in the QP match and no match conditions (4.12b and 4.12d) compared to the double match and name match conditions (4.12a and 4.12c). There are no other main effects or interactions in this region. In the final region there is again a main effect of QP in the rereading times, confirming that rereading times are longer in the name match and no match conditions (4.12c and 4.12d) compared to the double match and QP match conditions (4.12a and 4.12b). The trend for this pattern in the regression-path times and total viewing times was not confirmed statistically.

4.10 Discussion

This chapter has surveyed previous research on recency in pronoun resolution, and also the first-mention advantage. Four experiments were reported (three offline experiments and one online experiment) which tested for a recency preference in the processing and interpretation of pronouns. I will summarise and discuss the findings for the offline and online experiments below, before evaluating the implications of the findings in relation to previous experimental results (Experiments 4 and 5 from Chapter 3) and in relation to the notions of recency and first-mention in pronoun processing.

4.10.1 Summary of findings for the offline experiments

Experiment 6 was an offline forced choice questionnaire in which a DP antecedent and a QP antecedent were presented. The order in which the two pronouns appeared was manipulated. Both antecedents c-commanded the pronoun, which meant that they were both potential binders of the pronoun. Overall there were more DP responses (61%) than QP responses (39%), demonstrating a general preference for the DP antecedent. The proportion of DP responses was significantly different between the two conditions (75% when the DP was the most recent antecedent, 48% when the QP was the most recent antecedent), which shows that participants' preferences were also affected by the position of the antecedents. In sum, participants' offline choices were affected both by the recency of the antecedent (recent antecedents were preferred over first-mentioned antecedents), but also by the type of antecedent (DP antecedents were preferred over QP antecedents).

Both Experiments 7 and 8 compared QP antecedents with name antecedents. The order of the antecedents was manipulated by presenting two slightly differing structures which allowed the QP to c-command the pronoun (from both recent and distant positions) and in which the name did not c-command the pronoun (but see also the discussion on the position of the name in section 4.10.3). Experiment 7 items were fillers for a different offline task (reported as Experiment 1 in Chapter 2) and this was therefore a very small experiment. It differed from Experiment 6 in that an *either* option was offered, so participants could indicate that they found the pronoun ambiguous. The QP antecedent was chosen at a low rate, whether it was presented in a recent or distant position (23% in each). The name and the *either* option each received 38% of the responses overall, but these responses varied according to the item type: the name was chosen more often than *either* when the name was the recent antecedent (48% versus 28%), and *either* was chosen more often when the name was the distant antecedent (49% versus 28%). Like Experiment 6, Experiment 7 demonstrates that offline interpretations are affected by both the position of the antecedents and the antecedent type. However, when the QPs were

closer to the pronoun they did not receive more responses than when they were distant. The results from Experiment 7 must be interpreted with a certain degree of caution because the differences between the content of the two item types could have given rise to the effects, as the items were not systematic variations but two different groups of items.

Experiment 8 was a more extended version of Experiment 7, with more items and similar materials, although in this experiment there was a response scale allowing participants to express a degree of certainty or uncertainty about their interpretations which was reflected in the response scoring. The mean score for the QP first condition (where the name was the most recent antecedent) was 2.64 and for the name first condition (QP as the recent antecedent) was 3.56, with scores closer to five indicating a QP preference. Both these scores significantly differ from the mid-point of the scale. This shows that there were more QP responses when the QP was the most recent antecedent and more name responses when the name was the most recent antecedent. This result again highlights that offline interpretations are affected by the antecedent order, with a preference for the most recent antecedent.

4.10.2 Summary of findings for the online results

In Experiment 9 participants read sentences in four conditions in a match-mismatch paradigm. The gender match of the QP antecedent and the name antecedent was manipulated, and, unlike in Experiment 4, the QP was the most recent antecedent in all conditions. Main effects of both QP match and name match were found. The main effects in all cases were for shorter reading times in the match conditions compared to the mismatch conditions, reflecting difficulty when the antecedents mismatched in gender with the pronoun. No effects were found in the early measures (first-fixation durations and first-pass times), but in the regression-path times there was a marginal effect of QP in the pronoun region and a main effect of name in the prefinal region. There was also a marginal main effect of name and a significant interaction between name and QP in the precritical region, however this was difficult to interpret because it occurs before the pronoun was encountered; it is also difficult to attribute the

effects here to parafoveal processing because the pronoun is separated from the precritical region by the word *that*. However, since a different pattern of results is seen in the regression-path times in the following region, it can be assumed that the effects in the precritical region did not extend into the pronoun region. There was a main effect of QP in the rereading times in the precritical, pronoun (marginal effect), and final regions. This was also reflected in a main effect of QP in the total viewing times in the precritical and pronoun regions. These effects show an increased processing cost associated with a mismatching QP antecedent and a mismatching name antecedent. However, the QP effects are more numerous and appear in both earlier regions (precritical and pronoun) and the final region. The name effect occurs only in the precritical region and the total viewing times of the precritical region. In addition, the effect in the precritical region is likely to arise from the (marginal) effect regression-path times, occurring before the pronoun was encountered. Overall, the mismatching QP antecedent caused increased processing costs, with some cost also associated with a mismatching name antecedent.

4.10.3 Comparing results from Experiments 4 and 5 to Experiments 6-9

In the previous chapter Experiments 4 and 5 were reported, which tested the POB hypothesis. The POB hypothesis states that VB antecedents are always considered before CR antecedents when pronouns are processed. This hypothesis was tested by presenting a pronoun and two potential antecedents, a QP that c-commanded the pronoun (the VB antecedent), and a name which did not c-command the pronoun (the CR antecedent). There was no evidence from either experiment that VB antecedents are considered first. Experiment 4 demonstrated that a CR was considered, but that the VB antecedent appeared not to be considered at all during processing. Experiment 5 showed that both VB and CR antecedents were affected by a context manipulation, demonstrating that they were both considered during processing, and the effect of the context emerged in an earlier region for the CR antecedent compared to the VB antecedent. Thus both experiments provide evidence that is incompatible with the POB hypothesis. However, in both experiments the CR antecedent (the name)

was the most recent antecedent and the VB antecedent (the QP) was more distant. The position of the antecedents does not undermine the evidence against the POB hypothesis because antecedent position should not affect the search for a VB antecedent in this model. Nevertheless the linear order confound in the materials does raise the question of whether it was properties of the antecedents themselves or the linear order of the antecedents that gave rise to the effects in Experiments 4 and 5.

In Experiment 9 the linear order of the antecedents was reversed, without altering the route (VB or CR) by which the antecedents were linked to the pronoun. The QP was still in c-commanding position, but was closer to the antecedent than the name, which did not c-command the antecedent. At this point it should be noted that the structural relationship between the name and the pronoun in Experiment 9 (and in the name first conditions in Experiments 7 and 8) is questionable. Consider (4.15).

(4.15a) *QP first*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that he should wave as the parade passed.

(4.15b) *Name first*

The squadron paraded through town. It looked to James that every soldier was completely convinced that he should wave as the parade passed.

In (4.15a) *James* is inside a relative clause which modifies *every soldier*, making c-command of the pronoun from such a position impossible. In (4.15b) it is not so clear that c-command is impossible. It can be argued that c-command is, in fact, possible here.¹² One argument in favour of this perspective is that quantifiers can bind pronouns from such a position

¹² I am grateful to an anonymous reviewer of *Cunnings et al.* (accepted) for pointing out this possibility and providing the example in (16).

(4.16a) and pronouns in the same position violate condition C of Binding Theory (see Chapter 2 section 2.2) if they bind names lower in the structure (4.16b).

(4.16a) It seems to everyone that his mother is angry.

(4.16b) It seems to him_{*i/j} that John_i is angry.

This argument is somewhat problematic because of the possibility of a quantifier semantically binding a pronoun in the absence of c-command (see, for example Büring 2005 for a discussion of indirect binding and e-type pronouns). The availability of binding in (4.16a) and its prohibition in (4.16b) therefore does not entail a c-command relationship. Nevertheless, it should be acknowledged that it is not clear whether the route used to link *James* and *he* in (4.15b) is coreference or variable binding.

Returning to the comparison between Experiments 4 and 9, while POB predictions can truly be assessed in Experiment 4, it is rather more difficult to assess the POB predictions in Experiment 9 because of the uncertainty about the relationship between the name antecedent and the pronoun. But having found that the POB predictions were not borne out in Experiments 4 and 5, it was possible to separate a linear order preference from an antecedent-type preference in Experiment 9, which gives further insight into the results of Experiments 4 and 5.

There was clearly a preference for the QP antecedent during processing in Experiment 9, in comparison to the name preference in Experiment 4 and the earlier consideration of the name antecedent in Experiment 5. This shows that the effect in Experiments 4 and 5 is likely to have been driven, at least in part, by the position of the antecedents.

The offline experiments presented in this chapter reveal a preference not only for more recent antecedents over more distant ones, but also for DPs and names over QPs. This means that the advantage for the name in Experiments 4 and 5 might also have been boosted by the more favoured

antecedent type appearing in a more recent position. Indeed, this double advantage (position and type) for the name may go some way to explain why there were no effects for the more distant antecedent in Experiment 4 while there was an effect for the name antecedent when it was more distant, in Experiment 9. In Experiment 4 the name had the advantage of being a favoured antecedent type and it was in a favoured position, whereas in Experiment 9 the name did not have a favoured position, but was still a favoured antecedent type.

The offline disadvantage for the QP fits well with findings of Carminati, Frazier and Rayner (2002) and Koornneef (2008), discussed in Chapter 3, who both found increased processing costs for QP antecedents compared to DP antecedents, and they also fit in with Burkhardt's (2005) cost hierarchy (also discussed in Chapter 3). However, if QPs require more processing time because of their internal complexity, it is difficult to understand why this should affect offline interpretations when participants have time to consider their choices. As pointed out in the previous chapter, it is difficult to see how complexity avoidance is helpful to the parser during pronoun interpretation. If the parser seeks simpler antecedents over more complex ones, it will presumably end up with undesired interpretations a lot of the time. A more plausible explanation for the offline dispreference for QPs is the uncertainty about their semantic number and therefore their number match with the pronoun. In English, it is possible to refer to QPs such as *every boy* with the plural pronoun *they*, reflecting a group reading for the QP. A name such as *James* or a DP such as *the soldier*, on the other hand, is unambiguously singular. The uncertainty about the match between the singular pronoun *he* and the QP might have led to the name/DP preference in the offline tasks. It may also be a factor during processing, contributing to the name preference in the online experiments. If it is a factor it clearly does not completely rule out the QP from consideration, given that the QP was considered when it was closest to the pronoun (Experiment 9) and also when it was distant (Experiment 5).

Altogether the results of the offline and online experiments in this chapter, combined with those in the previous chapter give a very clear picture of a

recency preference during online processing and also in offline interpretation. There is also a general preference for name/DP antecedents over QP antecedents which may be driven by processing complexity online and uncertainty about semantic number offline and online.

4.10.4 A memory-driven recency effect?

Given the clear preference for a recent antecedent in the experiments reported in this and the previous chapter, the reason for such a preference should be considered. In the current chapter previous studies on recency in pronoun resolution were presented. These studies show an increased processing cost for antecedents that are more distant from the antecedent, although this might be moderated somewhat by a first-mention advantage, as Streb *et al.* (2004) suggest. In Clark and Sengul's (1979) study the pronoun was easier to process when the antecedent was presented one sentence back, but there was no significant difference between presentation of the antecedent two and three sentences away, suggesting a privilege for the most recent clause or sentence. All the previous studies presented in section 4.4 contained non-matching antecedents that intervened between the pronoun and the matching antecedent, at the same time as manipulating the distance, so it is not possible to determine whether distance on its own caused the effects. Nonetheless it is logical to assume that an antecedent that is further away from the pronoun increases the processing cost on the pronoun, all other things being equal. This is compatible to some extent with retrieval models such as Lewis and Vasishth (2005). While the direct access to items in memory via feature-match means that the order in which items appeared does not have an effect on their retrieval time, the activation levels of items in memory decays over time if they are not reactivated, and this does play a role in retrieval.¹³ This all points to recency not being a cue in pronoun processing, but recency effects manifesting as a by-product of the memory/retrieval system. For example it is possible that the memory representation of more

¹³ For a discussion of decay and interference effects (and experimental evidence) see Lewis and Vasishth (2005).

distant items is impaired by the encoding of newer items, as suggested in ‘feature-overwriting’ (e.g. Oberauer and Kliegl 2006) in which case the more distant antecedent is less likely to be retrieved.¹⁴

The evidence from my own experiments is rather different. Rather than simply suggesting a retrieval cost for the non-local antecedent, the results suggest a preference for the recent antecedent. In particular, the fact that the recent antecedent was favoured in the offline choices, as well as the similarity between the offline and online results, shows that participants are more likely to settle on a recent antecedent than a more distant one. A recency preference for the offline choices is difficult to explain by memory decay alone.

I suggest instead that the recency preference seen in this experiment is, at least in part, due to discourse considerations. The two structures may promote the recent antecedent as being more important in the upcoming discourse. For example, consider again the QP first structure, repeated here as (4.17a):

(4.17a) *QP first*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that he should wave as the parade passed.

James is within a relative clause that modifies *every soldier*. The importance of *James* is such that he divides the set of soldiers into two groups: those who knew that *James* was watching, and presumably, those who did not know that *James* was watching. It can therefore be assumed that the

¹⁴ I am grateful to an anonymous reviewer of *Cunnings et al.* (accepted) for suggesting feature-overwriting as a component in recency effects.

character *James* will play a further role in this story, and that can be achieved by assuming that *James* is the referent of the pronoun.

In the name first structure, repeated as (4.17b) below, the scenario is rather different.

(4.17b) *Name first*

The squadron paraded through town. It looked to James that every soldier was completely convinced that he should wave as the parade passed.

Here, we are given insight into the point of view of the first-mentioned character via the phrase *it looked to James that*. In James' view, a set of soldiers is involved in an event that is described. James is therefore distanced from the event being described, and the set of soldiers is more closely involved, leading to an expectation that the set of soldiers will be referred to again.

While it is not possible to assess whether the discourse structures of the two conditions promoted the more recent antecedent in both cases without further testing of the materials, a discourse-based explanation is in line with the fact that the recency preference was observed in both offline and online experiments. The discourse factors, of course, could have an effect alongside effects of memory.

4.10.5 Recency versus first-mention

A final consideration is how both memory-based and discourse-based recency effects can co-exist with a first-mention advantage, which seems to work in opposition to recency. First-mention effects appear in studies that have used materials such as the following (from Arnold *et al.* 2000):

(4.18)

Donald is bringing some mail to {Mickey/Minnie} while a violent storm is beginning.

{He's/She's} carrying an umbrella, and it looks like they're both going to need it.

The first-mentioned entity is a subject and has the thematic role of agent, which on their own are factors that increase discourse prominence. (While first-mention effects have also been found when the first-mentioned entity is not a subject, there are no studies, to my knowledge, in which the second-mentioned entity is also a subject or agent.) If there is a memory-based advantage for the most recent antecedent, this is clearly defeated by the expectation, hence increased activation, associated with first-mention and/or subjecthood/agenthood. However, in Experiments 4, 5, and 9, the more recent entity is a subject, and separate discourse structural factors may also give prominence to the more recent entity (see section 4.10.4 above). In this case the more recent entity has the advantage over the first-mentioned entity, even though the first-mentioned entity is a subject. Together, these processing preferences suggest that discourse prominence associated with first-mention is defeated by intervening antecedents that are equally prominent. First-mention is only beneficial to an antecedent as long as there is no competition from intervening antecedents. A recent antecedent that has a grammatical role associated with lower prominence may intervene, as in (4.18), without upsetting the first-mention preference. But an intervening antecedent that is a subject, for example, defeats the first-mention preference.

This suggests that not only is the pronoun resolution process very sensitive to discourse prominence, but it is also very locally concerned, in that previously prominent entities are overlooked if a more recent prominent entity is introduced. Recency preferences, then, seem to come about through an interaction of discourse and memory effects.

Nevertheless, there does appear to be a lingering advantage for a first-mentioned antecedent when the most recent antecedent proves to be

unsuitable on other grounds, such as feature match. This is demonstrated by Streb *et al.*'s (2004) ERP results. Recall that they found the distant condition was less costly than the medium distance condition:

(4.19a) *Far distance condition*

Context 1: **Lisa** strolls across a bazaar.

Context 2: Peter sells gems to tourists.

Context 3: The gems are cut excellently.

Target: Then **she** will buy a diamond from the trader.

(4.19b) *Medium distance condition*

Context 1: The weather is beautiful today.

Context 2: **Gerhart** is an experienced mountaineer.

Context 3: Anna wants to go on a walking tour.

Target: Then **he** shows the ascent to the tourist.

Both of these conditions have in common that the most recent antecedent, which is a subject (context 3) mismatches in gender/number features with the pronoun. In this case, *Lisa* in (4.19a) is identified more easily than *Gerhart* in (4.19b) as the suitable antecedent. This shows that, despite the intervening antecedents, there is a particular advantage to being first-mentioned that is not completely erased and can be helpful when the most recent antecedent is unsuitable.

Overall, the advantage that an entity obtains (in terms of expectancy/activation) from being first-mentioned cannot be discounted. The first-mention advantage is, however, severely compromised by an intervening antecedent that can compete in terms of discourse prominence. In sum, a recency versus a first-mention cue in pronoun processing can be better understood as the interaction of discourse-based expectancies and memory considerations; expectancies are particularly affected by grammatical or thematic role information which can promote a more recent antecedent over a first-mentioned one.

4.11 Summary and conclusions

In this chapter I considered the factor of recency in pronoun resolution. I considered previous literature on both first-mention and recency because of the emphasis on the first-mention advantage in much pronoun processing literature. I reported the results from one eye-tracking experiment and three questionnaire studies that follow up Experiment 4 from Chapter 3, exploring recency as a factor in the processing of pronouns. I demonstrated the existence of a recency preference during processing. This preference comes about not through memory factors alone but by an interaction of memory and discourse prominence factors. I propose that previously prominent entities (such as first-mentioned entities) are overlooked if a more recent prominent entity is introduced, but a first-mention advantage does linger and eases the recovery from a non-matching recent antecedent.

CHAPTER 5

Looking beyond the sentence: the role of prior discourse in pronoun resolution

5.1 Introduction

In Chapters 2 and 3 of this thesis, I was mainly concerned with constraints and cues on pronoun resolution within the sentence. In Chapter 2 I investigated the processing of pronouns when they are subject to the application of condition B of the Binding Theory (Chomsky 1981; 1986), and also pronouns that are exempt from condition B. Condition B only excludes potential antecedents for the pronoun which appear within the binding domain of the pronoun (or are co-arguments of the same predicate), so it is unconcerned with what happens outside the sentence. Similarly, in Chapter 3 the *Primitives of Binding* (POB) hypothesis (Koornneef 2008; see also Reuland 2001; 2011) was investigated, which proposes that variable binding (VB) antecedents are considered before coreference (CR) antecedents. Variable binding relies on a c-command relationship, which only applies within a sentence. Even at the level of the sentence, however, the influence of discourse prominence cannot be ignored. It was proposed in Chapter 4 that a recency preference is attributable to an interaction of memory and discourse factors which can promote one antecedent over another even within the sentence. It is almost impossible to study pronouns and not be concerned with discourse issues, and in the current chapter I examine one aspect of discourse prominence in more detail. I do this by looking again at the data from Experiment 5 that was originally reported in Chapter 3. In Experiment 5 two factors were manipulated: the pronoun's match with a VB or a CR antecedent, and the context provided by the lead-in sentence. The VB/CR manipulation has

already been discussed in Chapter 3, and in the current chapter I discuss the context manipulation.⁵²

5.1.1 Discourse prominence

There is widespread agreement that pronouns tend to refer to discourse prominent entities (e.g. Garnham 2001; Givón 1983; Grosz, Joshi and Weinstein 1995; Cowles, Walenski and Kluender 2007). This is often linked with the idea of accessibility and the accessibility hierarchy (Ariel 1990). The idea behind the accessibility hierarchy is that anaphoric forms are markers of discourse accessibility, so that the more informative the anaphoric form, the lower the discourse accessibility of its referent. Third-person pronouns, being in themselves quite uninformative (in comparison to names or definite descriptions, for instance) should have highly accessible referents. Speakers choose less informative forms to refer to highly accessible referents because they assume that the relevant information is already prominent. Discourse prominence has been widely discussed and is referred to by many different names, for example saliency, accessibility and topichood. All these labels refer to roughly the same notion: information that is assumed to be important and relevant in the current part of a text or discourse. With respect to pronouns and antecedents in particular, discourse prominent entities are those which are playing an important role in the current event or topic being described. For example, in both (5.1) and (5.2) the most likely referent for the **bold** pronoun is *Susan*.

- (5.1) Susan loved writing letters, in fact, she loved writing them so much that she tried to make time to write at least one letter every day of the week. **She** would often sit at the desk writing for several hours at a time.

⁵² Parts of this chapter have been published as Patterson (2013), which presents a reanalysis of the data reported in Patterson (2012).

(5.2) Susan loved writing letters, in fact, she loved writing them so much that she tried to make time to write at least one letter every day of the week. Last week she even wrote three letters to her friend Betty. **She** would often sit at the desk writing for several hours at a time.

Susan is the only referent introduced in the discourse in (5.1), while in (5.2) there is an additional potential referent in the discourse (*Betty*), but it is still assumed that *Susan* is the probable referent of the pronoun in the final sentence because the discourse seems to be about her. However, it is difficult to pinpoint precisely what makes *Susan* discourse prominent. Of course there are many possibilities: *Susan* is the subject of the first sentence, the first-mentioned entity, and she is referred to more than once. As discussed in the previous chapter, the structure-building approach (Gernsbacher and Hargreaves 1988) emphasises the prominence in memory of the first-mentioned entity. The expectancy hypothesis (Arnold 2001; Arnold, Brown-Schmidt and Trueswell 2007), too, places importance on first-mentioned entities but acknowledges the contribution of various other factors such as grammatical role and recency in increasing expectations of repeated reference. Subjecthood is an important factor in Centering Theory (CT) (Grosz *et al.* 1995) and may be separable to some degree from a first-mention preference (Järvikivi *et al.* 2005; cf Kaiser and Trueswell 2008). Elsewhere, the importance of a parallel function between pronoun and antecedent (i.e. whether or not they have the same grammatical role) has been claimed to work alongside a subjecthood preference (e.g. Stevenson, Nelson and Stenning 1995). Despite the identification of a number of such cues involved in pronoun processing, it has proved extremely difficult to bring these together to form a definitive definition for discourse prominence. Attempting to define such a notion can end up being rather circular: pronouns tend to refer to discourse prominent entities, so if an entity is the referent of a pronoun, that entity must be discourse prominent.

The particular aspect of discourse prominence that I focus on in this chapter is how *prior* discourse prominence can affect the current accessibility of potential antecedents. This was tested in Experiment 5,

which was reported in Chapter 3. In Experiment 5 two factors were manipulated: *antecedent type* (QP or DP, where the QP was a VB antecedent and the DP was a CR antecedent) and *context*. It is the context manipulation that is the main focus of this chapter, the antecedent type manipulation having already been discussed (see Chapter 3). Given the breadth of research in the general area of discourse factors and pronoun resolution, it is of course impossible to provide a comprehensive review of all aspects of this field. The notions that are directly relevant to Experiment 5 are *coherence* and *topichood*, so I begin by discussing relevant prior work on the influence of these factors in pronoun processing, looking in particular at the Centering Theory approach. I then present a summary of the results from Experiment 5, and discuss their implications regarding the effect that prior discourse prominence has in online pronoun resolution.

5.2 Topichood in pronoun resolution

Much prior work on discourse aspects of pronoun resolution has attempted to identify individual factors which increase the accessibility of certain antecedents such as: subjecthood (Crawley, Stevenson and Kleinman 1990); first-mention (Gernsbacher and Hargreaves 1988; Arnold *et al.*, 2007; Arnold 2001); implicit causality (Stewart, Pickering and Sanford 2000; Koornneef and Van Berkum 2006; Cozijn *et al.* 2011); parallelism (Stevenson *et al.* 1995); and clefting (Foraker and McElree 2007), among others. Some of these factors are useful for pronoun resolution because biases from verbs and connectors bring a certain entity into focus, as is the case for implicit causality (see also the coherence analysis account presented below). Other factors succeed in increasing the prominence of particular entities because they are (often) associated with topic status, as is the case for subjecthood.

While the discourse prominence of topics is not disputed, it has been difficult to arrive at a precise definition for the notion of topic; nevertheless it is particularly of interest for pronouns because topichood can operate on a sentence level and also on the level of the wider discourse. In terms of the sentence, topics can be thought of as “a sentence-initial constituent which is predicated about” (Schwabe and Winkler 2007, p. 2). Cowles *et al.* (2007)

suggest that this aboutness (see also Reinhart 1981) can also be applied to a longer discourse to define the discourse topic:

“The notion of discourse topic...can be defined in an analogous fashion to that of sentence topic, only at the level of sets of propositions rather than single propositions. Applying the aboutness condition to discourse topics, the discourse topic is the referent that a set of propositions is about.”

(Cowles *et al.* 2007, p. 5-6)

Cowles *et al.* (2007) suggest that both sentence topics and discourse topics are discourse prominent and affect online pronoun processing. They tested this in a cross-modal priming study, comparing the effects of sentence topic and discourse topic with the effects of contrastive focus on the resolution of pronouns. In their experiment they constructed three-sentence discourses, with the final sentence containing the pronoun of interest. I will focus on the sentence and discourse topic materials, and leave aside the question of clefted focus. Examples of the topic lead-in sentences are shown in (5.3) and (5.4) below, and (5.5) shows the target sentence. First-mention and second-mention were also manipulated in this experiment, as shown in the (a) and (b) versions of the lead-in sentences.

(5.3a) *Discourse topic: first-mention*

Anne wanted to see the new movie with Sarah.
So, Anne called Sarah.

(5.3b) *Discourse topic: second-mention*

Anne wanted to see the new movie with Sarah.
When Sarah came home, Anne called.

(5.4a) *Sentence topic: first-mention*

A new movie opened in town.
So, Anne called Sarah.

(5.4b) *Sentence topic: second-mention*

A new movie opened in town.

When Sarah came home, Anne called.

(5.5) *Target sentence*

But later that night, she couldn't go to the movie after all.

The discourse topic, for the purpose of the experiment, was defined as “a referent that has been a sentential topic for more than one sentence without any intervening sentence topics” (Cowles *et al.* 2007, p. 6). This means that in the discourse topic conditions, *Anne* (the matrix subject in both the first-mention and second-mention conditions) is the discourse topic. In the sentence topic conditions, neither referent is mentioned in the first sentence, but *Anne* is the topic of the second sentence. Before the cross-modal priming experiment was carried out, the experimental items were pretested by asking participants to decide who the pronoun referred to, and to rate on a scale how well the discourses fitted together as a whole. The discourses in the three conditions (including the focus condition not given here) were rated equally well for goodness of fit, but the sentence topic second-mention was rated worse than the sentence topic first-mention. The sentence/discourse topics were chosen more often than the non-topics, and at a rate of more than 50%, although the proportion of choices was modulated by the first-/second-mention manipulation with lower proportions for the second-mention topics compared to the first-mention topics. In the online cross-modal priming experiment, participants listened to the sentences and then responded to a visually presented word which was shown at the offset of the pronoun in the target sentence, by reading the word aloud. The visually presented word was either the topic (*Anne* in the above examples) or the non-topic (*Sarah*). Topics, if they have higher activation than non-topics as a result of their discourse prominence, should elicit faster responses (in this case naming latencies) to the visual presentation. The results confirmed this prediction for both sentence topics and discourse topics. Naming latencies were shorter for both the sentence topic and the discourse topic compared to the non-topic. There was no effect of first-/second-mention in the naming latencies.

Cowles *et al.*'s (2007) results suggest that topics, whether sentence topics or discourse topics, have higher activation (or are easier to retrieve, which is likely to be linked to activation) than other antecedents that are non-topics. One aspect of topichood that was not explored in the study was whether or not the activation of the topics was relevant prior to the pronoun. If discourse prominence increases a referent's activation and as such can be used predictively, as suggested by Arnold's expectancy hypothesis (Arnold 2001; Arnold *et al.* 2007), then the difference in activation should be visible at any point that reactivation is required. Since a cross-modal task measuring naming latencies necessarily involves reactivation, the same result might be found prior to the pronoun region. Even so, the results here are interesting because they seem to show that the retrieval involved in pronoun resolution is sensitive to prominence not just within the prior sentence but also within the preceding discourse. But one crucial question that remains unresolved from this study is how useful the preceding discourse actually is: note that the second sentence in both the discourse topic and the sentence topic conditions is the same. Given that both conditions elicited the expected result (i.e. more activation for the topic), it is not clear that the results were based on discourse topichood in one condition and sentence topichood in another, or if both were based on sentence topichood. As such, they do not provide evidence on the role of prior discourse prominence on current accessibility.

In order to account for the role of prior discourse in the resolution of pronouns, it may be necessary to have a more comprehensive view of how pronoun resolution relates to the unfolding discourse. Two frameworks have emerged that consider discourse coherence to be important in that respect: Kehler's coherence analysis, and Centering Theory. These approaches are outlined below.

5.3 Coherence in pronoun resolution

5.3.1 The coherence analysis

Work by Kehler on the involvement of discourse coherence in pronoun resolution (Kehler 2002; Rohde, Kehler and Elman 2006; Rohde, Kehler and Elman 2007; Kehler *et al.* 2008) seeks to move away from seeing individual discourse factors such as subjecthood as processing strategies or preferences for pronoun resolution. This is partly because it is unhelpful to assume that the parser must deal with a long list of potentially conflicting cues at the point of encountering a pronoun. Instead, the semantic contribution of the surrounding discourse is emphasised. Kehler *et al.* (2008) argue that the subject preference proposed by Crawley *et al.* (1990), for example, is not a mechanistic parsing preference but is rather the result of particular types of relationships that exist between sentences or clauses in a discourse. These relationships are called *coherence* relationships. For example, in both (5.6a) and (5.6b) (taken from Kehler *et al.* 2008) the second clause provides an explanation for the first:

(5.6a) The city council denied the demonstrators a permit because they feared violence.

(5.6b) The city council denied the demonstrators a permit because they advocated violence.

In (5.6) the 'Explanation' coherence relationship exists between the first and second clause, but *they* in (5.6a) is likely to be understood as referring to the city council whereas *they* in (5.6b) is understood as referring to the demonstrators. Establishing an Explanation relation between the two clauses involves using both semantic and world knowledge to assess how the first clause plausibly follows from the second clause. Establishing this coherence relationship in (5.6a), for example, involves knowing that a permit may enable demonstrators to cause violence, so denying a permit

plausibly follows if the city council fear such violence.⁵³ Resolving the pronoun is then “a by-product of establishing an Explanation relation” (Kehler *et al.* 2008, p. 5). Viewed in this way, pronoun resolution need not have its own preferences and heuristics and falls out of the process of understanding the discourse and, crucially, understanding the relationships between clauses and sentences. Kehler puts forward a number of discourse coherence relations, such as Explanation, Occasion, Parallel, and Result. While functions well for explaining the production of coreference, where the speaker already knows which coherence relations will exist between clauses, it is not sufficient to explain how hearers begin to process pronouns online when they may not already know how one clause or sentence will relate to the other. Kehler *et al.* (2008) therefore extend the coherence approach to make it more relevant for online processing. In doing so they follow a similar line of argument as in Arnold’s expectancy hypothesis. They propose that comprehenders build up expectations about which entities will be mentioned again based on (i) the likelihood of a particular coherence relation and (ii) the likelihood of a particular referent (e.g. grammatical subject) for a pronoun within that coherence relation. Put together with general expectancy notions (the idea that, for example, first-mentioned entities, subjects or topics are likely to be rementioned in a discourse), the coherence approach does away with cues and preferences and recasts them in terms of expectations which can build up before the pronoun is encountered. However, while the coherence analysis takes into account expectations built on the prior discourse, it does not examine specifically whether discourse topichood or prior discourse prominence increases the accessibility of entities later in the discourse.

⁵³ For a more detailed explanation involving the binding of semantic variables, see Kehler *et al.* (2008).

5.3.2 Centering Theory: an outline

The notions of coherence and accessibility or prominence are also central to Centering Theory (CT) (Grosz *et al.* 1995).⁵⁴ This is a framework for tracking prominence as a discourse unfolds. It was originally developed in the field of computational linguistics, but its ideas have proved to be useful in psycholinguistics as a testable mechanism for pronoun resolution preferences based on discourse prominence. CT makes use of the factors that are known to be important in discourse prominence, in particular subjecthood and first-mention, to provide a ranking of the discourse entities at different points throughout a discourse. Previous discourse status has also been suggested as a factor that could contribute to discourse prominence (Gordon, Grosz and Gilliom 1993). The basic mechanisms of CT are outlined below.

CT assumes that discourses can be split into a series of utterances, with prominence rankings defined within each utterance. Prominence rankings are assumed to relate to the accessibility of a particular entity within a discourse model, so entities that are highly ranked in terms of prominence are more accessible within the discourse model than those with lower rankings. One disadvantage of CT at present is that there is no consensus about what constitutes an utterance, although this is normally assumed to be a sentence or a clause (see Kameyama 1998 and Poesio *et al.* 2004 for a discussion of the definition of an utterance). Two key concepts in CT are backward-looking centers (Cbs) and forward-looking centers (Cfs). These centers (entities) make links to previous and upcoming utterances, creating coherence. It is normally assumed that there is one Cb per utterance (and no Cb in the first utterance of a discourse). The Cb is the entity which links back to the previous utterance, usually by appearing in some form in both the current and the previous utterance. The first utterance of a discourse cannot contain a Cb because there is no previous utterance to link back to. The set of Cfs are the entities that are ranked in terms of prominence, which

⁵⁴ Grosz *et al.* (1995) is the first comprehensive outline of CT, but it was developed much earlier and parts of it were presented in earlier publications.

will then (to some extent) determine their importance in the upcoming discourse. An example (derived from Gordon *et al.* 1993) of how this plays out as a discourse unfolds is given below.

In (5.7) there are three entities that can be ranked in terms of their prominence. The entities are listed below in the example along with the elements that contribute to their ranking:

(5.7)

Utterance 1: Susan gave Betsy a pet hamster.

Cf ranking:

SUSAN=subject, first-mention

BETSY=indirect object, second-mention

HAMSTER=direct object, third-mention

(5.7) shows the first utterance of a discourse and therefore there is no Cb. There is a set of three Cfs: *Susan*, *Betsy* and *Hamster*, entities introduced or mentioned in the utterance. *Susan* is both the grammatical subject and the first-mentioned entity, and both of these factors boost the prominence ranking of *Susan*. *Susan* is therefore the top-ranked, most prominent, entity in the first utterance.

(5.8)

Utterance 1: Susan gave Betsy a pet hamster.

Utterance 2: She reminded her that such hamsters were quite shy.

Cb=SUSAN

Cf ranking:

SUSAN=subject, first-mentioned

BETSY=indirect object, second mention

HAMSTER=direct object, third mention

(5.8) shows the Cb, and also the Cf prominence rankings for Utterance 2. *Susan* is the Cb. This status is given in Utterance 2 because *Susan* is the most highly ranked entity that appears in both Utterance 1 and Utterance 2. Similarly, CT assumes that if any entity is referred to using a pronoun, it

should be the top-ranked entity from the previous utterance. The top-ranked entity from the previous utterance (*Susan*) is assumed to be the referent of the pronoun *she*.

One further aspect of CT is that it defines the way that one utterance links to another via the relative status of the Cb. These links are defined by Grosz *et al.* (1995) as *continue*, *retain* and *shift*, although several other transition types have been suggested (see Poesio *et al.* 2004 for a discussion). I focus here on *continue* and *shift* as they are more relevant to Experiment 5, and also because the behavioural predictions for the *retain* transition have not been tested to my knowledge. Put simply, in a *continue* transition between two utterances the Cb of the first utterance is the same as the Cb of the second utterance. In a *shift* transition, the Cb of the first utterance is not the same as the Cb of the second utterance. The different transitions have consequences for processing. It is assumed that, all else being equal, a *shift* is more costly to process than a *continue*. This is because a *shift* entails a change of assumed topic.

5.3.3 The testing of CT assumptions in relation to pronoun processing

The processing consequences for certain claims of CT have been tested in psycholinguistic experiments. Perhaps the most well known study is Gordon *et al.* (1993), who uncovered the repeated-name penalty (see below). They present a series of five experiments that test various aspects of CT: experiment 1 tests the realisation of a Cb as a pronoun versus a name, testing the CT assumption that the top-ranked entity of a previous utterance should preferentially be realised as a pronoun (if anything in the current utterance is realised as a pronoun); experiments 2 and 3 used the repeated-name penalty as a metric for determining more precisely how a Cb could be defined, testing in particular whether surface-initial position is as important as subjecthood. Experiments 4 and 5 use the transitions *continue* and *shift* to further narrow down the contribution of prior discourse status to the repeated-name penalty. I will focus on experiments 1 and 4 here.

Experiment 1 used four-sentence discourses which introduced two characters and manipulated whether subsequent reference to the characters was achieved using pronouns or names. One character was introduced in the first sentence and was always the subject of the following sentences. The second character was introduced in object position in the second sentence. The materials were as follows in (5.9); I have highlighted the introduction of each character in bold.

(5.9)

1. **Bruno** was the bully of the neighbourhood
2. {He/Bruno} chased **Tommy** all the way home from school one day.
3. {He/Bruno} watched {him/Tommy} hide behind a big tree and start to cry.
4. {He/Bruno} yelled at {him/Tommy} so loudly that all the neighbours came outside.

The pronominalisation of subsequent mentions of both characters was manipulated (as shown by the material in the curly brackets) to create three conditions. In the name-name condition both characters were referred to using names in subsequent mentions. In the pro-pro condition, both characters were realised as pronouns in all subsequent mentions. In the pro-name condition, the first character (*Bruno* in the example above) was realised as a pronoun in subsequent mentions and the second character (*Tommy* in the example above) was realised as a name in subsequent mentions. The sentences were presented as a phrase-by-phrase self-paced reading task, and reading times were measured for each sentence 1-4. Reading times were significantly higher when all subsequent mentions of both characters were names (the name-name condition), compared to the other two conditions. There was no difference between the pro-pro condition and the pro-name condition. Even though the reading-time was measured at the end of the sentence rather than on the pronoun, the similarity in reading times between the pro-pro and pro-name conditions, and the reading-time cost for the name-name condition, show two things. Firstly, there is a cost associated with realising previously-mentioned

entities as names instead of pronouns, and secondly, the cost is restricted to more prominent entities (i.e. *Bruno* in the example above). This confirmed the CT assumption that the top-ranked entity of a previous utterance should preferentially be realised as a pronoun. Gordon *et al.* (1993) claim that the repeated-name penalty is applicable to all definite descriptions, not just names. It should be noted, however, that most experimental work in this area has focussed on names (e.g. Gordon *et al.* 1993; Gordon and Scearce 1995; Rose 2007; Fukumara and Van Gompel 2010; Gelormini-Lezama and Almor 2011). This leaves open the question of whether the repeated-name penalty would be seen if definite descriptions were repeated. Secondly, Gordon and Hendrick (1998) claim that in complex constructions, the repeated-name penalty is reduced. This is certainly a logical expectation. If the reader is processing a complex discourse, processing resources for maintaining the topic of an utterance will be more limited. Therefore, reminding the reader/hearer more explicitly about the topic of an utterance by repeating a name might be more helpful than disruptive in these cases.

Experiment 4 in Gordon *et al.* (1993) explored the influence of prior discourse prominence on the repeated-name penalty by using continue and shift transitions. Recall that shift transitions are assumed to be more costly to process than continue transitions. Pairs of sentences were presented as follows:

(5.10) *Initial sentence pair*

George jumped out from behind a tree and frightened Debbie.
He was surprised at her hysterical reaction.

(5.11) *Continue condition*

{He/George} never thinks about how others might feel.
Practical jokes are not always fun for everyone.

(5.12) *Shift condition*

{She/Debbie} screamed loudly and ran away.
Practical jokes are not always fun for everyone.

The reading times were measured at the end of the critical sentence and the end of the final sentence. There was no main effect of the continue/shift manipulation, but there was an interaction between continue/shift and pronominalisation in reading times for the critical sentence. In the continue condition, there were longer reading times when the name was presented compared to the pronoun, i.e. a repeated-name penalty. There was no such penalty in the shift condition. In addition, the critical sentence was read numerically more slowly in the shift condition with a pronoun compared to the continue condition with a pronoun, but the difference did not reach significance. In the reading times for the final sentence, however, there was an effect of the continue/shift manipulation, with slower reading times for the shift condition. In all, this experiment shows again that the repeated-name penalty is restricted to previously prominent discourse entities, and also that there is a cost associated with shift compared to continue transitions, although this was only significant in the final sentence.

Further testing of the continue and shift transitions has been carried out by Gordon and Scearce (1995). They report two self-paced reading experiments which investigated how the local discourse structure affects pronoun resolution, in particular how a default assumption for the likely referent of a pronoun is set up by the preceding discourse and can then be overridden by what they term knowledge-based processes (which is similar to an assessment of plausibility, based on information from the verb). They manipulated the discourse biases by exploiting the continue and shift transitions and the processing cost involved in a repeated-name penalty. They presented short texts composed of four sentences each. The first two sentences and the final sentence were the same in all conditions, as in (5.13) and (5.15), and the transition and repeated name was manipulated to create four conditions, shown in (5.14). The symbol | represents the division of the sentences into frames for presentation, which I have labelled (i) and (ii) for ease of reference.

(5.13) *Introductory sentences*

Bill wanted John to look over | some important papers.

He had to mail him | the documents by Monday.

(5.14a) *Continue, pronoun*

(i) Unfortunately, he | (ii) never sent the papers.

(5.14b) *Continue, name*

(i) Unfortunately, Bill | (ii) never sent the papers.

(5.14c) *Shift, pronoun*

(i) Unfortunately, he | (ii) never received the papers.

(5.14d) *Shift, name*

(i) Unfortunately, John | (ii) never received the papers.

(5.15) *Final sentence*

As a result, the whole deal | fell behind schedule.

Having established a discourse topic in the first two sentences via subjecthood and first-mention, the next sentence either continues with the same topic (a continue transition) or changes the topic (a shift transition). The authors do not give a set of predictions for experiment 1 in their paper, so the predictions I have derived below are based on the assumptions that they intended to test, i.e. differences between shift and continue transitions as predicted by standard approaches to CT, the repeated-name penalty for discourse prominent entities, and the influence of knowledge-based processes (semantic information from the verb). My derived predictions are as follows: the repeated name in the continue condition should bring about a processing cost in comparison to a pronoun, because of the repeated-name penalty. The repeated name in the shift condition should not result in a penalty because the new discourse topic (*John* in the example) was not most prominent in the previous Cf. However, when reading part (i) these effects should be modulated by the unavailability of the information from the verb, which only appears in (ii). Participants should process the name/pronoun based only on the discourse biases set up in (5.13) and not the plausibility information available in (5.14*ii*). In (5.14*i*), participants should assume that the pronoun *he* refers to *Bill* in both the continue and shift conditions (5.14*ai* and 5.14*ci*). In the continue condition the presentation of the name should result in a repeated-name penalty for (5.14*bi*). In the shift condition (5.14*di*) the presentation of the

name *John* represents a shift, which should also result in a processing cost in comparison to the pronoun. Thus, (5.14bi) and (5.14di) will incur a cost compared to (5.14ai) and (5.14ci). When the second half of the critical sentence is presented, there should be no disruption in either (5.14aii) and (5.14bii) because the verb plausibility information suggests that *Bill* (the assumed discourse topic) is the subject of *sent*. In (5.14cii) and (5.14dii), the plausibility information suggests that *John* is the subject of *received*. In (5.14dii) this may not increase the reading time because the name *John* has already been processed and a topic shift has taken place. But in (5.14cii) the information from the verb clashes with the assumption that the pronoun *he* referred to *Bill*, triggering a re-analysis thus increasing the processing time required.

The results reflected the predictions I derived from CT assumptions, with one exception. Correctly predicted were reading times for (5.14i), which were longer for names than for pronouns (5.14bi and 5.14di longer than 5.14ai and 5.14ci). In (5.14ii) there was an interaction between transition type and expression type (pronoun or name), such that (5.14dii) had shorter reading times than (5.14cii), as predicted. But in the continue conditions, (5.14aii) had shorter reading times than (5.14bii), which is not predicted from the assumed interaction of discourse bias and plausibility information. It appears that the repeated name penalty effect found in (5.14bi) had not disappeared when the next region was processed.

In Gordon and Scearce's second experiment, the verb is presented before the name/pronoun, in order to examine whether the semantic/plausibility information completely overrides discourse structure assumptions. This is an interesting question in itself but is not of direct relevance to the current discussion. I focus instead on their first experiment because I disagree with the authors' interpretation of the results and I suggest that their findings actually confirm the cost associated with a shift transition, which is relevant to my own experiment (Experiment 5) discussed below. Gordon and Scearce suggest that their results show that shift transitions do not in fact have a cost compared to continue transitions. They also refer to the weak evidence regarding shift and continue transitions in Gordon *et al.*

(1993), and suggest that the CT predicted cost for shifts may be associated with larger transitions than sentence pairs. This is based on the fact that the reading times for (5.14a) and (5.14d) did not differ when the reading times across the two frames are combined. However, this is misleading. The cost for shift, I would argue, occurs at (5.14di), not across the two frames (5.14di) and (5.14dii). This is because *Bill* is the most prominent entity in the Cf of (5.13). Part *i* of (5.14a)-(5.14c) all represent a continue transition at this point, because the pronoun is assumed to refer to *Bill* in (5.14a) and (5.14c) and the name overtly refers to *Bill* in (5.14b). (5.14di), in referring overtly to *John*, makes *John* the most discourse prominent entity at that point, representing a shift from the previous sentence. This incurs a cost, as I set out in my predictions, and that cost was found in the results. The cost for the name in (5.14di) cannot be attributed to a repeated-name penalty because *John* is not the most discourse prominent entity in the previous sentence. In contrast to the authors, then, I suggest that there *is* a visible processing cost associated with shift transitions in comparison to continue transitions.

In summary, there is processing evidence for certain assumptions of the CT framework. The repeated-name penalty confirms that rementioning a prominent entity from the previous discourse is costly when a name is used, in comparison to a pronoun. However, there is no evidence so far that this applies to entities that have been introduced with definite descriptions, and there is a suggestion that the repeated-name penalty may not be so applicable in more complex discourses. There is also evidence for the CT assumption that shift transitions are more costly to process than continue transitions: this suggests that maintaining the same topic over several utterances leads to a more coherent discourse.

5.4 Experiment 5: CT and the influence of the prior discourse

The repeated-name penalty and the cost of shift transitions represent two ways in which CT captures the influence of discourse prominence on the processing of pronouns. However, these do not directly address the

question posed at the start of this chapter: how does prior discourse prominence affect the current accessibility of potential antecedents for a pronoun? Gordon *et al.* (1993) do make a suggestion on this issue, albeit untested, which can be captured in the CT framework. Recall that prominence ranking in the Cf is thought to be influenced by subjecthood and first-mention, although in principle other factors are not excluded from influence. Gordon *et al.* question whether Cb status may also influence the rankings in the Cf. A Cb in an utterance is likely to have been highly ranked in the previous utterance, because the Cb is the highest ranked entity from the previous utterance that is mentioned in both the previous and the current utterance. As such, an entity that has previously been prominent is more likely to play an important role in the upcoming discourse.

Experiment 5 tests the possibility that Cb status influences the prominence of entities in the current discourse. It investigates whether a discourse referent retains the increased activation from its previous discourse status in the text, making Cbs more accessible, all other things being equal, than entities which are not Cbs. This can also be described in more theory-neutral terms, by referring to topichood: are entities that were previously topics more accessible, all else being equal, than those entities that were not previously topics? An additional question is how the previous discourse status interacts with other preferences: for example, can a boost in prominence from Cb status overcome the recency preference observed in Experiments 4 and 9 (reported in Chapters 3 and 4)?⁵⁵

Experiment 5 is reported in this chapter as follows. I first describe the materials in full (also described in Chapter 3, section 3.6.1), followed by a

⁵⁵ The latter research question has been changed since this experiment was previously reported in the Patterson (2012) paper, because of the recency preference that was subsequently revealed in Experiments 4 and 9. Instead of investigating “pronoun resolution preferences with two potential subject antecedents” (Patterson 2012, p. 101), the question has been reframed as the recency preference interacting with cues from the prior discourse.

description of how they could be realised in a CT framework. I also note several problems with applying CT assumptions to the materials. The predictions are then outlined. The results for Experiment 5 are summarised rather than given in full, because full details can be found in Chapter 3 (sections 3.6.3 and 3.6.4).

5.4.1 Experiment 5 materials

The materials are adapted from the materials in Experiment 4 (see Chapter 3, section 3.5). They were comprised of 24 short scenarios, of three sentences each. The first sentence was the context sentence, the second was the critical sentence and the third was the wrap-up sentence, as shown in (16) below. The context sentence mentions one of the two potential antecedents: this was systematically varied. The critical sentence contained the pronoun and both potential antecedents. The first potential antecedent in the critical sentence was a QP (e.g. *every soldier*), and the second was a DP (e.g. *the queen*). All the nouns in the QPs were stereotypical role names. All the DPs were nouns with definitional gender. The overall set of QPs did not differ in length (letters) or frequency from the set of DPs, although it was not possible to match length and frequency in each individual item. To avoid biases, the texts were composed so that either antecedent could plausibly be the referent of the pronoun, but the gender cues determined the correct antecedent. (5.16) shows the four conditions for Experiment 5:

(5.16a) *DP antecedent, context match*

Context sentence: The queen felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed.

Wrap-up sentence: The entire town had come out to watch.

(5.16b) *DP antecedent, context mismatch*

Context sentence: The soldiers felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was

watching intently was absolutely convinced that she should wave as the parade passed.

Wrap-up sentence: The entire town had come out to watch.

(5.16c) *QP antecedent, context match*

Context sentence: The soldiers felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed.

Wrap-up sentence: The entire town had come out to watch.

(5.16d) *QP antecedent, context mismatch*

Context sentence: The queen felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed.

Wrap-up sentence: The entire town had come out to watch.

Two factors were systematically varied to create the four experimental conditions shown above. These factors were *antecedent* (DP or QP) and *context* (match or mismatch). *Antecedent* refers to whether the pronoun matches in (stereotypical) gender with the QP antecedent or the DP antecedent. *Context* refers to whether the entity introduced in the context sentence matches or mismatches with the pronoun. It is the context manipulation that is of primary interest in this chapter.

It has already been shown in Experiment 4 (and confirmed by Experiment 9) that, with a neutral context sentence, there is a preference for the most recent antecedent in the critical sentence, which is in this case the DP. Note, though, that the design of Experiment 5 differs from Experiment 4 in not presenting a match/mismatch paradigm (with a double match, two single match and a no match condition). In addition, Experiment 5 uses DPs instead of names, and of course another critical difference is the content of the context sentence which was neutral in Experiment 4. Possible

consequences of the design differences between Experiments 4 and 5 are considered in Chapter 3, section 3.6.1.

5.4.2 Applying Centering Theory assumptions to the materials for Experiment 5

5.4.2.1 QPs in Centering Theory

There are some challenges in applying CT assumptions directly to the materials for Experiment 5. The first challenge is the use of QPs such as *every soldier*. As discussed in Chapter 3 (see section 3.2 for a full discussion), a link between the pronoun and the QP cannot be established via coreference because the QP is thought to be a non-referring expression; a link can be established through variable binding because the QP c-commands the pronoun. A link between the DP and the pronoun can only be established via coreference, since the DP does not c-command the pronoun. One of the reasons why two different mechanisms for pronoun resolution have been proposed is that QPs, being non-referring expressions, cannot establish or refer to an individual in the discourse representation. In CT this makes it unclear what kind of representation of the QP appears in the ranking of Cf, since discourse entities are normally assumed here. The prominence hierarchy in CT models the state of the hearer's/reader's attention (Grosz *et al.* 1995). Whether we accept that the mental representation is separable from the semantic representation (e.g. mental models theory, Johnson-Laird 1983; as applied to anaphora, Garnham and Oakhill 1992; Garnham 2001), or whether mental representations are propositional in nature, there is general agreement that a QP such as *every soldier* does not invoke an individual discourse referent. What must be assumed, at least, is that a set of soldiers is invoked or introduced (or sets of sets according to traditional semantic notation) which may be represented along with its verbal argument (e.g. in a tripartite structure of operator, restrictor and nuclear scope). The question is whether such representations can be ranked as more or less salient in a set of Cfs in the same way as a normal discourse referent.

I believe that Experiments 4 and 9 (reported in Chapters 3 and 4) actually provide evidence that a QP antecedent can indeed be influenced by factors contributing to prominence in the same way as a DP antecedent. If a QP antecedent in the position of a variable binder is completely unaffected by discourse prominence considerations, then it is difficult to explain why there was no evidence that the QP antecedent was considered during processing in Experiment 4 but was considered in Experiment 9, and also why in offline tasks QP antecedent choices were affected by linear order (Experiments 6 and 8, reported in Chapter 4). As it stands, it appears that QPs *are* affected by discourse considerations which can make them more or less prominent. As such, an attempt to incorporate QP antecedents into CT is warranted, despite the fact that so far they have not been explicitly included. For the purpose of this experiment I will assume that a set is subject to certain prominence factors and can therefore be represented and ranked in the Cf. In so doing I do not claim that QPs are equivalent in all aspects to DPs in terms of their representation. They are clearly different syntactically, semantically and even in their information structural status. Instead I claim that, as far as QP antecedents can be influenced by recency, which may in turn be influenced by discourse and memory considerations, QPs can be considered to be subject to certain prominence constraints and thus should be included under a CT approach.

A related question is whether the DP forms part of the QP set because it is inside the relative clause, and therefore might not warrant a separate discourse entity in itself. Again, because of evidence that the DP in this position is prominent enough to lead to a recency preference, I will consider that the DP should be represented separately in the discourse as a normal entity.

5.4.2.2 Consideration of utterances in Centering Theory

The second challenge in applying CT to the current materials is that there is no consensus about what constitutes an utterance. As pointed out above, while much debate and empirical testing has addressed the notion of utterances as sentences versus utterances as clauses (Kameyama 1998), intervening relative clauses such as those in the current materials have not

been fully addressed (Poesio *et al.* 2004). I will therefore proceed as follows: the context sentence in the current experiment is a single clause and can be straightforwardly assigned to utterance 1 (U1). The second sentence is long and somewhat complex and will be split into two utterances (U2 and U3). I will keep the QP, the intervening relative clause and the main verb together in U2, and start U3 at the CP following the main verb. This splits a complex sentence into coherent parts and crucially keeps the QP together with the verb, while allowing the complement of the main verb (which is a finite clause) to be separated.

5.4.2.3 Centering Theory analysis of the materials

The CT analysis from which I derive my predictions is as follows (the ordering of the Cfs in the curly brackets represents the prominence hierarchy, with more prominent entities first):

(5.17a) DP antecedent, context match

U1: The queen felt really quite uneasy about the squadron parade.
No Cb. Cf={queen}

U2: Every soldier who knew that the queen was watching intently
was absolutely convinced
Cb=queen. Cf={queen, soldiers}

U3: that she should wave as the parade passed (she=queen)
Cb=queen.

(5.17b) DP antecedent, context mismatch

U1: The soldiers felt really quite uneasy about the squadron
parade.
No Cb. Cf={soldiers}

U2: Every soldier who knew that the queen was watching intently
was absolutely convinced
Cb=soldiers. Cf={soldiers, queen}

U3: that she should wave as the parade passed (she=queen)
Cb=queen.

(5.17c) *QP antecedent, context match*

U1: The soldiers felt really quite uneasy about the squadron parade.

No Cb. Cf={soldiers}

U2: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb=soldiers. Cf={soldiers, queen}

U3: that he should wave as the parade passed (he=soldiers)

Cb=soldiers.

(5.17d) *QP antecedent, context mismatch*

U1: The queen felt really quite uneasy about the squadron parade.

No Cb. Cf={queen}

U2: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb=queen. Cf={queen, soldiers}

U3: that he should wave as the parade passed (he=soldiers)

Cb=soldiers.

This analysis predicts that the Cb status in U2 is enough to bring that entity to the top of the Cf rankings. This assumes that the importance of Cb status outweighs the prominence factors that made the most recent antecedent more favoured in Experiments 4 and 9, and also that it outweighs first-mention, which (as noted above) is often cited as important in the Cf rankings. I propose that first-mention can be discounted in these materials because, as the results of Experiments 4 and 9 showed, it did not influence the pronoun preferences in almost identical materials. Subjecthood can also be discounted because there are two subjects in U2. The question remains as to whether recency or Cb status will be more important in the rankings of U2. Specific predictions are made in the section below.

5.4.3 Topichood in the materials for Experiment 5

Outside of CT, the materials can be thought of in terms of topichood. In the context sentence, either *the soldiers* or *the queen* are the sentence topics.

However it is questionable whether they are also discourse topics. Under the definition of Cowles *et al.* (2007), discourse topic status depends on whether an intervening topic is found in the following sentence. In the critical sentence of the materials, which is the sentence containing the pronoun, there is a QP in the matrix clause and a DP in the relative clause. It could be argued from strict information structural grounds that neither the QP nor the DP in the critical sentence count as intervening topics. QPs are thought to be unsuitable as topics because topics should be referring expressions (see Hinterwimmer and Endriss 2009 for a discussion). The DP would be an unsuitable topic because of its position in the embedded clause; sentence topics are usually found in the matrix clause. Under these rather strict definitions the subject of the context sentence also has discourse topic status. However, this is somewhat undermined by the observation made in Chapter 4 that grammatical subjects, even when not in first-mention position, are discourse prominent enough to disrupt the first-mention advantage, contributing to the discourse prominence of the most recent antecedent in Experiments 4 and 9. In light of this I will not assume that the entity introduced in the context sentence is a discourse topic, and instead will assume that it is just a sentence topic. This still allows for investigation of how prior discourse prominence (sentence topichood in this case) influences an entity's current accessibility.

5.4.4 Experiment 5 predictions

It is unclear, from the point of view of both CT and topichood, whether prior discourse prominence (Cb status) should override the recency preference in the critical sentence of Experiment 5. Furthermore manipulating the context sentence may in fact disrupt the coherence of the texts by introducing competing prominence cues for the entities. For simplicity, I will restrict my predictions to (i) those in which the prior discourse prominence overrides the recency preference, and (ii) those in which it does not.

Under a CT analysis, if Cb status boosts the prominence of an entity in the Cf of U2 to the top of the ranking, the processing of the pronoun will be easier when the pronoun matches the top-ranked entity as opposed to

when it mismatches. This will lead to a main effect of context with longer reading times in the mismatching context conditions compared to the matching context conditions.

(The same outcome is expected if prior (sentence) topichood increases the accessibility of an entity in the following sentence at the point of processing the pronoun.)

Alternatively, if Cb status/prior topichood does not override the recency preference, there should be an effect of antecedent such that the DP conditions have shorter reading times than the QP conditions.

5.4.5 Experiment 5 results summary

Experiment 5 was an eye-tracking experiment conducted with 35 native speakers of English. Details of the method, participants, procedure and data analysis can be found in Chapter 3, section 3.6.3.

The critical sentence for each item was divided into the following regions of interest for the analysis:

(5.18)

| Every soldier | *who knew that* | the queen | *was watching intently*
was absolutely | convinced | that she | should wave | as the |
parade passed. |

- **[Every soldier]**: QP region, containing the QP (post-hoc analysis only)
- **[the queen]**: DP region, containing the DP (post-hoc analysis only)
- **[convinced]**: precritical region, containing one word preceding the pronoun; this was always the past participle of a verb whose subject was the QP
- **[that (s)he]**: pronoun region, containing the complementizer *that* and the pronoun
- **[should wave]**: spillover region, containing two words following the pronoun

- **[as the]**: prefinal region, containing two words following the spillover region
- **[parade passed]**: final region, containing two words following the prefinal region, always the last two words of the sentence.

Table 5.1 (Table 3.1 from Chapter 3) shows the mean first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times in the regions from precritical to final. Table 5.2 (Table 3.2 from Chapter 3) shows the outcome of the mixed-effect models analysis described in Chapter 3, section 3.6.4. The results are summarised per region of interest (adapted from Chapter 3, section 3.7.1).

Table 5.1. Means in all conditions in milliseconds (with standard deviations shown in brackets) for first fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions from precritical to final for Experiment 5.

	First fixation mean	First-pass mean	Regression- path mean	Rereading mean	Total mean
Precritical region					
DP, context match	211 (68)	244 (124)	312 (253)	126 (251)	370 (298)
DP, context mismatch	233 (87)	262 (126)	359 (424)	113 (226)	375 (254)
QP, context match	222 (73)	240 (90)	379 (424)	124 (259)	364 (273)
QP, context mismatch	210 (62)	231 (96)	330 (315)	128 (260)	359 (273)
Pronoun region					
DP, context match	199 (53)	258 (150)	358 (378)	163 (212)	421 (248)
DP, context mismatch	212 (80)	265 (148)	329 (277)	153 (279)	418 (311)
QP, context match	206 (60)	273 (124)	404 (430)	203 (362)	476 (388)
QP, context mismatch	213 (75)	250 (118)	320 (316)	188 (296)	438 (327)
Spillover region					
DP, context match	204 (80)	292 (188)	430 (459)	243 (336)	536 (380)
DP, context mismatch	195 (58)	273 (172)	340 (270)	184 (300)	457 (324)
QP, context match	198 (71)	291 (184)	447 (535)	214 (350)	505 (434)
QP, context mismatch	204 (68)	288 (157)	436 (471)	232 (317)	520 (347)
Prefinal region					
DP, context match	211 (61)	281 (144)	438 (371)	182 (289)	464 (314)
DP, context mismatch	211 (64)	273 (138)	492 (721)	173 (278)	446 (306)
QP, context match	206 (60)	285 (152)	417 (458)	140 (227)	425 (259)
QP, context mismatch	217 (72)	311 (161)	448 (340)	205 (335)	516 (377)
Final region					
DP, context match	224 (79)	312 (180)	991 (1573)	129 (220)	441 (271)
DP, context mismatch	202 (67)	310 (177)	769 (1019)	126 (280)	436 (345)
QP, context match	203 (73)	297 (165)	825 (1331)	140 (256)	437 (314)
QP, context mismatch	213 (75)	312 (199)	812 (986)	134 (237)	446 (329)

Table 5.2. Model outcomes for each region and measure, Experiment 5. The estimate for the adjustment for each factor in the model is given in the 'estimate' column, and the standard error of the estimate is shown in brackets. The t-value for each estimate is given with the significance level indicated: () = $p < .10$, * = $p < .05$, ** = $p < .001$. The baseline for antecedent was DP and the baseline for context was match. Coefficient estimates and t-values are not given (indicated by '-') when the outcome of the modelling process was that the empty model was the best model*

Table 5.2

	First-fixation durations		First-pass times		Regression-path times		Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
Precritical region										
Antecedent (match)	0.05 (0.04)	1.27	0.02 (0.04)	0.55	0.14 (0.08)	1.72(*)	-	-	-	-
Context	0.10 (0.04)	2.41*	0.10 (0.05)	1.81(*)	0.12 (0.08)	1.52	-	-	-	-
Antecedent x Context	-0.14 (0.05)	-3.03**	-0.13 (0.06)	-2.43*	-0.20 (0.08)	-2.39*	-	-	-	-
Context (antecedent=QP)	-0.04 (0.04)	-0.98	-0.04 (0.05)	-0.66	-0.09 (0.08)	-1.13	-	-	-	-
Pronoun region										
Antecedent (match)	-	-	-	-	-	-	-	-	-	-
Context	-	-	-	-	-0.10 (0.05)	-2.17*	-	-	-	-
Antecedent x Context	-	-	-	-	-	-	-	-	-	-
Spillover region										
Antecedent (match)	-	-	-	-	-0.01 (0.07)	-0.07	-35.26 (34.60)	-1.02	-0.12 (0.07)	-1.70(*)
Context	-	-	-	-	-0.14 (0.07)	-1.86(*)	-64.28 (32.04)	-2.01*	-0.14 (0.06)	-2.29*
Antecedent x Context	-	-	-	-	0.17 (0.08)	1.95(*)	78.57 (43.94)	1.79(*)	0.24 (0.08)	2.87**
Context (antecedent=QP)	-	-	-	-	0.03 (0.07)	0.42	14.28 (31.58)	0.45	0.09 (0.06)	1.42
Prefinal region										
Antecedent (match)	-	-	0.01 (0.05)	0.24	-	-	-	-	-0.05 (0.06)	-0.86
Context	-	-	<-0.01 (0.05)	-0.10	-	-	-	-	-0.04 (0.06)	-0.61
Antecedent x Context	-	-	0.11 (0.06)	1.79(*)	-	-	-	-	0.19 (0.10)	1.79(*)
Context (antecedent=QP)	-	-	0.11 (0.05)	2.24*	-	-	-	-	0.15 (0.07)	2.22*
Final region										
Antecedent (match)	-0.10 (0.03)	-2.82**	-	-	-	-	-	-	-	-
Context	-0.10 (0.03)	-2.83**	-	-	-	-	-	-	-	-
Antecedent x Context	0.14 (0.05)	2.90**	-	-	-	-	-	-	-	-
Context (antecedent=QP)	0.04 (0.03)	1.27	-	-	-	-	-	-	-	-

Precritical and pronoun regions

For a detailed discussion of the findings from the precritical region, see Chapter 3, section 3.7.1. In the pronoun region there was a main effect of context in the regression-path times; readers took longer to exit the pronoun region to the right in the context match conditions compared to the context mismatch conditions. This suggests an elevated processing cost for the matching contexts. However, a closer look at the precritical and pronoun regions together shows that, while the pattern in the two DP antecedent conditions reverses from one region to the next, the pattern for the QP conditions remains unchanged from the precritical region to the pronoun region. If the precritical region is taken into account, the main effect of context in the pronoun region could be driven by a change in pattern for the DP conditions and no change in pattern for the QP conditions. These effects can be seen clearly in the following figures which show the regression-path times in the precritical region (Figure 5.1) and the pronoun region (Figure 5.2). The result of this change in the DP antecedent conditions is that the DP and QP conditions pattern together in the pronoun region, leading to an effect of context with higher regression-path times for the context match conditions compared to context mismatch conditions.

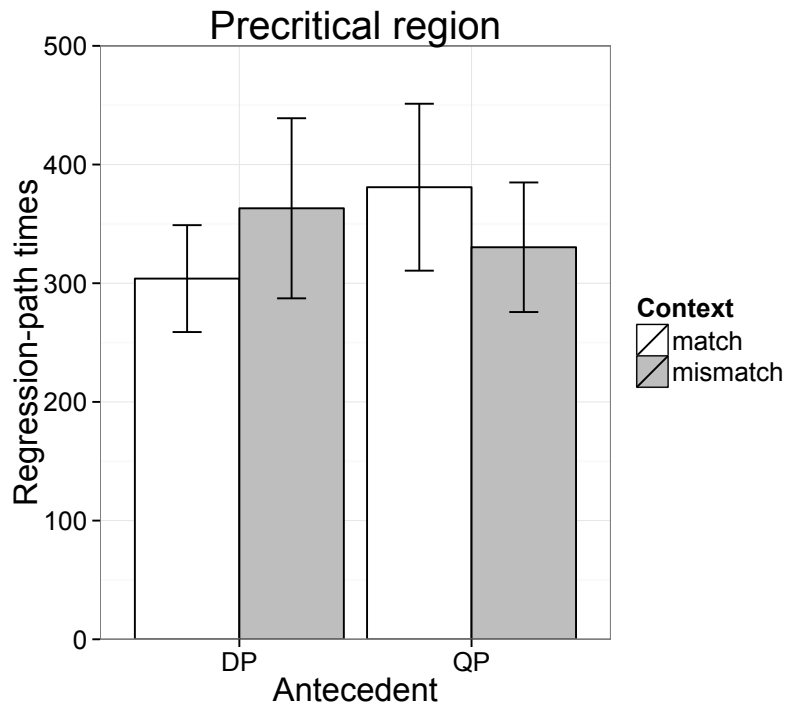


Figure 5.1. Mean regression-path times (in milliseconds) per condition in the precritical region for Experiment 5. The error bars are 95% confidence intervals.

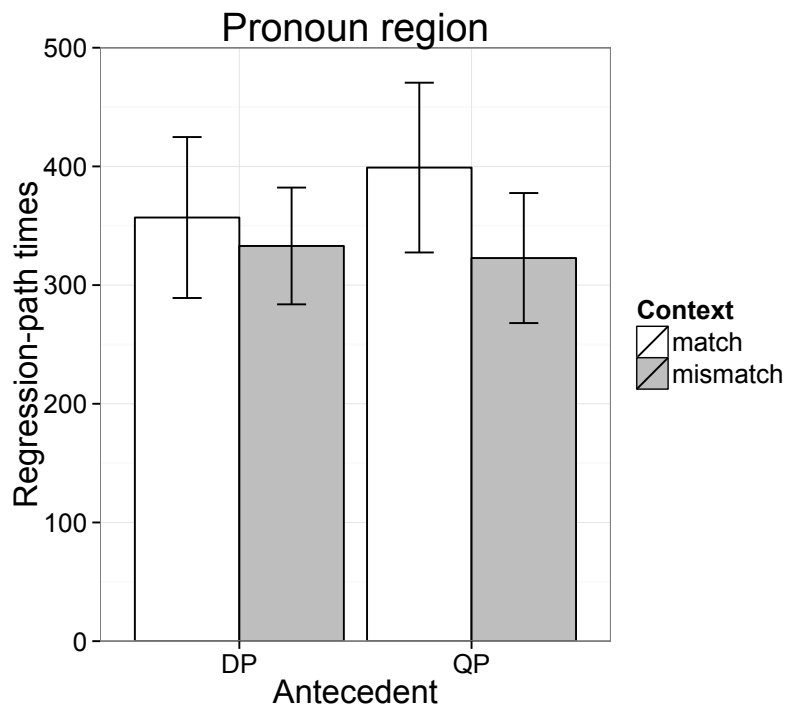


Figure 5.2. Mean regression-path times (in milliseconds) per condition in the pronoun region for Experiment 5. The error bars are 95% confidence intervals.

So, I suggest that the effects in the pronoun region should not be viewed in isolation. The QP antecedent conditions show the same pattern in the precritical and the pronoun region, so the difference between match and mismatch conditions seen here should not be interpreted as an effect of processing the pronoun. Looking at just the DP antecedent conditions, there is a clear change between the precritical and the pronoun regions, with numerically longer regression-path times in the context match condition compared to the context mismatch condition.

Spillover region

The pattern that emerged for the DP antecedent conditions in the pronoun region (numerical trend only) is seen more clearly in the spillover region. The rereading and total viewing times are significantly longer in the DP antecedent, context match (DPcm) condition than the DP antecedent, context mismatch (DPcmm) condition, as shown by the estimate for context. This is also seen in the regression-path times but the effect is marginally significant. There are also differences emerging between the DP antecedent and the QP antecedent conditions, visible in the total viewing times; total viewing times are longer in the QP antecedent, context mismatch (QPcmm) condition than the DPcmm condition, and this is also marginally significant in the regression-path and rereading times. There are also marginally longer total viewing times in the DPcm condition compared to the QP antecedent, context match (QPcm) condition. The results in this region show that the DP antecedent and the QP antecedent are responding differently to the context manipulation. However, while there is a significant difference between the context match and mismatch conditions for the DP antecedent, the difference between the context match and mismatch conditions for the QP antecedent is not significant here.

Prefinal region

In the prefinal region there are effects in the first-pass times and the total viewing times. In both measures the QPcmm condition is significantly longer than the QPcm condition, and also the QPcmm condition is marginally longer than the DPcmm condition. The QP antecedent in this region is affected differently by the context manipulation than the DP was

in the region before. For the QP, there is an advantage for the matching context rather than a cost. The DP in this region does not show effects of the manipulation of context.

Final region

There are significant effects in the first-fixation durations in the final region. First-fixation durations are significantly longer in the DPcm condition compared to the DPcmm condition, and also compared to the QPcm condition. The QPcmm condition has significantly longer first-fixation durations than the DPcmm condition. This is very similar to the pattern of results in the spillover region, however, it is difficult to interpret these results for two reasons. Firstly, the final region is far away from the pronoun and, being at the end of the sentence, is subject to general sentence wrap-up effects which may not be anything to do with pronoun processing. Secondly, the effect is found here in first-fixation durations only, so it is a very fleeting pattern. For this reason the effects in this region will not be discussed any further.

Overall pattern

The overall pattern of results is that the context and antecedent factors interact throughout. Beginning as a trend in the pronoun region, the context match is more costly than the context mismatch in the DP antecedent conditions, and beginning as a trend in the spillover region and reaching significance in the prefinal region, the context match is less costly than the context mismatch in the QP antecedent conditions. Thus, the effect of context emerges in a later region for the QP antecedent than the DP antecedent, and the effect of context goes in the opposite direction for the QP antecedent and the DP antecedent. The only simple main effect of context (that was not in the presence of an interaction) was in the pronoun region, and I argue above that this results from a change in response pattern from the precritical region in the DP antecedent conditions, and no change in response pattern for the QP antecedent conditions. The implications of these findings are discussed below.

5.5 Discussion

The context manipulation for Experiment 5 set out to test whether the prior discourse prominence of an entity can affect the current accessibility of the potential antecedents for a pronoun. The suggestion that Cb status may be a factor that influences the prominence ranking in the Cf is one way in which the influence of prior discourse status could be operationalised. Outside of CT, Cowles *et al.*'s (2007) claim that discourse topichood as well as sentence topichood could play a role in pronoun processing was also informative, even though their experiment did not conclusively demonstrate that the effects of discourse topichood and sentence topichood were separable. To be clear, Cowles *et al.* do not suggest that the prior discourse prominence of an entity can affect its current accessibility, however, the prominence of discourse topics is suggestive of an important role for discourse cues that is not limited to one or two sentences.

It was predicted, under a CT analysis, that if Cb status boosts the prominence of an entity in the Cf of U2 to the top of the ranking, processing the pronoun should be easier when the pronoun matches the top-ranked entity as opposed to when it mismatches, leading to longer reading times in the mismatching context conditions compared to the matching context conditions. This prediction was only partially borne out. In the QP antecedent conditions, the expected effect was seen, beginning in the spillover region as a trend and becoming significant in the precritical region. However, for the DP conditions the effect was in the reverse direction, beginning in the pronoun region and reaching significance in later measures of the spillover region (it can be seen as a trend in earlier measures in the spillover region). The fact that the effect was not uniform for both antecedent types is problematic for the hypothesis that Cb status (or prior discourse prominence) straightforwardly increases the later prominence of an entity in the discourse.

It was also not the case that the recency preference observed in Experiments 4 and 9 was overruled by the prior context effect, nor did the recency effect completely overrule any effects of the context sentence which would have led to an overall advantage for the DP antecedent. The DP *does*

appear to have an advantage over the QP in that the effects of context for the QP antecedent emerged in a later region than the effects of context for the DP. This is likely to reflect some extra processing cost involved in recovering the QP antecedent compared to the DP antecedent because, not being the most recent antecedent, it was dispreferred. This is further discussed in Chapter 3.

5.5.1 Assessing antecedent type and the influence of the repeated-name penalty

The differing antecedent types, DP and QP, might have led to the diverging effects of the context manipulation. The QP in the critical sentence, not being a referring expression, does not project an individual into the discourse and therefore cannot corefer with a pronoun. Rather than having an advantage as is expected under the modular POB approach, QPs in general appear to be somewhat dispreferred as antecedents (even though they are preferred when they are the most recent antecedent). But when the context sentence introduces a set, as in the QP context match condition (*The soldiers...*), the QP antecedent is processed more easily. The prior establishment of a set at the level of discourse, then, could aid the discourse prominence of the QP's representation. As discussed above, it is not clear exactly what the representation of the QP should be. It is unlikely that the pronoun corefers directly with the set established by the context instead of entering a binding relation with QP, because that would lead to a number mismatch. But the prior introduction of a set might in some way improve the discourse representation of the QP which in turn leads to easier processing at the pronoun.

For the DP, on the other hand, prior mention is unhelpful. One potential explanation for this result, also discussed in Patterson (2012), is that two mentions of the DP, one in the context sentence and one in the critical sentence, leads to the repeated-name penalty. This is a somewhat unsatisfactory explanation, for two reasons. Firstly, the repeated-name penalty is usually seen at the point of the repeated name; there is no evidence to suggest that the penalty extends to a later occurrence of the pronoun, which is the point at which the DP antecedent conditions exhibit

a cost for the matching context. As such, it would be an extension of what is currently known about the repeated-name penalty to make this suggestion. Secondly, as noted above, evidence for the repeated-name penalty comes mainly from studies that have used names, despite the claim that it should apply equally to DPs. Gordon and Hendrick (1998) also suggest that the repeated-name penalty might be reduced in more complex discourses; previous studies tend to use simple declarative sentences composed of one or two clauses, and so the evidence for extending such a penalty to more complex discourses is so far absent. Names have particularly stable discourse representations (Garrod and Sanford 1990), so it seems more likely that such a penalty would be stronger for names than it would for DPs such as *the queen*. But more importantly, the repeated-name penalty explanation is contradicted by the findings of the post-hoc analysis of the DP and QP regions (described in Chapter 3, section 3.7.1). First-pass reading times in the DP region were shorter when the context sentence contained *the queen* compared to when it contained *the soldiers*, making it very unlikely that the cost of repeated mention emerged later to have a detrimental effect on pronoun processing.

So, while there is a possible explanation for the advantage of the matching context condition for the QP antecedent, an explanation for the disadvantage for the matching context in the DP antecedent conditions is lacking. Below, I consider a different explanation that is able to account for the differing direction of the effects.

5.5.2 The cost of a topic shift and textual coherence: effects on pronoun resolution

In searching for an explanation for the rather curious finding that prior discourse prominence is helpful for one type of antecedent but unhelpful for another, it is perhaps useful to consider the fundamental concept that underpins CT and other frameworks for the analysis of discourse: coherence. In CT, coherence is in part defined by the transitions between utterances, as exemplified by the processing cost assumptions about continue, retain and shift transitions. The principle underlying this idea of coherence is that “frequent shifting leads to a lack of local

coherence" (Grosz *et al.* 1995, p. 215). However, the transitions are implemented in CT in such a way that no transition takes place between the first and second utterance. This is because shifting is concerned with maintenance of the Cb, which can only be established at the second utterance. As such, it is impossible to explain the coherence between the context sentence and the critical sentence in CT terms. But the notion of frequent shifting disrupting coherence can usefully be applied to explain the results of the current experiment in the following way. Consider the first two sentences of the DP antecedent conditions, repeated as (5.19) and (5.20):

(5.19) *DP antecedent, context mismatch*

Context sentence: The soldiers felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed.

(5.20) *DP antecedent, context match*

Context sentence: The queen felt really quite uneasy about the squadron parade.

Critical sentence: Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed.

In (5.19) the context sentence begins with the establishment of a set (*the soldiers*); it can be assumed that a set of soldiers is established in the discourse representation at that point, making it easier to process the subsequent QP *every soldier* (as evidenced in the post-hoc analysis: see Chapter 3, section 3.7.1). The attention of the reader is still on the set of soldiers until *the queen* is encountered in the critical sentence. Recall that in the previous chapter, it was argued that entities in that are grammatical subjects are important enough to disrupt a first-mention advantage, and Experiment 4 has already shown that there is an advantage for the

antecedent occupying that slot in this particular structure. Encountering *the queen* in the critical sentence, then, is likely to move the reader's attention away from the set of soldiers and towards *the queen*, i.e. something like a topic shift takes place at this point. (I use the term topic shift here in a very loose sense; in information structural terms *the queen* might not be considered a topic, as noted above, and under CT it cannot be said that a true shift has taken place.)

Now compare the situation for (5.19) to that in (5.20), when the context sentence begins with *the queen*. A discourse referent is established for *the queen*. Then in the critical sentence, the QP is encountered which means that establishing some kind of representation for a set of soldiers is necessary (I leave aside here the nature of this representation). This establishment switches the reader's attention to the set of soldiers. Next, the DP *the queen* is encountered in the critical sentence, again in subject position. As in (5.19), encountering *the queen* leads to another topic switch. Comparing (5.19) to (5.20), one topic switch can be said to have taken place before the pronoun is encountered in (5.19), and two topic switches take place in (5.20). Overall, there are more potential topic switches in the DP antecedent context match and the QP antecedent context mismatch conditions (both begin with *the queen* in the example materials) compared to the DP antecedent context mismatch and the QP antecedent context match conditions.

If "frequent shifting leads to a lack of local coherence" (Grosz *et al.* 1995, p. 215), then the results of Experiment 5 support this proposal. More topic shifts led to longer reading times at the pronoun (and beyond) for both the DP and the QP antecedents. Since pronouns are in some ways markers of discourse coherence, it makes sense that a more coherent discourse makes the establishment of pronoun reference easier. This fits well with the results of the current experiment, and explains why the context sentence has a different effect on the DP antecedent conditions than it did on the QP antecedent conditions.

5.6 Conclusion

In this chapter I questioned how prior discourse prominence can affect the later accessibility of entities in the discourse, thereby affecting pronoun resolution. I considered the issue of discourse topichood which appears to facilitate pronoun resolution, but the online evidence for this so far is incomplete. I then considered two approaches to the evaluation of discourse prominence: coherence analysis and CT. CT suggested a specific way in which prior discourse prominence could affect pronoun resolution, through the effect of Cb status on the prominence rankings in the Cf. I tested this proposal through the manipulation of the context sentence in Experiment 5. It was found that prior discourse status clearly had an effect on pronoun resolution, but not in the straightforward way that was predicted. The previous context was not always facilitative; I propose that this is due the number of topic switches that a reader must make; more switches upsets the local coherence of the discourse. The resulting lack of coherence has a detrimental effect on pronoun resolution, a process which relies on finding a prominent entity in the prior discourse and in some way acts as a marker of discourse coherence.

An attempt was made to understand the mechanisms of coherence in Experiment 5 using CT. This was not without problems, however, because of the uncertain nature of QP representations and the complexity of the clausal structure. Since it has been shown in my previous experiments that QP antecedents are subject to at least some of the same salience factors as DP antecedents, further efforts should be made to understand how CT can apply in these cases. Further research is needed to determine how to accommodate intervening relative clauses into CT utterances, and this paper suggests one way in which this could be achieved. Furthermore, CT contains some extremely useful vocabulary for understanding coherence relations. Whether or not the framework can be extended to include transitions between the first and second utterance, or indeed switches of attention within one utterance, remains a matter for future research. Insofar as attention switches appear to play an important role in the coherence of short but complex texts such as those in the current experiment, they should be accounted for in models of coherence and pronoun resolution.

CHAPTER 6

Conclusions

The aim of this thesis was to investigate how the parser handles multiple antecedents in order to understand more about how certain information sources play a role during resolution. A large number of cues and constraints have previously been identified as playing a role in pronoun resolution. I considered how both structural information and information provided by prior discourse is used in online processing. In this final chapter I first summarise my experimental findings and then discuss the main implications of these findings for our understanding of pronoun resolution processes.

6.1 Summary of experimental findings

In Chapter 2, one offline and two online experiments were reported in which the processing of pronouns in canonical condition B environments and SDP environments is investigated. In Experiment 1 (offline), participants chose the accessible antecedent almost exclusively for pronouns in condition B environments. Pronouns in SDP environments, in contrast, were interpreted as ambiguous in the majority of cases, but the local antecedent and non-local antecedent options were chosen in some cases; no preference for the local or the non-local antecedent was shown. In Experiment 2 (online), participants were sensitive to the gender of the accessible antecedent but not the inaccessible antecedent in condition B environments. In the SDP environments tested in Experiment 3 there was a (non-significant) tendency for participants to be sensitive to the gender of the local antecedent. These results show differing sensitivity to local and non-local antecedents depending on the syntactic environment, and also demonstrates faithfulness to the condition B constraint at all stages of processing.

Experiments 4 and 5 in Chapter 3 investigated the POB hypothesis, which claims that variable-binding (VB) antecedents are considered before

coreference (CR) antecedents. No evidence for the POB hypothesis was found in either Experiment 4 or Experiment 5. Instead, Experiment 4 demonstrated that the CR antecedent was considered while it seemed that the VB antecedent was not considered during online processing. In Experiment 5 a different experimental design brought out that both VB and CR antecedents can be considered, but the effects emerged in an earlier region for the CR antecedent compared to the VB antecedent. Both these results are incompatible with a search mechanism that initially ignores antecedents that can be linked to a pronoun via a discourse-mediated route, seeking instead those antecedents that can be linked to a pronoun at the level of semantics. In the absence of such a search mechanism, the preference for (or earlier consideration of) the CR antecedent in both experiments warranted further investigation, in particular because the CR antecedent was always closer to the pronoun than the VB antecedent. The effects of the order of antecedents was explored further in Chapter 4.

In Chapter 4, the recency preference found in Experiments 4 and 5 was explored further through three offline experiments and one online experiment. Experiments 6, 7 and 8 were offline questionnaire tasks in which the antecedent closest to the pronoun was systematically varied. In Experiment 6, which was a forced choice task, participants' interpretations were affected both by recency and the type of antecedent. Recent antecedents were preferred over more distant antecedents, and DP antecedents were preferred over QP antecedents.

Both Experiments 7 and 8 compared the effect of recency on QP and name antecedents. Experiment 7, like Experiment 6, showed that offline interpretations were affected by both the type of antecedent and an antecedent's recency. In Experiment 8 there were more QP responses when the QP was the most recent antecedent and more name responses when the name was the most recent antecedent, again highlighting that offline interpretations were affected by the antecedent order, with a preference for the most recent antecedent. Experiment 9 was an online follow-up experiment to Experiment 4. In all conditions the QP was presented closer to the pronoun than the name antecedent. There was an increased

processing cost associated with both a gender-mismatching QP antecedent and a gender-mismatching name antecedent compared to the match conditions. However, the QP effects were more numerous and appeared in earlier regions than the name effects. Like Experiment 4, this demonstrates more sensitivity to the recent antecedent during online processing than a more distant antecedent. Unlike Experiment 4 though, there was evidence that the more distant antecedent was also considered.

Finally, in Chapter 5 the effect of prior discourse prominence on pronoun resolution was investigated via the context manipulation in Experiment 5. The critical sentence in this experiment contained a QP, a DP and a pronoun; either the QP or the DP matched in gender with the pronoun. In a previous context sentence, the prior introduction of one of the antecedents was also manipulated to create conditions where the entity introduced in the context sentence either matched or mismatched the pronoun. In Centering Theory (CT) terms, I tested the hypothesis that Cb status affects the current prominence rankings of discourse entities, which in turn should affect pronoun resolution. The results showed that, while prior discourse status did have an effect on pronoun resolution, it was not a uniform effect because the previous context was not always facilitative. I proposed that this is due the number of topic switches that a reader must make; more switches upsets the local coherence of the discourse, and the resulting lack of coherence had a detrimental effect on pronoun resolution.

In the following sections I discuss the main implications of these findings for our understanding of pronoun resolution processes.

6.2 A structurally sensitive, non-modular, search mechanism

The parser demonstrates structural sensitivity during pronoun resolution in that it is able to exclude antecedents that are structurally inaccessible according to condition B of the Binding Theory (Chomsky 1981; 1986), even during early stages, and is sensitive to the details of syntactic structure that allow or disallow local antecedents (condition B versus SDPs). This finding

is compatible with previous findings that the parser is faithful to condition A governing reflexives (Xiang, Dillon and Phillips 2009; Felser and Cunnings 2012 for native speakers), other findings for condition B (Clifton, Kennison and Albrecht 1997; Clifton, Frazier and Deevy 1999; Lewis and Chow 2012), and emerging evidence regarding condition D on QPs (Cunnings, Patterson and Felser 2013). While this seems to support Nicol and Swinney's (1989) BAIF model, which assumes a distinction between an early grammatical stage (during which binding conditions are applied) and later stages, there is in fact no evidence yet that discourse or pragmatic information is systematically withheld relative to structural cues, and pragmatic reasoning may even be involved in the deployment of condition B. It is also difficult to bring together previous conflicting findings under this hypothesis. On the other hand Badecker and Straub's (2002) model and later similar proposals appear to underplay the important role that structural constraints can play. Because of the difficulty of separating cue types into distinct stages, I suggest that two simple adjustments to the weighting of cues in a parallel constraints model is more appropriate, until evidence is produced that discourse and pragmatic information is systematically excluded from early stages of processing. By increasing the cue weight of structural information in a parallel constraints model, the findings for and against interference can (mostly) be accounted for. The effective implementation of the cues may be impeded when the memory is taxed because of additional tasks, giving rise to interference effects (e.g. Badecker and Straub 2002; Clackson, Heyer and Clahsen 2012) which under normal circumstances can be suppressed. This also fits in with findings on pronoun resolution from other populations in which processing resources may be taxed, such as children (Clackson, Felser and Clahsen 2011; van Rij 2012) and non-native speakers (Patterson, Trompelt and Felser in prep.).

The sensitivity of the parser to structural cues does not entail that cue types can be easily separated into distinct sequential stages, such as 'grammatical' and 'everything else' (BAIF) and 'semantics' and 'discourse' (POB). There is as yet no evidence that discourse information is systematically suppressed until after an initial structural or semantic stage has been completed. Similarly, the earlier consideration of non-c-

commanding antecedents cannot be accounted for in a modular parser that initially searches for c-commanding antecedents. Instead, my results support a search mechanism that is structurally *sensitive* but is not *modular*.

6.3 QPs as antecedents

6.3.1 Avoidance of QP antecedents

Despite their importance in semantic theory and certain syntactic approaches to pronouns, QPs have rarely been used as antecedents in pronoun processing studies. Evidence from both Carminati, Frazier and Rayner (2002) and Koornneef (2008) shows that QP antecedents can incur a processing cost in comparison to non-quantified DPs, which could be a reflection of their internal complexity. But a processing cost in itself does not entail that QPs are dispreferred compared to other types of antecedent, nor does it mean that antecedent type is taken into consideration by the parser during processing. As a cue for the parser to take into account during processing, antecedent type would be fairly unhelpful because it does not make an entity more or less likely to be an antecedent. This is strongly supported by the fact that QP antecedents *are* considered during processing when they are more recent. However, the results from the offline questionnaires do suggest that QP antecedents are avoided during antecedent choice tasks. This reluctance to pick a QP antecedent could stem from uncertainty about number features, given that QPs such as *every soldier* can give rise to a group reading, and can then be referred to subsequently with a plural pronoun. This is likely to be the result of conscious reflection rather than automatic processing. On this basis a separate cue for antecedent complexity does not need to be proposed, but pronoun processing can be affected by the complexity of certain antecedent types.

6.3.2 The involvement of discourse in binding relations with QPs

It is very clear from the experimental results that the processing of dependencies between pronouns and QPs is influenced by discourse

factors. This means that the search process does not initially take place at a purely semantic level, before discourse considerations are taken into account as suggested by the POB hypothesis. This calls into question the precise role that variable binding and coreference play during pronoun processing. There is much syntactic evidence for making a distinction between referring and non-referring expressions that has implications for pronoun-antecedent relations, but the sequential division of labour into semantic and discourse levels in the way that Koornneef (2008) and Reuland (2011) suggest is not currently workable given the processing evidence. The evidence points more towards an approach that combines syntax and discourse elements, such as put forward by Burkhardt (2005). An alternative would be to suggest that search and integration processes were somehow separate from the actual formation of a dependency; candidate antecedents could be searched for and evaluated on the basis of discourse and other factors, and then the dependency itself is formed subsequently. Such an approach would have to clarify the processing implications for the separation of search and integration on one hand, and dependency formation on the other.

In the current thesis only QPs with the quantifier *every* were tested. There is clearly a need for further research with different types of QP antecedents, particularly in view of the suggestion that QPs such as *every soldier* are in fact referential (Burkhardt 2005). Further testing of the influence of discourse factors on QP-pronoun dependencies would also be welcome to extend the current findings.

The POB hypothesis was tested because of strong predictions it made for processing. In syntax there is a wide variety of approaches to binding and non-referential NPs, but this has largely not been taken up in the pronoun processing literature. There is much insight to be gained from testing the processing implications of theoretical accounts, but this also depends on the formulation of clear processing predictions. This underscores the need for researchers to work together for the benefit of both processing models and theoretical accounts.

6.4 Recency as a cue in pronoun processing

Recency has been recognised as a factor in pronoun processing for some time, but it has received comparatively less attention than other factors (in particular the first-mention advantage). The nature of the advantage gained from recency is somewhat different from first-mention. While a first-mentioned antecedent can be attributed a degree of importance at the point at which it is encountered, leading to a higher degree of activation for the first-mentioned entity, recency can only be judged retrospectively because it is not clear at the point of encountering an entity that it will be recent relative to the pronoun. This only becomes apparent when the pronoun is encountered. The recency advantage, then, appears to arise from an interaction between the properties of the memory system and discourse factors that determine the activation of an antecedent. As such, it does not need to be stipulated as a separate pronoun processing cue on its own. If discourse factors and memory combine to promote a recent antecedent, recency effects will be apparent.

6.5 Prior discourse and coherence

Online pronoun resolution is clearly affected by discourse cues that make certain entities more prominent than others. Prior discourse prominence also plays a role, but not in a uniform manner. The way in which prior discourse prominence is involved in later pronoun resolution suggests, firstly, that coherence is very important; maintaining a topic is more important than prior prominence on its own, and frequent switching is disruptive to pronoun resolution. Secondly, I propose that pronoun resolution is concerned more with local entities (those appearing in the same or last sentence) than more distant discourse topics, unless there are no intervening entities. I would therefore question the role of discourse topichood and more global factors in pronoun resolution. Further processing research on the delicate interplay of discourse prominence factors that affect pronoun resolution is needed.

6.6 Conclusion

The aim of this thesis was to investigate how the parser handles multiple antecedents during processing. A large number of cues and constraints has previously been identified as playing a role in this process, and I have found evidence for the involvement of both structure and discourse. More specifically, I propose that pronoun resolution is a structurally sensitive process, but not a modular one. It is difficult to uphold models that advocate a strict separation of structural or semantic factors from discourse factors in the timecourse of pronoun processing. My findings support a parallel constraints model of pronoun resolution that is sensitive to the activation of potential antecedents based on discourse factors, but one in which structural cues are strongly weighted.

References

- Alcocer, P. and Phillips, C. (submitted). Using relational syntactic constraints in content-addressable memory architectures for sentence parsing.
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. New York: Routledge.
- Arnold, J.E. (1998). *Reference Forms and Discourse Patterns*. Doctoral dissertation, University of Stanford.
- Arnold, J.E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* **31**: 137-162.
- Arnold, J.E., Brown-Schmidt, S. and Trueswell, J.C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes* **22**: 527-565.
- Arnold, J.E., Eisenband, J.G., Brown-Schmidt, S. and Trueswell, J.C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition* **76**: B13-B26.
- Avrutin, S. and Thornton, R. (1994). Distributivity and binding in child grammar. *Linguistic Inquiry* **25**: 165-171.
- Baayen, H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Badecker, W. and Straub, K. (1994). (Cited in Badecker and Straub 2002). *Evidence that binding principles participate in a constraint satisfaction process*. Presented at: "Seventh Annual CUNY Sentence Processing Conference", New York, USA.
- Badecker, W. and Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28**: 748-769.
- Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**: 255-278.
- Bartek, B., Lewis, R.L., Vasishth, S. and Smith, M.R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition* **37**: 1178-1198.
- Bates, D.M., Maechler, M. and Bolker, B. (2012). lme4 package for R, version 0.999999-0, published 23.06.2012.
- Bergmann, C., Paulus, M. and Fikkert, P. (2012). Preschoolers' comprehension of pronouns reflexives: the impact of the task. *Journal of Child Language* **39**: 777-803.
- Bosch, P. (1983). *Agreement and Anaphora: a Study of the Role of Pronouns in Syntax and Discourse*. London: Academic Press.

- Boston, M.F., Hale, J., Kliegl, R., Patil, U. and Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* **2**: 1, 1-12.
- Brown-Schmidt, S., Byron, D.K. and Tanenhaus, M.K. (2005). Beyond salience: interpretation of personal and demonstrative pronouns. *Journal of Memory and Language* **53**: 292-313.
- Büring, D. (2005). *Binding Theory*. Cambridge, UK: Cambridge University Press.
- Burkhardt, P. (2005). *The Syntax–Discourse Interface: Representing and Interpreting Dependency*. Amsterdam: John Benjamins Publishing Company.
- Callahan, S.M. (2008). Processing anaphoric constructions: insights from electrophysiological studies. *Journal of Neurolinguistics* **21**: 231-266.
- Carminati, M., Frazier, L. and Rayner, K. (2002). Bound variables and c-command. *Journal of Semantics* **19**: 1-34.
- Chien, Y.-C. and Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* **1**: 225-295.
- Chomsky, N. (1981). *Lectures in Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clackson, K., Felser, C. and Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: evidence from eye-movements during listening. *Journal of Memory and Language* **65**: 128-144.
- Clackson, K., Heyer, V. and Clahsen, H. (2012). *Online application of Binding Principles A and B: evidence from eye movements during listening*. Presented at: "Generative Linguistics in the Old World (GLOW) Workshop 1: The timing of grammar: experimental and theoretical considerations", Potsdam, Germany.
- Clark, H.H. and Sengul, C.J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition* **7**: 35-41.
- Clifton, C. Jr, Frazier, L. and Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica* **11**: 11-39.
- Clifton, C. Jr, Kennison, S.M. and Albrecht, J.E. (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.
- Conroy, A., Takahashi, E., Lidz, J. and Phillips, C. (2009). Equal treatment for all antecedents: how children succeed with Principle B. *Linguistic Inquiry* **40**: 446-486.

- Cowles, H.W., Walenski, M. and Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi* **26**: 3-18.
- Cozijn, R., Commandeur, E., Vonk, W. and Noordman, L.G.M. (2011). The time course of the use of implicit causality information in the processing of pronouns: a visual world paradigm study. *Journal of Memory and Language* **64**: 381-403.
- Crawley, R.A., Stevenson, R.J. and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* **19**: 245-264.
- Cunnings, I. and Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes* **28**: 188-219.
- Cunnings, I., Patterson, C. and Felser, C. (2013). *Online sensitivity to structural constraints on bound variable anaphora*. Presented at: "26th Annual CUNY Conference on Human Sentence Processing", Columbia, USA.
- Cunnings, I., Patterson, C. and Felser, C. (accepted). Variable binding and coreference in sentence comprehension: evidence from eye movements.
- Dillon, B.W. (2011). *Structured Access in Sentence Comprehension*. Doctoral dissertation, University of Maryland.
- Dillon, B.W., Mishler, A., Sloggett, S. and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modelling evidence. *Journal of Memory and Language* **69**: 85-103.
- Ehrlich, K. and Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior* **22**: 75-87.
- Elbourne, P. (2005). *Situations and Individuals*. Cambridge, MA: MIT Press.
- Evans, G. (1977). Pronouns, quantifiers and relative clauses (I). *Canadian Journal of Philosophy* **7**: 467-536.
- Felser, C. and Cunnings, I. (2012). Processing reflexives in English as a second language: the role of structural and discourse-level constraints. *Applied Psycholinguistics* **33**: 571-603.
- Foraker, S. and McElree, B. (2007). The role of prominence in pronoun resolution: active versus passive representations. *Journal of Memory and Language* **56**: 357-383.
- Frazier, L. and Clifton, C. Jr (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy* **29**: 315-346.
- Frazier, L. and Fodor, J.D. (1978). The sausage machine: a new two-stage parsing model. *Cognition* **6**: 291-325.
- Fukumura, K. and Van Gompel, R.P.G. (2010). Choosing anaphoric expressions: do people take into account likelihood of reference? *Journal of Memory and Language* **62**: 52-66.

- Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora*. Essays in Cognitive Psychology. Hove: Psychology Press.
- Garnham, A. and Oakhill, J. (1992). Discourse processing and text representation from a "Mental Models" perspective. *Language and Cognitive Processes* 7: 193-204.
- Garrod, S. and Sanford, A. (1990). Referential processes in reading: focusing on roles and individuals. In: G.B. Flores d'Arcais, K. Rayner and D. Balota (eds), *Comprehension Processes in Reading*. Hillsdale, NJ: LEA, 465-484.
- Garrod, S. and Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution. *Journal of Memory and Language* 42: 526-544.
- Gelormini-Lezama, C. and Almor, A. (2011). Repeated names, overt pronouns, and null pronouns in Spanish. *Language and Cognitive Processes* 26: 437-454.
- Gernsbacher, M.A. (1989). Mechanisms that improve referential access. *Cognition* 32: 99-156.
- Gernsbacher, M.A. and Hargreaves, D.J. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language* 27: 699-717.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68: 1-76.
- Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In: Y. Miyashita, A. Marantz and W. O'Neil (eds), *Image, Language, Brain*. Cambridge, MA: MIT Press, 95-126.
- Givón, T. (ed.) (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins Publishing Company.
- Gordon, P.C., Grosz, B.J. and Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17: 311-347.
- Gordon, P.C. and Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science* 22: 389-424.
- Gordon, P.C., Hendrick, R. and Foster, K.L. (2000). Language comprehension and probe-list memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 766-775.
- Gordon, P.C. and Scarce, K.A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory and Cognition* 23: 313-323.
- Greene, S.B., McKoon, G. and Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18: 266-283.

- Grosz, B.J., Joshi, A.K. and Weinstein, S. (1995). Centering: a framework for modelling the local coherence of a discourse. *Computational Linguistics* **21**: 203-226.
- Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Doctoral dissertation, University of Massachusetts, Amherst.
- Heim, I. (1998). Anaphora and semantic interpretation: a reinterpretation of Reinhart's approach. In: U. Sauerland and O. Percus (eds), *The Interpretive Tract*. MIT Working Papers in Linguistics **25**. Cambridge, MA: MITWPL, 205-246.
- Hendriks, P. and Spender, J. (2006). When production precedes comprehension: an optimization approach to the acquisition of pronouns. *Language Acquisition* **13**: 319-348.
- Hestvik, A. (1991). Subjectless binding domains. *Natural Language and Linguistic Theory* **9**: 455-496.
- Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics* **3**: 211-237.
- Higginbotham, J. (1983). Logical form, binding and nominals. *Linguistic Inquiry* **14**: 395-420.
- Higginbotham, J. (1985). On semantics. *Linguistic Inquiry* **16**: 547-593.
- Hinterwimmer, S. and Endriss, C. (2009). Indefinites as direct and indirect aboutness topics. In: M. Zimmerman and C. Fery (eds), *Information Structure*. Oxford: Oxford University Press, 89-115.
- Huang, Y. (1994). *The Syntax and Pragmatics of Anaphora: A Study With Special Reference to Chinese*. Cambridge, UK: Cambridge University Press.
- Huang, Y. (2000). *Anaphora: A Cross-Linguistic Approach*. Oxford: Oxford University Press.
- Järvikivi, J., Van Gompel, R.P.G., Hyönä, J. and Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject-preference accounts. *Psychological Science* **16**: 260-264.
- Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Just, M.A. and Carpenter, P.A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review* **87**: 329-354.
- Kaiser, E. (2003). *The Quest for a Referent: a Cross-Linguistic Look at Reference Resolution*. Doctoral dissertation, University of Pennsylvania.
- Kaiser, E. (2005). When salience isn't enough: pronouns, demonstratives and the quest for an antecedent. In: R. Laury (ed.), *Minimal reference in Finnic: The Use and Interpretation of Pronouns and Zero in Finnish and Estonian Discourse*. Helsinki: Suomalaisen Kirjallisuuden Seura, 135-162.

- Kaiser, E., Runner, J.T., Sussman, R.S. and Tanenhaus, M.K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* **112**: 55-80.
- Kaiser, E. and Trueswell, J.C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* **23**: 709-748.
- Kameyama, M. (1998). Intrasentential centering: a case study. In: M.A. Walker, A.K. Joshi and E.F. Prince (eds), *Centering Theory in Discourse*. Oxford: Clarendon Press, 89-112.
- Kamp, H. (1981). A theory of truth and semantic representation. In: J.A.G. Groenendijk, T.M.V. Janssen and M.B.J. Stokhof (eds), *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum, 277-322.
- Kayne, R. (2002). Pronouns and their antecedents. In: S.D. Epstein and T.D. Seely (eds), *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell, 133-166.
- Kehler, A. (2002). *Coherence, Reference and the Theory of Grammar*. Stanford: CSLI Publications.
- Kehler, A., Kertz, L., Rohde, H. and Elman, J.L. (2008). Coherence and coreference revisited. *Journal of Semantics* **25**: 1-44.
- Kennison, S.M. (2003). Comprehending the pronouns *her*, *him*, and *his*: implications for theories of referential processing. *Journal of Memory and Language* **49**: 335-352.
- Kennison, S.M. and Trofe, J.L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research* **32**: 355-378.
- Koornneef, A. (2008). *Eye-Catching Anaphora*. Doctoral dissertation, University of Utrecht.
- Koornneef, A. and Van Berkum, J.J.A. (2006). On the use of verb-based implicit causality in sentence comprehension: evidence from self-paced reading and eye tracking. *Journal of Memory and Language* **54**: 445-465.
- Koornneef, A., Wijnen, F. and Reuland, E. (2006). Towards a modular approach to anaphora processing. In: R. Artsein and M. Poesio (eds), *Ambiguity in Anaphora Workshop Proceedings*. Malaga: ESLLI, 65-72.
- Kreiner, H., Sturt, P. and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *Journal of Memory and Language* **58**: 239-261.
- Lewis, R.L. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* **29**: 375-419.
- Lewis, R.L., Vasishth, S. and Van Dyke, J.A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* **10**: 447-454.

- Lewis, S. and Chow, W.Y. (2012). *Structural constraints on pronoun resolution: distinguishing early and late sensitivity to illicit antecedents*. Presented at: "Generative Linguistics in the Old World (GLOW) Workshop 1: The timing of grammar: experimental and theoretical considerations", Potsdam, Germany.
- Moxey, L.M. and Sanford, A.J. (1993). *Communicating Quantities: A Psychological Perspective*. Hove: Lawrence Erlbaum Associates Ltd.
- Nicol, J. (1988). *Coreference Processing During Sentence Comprehension*. Doctoral dissertation, Massachusetts Institute of Technology.
- Nicol, J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* **18**: 5-19.
- Oberauer, K. and Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language* **55**: 601-626.
- Partee, B.H. (2013). On the history of the question of whether natural language is "illogical". *History and Philosophy of the Language Sciences blog* (<http://hiphilangsci.net/2013/05/01/on-the-history-of-the-question-of-whether-natural-language-is-illogical/>). Accessed 06.06.2013.
- Patil, U., Vasishth, S. and Lewis, R.L. (2011a). *Early retrieval interference in syntax-guided antecedent search*. Presented at: "24th Annual Meeting of the CUNY Conference on Human Sentence Processing", Stanford, USA.
- Patil, U., Vasishth, S. and Lewis, R.L. (2011b). *Are grammatical constraints immune to retrieval interference?* Presented at: "10th International Symposium of Psycholinguistics", Donostia - San Sebastián, Spain.
- Patil, U., Vasishth, S. and Lewis, R.L. (2012). *Early effect of retrieval interference on reflexive binding*. Presented at: "Generative Linguistics in the Old World (GLOW) Workshop 1: The timing of grammar: experimental and theoretical considerations", Potsdam, Germany.
- Patterson, C. (2012). The effect of local discourse coherence on pronoun resolution: an eye-tracking study. *Essex Graduate Student Papers in Language and Linguistics* **13**: 96-119.
- Patterson, C. (2013). Discourse coherence in pronoun resolution. *Discours* **12**. (<http://discours.revues.org/8820>).
- Patterson, C., Trompelt, H. and Felser, C. (in prep.). The online application of binding condition B in native and non-native pronoun resolution.
- Perovic, A., Modyanova, N. and Wexler, K. (2013). Comprehension of reflexive and personal pronouns in children with autism: a syntactic or pragmatic deficit? *Applied Psycholinguistics* **34**: 813-835.
- Pesetsky, D. (1987). Wh-in-situ: movement and unselective binding. In: E. Reuland and A. Meulen (eds). *The Representation of (in)Definiteness*. Cambridge, MA: MIT Press, 98-129.

- Phillips, C., Wagers, M.W. and Lau, E.F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In: J.T. Runner (ed.), *Experiments at the Interfaces*. Bingley: Emerald Publications, 153-186.
- Pickering, M.J., Frisson, S., McElree, B. and Traxler, M.J. (2004). Eye movements and semantic composition. In: M. Carreiras and C. Clifton, Jr (eds), *The Online Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*. New York/Hove: Psychology Press, 33-50.
- Poesio, M., Stevenson, R., Di Eugenio, B. and Hitzeman, J. (2004). Centering: a parametric theory and its instantiations. *Computational Linguistics* **30**: 309-363.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* **124**: 372-422.
- Reinhart, T. (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* **27**: 53-94.
- Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. London: Croom Helm.
- Reinhart, T. and Reuland, E. (1993). Reflexivity. *Linguistic Inquiry* **24**: 657-720.
- Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry* **32**: 439-492.
- Reuland, E. (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.
- Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy* **12**: 683-721.
- Rohde, H., Kehler, A. and Elman, J.L. (2006). Event structure and discourse coherence biases in pronoun interpretation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 697-702.
- Rohde, H., Kehler, A. and Elman, J.L. (2007). Pronoun interpretation as a side effect of discourse coherence. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 617-622.
- Rooryck, J. and Vanden Wyngaerd, G. (2011). *Dissolving Binding Theory*. Oxford: Oxford University Press.
- Rose, R. (2007). Pronoun resolution and the influence of syntactic and semantic information on discourse prominence. In: A. Branco (ed.), *Anaphora: Analysis, Algorithms and Applications*. 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos, Portugal, Selected Papers. Berlin/London: Springer, 28-43.
- Runner, J.T., Sussman, R.S. and Tanenhaus, M.K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition* **89**: B1-B13.

- Runner, J.T., Sussman, R.S. and Tanenhaus, M.K. (2006). Assigning reference to reflexives and pronouns in picture noun phrases. *Cognitive Science* **30**: 1-49.
- Schwabe, K. and Winkler, S. (2007). On information structure, meaning and form: generalizations across languages. In: K. Schwabe and S. Winkler (eds), *On Information Structure, Meaning and Form: Generalizations across Languages*. Amsterdam: John Benjamins Publishing Company, 1-29.
- Sekerina, I.A., Stromswold, K. and Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* **31**: 123-152.
- Sportiche, D. (2013). Binding theory – structure sensitivity of referential dependencies. *Lingua* **130**: 187-208.
- Stevenson, R.J., Nelson, A.W.R. and Stenning, K. (1995). The role of parallelism in strategies of pronoun resolution. *Language and Speech* **38**: 393-418.
- Stewart, A.J., Pickering, M.J. and Sanford, A.J. (2000). The time course of the influence of implicit causality information: focusing versus integration accounts. *Journal of Memory and Language* **42**: 423-443.
- Streb, J., Hennighausen, E. and Rösler, F. (2004). Different anaphoric expressions are investigated by event-related-potentials. *Journal of Psycholinguistic Research* **33**: 175-201.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* **48**: 542-562.
- Sturt, P., Scheepers, C. and Pickering, M.J. (2002). Syntactic ambiguity resolution after initial misanalysis: the role of recency. *Journal of Memory and Language* **46**: 371-390.
- Tabor, W., Galantucci, B. and Richardson, D. (2004). Effects of merely local coherence on sentence processing. *Journal of Memory and Language* **50**: 355-370.
- Tenny, C. (2004). *Pronoun binding and the grammar of sentience*. Presented at: “Workshop on Semantic Approaches to Binding Theory”, ESSLLI, Université Henri Poincaré, Nancy.
- Van Dyke, J.A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **33**: 407-430.
- Van Dyke, J.A. and Lewis, R.L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* **49**: 285-316.
- Van Gompel, R.P.G. and Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition* **90**: 255-264.

- Wagers, M.W., Lau, E.F. and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language* **61**: 206-237.
- Warren, T. (2003). The processing complexity of quantifiers. In: L. Alonso-Ovalle (ed.), *On Semantic Processing*. UMass Occasional Papers in Linguistics **27**. Amherst, MA: GLSA, 211-237.
- Wexler, K. and Chien, Y.-C. (1985). The development of lexical anaphors and pronouns. *Papers and Reports on Child Language Development* **24**: 138-149.
- Xiang, M., Dillon, B. and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* **108**: 40-55.

Appendix 1

Materials for Experiment 1.

Condition B environments

1. The boy remembered that Matthew had bought him a new computer game.
2. Michael was unaware that Mr. Smith had made him a huge sandwich for lunch.
3. William was relieved that the chairman had reminded him about the appointment.
4. The prince was surprised that the king had rewarded him with a new palace.
5. Mr. Jones was shocked that Ian had added him to the guest-list without permission.
6. Pamela noticed that Angela had poured her yet another cup of tea.
7. The new girl heard that Caroline had blamed her for the recent accident.
8. Katherine was impressed that the waitress had baked her fresh bread rolls for breakfast.
9. Mrs. Wells was relieved that Margaret had taught her a few words in French.
10. The air-hostess realised that Sophie had given her the early shift again.

SDP environments

1. At the post office, Ernie watched the boy pick up the parcel next to him.

2. Harry heard William pull the curtain around him in the quiet hospital ward.
3. In the smoky bar, Mr. Smith glimpsed the waiter tap the window next to him.
4. The king noticed the prince place a sword behind him during the exquisite banquet.
5. Ian saw Mr. Jones move the papers towards him at the office meeting.
6. Angela heard the sportswoman drag the luggage behind her in the hotel lobby.
7. Caroline noticed the new girl watch the seat next to her nervously.
8. During the train journey, Monica noticed Katherine grab the handle beside her.
9. At the theatre, Margaret saw the headmistress survey the seats around her.
10. In the cafe, Sophie watched the waitress clear all the plates near her.

Sentences containing QPs (also reported as Experiment 7)

1. Every firefighter who heard that Adam was calling for help was advised that he should remain calm.
2. Every pilot who knew that Peter was working at the control tower today hoped that he would arrive on time.
3. Every florist who noticed that Suzy was on the committee feared that she might not know all the flowers' Latin names.
4. Every babysitter who said that Anna was lazy was told that she would not get another job.
5. Every secretary who thought that Karen looked a bit pale believed that she should have more time off work.

6. It appeared to Paul that every boxer in the changing room thought that he should fight a lot harder.
7. It looked to Matthew that every mechanic at the local garage was certain that he had identified the problem correctly.
8. It seemed to Lisa that every housekeeper sent by the agency believed that she was capable of running the place without help.
9. It appeared to Sarah that every nurse on the hospital ward expected that she would take over some of the night shifts.
10. It looked to Caroline that every fashion model at the show hoped that she would really impress the audience.

Appendix 2

Materials for Experiments 2 and 3. Manipulation of the name-pronoun match is shown in parentheses.

Experiment 2

1. Band practice was beginning to get rather dull. {John/Jane} remembered that {Mark/Jane/John} had taught him a new song on the guitar. That really lifted everyone's spirits!
2. The day of the marathon was fast approaching. {Bob/Ann} recalled that {Ben/Ann/Bob} had told him that the training would be easy. In fact this was a lie, but at least it was all for a good cause.
3. Everyone wanted to know how the project was going. {Ben/Sue} reported that {Bob/Sue/Ben} had criticised him for big delays at the start. Since then, things had gone much more smoothly.
4. Work experience was gradually getting better. {Roger/Laura} thought that {David/Laura/Roger} had treated him a lot more fairly that day. By the end it might even become enjoyable!
5. The most glamorous event in the publishing calendar was nearly over. {Mark/Ruth} wished that {John/Ruth/Mark} had introduced him to the stars of the show. It would have been wonderful to get some books signed.
6. The last cleaning job had been a disaster from start to finish. {David/Jenny} suspected that {Roger/Jenny/David} had blamed him for the mud on the carpet. A large discount had to be offered as a result.
7. It had been a marvellous but tiring trip to London. {Simon/Helen} was happy that {Peter/Helen/Simon} had reserved him a seat on the train home. There were a lot of passengers and not very much space!

8. Moving day had arrived and there was plenty of confusion. {Peter/Susan} realised that {Simon/Susan/Peter} had locked him in the new place by mistake. Without any boxes to unpack, there was nothing to do but wait.
9. The atmosphere at the staff meeting had become rather awkward. {Gary/Lisa} heard that {Adam/Lisa/Gary} had described him as dull but good at the job. It was felt to be a somewhat personal comment.
10. Company profits had increased slightly compared to last year. {Adam/Mary} was stunned that {Gary/Mary/Adam} had rewarded him with a very big pay rise. Perhaps it was an incentive for next year!
11. Some people find it difficult to concentrate in exams. {Paul/Kate} was sure that {Matt/Kate/Paul} had poked him in the arm with a pen. The wisest thing to do would be to ignore it.
12. The school sports day was about to finish. {Matt/Beth} wished that {Paul/Beth/Matt} had warned him about the egg and spoon race. There was a lot of cheating going on, especially by the parents.
13. It had been a stressful week of work and study. {Jane/John} was grateful that {Ruth/John/Jane} had cooked her a meal as a special treat. It was time to relax at last.
14. The end of year test was about to start. {Ruth/Mark} remembered that {Jane/Mark/Ruth} had bought her a new pen for good luck. If only it could magically write the correct answers!
15. The tour of the old castle took place on a moonlit night. {Ann/Bob} was cross that {Sue/Bob/Ann} had reminded her about the ghost in the tower. After all, it was only a story, wasn't it?
16. The university reunion got off to a bad start. {Laura/Roger} was sure that {Jenny/Roger/Laura} had owed her some money for rent and bills. The evening was quite tense after that argument.

17. Auditions at the drama society had been busier than usual. {Sue/Ben} was glad that {Ann/Ben/Sue} had assigned her the biggest role in the play. Anything less would have been unthinkable.
18. It had been a perfect Saturday for doing jobs in the garden. {Jenny/David} wished that {Laura/David/Jenny} had helped her with the weeds along the path. Some people don't like to get their hands dirty!
19. Waiting for hospital treatment can make some people nervous. {Lisa/Gary} dreamt that {Mary/Gary/Lisa} had discharged her on the day of the operation. Luckily that is not what happened in reality.
20. It was the first day of the new weight-loss plan. {Mary/Adam} was sorry that {Lisa/Adam/Mary} had baked her a huge sponge cake with icing. The diet would have to start the next day instead.
21. Not everyone enjoyed singing the national anthem before the match. {Kate/Paul} recalled that {Beth/Paul/Kate} had taught her a new verse with silly words. It had certainly been hard to keep a straight face.
22. The hotel inspection was extremely thorough. {Helen/Simon} knew that {Susan/Simon/Helen} had served her the wrong wine with the meal. But that had been the only mistake all day.
23. The first day of work can be daunting. {Beth/Matt} thought that {Kate/Matt/Beth} had educated her on the pitfalls of being new. But there were plenty of difficulties that hadn't been mentioned.
24. The annual village fundraiser was surprisingly competitive. {Susan/Peter} hoped that {Helen/Peter/Susan} had awarded her the prize for the best pie. All the other entries had been mysteriously damaged.

Experiment 3

1. Suddenly the lights went on and there were police everywhere. {Barry/Megan} saw {Gavin/Megan/Garry} place a gun near him on the ground with great care. The robbery was definitely over now.
2. It was rush hour and the carriage was full of commuters. {Gavin/Diane} saw {Barry/Diane/Gavin} press the alarm beside him on the wall of the train. There was a loud ringing and the train came to a halt.
3. A game of boules can get very competitive. {Hugh/Dawn} spotted {Mike/Dawn/Hugh} inch the ball towards him with a flick of the toe. That would make a difference to the scoreline.
4. The dress rehearsal was about to start. {Mike/Cath} heard {Hugh/Cath/Mike} drag a chair behind him on the newly painted stage. The rest of the cast were late as usual.
5. The atmosphere at the protest was getting more and more tense. {Steve/Lynne} glimpsed {Scott/Lynne/Steve} spray gas near him at the edge of the crowd. Things were beginning to get ugly.
6. Camping is fun but it can get very cold in a tent. {Scott/Clare} felt {Steve/Clare/Scott} pull the rug round him in the middle of the night. Some extra layers of clothing would have been more effective.
7. The rock climbers were in serious danger with the storm moving in. {Colin/Linda} sensed {Nigel/Linda/Colin} tug the rope below him as the rain poured down. The sooner they were off the mountain, the better.
8. The building project was finally in the finishing stages. {Nigel/Sally} watched {Colin/Sally/Nigel} fix a lamp above him with great skill and care. Lucky that someone knew what they were doing!

9. It was another Monday morning on the school bus. {Megan/Barry} saw {Diane/Barry/Megan} slide a bag towards her bit by bit under the seat. Something suspicious was going on!
10. The prison was usually quiet at night. {Diane/Gavin} heard {Megan/Gavin/Diane} tap the wall next to her in the dark of the cell. Perhaps it was a coded message?
11. All the players were still standing in a circle. {Dawn/Hugh} glimpsed {Cath/Hugh/Dawn} place the flag behind her at the end of the game. It hadn't been so hard to win after all.
12. It was the first day of hockey practice. {Cath/Mike} watched {Dawn/Mike/Cath} move a ball in front of her with no effort at all. Not everyone was going to find it that easy!
13. A whole shelf of jam had crashed to the floor in the supermarket. {Lynne/Steve} saw {Clare/Steve/Lynne} check the area round her for any stray bits of glass. An army of staff with mops and buckets soon arrived.
14. The cinema had a distinct and unpleasant odour of old trainers. {Clare/Scott} heard {Lynne/Scott/Clare} spray the air near her with a bottle of deodorant. The heavy chemical scent was almost as overpowering.
15. The concert hall was filling up quickly. {Linda/Colin} watched {Sally/Colin/Linda} wipe the seat beside her with a clean new tissue. Some people can be quite fussy.
16. Everyone took their places for choir practice. {Sally/Nigel} noticed {Linda/Nigel/Sally} put a glass next to her full of a murky liquid. What on earth was it for?
17. After the picnic, everyone was encouraged not to leave any litter. {Mandy/Kevin} spotted {Vicky/Kevin/Mandy} throw a can behind her into a nearby rubbish bin. Good shot!

18. The town meeting was about to start. {Andy/Lucy} noticed {Ryan/Lucy/Andy} place a chair next to him at the front of the hall. An important visitor was expected.

Appendix 3

Items for Experiment 4. Manipulations are shown in parentheses.

1. The squadron paraded through town. Every soldier who knew that {James/Helen} was watching was convinced that {he/she} should wave as the parade passed. The entire town was extremely proud that day.
2. The new construction site was busy. Every builder who saw that {Steven/Joanna} was at the site was told that {he/she} should complete a full day's work. There was no time for any slacking off!
3. The building was unsafe. Every firefighter who believed that {John/Jane} was inside was informed that {he/she} must wait for more help to arrive. Luckily nobody was seriously injured that day.
4. It was another day of training. Every footballer who saw that {Roger/Sarah} was on the pitch was told that {he/she} had to practice for three hours. It was going to be a tough workout for sure.
5. It was very busy at the airport. Every pilot who knew that {David/Diana} was passing through was relieved that {he/she} did not get at all delayed. Air traffic control were running things very efficiently.
6. The boxing night was always popular. Every boxer who heard that {Adam/Emma} was at the event knew that {he/she} needed to really impress the crowd. It was certainly not going to be very easy.
7. A farming trade fair was in town. Every butcher who heard that {Bob/Ann} was at the fair was informed that {he/she} would be able to make a speech. The day was considered to be a great success by all.

8. The large passenger ferry left port. Every sailor who saw that {Mark/Mary} was on deck was informed that {he/she} should be ready for the trip ahead. There was a chance that the sea was going to be rough.
9. The garage was full of cars. Every mechanic who knew that {Peter/Susan} was being lazy was surprised that {he/she} could have a long lunch break. Some people seem to have all the luck!
10. The church had a big congregation. Every priest who heard that {Paul/Katy} was newly ordained agreed that {he/she} should talk to the local people. It was important that everybody knew each other.
11. The mining village was a lively place. Every miner who knew that {Bill/Lucy} was hard at work was told that {he/she} could get some time off next week. It was something worth looking forward to.
12. Good training is very important. Every electrician who heard that {Ben/Amy} was on a course thought that {he/she} should work extra hours each week. It is essential to get as much work experience as possible.
13. People can be superstitious. Every fortune teller who knew that {Helen/James} was a believer thought that {she/he} should make plans about the future. It is always important to try and think ahead.
14. The delivery was due very soon. Every florist who knew that {Joanna/Steven} was visiting the shop was told that {she/he} would like all the new stock. It is important to have a variety of products on sale.
15. Some jobs are not well paid. Every babysitter who believed that {Jane/John} was underpaid was reassured that {she/he} would be given a pay rise soon. That was certainly some good news to hear.

16. It was another day on the ward. Every midwife who noticed that {Sarah/Roger} was at the hospital thought that {she/he} should prepare for a busy day. There is always much to be done at the maternity unit.
17. Life at work can be hectic. Every typist who understood that {Diana/David} was in a hurry decided that {she/he} should try and work a little faster. It was getting late and everyone wanted to go home.
18. The office was very competitive. Every secretary who heard that {Emma/Adam} was visiting was reminded that {she/he} could get a promotion very soon. It was going to be an interesting month at work.
19. The country mansion was beautiful. Every housekeeper who saw that {Ann/Bob} appreciated the garden hoped that {she/he} would be able to enjoy the day. The weather did look like it was going to be warm and sunny.
20. The beauty salon was very popular. Every beautician who saw that {Mary/Mark} was passing by was pleased that {she/he} could have a chat about work. It made the day a little bit more interesting.
21. It was very busy at the hospital. Every nurse who noticed that {Susan/Peter} was on the ward was relieved that {she/he} could go home an hour early. It had been a very long and tiring day.
22. It was a competitive business. Every fashion model who saw that {Katy/Paul} was new was reassured that {she/he} would not be given special treatment. There was definitely some tension in the dressing room.
23. It was a tough schedule. Every cheerleader who noticed that {Lucy/Bill} was having trouble said that {she/he} needed some time off from training. The chance to take a rest was certainly welcome.

24. There was much housework to be done. Every cleaner who heard that {Amy/Ben} was on holiday was informed that {she/he} would work more hours next week. It was the busiest time of the entire year.

Appendix 4

Items for Experiment 5. Manipulations are shown in parentheses.

1. {The queen/The soldiers} felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that {he/she} should wave as the parade passed. The entire town had come out to watch.
2. {The businesswoman/The builders} disliked having to attend the site inspections. Every builder who saw that the businesswoman had arrived for the inspection was told that {he/she} needed to get round the site quickly. This was no time for slacking off!
3. {The policewoman/The firefighters} had been first to arrive at the burning building. Every firefighter who thought that the policewoman was perhaps in danger was told that {he/she} must stop people entering the area. Luckily the situation was dealt with quickly.
4. {The headmistress/The footballers} {was/were} delighted with the team's obvious progress. Every footballer who noticed that the headmistress was shouting and cheering wildly thought that {he/she} could further upset the rival team. This was turning into an interesting match.
5. {The air-hostess/The pilots} had once again been running late. Every pilot who knew that the air-hostess had just about made it was relieved {he/she} did not get at all delayed. This schedule was a bit too tight for everyone.
6. {The showgirl/The weight-lifters} started a thorough warm up session backstage. Every weight-lifter who heard that the showgirl was performing in tonight's show knew that {he/she} needed to

really impress the crowd. Audiences can get rowdy after a drink or two.

7. {The mayoress/The butchers} liked to support the local food fair. Every butcher who heard that the mayoress would be at the baking contest was told that {he/she} might get to taste the pies. There should be enough to go round.
8. {The princess/The sailors} {was/were} keen to inspect the newly refurbished ship. Every sailor who saw that the princess was walking around on the deck was told that {he/she} should prepare for a long trip. It would be weeks before the ship docked again.
9. {The saleswoman/The mechanics} hated working at the garage on hot days. Every mechanic who knew that the saleswoman was in a really bad mood was surprised that {he/she} could have a long lunch break. Some fresh air would be good for everyone.
10. {The bride/The priests} {was/were} anxious about the arrangements for the wedding. Every priest who heard that the bride had sent a very long list of demands agreed that {he/she} should try hard to be reasonable. Weddings can be difficult to organise.
11. {The chairwoman/The miners} {was/were} concerned about the mining company. Every miner who knew that the chairwoman was planning to carry out an inspection was told that {he/she} would need to check the timesheets. Nothing should be missed!
12. {The waitress/The electricians} had seen sparks coming from the wiring system. Every electrician who noticed that the waitress was beginning to get rather worried knew that {he/she} could not just ignore the problem. It's better to fix things straight away.
13. {The emperor/The fortune-tellers} pondered the rather delicate situation. Every fortune-teller who knew that the emperor was a true believer definitely thought that {she/he} should be as honest as possible. It's not always easy to admit the truth.

14. {The bridegroom/The florists} had not been involved in all the decisions. Every florist who knew that the bridegroom was visiting the shop today was told that {she/he} would need to be quite flexible. There might be some disagreement about the flowers.
15. {The prince/The babysitters} {was/were} very unhappy about the new nursery arrangements. Every babysitter who believed that the prince was extremely fussy and spoiled was assured that {she/he} should be treated with some respect. The next few months were going to be difficult.
16. {The king/The midwives} {was/were} waiting anxiously inside the palace. Every midwife who noticed that the king was beginning to get really nervous thought that {she/he} should prepare for a long night. Crowds were already gathering to wait for news.
17. {The businessman/The typists} wanted to leave early on that particular day. Every typist who understood that the businessman had a lot of calls to make thought that {she/he} could try to work slightly faster. Perhaps a cup of tea would help.
18. {The headmaster/The secretaries} needed to be thorough about paperwork. Every secretary who heard that the headmaster was back at the school was reminded that {she/he} had lots to catch up on. The holidays were definitely over now.
19. {The butler/The housekeepers} had been preparing for the inspection. Every housekeeper who was certain that the butler appreciated proper hard work hoped that {she/he} would be able to relax later. It had been a busy couple of weeks.
20. {The postman/The beauticians} strolled up the road towards the salon. Every beautician who noticed that the postman was whistling as usual was pleased that {she/he} could enjoy the early summer mornings. It was going to be a beautiful day.
21. {The policeman/The nurses} {was/were} standing in the car park outside the clinic. Every nurse who noticed that the policeman had

already started to investigate thought that {she/he} should note which drugs were missing. It's important not to miss any details.

22. {The waiter/The fashion models} {was/were} not enjoying the party at all. Every fashion model who saw that the waiter had nearly run out of pricey champagne wished that {she/he} had brought along one more bottle. Soon only the cheap wine would be left.
23. {The sportsman/The cheerleaders} had done well all through the season. Every cheerleader who noticed that the sportsman really needed a proper rest wished that {she/he} could cancel the next training session. There never seemed to be any time off!
24. {The statesman/The cleaners} {was/were} running late again that morning. Every cleaner who knew that the statesman had a bad temper in the mornings wished that {she/he} could work in a different office. It would be better for everyone.

Appendix 5

Items for Experiment 6. Manipulations are shown in parentheses.

1. {Every soldier/The policeman} believed that {the policeman/every soldier} was convinced that he should control the crowd.
2. {Every pilot/The prince} was sure that {the prince/every pilot} thought that he could go a bit faster.
3. {Every priest/The king} learned that {the king/every priest} thought that he should be suspicious of parliament.
4. {Every butcher/The businessman} realised that {the businessman/every butcher} thought that he could adjust the price of the stock.
5. {Every mechanic/The sportsman} saw that {the sportsman/every mechanic} was sure that he could hide the scratch on the car.
6. {Every sailor/The statesman} knew that {the statesman/every sailor} thought that he could swim over to the island.
7. {Every cheerleader/The showgirl} sensed that {the showgirl/every cheerleader} thought that she should follow the team off the field.
8. {Every secretary/The chairwoman} knew that {the chairwoman/every secretary} heard that she might get a new computer.
9. {Every housekeeper/The princess} heard that {the princess/every housekeeper} was saying that she had a lucky escape from the fire.
10. {Every fashion model/the saleswoman} believed that {the saleswoman/every fashion model} knew that she could help the clothing company.
11. {Every nurse/The queen} understood that {the queen/every nurse} feared that she might be criticised in the press.

12. {Every florist/The mayoress} realised that {the mayoress/every florist} hoped that she could do an outstanding job.

Appendix 6

Items for Experiment 9. Manipulations are shown in parentheses.

1. The squadron paraded through town. It looked to {James/Helen} that every soldier was completely convinced that {he/she} should wave as the parade passed. The entire town was extremely proud that day.
2. The new construction site was busy. It appeared to {Steven/Joanna} that every builder was emphatically told that {he/she} should complete a full day's work. There was no time for any slacking off!
3. The building was unsafe. It appeared to {John/Jane} that every firefighter was very quickly informed that {he/she} must wait for more help to arrive. Luckily nobody was seriously injured that day.
4. It was another day of training. It appeared to {Roger/Sarah} that every footballer was very specifically told that {he/she} had to practice for three hours. It was going to be a tough workout for sure.
5. It was very busy at the airport. It seemed to {David/Diana} that every pilot was certainly very relieved that {he/she} did not get at all delayed. Air traffic control were running things very efficiently.
6. The boxing night was always popular. It looked to {Adam/Emma} that every boxer almost definitely knew that {he/she} needed to really impress the crowd. It was certainly not going to be very easy.
7. A farming trade fair was in town. It seemed to {Bob/Ann} that every butcher was quite happily informed that {he/she} would be able to make a speech. The day was considered to be a great success by all.

8. The large passenger ferry left port. It appeared to {Mark/Mary} that every sailor was dutifully informed that {he/she} should be ready for the trip ahead. There was a chance that the sea was going to be rough.
9. The garage was full of cars. It looked to {Peter/Susan} that every mechanic was most happily surprised that {he/she} could have a long lunch break. Some people seem to have all the luck!
10. The church had a big congregation. It seemed to {Paul/Katy} that every priest almost immediately agreed that {he/she} should talk to the local people. It was important that everybody knew each other.
11. The mining village was a lively place. It appeared to {Bill/Lucy} that every miner was enthusiastically told that {he/she} could get some time off next week. It was something worth looking forward to.
12. Good training is very important. It looked to {Ben/Amy} that every electrician understandably thought that {he/she} should work extra hours each week. It is essential to get as much work experience as possible.
13. People can be superstitious. It seemed to {Helen/James} that every fortune teller most definitely thought that {she/he} should make plans about the future. It is always important to try and think ahead.
14. The delivery was due very soon. It looked to {Joanna/Steven} that every florist was quite assuredly told that {she/he} would like all the new stock. It is important to have a variety of products on sale.
15. Some jobs are not well paid. It seemed to {Jane/John} that every babysitter was delightfully reassured that {she/he} would be given a pay rise soon. That was certainly some good news to hear.
16. It was another day on the ward. It looked to {Sarah/Roger} that every midwife almost undoubtedly thought that {she/he} should

- prepare for a busy day. There is always much to be done at the maternity unit.
17. Life at work can be hectic. It appeared to {Diana/David} that every typist very promptly decided that {she/he} should try and work a little faster. It was getting late and everyone wanted to go home.
 18. The office was very competitive. It seemed to {Emma/Adam} that every secretary was continually reminded that {she/he} could get a promotion very soon. It was going to be an interesting month at work.
 19. The country mansion was beautiful. It appeared to {Ann/Bob} that every housekeeper very wholeheartedly hoped that {she/he} would be able to enjoy the day. The weather did look like it was going to be warm and sunny.
 20. The beauty salon was very popular. It looked to {Mary/Mark} that every beautician was really rather pleased that {she/he} could have a chat about work. It made the day a little bit more interesting.
 21. It was very busy at the hospital. It seemed to {Susan/Peter} that every nurse was quite thoroughly relieved that {she/he} could go home an hour early. It had been a very long and tiring day.
 22. It was a competitive business. It seemed to {Katy/Paul} that every fashion model was quickly reassured that {she/he} would not be given special treatment. There was definitely some tension in the dressing room.
 23. It was a tough schedule. It appeared to {Lucy/Bill} that every cheerleader very emphatically said that {she/he} needed some time off from training. The chance to take a rest was certainly welcome.
 24. There was much housework to be done. It seemed to {Amy/Ben} that every cleaner was rather quietly informed that {she/he} would work more hours next week. It was the busiest time of the entire year.

Erklärung

Gemäß §4 (2) 8. der Promotionsordnung der Humanwissenschaftlichen Fakultät der Universität Potsdam vom 17.10.06

Hiermit erkläre ich, Clare S. Patterson, geb. 30.08.1979, dass die ich Arbeit selbständig und ohne unzulässige Hilfe Dritter verfasst habe und bei der Abfassung nur die in der Dissertation angegebenen Hilfsmittel benutzt, sowie alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Potsdam, den _____

Clare S. Patterson