

# ABSTRACT

Title of document:           STRUCTURED ACCESS IN SENTENCE  
COMPREHENSION

Brian W. Dillon  
Doctor of Philosophy, 2011

Directed by: Professor Colin Phillips  
Department of Linguistics

Abstract: This thesis is concerned with the nature of memory access during the construction of long-distance dependencies in online sentence comprehension. In recent years, an intense focus on the computational challenges posed by long-distance dependencies has proven to be illuminating with respect to the characteristics of the architecture of the human sentence processor, suggesting a tight link between general memory access procedures and sentence processing routines (Lewis & Vasishth 2005; Lewis, Vasishth, & Van Dyke 2006; Wagers, Lau & Phillips 2009). The present thesis builds upon this line of research, and its primary aim is to motivate and defend the hypothesis that the parser accesses linguistic memory in an essentially structured fashion for certain long-distance dependencies. In order to make this case, I focus on the processing of reflexive and agreement dependencies, and ask whether or not non-structural information such as morphological features are used to gate memory access during syntactic comprehension. Evidence from eight experiments in a range of methodologies in English and Chinese is brought to bear on this question, providing arguments from interference effects and time-course effects that primarily syntactic information is used to access linguistic memory in the construction of certain long-distance dependencies. The experimental evidence for structured access is compatible

with a variety of architectural assumptions about the parser, and I present one implementation of this idea in a parser based on the ACT-R memory architecture. In the context of such a content-addressable model of memory, the claim of structured access is equivalent to the claim that only syntactic cues are used to query memory. I argue that structured access reflects an optimal parsing strategy in the context of a noisy, interference-prone cognitive architecture: abstract structural cues are favored over lexical feature cues for certain structural dependencies in order to minimize memory interference in online processing.

STRUCTURED ACCESS IN SENTENCE COMPREHENSION

by

Brian William Dillon

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2011

Advisory Committee:  
Professor Colin Phillips, Chair  
Professor Norbert Hornstein  
Professor William Idsardi  
Professor Jeffrey Lidz  
External: Professor Robert DeKeyser, SLA

© Copyright by  
Brian William Dillon  
2011



## Acknowledgments

I'd like to thank first and foremost Colin Phillips for all the support he's given me over the last six years. Colin has spent a significant amount of time patiently listening to me and my half-baked ideas week after week, helping me to sharpen those ideas while simultaneously teaching me how to be a responsible and engaged scientist. There's no question that he has really been an all-around top-notch advisor. I'm still puzzled as to why he thought I was qualified to run an EEG lab way back when, but I'm thankful that he gave me the chance; I wouldn't be where I am today if he hadn't thought so. My time at Maryland has been an extremely frustrating and extremely rewarding experience that I wouldn't trade for anything. For all his time and energy that he's given me over the years, I owe a great debt to him that I really can't sum up in a paragraph. So I'll just leave it this: thank you, Colin!

I've also been very lucky to work with Bill Idsardi during my time at Maryland. I'm thankful for his all-in support and encouragement in pursuing my research ideas, and his incredibly diverse approach to research questions in cognitive science has been an inspiration along the way. From the highest-level discussions of our work to the minute details of hierarchical clustering, it seems there was nothing that I couldn't talk to Bill about, and he was always willing to lend an ear.

Jeff Lidz was also a huge help through the years. His excitement for language research was an important source of encouragement in frustrating times. I benefitted from his insight on too many occasions to count, and he was never too busy to find the time to talk (or, if he was, he didn't let on).

Of course, many thanks are due to Norbert Hornstein for his daily afternoon cookie deliveries, but more importantly, I'm thankful for his many non-cookie related visits to 1413 H. I've really enjoyed and learned a lot from our discussions over the years, and I'm going to miss them.

So many people deserve thanks for the help and friendship they've given me along the way. Thanks to Ming Xiang, who has been a good friend and colleague since the beginning. Ming has a special talent for keeping things in perspective and it's been great to work with her over the years. Thanks to Matt Wagers, who I've learned a lot from over the years, who very patiently taught me how to run SAT, and who's been a good friend to boot.

Thanks also to my good friends and classmates: Pedro Alcocer, Annie Gagliardi, and Shannon Hoerner have helped me time and time again to relax and not take things so seriously, and Alex Drummond, Dave Kush, and Terje Lohndal have given me many impromptu syntax lessons over the years. Thanks to Ewan Dunbar for non-stop math fun. Thanks to Wing Yee Chow for all her help and discussion over the years; a good deal of the research in this thesis would not have been possible without her help.

I also feel lucky to have been part of a phenomenal lab during my time at Maryland, and I'm going to miss everyone from the UMD CNL lab, past and present. Thanks to everyone, seriously.

The ideas presented in this thesis have benefitted from discussions with many, many people. In particular I'd like to thank Rick Lewis and Shravan Vasishth, who have both given me a lot of support and helpful feedback on this work. The computational modeling in this paper would not have been possible without Rick's guidance. Additionally, I am grateful to Taomei Guo, who very generously provided me with support for running the Chinese experiments reported here. Special thanks are also due to a number of amazing researchers who have helped me develop ideas or given me helpful guidance at several stages in this thesis: Rajesh Bhatt, Lyn Frazier, Roger Levy, Brian McElree, Adrian Staub, and Amy Weinberg.

The research I report here was supported by a number of outstanding research assistants who I've had a lot of fun working with. Many thanks to Peiyao Chen, Fengqin Liu, Alan Mishler, Mike Shvartsman, Shayne Sloggett, and Angela Stanley.

Last but definitely not least: thank you, Jorge, for being my best friend throughout all of this and for being so supportive of my choices over the years. I'm incredibly lucky to have you in my life, and I hope to be so lucky for a long time to come.

## Table of Contents

<b>Acknowledgments</b>	<b>ii</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
<i>Models of memory and syntactic representation</i>	9
<i>Cue-based parsing and the psycholinguistic enterprise</i>	15
<i>Outline of the dissertation</i>	19
<b>Chapter 2: The argument from interference: English agreement and reflexives</b>	<b>24</b>
<i>The argument from interference</i>	26
<i>Partial-match interference in subject-verb agreement</i>	31
<i>Lack of interference in reflexive dependencies</i>	40
<i>Experiment 1: Direct comparison of agreement and reflexives</i>	47
Participants	49
Stimuli	49
Offline judgments	51
Procedure	53
Data Analysis	55
Results: Agreement	58
Results: Reflexives	61
Direct comparison of interference effect	63
Discussion of Experiment 1	64
<i>Experiment 2: Agreement revisited</i>	68
Participants	69
Materials	70
Procedure	70
Data Analysis	71
Results	72
Discussion	76
<i>Experiment 3: Reflexives revisited</i>	80
Participants	80
Materials	81
Data Analysis	81
Results	81
Discussion	85
<i>Overview of Experiments 1-3</i>	87
<i>Discussion</i>	90
<i>Reflexive interpretation</i>	96
<i>Attraction in reflexive production</i>	99
<i>The difference between agreement and reflexives</i>	107
<i>Conclusion</i>	113
<b>Chapter 3: Revisiting the interference argument: optimal information retrieval</b>	<b>116</b>
<i>Rational memory access</i>	119
<i>Implementing rational retrieval: the model</i>	124
<i>Modeling reflexive and agreement dependencies</i>	129
The relation between model predictions and experimental findings	129

Modeling feature-based and structured access	133
The model	135
<i>Experiment 4: comparison of interference effect for agreement and reflexives</i>	138
<i>Experiment 5: comparison of access strategies for reflexives</i>	146
<i>Interim conclusions</i>	152
<i>The predictions of rational memory access models</i>	153
<i>Relation to previous work</i>	164
<i>Conclusion</i>	173
<b>Chapter 4: Processing long-distance reflexives in Mandarin Chinese</b>	<b>175</b>
<i>Linking structured access and structured search</i>	178
<i>The argument from time course</i>	180
<i>Chinese long-distance anaphors</i>	192
<i>Experiment 6: SAT Evidence</i>	200
Participants	201
Materials	202
Procedure	205
Data Analysis	207
Empirical Accuracy Analysis	210
Model Selection Analysis	211
Parameter Estimation Analysis	213
Discussion	222
<i>Experiment 7: ERP Evidence</i>	226
Participants	231
Materials	231
Procedure	233
EEG Recording	234
EEG Analysis	234
Results: Behavioral Data	236
Results: ERP Data	237
Discussion	239
<i>General Discussion</i>	244
<i>Locality bias in ziji dependencies</i>	246
<i>Alternative accounts of the data</i>	250
<i>Linguistic and discourse antecedents for ziji</i>	253
<i>Conclusion</i>	255
<b>Chapter 5: Structured access across dependency types</b>	<b>257</b>
<i>A puzzle for the hypothesis of structured access in comprehension</i>	260
<i>Revisiting Chinese anaphors</i>	265
<i>The scope of structured access: contrasting ziji &amp; ta-ziji</i>	270
<i>Experiment 8: Ziji and sub-commanding antecedents</i>	276
Participants	276
Stimuli	276
Procedure	278
Offline judgments	279
Data Analysis	280
Results	282
<i>Experiment 9: Ta-ziji and sub-commanding antecedents</i>	283
Participants	283
Stimuli	284

Procedure	285
Offline judgments	285
Data Analysis	287
Results	287
Discussion	289
<i>Structured access as syntactic parsing</i>	294
Agreement as uninterpreted syntax	301
<i>Structured access as an optimal access strategy</i>	304
<i>Structured access and Mandarin anaphors</i>	312
<i>The footprint of structured access</i>	314
<i>Conclusion</i>	320
<b>Chapter 6: Conclusion</b>	<b>322</b>
<i>No consideration of illicit antecedents: Experiments 1-5</i>	323
<i>No immediate access to distant but accessible antecedents: Experiments 6-9</i>	324
<i>Interpreting interference effects</i>	325
<i>Structured access and the architecture of comprehension</i>	326
Structured access as an optimal adaptation	327
Blocking effects	329
Negative constraints	330
<i>Conclusion</i>	332
<b>Appendix A: Retrieval schedules for models in Chapter 3</b>	<b>333</b>
<i>Agreement conditions:</i>	333
<i>Reflexive conditions:</i>	334
<b>References</b>	<b>335</b>

## List of Tables

<b>Table 2.1:</b> Summary of agreement conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined. ....	50
<b>Table 2.2:</b> Summary of reflexive conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined. ....	50
<b>Table 2.3:</b> Mean judgments and standard error by subjects for Experiment 1 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable. ....	52
<b>Table 2.4:</b> Table of means (in ms where applicable) for agreement conditions for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	58
<b>Table 2.5:</b> Table of means (in ms where applicable) for reflexive conditions for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	60
<b>Table 2.6:</b> Summary of fixed effects for best-fit models on agreement conditions at the critical agreeing verb region, including <i>t</i> -values ( <i>z</i> -values for first-pass regression probability data). An asterisk (*) indicates significance at $\alpha = 0.05$ , while a cross (†) indicates significance at $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds. ....	62
<b>Table 2.7:</b> Summary of fixed effects for best-fit models on reflexive conditions at the critical reflexive region, including <i>t</i> -values ( <i>z</i> -values for first-pass regression probability data). An asterisk (*) indicates significance at $\alpha = 0.05$ , while a cross (†) indicates significance at $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds. ....	63
<b>Table 2.8:</b> Summary of agreement conditions in Experiment 2. Regions included in the analysis are underlined. ....	70
<b>Table 2.9:</b> Table of means (in ms where applicable) for Experiment 2, agreement conditions with a singular head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	71
<b>Table 2.10:</b> Table of means (in ms where applicable) for Experiment 2, agreement conditions with a plural head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	72
<b>Table 2.11:</b> Summary of fixed effects for best-fit models at the critical agreeing verb region in Experiment 2, including <i>t</i> -values ( <i>z</i> -values for first-pass regression probability data). An asterisk (*) indicates significance at $\alpha = 0.05$ ; a cross (†) indicates significance at $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds. ....	74
<b>Table 2.12:</b> Summary of reflexive conditions in Experiment 3. Regions included in the analysis are underlined. ....	81
<b>Table 2.13:</b> Table of means (in ms where applicable) for Experiment 3, reflexive conditions with a singular head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	82
<b>Table 2.14:</b> Table of means (in ms where applicable) for Experiment 3, reflexive conditions with a plural head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses. ....	84

<b>Table 2.15:</b> Summary of fixed effects for best-fit models at the critical reflexive region in Experiment 3, including <i>t</i> -values ( <i>z</i> -values for first-pass regression data). An asterisk (*) indicates significance at $\alpha = 0.05$ ; a cross (†) indicates significance at $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds..	84
<b>Table 3.1:</b> Distribution of NP feature match and match of the inaccessible NP to retrieval cues across experimental conditions.....	154
<b>Table 4.1:</b> Summary of conditions in experiment: Critical <i>ziji</i> conditions. Critical conditions are 1-2 and 4-5; conditions 3 and 6-9 were included for purposes of <i>d'</i> scaling (see text). .....	207
<b>Table 4.2:</b> Difference in parameter estimates between local and long-distance configurations on the critical <i>ziji</i> and control conditions. Values greater than 0 indicate a processing advantage for local antecedents. Standard errors by subject are in parentheses. ....	222
<b>Table 4.3:</b> Summary of conditions in Experiment 2. ....	233
<b>Table 4.4:</b> Table of experimental fixed effects (coefficients in $\mu V$ , with standard error). Experimental fixed effects only shown if the best-fit model included a significant interaction of experimental effect with <i>anteriority</i> and <i>laterality</i> . † = $p < 0.1$ , * = $p < 0.05$ , ** = $p < 0.01$ . ....	239
<b>Table 5.1:</b> Summary of interference properties of long-distance dependencies. ....	262
<b>Table 5.2:</b> Critical conditions from Experiment 8. Region breaks are indicated by slashes. ....	277
<b>Table 5.2:</b> Mean judgments and standard error by subjects for <i>ziji</i> Experiment 8 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable. ....	279
<b>Table 5.4:</b> Critical conditions from Experiment 9. Region breaks are indicated by slashes. ....	285
<b>Table 5.5:</b> Mean judgments and standard error by subjects for <i>ziji</i> Experiment 8 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable. ....	286
<b>Table 5.6:</b> Summary of fixed effects for best-fit models at the critical <i>ziji</i> and spillover regions, including <i>t</i> -values .....	288
<b>Table 5.7:</b> Summary of access properties of long-distance dependencies. ....	295
<b>Table A.1:</b> Constituent creation times and feature makeup for agreement conditions. .	333
<b>Table A.2:</b> Schedule of retrievals and cue sets. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = critical retrieval to attach DP1 to BE.....	333
<b>Table A.3:</b> Constituent creation times and feature makeup for agreement conditions. .	334
<b>Table A.4:</b> Schedule of retrievals and cue sets for the structured access reflexive model. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = attachment of DP1 to VP2; R4 = attachment of REFL to VP2; R5 = critical retrieval of REFL's antecedent. ....	334
<b>Table A.5:</b> Schedule of retrievals and cue sets for the feature-based access reflexive model. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = attachment of DP1 to VP2; R4 = attachment of REFL to VP2; R5 = critical retrieval of REFL's antecedent. ....	334

## List of Figures

<b>Figure 1.1:</b> Hypothetical processes for processing a reflexive pronoun, demonstrating different modes of antecedent activation. ....	6
<b>Figure 1.2:</b> Structured access mechanisms use a narrow syntactic set of cues to access the reflexive's antecedent (left panel). Feature-based access mechanisms deploy a wider range of cues that to access the representation of the antecedent (right panel).....	7
<b>Figure 2.1:</b> Target and distractor NPs when processing a reflexive pronoun in English. ....	27
<b>Figure 2.2:</b> Multiple match and partial match interference configurations. For multiple match interference, the target memory is a perfect match to the search cues, but the distractor overlaps in some feature content with the target. In the partial match situation, neither target nor distractor is a perfect match to the search cues. ....	29
<b>Figure 2.3:</b> Interference effects in Experiments 1-3 in Pearlmutter et al (1999). The interference effect is the difference in RTs at the critical region that is due to manipulating the feature content of the distractor, as shown.....	35
<b>Figure 2.4:</b> Interference effects in Wagers et al (2009). ....	37
<b>Figure 2.5:</b> Interference effects for early (first-pass) measures in Sturt (2003).....	42
<b>Figure 2.6:</b> Interference effects for late (total time) measures in Sturt (2003). ....	43
<b>Figure 2.7:</b> Mean first-pass reading time at the critical region in Experiment 1. Error bars show standard error by participants. ....	65
<b>Figure 2.8:</b> Mean total reading time at the critical region in Experiment 1. Error bars show standard error by participants. ....	66
<b>Figure 2.9:</b> Mean first-pass reading time at the critical region in Experiment 2. Error bars show standard error by participants. ....	77
<b>Figure 2.10:</b> Mean total reading time at the critical region in Experiment 2. Error bars show standard error by participants. ....	78
<b>Figure 2.11:</b> Mean first-pass reading time at the critical region in Experiment 3. Error bars show standard error by participants. ....	85
<b>Figure 2.12:</b> Mean total reading time at the critical region in Experiment 3. Error bars show standard error by participants. ....	86
<b>Figure 2.13:</b> Interference effects (in ms) observed for in total time measures across Experiments 1-3. Error bars reflect 95% CI by participants.....	88
<b>Figure 2.14:</b> Total reading times (ms) at critical region, combining similar conditions across Experiments 1-3. Error bars are standard error by participants. ....	89
<b>Figure 3.1:</b> Average activation for target (black) and distractor (red) NPs for a sentence that shows partial-match interference at the agreeing verb. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural verb <i>were</i> . ....	131
<b>Figure 3.2:</b> Average activation for target (black) and distractor (red) NPs for a sentence that shows partial-match interference at the reflexive. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural reflexive <i>themselves</i> .....	133



<b>Figure 3.3:</b> Percentage of retrieval of distractor NP, for all parameterizations ( $n=324$ ), at critical probe position for agreement and feature-based reflexive models. ....	141
<b>Figure 3.4:</b> Difference in interference error between agreement and feature-based reflexive models ( $n=324$ ). Error bars indicate 95% confidence intervals. ....	142
<b>Figure 3.5:</b> Predicted interference effect ([+intr]-[-intr] conditions) for agreement and feature-based reflexive models ( $n=324$ ). Error bars indicate 95% confidence intervals. ....	144
<b>Figure 3.6:</b> Difference in predicted interference effect (agreement-reflexive conditions) for all models ( $n=324$ ). Error bars indicate 95% confidence intervals. ....	145
<b>Figure 3.7:</b> Percentage of incorrect retrievals for reflexive conditions for feature-based and structured access models ( $n = 324$ ). ....	148
<b>Figure 3.8:</b> Comparison of predicted interference effects (solid) for reflexive conditions ( $n = 324$ ) and observed reflexive interference effects from Experiments 1 and 3 (by participants, $n = 72$ ). Error bars indicate 95% CI. ....	149
<b>Figure 3.9:</b> Relationship between interference error and interference effect on average retrieval latency. Blue points indicate comparisons between grammatical conditions, and red points are ungrammatical comparisons, for each of 324 parameterizations. ....	157
<b>Figure 3.10:</b> Effect of interference in multiple match and partial match comparisons. The inhibition in multiple match interference is driven by decreased retrieval latencies on the target noun, due to feature overlap. The facilitation in partial match interference is an increased overlap in the distribution of target and distractor retrieval distributions. The race aspect of the retrieval process leads to an overall facilitation effect, which unambiguously indicates that incorrect access has occurred online. ....	160
<b>Figure 4.1:</b> Architecture of a content-addressable memory, from Gallistel & King (2009). Memories consist of three bits, and each bit is probed in parallel for a match. In the present case, all memories that contain a 0 in third position are returned in response to a retrieval query. ....	181
<b>Figure 4.2:</b> Hypothetical SAT curves showing a) two processes that differ in asymptotic accuracy only (top panel) and b) two processes that differ in processing speed only (bottom panel) (figure from Öztekin & McElree 2010). Vertical and horizontal lines indicate that point at which each curve is at 50% of asymptotic accuracy. ....	186
<b>Figure 4.3:</b> Example of a structured search process for finding <i>ziji</i> 's antecedent in the sentence <i>Lisi shuo fengbao hai-le ziji</i> "Lisi said the storm harmed him". The hypothetical structural cues do not allow comprehenders to rule out consideration of the local subject <i>fengbao</i> "storm". Thus comprehenders must evaluate multiple subject positions in the search for the correct antecedent <i>Lisi</i> . This structured access predicts that processing time should grow with the number of subject positions that need to be evaluated. ....	199
<b>Figure 4.4:</b> Example of feature-based access for finding <i>ziji</i> 's antecedent in the sentence <i>Lisi shuo fengbao hai-le ziji</i> "Lisi said the storm harmed him". The mixture of structural and semantic cues allow direct access to the correct	

antecedent <i>Lisi</i> . Feature-based access predicts that processing time should be constant with the number of subject positions that need to be evaluated. ....	200
<b>Figure 4.5:</b> SAT functions for LD and local antecedent <i>ziji</i> conditions with fully saturated models ( $2\lambda-2\beta-2\delta$ ), over average data (not averaged parameters). Accuracy is scaled to show proportion of asymptote; vertical bars indicate time point at which 50% accuracy is reached. ....	215
<b>Figure 4.6:</b> SAT functions for LD and local control conditions with fully saturated models ( $2\lambda-2\beta-2\delta$ ), over average data (not averaged parameters). Accuracy is scaled to show proportion of asymptote; vertical bars indicate time point at which 50% accuracy is reached. ....	216
<b>Figure 4.7:</b> Average asymptotic accuracy ( $\lambda$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show $\pm 1$ SE, corrected for between-participant variance. ....	217
<b>Figure 4.8:</b> Average rate ( $\beta$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show $\pm 1$ SE, corrected for between-participant variance. ....	218
<b>Figure 4.9:</b> Average intercept ( $\delta$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show $\pm 1$ SE, corrected for between-participant variance. ....	219
<b>Figure 4.8:</b> Average speed ( $\delta+\beta^{-1}$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show $\pm 1$ SE, corrected for between-participant variance. ....	220
<b>Figure 5.1:</b> Structured search for <i>ziji</i> forces the parser to consider the local antecedent position before the sub-commanding antecedent. ....	272
<b>Figure 5.2:</b> Direct access for <i>ta-ziji</i> based on semantic or discourse prominence allows the parser to immediately access the sub-commanding antecedent. ....	273
<b>Figure 5.3:</b> Region-by-region mean log reading times for Experiment 8. Error bars represent $\pm 1$ standard error, by participants, corrected for between-participant variance. ....	283
<b>Figure 5.4:</b> Region-by-region mean log reading times for Experiment 9. Error bars represent $\pm 1$ standard error, by participants, corrected for between-participant variance. ....	288
<b>Figure 5.5:</b> Effect of embedded animate (embedded [+animate] subject – embedded [-animate] subject) on spillover reading times in Experiments 8 ( <i>ziji</i> ) and 9 ( <i>ta-ziji</i> ). Error bars represent 95% confidence interval, by participants. ....	290
<b>Figure 5.6:</b> Rate of occurrence of gender attraction errors across languages. Figure from Lorimor, Bock, Zalkind & Sheyman (2008). ....	311

## **Chapter 1: Introduction**

In online language comprehension, the information contained in a sentence unfolds over time. In order to successfully understand a sentence, a comprehender must have a mechanism for maintaining and combining the information contained in each of the words of the sentence. Because language comprehenders perceive linguistic input in a sequential, left-to-right order, the basic act of understanding a sentence must make use of working memory to manage the information conveyed by the incoming speech.

It is tempting to view the working memory system for linguistic comprehension as carrying out the relatively straightforward task of combining adjacent words into higher-order units of meaning and syntactic structure. However, this simple picture is rapidly complicated by the fact that human language is full of ‘long-distance’ dependencies between words. These are relationships between two non-adjacent, and possibly quite distant, words in a sentence. One clear example of

such a dependency is the relationship between a pronoun (an *anaphor*) and its referent (the *antecedent*). Other common examples include subject-verb agreement and the relationship between fronted *wh*-words and the verbs they combine with. Long-distance dependencies present unique computational challenges to the parser, and suggest the need for sophisticated methods of information storage and retrieval that are flexible enough to handle the range of these dependencies that comprehenders are bound to encounter in every conversation. The nature of these dependencies, and the fact that comprehenders on average have little trouble understanding them, lead to important questions about the relationship between linguistic structure and working memory mechanisms. In this work I will attempt to address the following theoretical question: *how do linguistic representations and working memory processes interact to allow the construction and interpretation of long-distance dependencies?*

In the present work, I attempt to articulate and defend the hypothesis that syntactic structure provides the crucial information that aids comprehenders in organizing and retrieving information in linguistic working memory stores. My primary claim is that for certain long-distance linguistic dependencies comprehenders employ a *structured access* mechanism. For these dependencies, comprehenders access linguistic memory by deploying uniquely structural information, selectively attending to these features over otherwise useful morphological and semantic information.

This claim may appear unremarkable, as there is no shortage of psycholinguistic research that suggests that the grammar is deployed rapidly online to structure incoming material (Frazier 1998; Phillips, Wagers & Lau 2010). However,

in recent years this view has been challenged on several fronts. A number of researchers have suggested that grammatical relations are at best deployed as violable constraints online alongside morphological and semantic constraints (Tabor, Galantucci & Richardson, 2004; Lewis & Vasishth 2005; Van Dyke 2007); at worst, they are not at all deployed in initial parsing (Townsend & Bever 2001; Ferreira, Bailey & Ferraro 2002). Ferreira & Patson (2007) provide a useful summary and a clear articulation of the opposite position. These challenges reflect very different architectural commitments, ranging from claims about the subsymbolic nature of online linguistic computation (Tabor et al 2004), to the primacy of heuristic strategies in parsing (Townsend & Bever 2001; Ferreira et al 2002), or constraints on the representation of linguistic information in working memory (Lewis & Vasishth 2005; Van Dyke 2007).

This last challenge, the difficulty of representing structured syntactic relations in online working memory, is the focus of the present work. There is an emerging consensus that the computational properties of the sentence processor's memory architecture mirror those found in domain-general working memory (McElree 2000; Gordon, Hendrick & Johnson 2001; Gordon, Hendrick & Levine 2002; McElree, Foraker & Dyer 2003; Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006; Wagers 2008). One claim that has come to be associated this view is that the parser forced to construct syntactically illicit representations because of the constraints that the memory architecture places on memory access (Van Dyke 2007; Vasishth, Brüssow, Drenhaus & Lewis 2008). However, this claim of structural fallibility depends more on specific hypotheses about the type of information used to access

memory than hypotheses about the computational properties of the memory architecture itself. The primary goal of this thesis is to argue that although the memory architecture does place interesting constraints on representation of linguistic information during online comprehension, comprehenders are nonetheless able to engage structured access mechanisms that effectively target and access specific syntactic positions during parsing.

It should be clear from the outset that in arguing for structured access, I am making a claim about the type of information recruited to access working memory in parsing long-distance dependencies, rather than an endorsement of any particular memory architecture; “structured access” is intended as a general term for strategies that privilege structural information in accessing memory, without implying any commitment to a particular theory of memory. Thus the claim of structured access stands in contrast to mechanisms that use a wider range of morphological, syntactic and semantic features in parallel to access working memory, as has been advocated in recent years by a number of authors (Van Dyke & Lewis 2006; Van Dyke 2007; Vasishth et al 2008; Wagers, Lau & Phillips 2009).

To see the difference between the two sorts of account, consider the problem of finding an antecedent for a bound reflexive pronoun. There are a number of syntactic constraints on which structural positions a reflexive’s antecedent can occupy (Chomsky 1981), in addition to a formal requirement of feature concord between the reflexive pronoun and its antecedent in English. Upon encountering a reflexive pronoun in English, the processor’s task is to construct a legal binding chain, which presumably starts by accessing potential antecedents. There are a number of ways that

antecedent reactivation might occur. For example, the parser might employ the full range of the information in the reflexive to find the correct antecedent, using morphological, syntactic, and semantic cues in parallel in a *feature-based access* mechanism (e.g. Badecker & Straub 2002; Patil, Vasishth & Lewis 2011). Alternatively, it might engage a *structured access mechanism* that selectively attends to syntactic information in retrieving potential antecedents (Nicol 1988; Sturt 2003a; Xiang, Dillon & Phillips 2009). The primary difference between these views is their prediction about the impact of structurally inaccessible, but feature matched nouns in the process of resolving the reflexive's reference. A comparison of two sorts of process is considered in Figures 1.1 and Figure 1.2. One plausible procedure for processing a reflexive involves activating a set of candidate antecedents based on a mixture of morphological, semantic, and syntactic feature information, and then constructing the binding dependency with one of the activated antecedents. This amounts to a claim of temporary, but spurious ambiguity: the parser is temporarily considering two antecedents for the reflexive, even though only one ends up grammatically licensed. This temporary ambiguity that arises in the feature-based account is schematized in the right panel of Figure 1.2. On a structured access account, however, antecedent reactivation proceeds using only syntactic information, and so the antecedent selection process is blind to the feature content. Only structurally licit antecedents are considered, as in the left panel of Figure 1.2. The syntactic information is deployed as a hard constraint, and ambiguity does not arise at any level of processing.

The man [<sub>CP</sub> who saw John] hurt **himself** ...

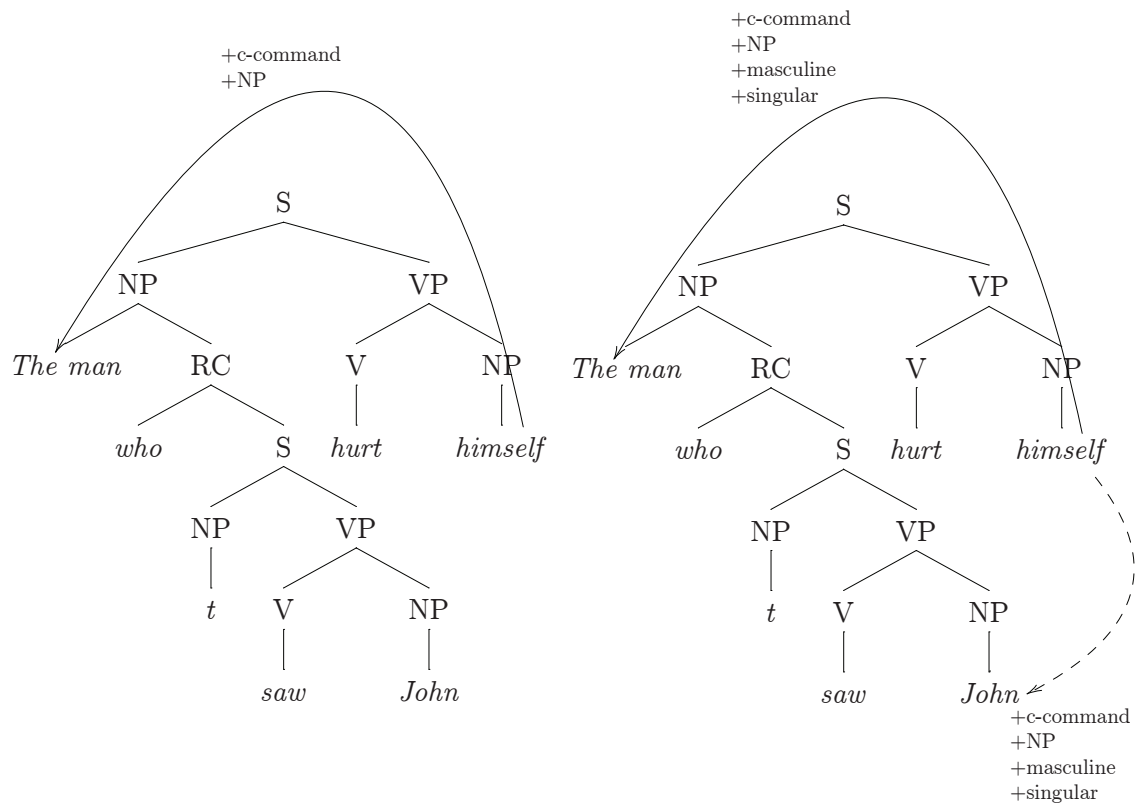
<u>Structured access</u>	<u>Feature-based access</u>	
{the man}	{the man, John}	<i>Antecedent activation</i>
<the man=himself>	<the man=himself>	<i>Binding</i>

**Figure 1.1:** Hypothetical processes for processing a reflexive pronoun, demonstrating different modes of antecedent activation.

Interestingly, for at least some long-distance dependencies, the feature-based approach to memory access suggested by the right panel of Figure 1.2 appears to be correct. For example, in subject-verb agreement dependencies, morphological features appear to be used in the generation of candidates for the agreement relation, leading to spurious illusions of grammaticality in agreement formation (Clifton, Frazier & Deevy 1999; Pearlmutter, Garnsey & Bock 1999; Wagers et al 2009). This finding is expected given recent models of linguistic working memory, and this has led some authors to propose that such feature-based access is the primary manner of memory access (Lewis et al 2006; Van Dyke 2007; Vasishth et al 2008; Patil et al 2011).

The generality of feature-based access in online parsing remains unclear, however. A number of results appear to be suggestive of structured, rather than feature-based access mechanisms online (Nicol 1988; Sturt 2003a; Xiang et al 2009),





**Figure 1.2:** Structured access mechanisms use a narrow syntactic set of cues to access the reflexive's antecedent (left panel). Feature-based access mechanisms deploy a wider range of cues that to access the representation of the antecedent (right panel).

the possibility that there are distinct strategies that comprehenders employ to organize and access information during parsing. This thesis explores this possibility and builds support for a uniquely structural access mechanism in comprehension. The main prediction of a structured access mechanism is that whenever it is engaged, comprehenders should selectively retrieve information based on its syntactic position, rather than its feature content. There are two types of empirical evidence that I offer for this claim. In chapters 2 and 3, I present experimental evidence that feature-matching but structurally inaccessible antecedents do not impact early memory access for English reflexives. Chapters 4 and 5 demonstrate the converse situation: even in

the presence of structurally accessible feature-matching candidates, feature-mismatching structurally accessible antecedents are accessed in the construction of long-distance reflexive dependencies in Mandarin Chinese. Across all studies, comprehenders appear to reactivate particular syntactic positions during comprehension, in the face of both inaccessible (Chapters 2-3) and accessible (Chapters 4-5) feature-matching material. These two sources of evidence confirm the central prediction of a structured access theory of memory access: it is primarily syntactic position, rather than feature match, that guides access to linguistic memory for the dependencies considered. In addition to experimental evidence, I also present evidence from explicit computational models that further supports the conclusion that memory access proceeds in a structured fashion. Lastly, I will turn to a critical assessment of claims that appear to run counter to my central argument, showing that the empirical support for feature-based access is actually rather limited.

Although it is likely that short-term working memory places interesting constraints on representing syntactic hierarchy during parsing, the arguments presented here stress that these constraints do not inhibit the parser's ability to engage in structured access. Rather, the constraints on representation that may accompany a noisy content-addressable cognitive architecture may in fact provide the key to understanding the role of structured access in parsing: in Chapter 5 I argue that structured access reflects an optimal strategy for an interference-prone parser. In particular, by limiting search cues to the most predictive and minimal set of cues, disruptive interference can be minimized. In general, deploying uninformative and superfluous cues to memory access, such as morphological features in a feature-based

access account, increases the risk of memory interference with no countervailing benefit for memory access. If this claim is correct, then for the dependencies considered here, structured access is a rational strategy for the parser to pursue. In this way, the disruptive effect of memory interference actually provides functional pressure for abstraction in parsing.

### *Models of memory and syntactic representation*

The role of linguistic structure in memory access has been an active area of research for recent work that has focused on the architecture of linguistic memory in sentence comprehension. There has been an increasing amount of interest in the fine computational details of the memory architecture of the parser, and research on this front has become influential in thinking about the relationship between syntactic competence and online patterns of processing difficulty.

One major goal of this recent line of research is an explicit characterization of the computational characteristics of the parser's working memory architecture. The starting point for this work was the intuition that insights from research on working memory processes from other domains of cognition could be ported in a fairly straightforward way to model the memory architecture of the language processor. One explicit characterization of this intuition is given by Lewis & Vasishth, who state that the "goal... is to explain as much detailed psycholinguistic phenomena as possible with independent principles of cognitive processing" (2005: pp. 377). The guiding hypothesis of this research is that a small set of general computational

principles govern memory access in sentence comprehension, just as they do in any cognitive task that requires retention of information in a short-term memory store (McElree 2000; Gordon et al 2001; Gordon et al 2002; McElree, Foraker & Dyer 2003; Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006; Wagers 2008; see also Greene, McKoon & Ratcliff 1992 for similar ideas in the domain of reference resolution). This approach has had considerable empirical purchase and has been supported by successful computational models (Lewis & Vasishth 2005; Vasishth et al 2008; Wagers 2008).

Parsing models based on this idea form a heterogeneous group of sentence processing theories that are collectively referred to as *cue-based approaches* to parsing. A great deal of the excitement that surrounds these frameworks is the promise that whatever principles govern the parser's behavior are the same general principles that are thought to govern information processing across cognitive domains. Though memory considerations have long been used to motivate parsing principles (Yngve 1960; Miller & Chomsky 1963; Kimball 1973; Frazier & Fodor 1978), this line of research actually makes the stronger claim that the memory systems that enable sentence processing are essentially identical to those recruited for more general working memory tasks, with no role for linguistically specialized memory mechanisms such as hold cells or stacks (Wanner & Maratsos 1978; Marcus 1980). This highlights the excitement inherent in these approaches; processing principles that have long been established and debated could in principle be a simple reflex of 'memory limitations' in a broad sense (see also Bever 1970), even if one

maintains that the memory systems for language are separate from those seen in other cognitive domains (as in Caplan & Waters 1998; Lewis & Vasishth 2005).

Although I review the main empirical arguments for cue-based approaches to parsing in Chapter 3, as well as their formal characterization, it is worthwhile to briefly summarize the main theoretical commitments of this approach to frame the discussion that follows. It is important to bear in mind that each of these theoretical commitments are independently motivated insofar as they are drawn from theories of short-term memory access in more ‘domain-general’ areas of cognition (i.e. list memory, McElree & Doshier 1989). Lewis, Vasishth and Van Dyke (2006) present an explicit and succinct characterization of the relevant computational principles assumed across various implementations of the cue-based parsing view. The first, and arguably the most crucial, is the assumption of a content-addressable memory architecture. In a content-addressable architecture, stored pieces of information (*memories*) are indexed and retrieved according to the content of their representation, rather than their location in memory (Kohonen 1980). For example, rather than storing a *wh*-filler in a special hold cell (Wanner & Maratsos 1978) for later retrieval, in a content-addressable architecture it simply suffices to mark the filler with [+*wh*] content. When needed, the *wh*-filler can be accessed in light of bearing the crucial [+*wh*] feature in its representation, and it need not be stored in any particular location or cell in memory.

There are several consequences of adopting this manner of indexing and retrieving memories. The first is the *direct access* property of these architectures (McElree & Doshier 1989). Direct access refers to the fact that memories with the

target content are immediately accessed in memory access, without a need to first traverse or check memories that do not match the desired content. This means that memories are retrieved in constant time relative to the size of the search space, an important point that forms the basis for discussion in Chapter 4. Another consequence of adopting a content-addressable architecture is that of retrieval or encoding interference (Kohonen 1980). When multiple memories contain the target content (in full or in part), then there is a possibility that something other than the desired memory will be retrieved, or that the desired memory will be more difficult to recover. Put differently, the degree to which a memory has unique content in its representation is the degree to which it may be seen as having a unique memory location. The less unique its content-defined ‘address’ is, the less reliable access to that memory will be. These two computational properties form the basis for the two main empirical arguments—arguments from time course and arguments from interference—that have been offered for a content-addressable architecture.

The second main theoretical commitment of a cue-based parsing approach is that of a limited focus of attention. The adoption of a limited focus of attention for concurrent processing of elements is well supported in other cognitive domains (McElree & Doshier 1989; Lewis et al 2006), but somewhat less is known about the size or character of focal attention in sentence processing, and this question is very much a focus of current research (Wagers & McElree 2009). In cue-based parsing approaches, the assumption of a limited focus of attention attributes a good deal of explanatory power to our characterization of memory representation and access. This is because, by hypothesis, a limited focus of attention for concurrently processing

elements entails that sentence processing involves a good deal of passing information between the active processing state and the more passive memory representation state (Lewis & Vasishth 2005; Wagers 2008). If sentence-processing crucially relies on a skillful shunting of information into and out of active and passive processing states, then the manner of retrieving that information and restoring it to attention takes on a central role in the characterization of the parsing process. The commitment to a limited focus of attention will not be directly addressed in this thesis, although insofar as it foregrounds the role for memory access and information retrieval, it is an important assumption that underlies the arguments presented here.

With these two theoretical commitments in mind, a cue-based approach to parsing maintains that the normal process of constructing grammatical representations of speech input proceeds by storing the pieces of structure in a ‘passive’ content-addressable memory store, and carrying out targeted retrievals of structure in order to engage processes related to the retrieved structure. In the remainder of this thesis, the terms *memory retrieval* and *memory access* are used interchangeably to refer to the process by which a given piece of structure is restored from a passive storage state to a state that is active for processing. In cue-based models of parsing, the generation of structure is by hypothesis parasitic on the retrieval process (Lewis & Vasishth 2005). Thus in order to draw more explicit parallels with previous work, it is useful to occasionally refer to memory retrieval as the *generation of structure*. I will use the term *cue* to describe the information that is used to access memory (or, equivalently, the information that is used to generate linguistic structure). A cue in the context of linguistic processes will generally refer to an atomic feature of the representation used

for a given piece of structure, such as a distinguishing semantic or morphological feature of a lexical item. A formal model of the retrieval process, and the relationship of retrieval cues to the properties of memory retrieval will be presented in Chapter 3. It can be seen that the name ‘cue-based parsing’ is fairly transparent: the approach maintains that the cues used to access memory representations represent a major informational bottleneck in the parsing process.

The cue-based parsing approach is compatible with a number of different implementations (see, e.g. Lewis & Vasishth 2005 versus Van Dyke 2007), and there are a number of theoretical questions that become important upon adopting this general framework for understanding parsing. One is the relationship between prospective structure-building processes (i.e. ‘active parsing’) and retrospective memory access processes, a topic that I will not address in detail here. A second theoretical question is what exactly the nature of the cues used in parsing is, and what their relation to a speaker’s grammatical knowledge is. An in-depth treatment of this question is one of the primary goals of this thesis. If we take the parallel between the information used to access memory and the information used to generate structure seriously, it can be seen that this question is an alternative way of understanding familiar questions of information encapsulation in the parsing process: what information is used to build structure, and when (Frazier & Clifton 1996)? As such, the substantive content of the cue set should provide an important point of debate in the context of cue-based parsing frameworks. For these models, the claim of structured access may be understood as the claim that the cues used in the initial stages of parsing are primarily syntactic in nature.



Because the various cue-based parsing models represent the most explicit attempts to integrate explicit models of short-term memory and parsing processes, much of the discussion in the rest of the thesis will be cast in this framework. Likewise, because of the existence relatively explicit computational models, I also explore models of my results using one implementation of a cue-based parser to support the experimental findings presented here. As stated above, the claim of structured access in the context of cue-based parser amounts to the claim that primarily syntactic cues are used in memory retrieval, though the theoretical status of a relational syntactic ‘cue’ in this framework is unclear. Again, however, the central claim of structured access is independent of the particular implementation I adopt. It is important to note that even though I present a model of structured access in a cue-based framework, these mechanisms are compatible with a much wider range of assumptions about the computational character of the parser. More broadly understood, the claim is that only syntactic information drives the generation of linguistic structure for a certain set of linguistic dependencies.

#### *Cue-based parsing and the psycholinguistic enterprise*

Even if one does not subscribe to the main theoretical commitments of cue-based parsing, these models provide a very interesting way of understanding familiar questions in psycholinguistics, and new ways of asking and testing these questions. One important contribution of this line of work is that it has compelled researchers to consider *computationally complete* models of parsing (Lewis & Vasishth 2005, p.

377). As Lewis & Vasishth put it, computationally complete models of parsing are those that give a joint characterization of the processes, memories, and control structures used by the parser. *Memories* are the temporary pieces of information that are relevant for a given parsing task, *processes* refer to operations over that information, and *control structures* may be understood as the decision principles that determine when different processes are applied. From this point of view, it can be seen that much work in sentence processing may be understood as characterizing the control structure of the parser. For instance, research on the behavior of the parser in the face of ambiguous input (e.g. Kimball 1973, Frazier 1978, et seq) is by definition research on the control structure of the parser, being concerned with the principles that govern actions taken at choice points in the parse. This work can in principle be carried out with relatively minimal commitments to the specifics of memories and processes involved in syntactic processing.

In the context of the broader goal of developing a computationally complete parser, the work on memory architecture in sentence processing may be seen as the complement to work on the control structure of parsing. The focus of this line of work, and the focus of this thesis, is on characterizing the nature of the parser's memories (i.e. temporary linguistic representations) and processes (i.e. computations over that information). This is in no sense a new concern for psycholinguists, as many researchers have made substantive claims about memories and process in the parser (Fodor, Bever & Garrett 1974; Levelt 1974; Frazier & Fodor 1978; Wanner & Maratsos 1978; Marcus 1980; Berwick & Weinberg 1984; Prichett 1993; Sturt, Pickering & Crocker 2000, among many others). Put this way, the aim of the present

thesis and its relation to prior work can be made clearer. The claim of a structured access mechanism makes the claim that linguistic working memory encodes detailed syntactic information, and that the parser can selectively attend to this information in performing online parsing operations. This stands in contrast to claims that the fine-grained information is either not encoded (Townsend & Bever 2001; Ferreira & Patson 2007), or that all linguistic information is deployed in parallel during parsing (e.g. MacDonald, Pearlmutter & Seidenberg 1994; Vasishth et al 2008).

Just as in psycholinguistic work that has aimed to characterize the generation process, one interesting question for cue-based approaches to parsing concerns the generality of the interaction between representation and memory access. On one interpretation of this literature (e.g. Greene et al 1992; Van Dyke 2007), it can be understood as making the strong claim that a general-purpose memory access mechanism and a domain-specific representation are the only components of the human sentence processor (see also Ratcliff & McKoon 2008). On this view, the role of the grammar is simply to provide declarative representations that are manipulated by general cognitive mechanisms. That is, linguistic representation only provides domain-specific memories, but no domain-specific processes. This is a view that a number of separate research traditions appear to be converging on (for similar conclusions from a different point of view, see McDonald et al 1994; Pearlmutter & McDonald 1995; Jurafsky 1996; Levy 2008). This stands in contrast to theories that posit specialized mechanisms or principles that are operative only in linguistic processing (Kimball 1973; Wanner & Maratsos 1978; Marcus 1980; Frazier, Clifton & Randall 1983; Berwick & Weinberg 1984). As before, the intuition pursued here is

straightforward: if there are existing computational principles that have explanatory purchase across multiple domains of cognition (content-addressability, rational inference, etc.), then theories that make use of independently motivated mechanisms have an advantage over those that invoke more specialized mechanisms.

In comparing cue-based approaches to the broader psycholinguistic enterprise, it can be seen that familiar psycholinguistic questions remain of central interest. These involve the types of information that are recruited in online processing, as well as the domain-specificity of the parser's processing routines. The adoption of a content-addressable memory architecture does not in and of itself provide an answer to the question of what information is used to guide the parse. However, these architectures are often associated with a corollary claim that a wide range of linguistic features, including morphological and semantic features, are always deployed to access memory during parsing, a claim that is also endorsed in constraint-based models of parsing (e.g. MacDonald et al 1994). The structured access claim I pursue here rejects this as an inherent feature of the parser: if there exists structural access mechanisms deployed by the parser, this claim cannot be globally true.

As for questions of domain-specificity, one might be tempted to view the central claim of this thesis as a claim about domain-specific processes that apply only to linguistic processing. This is not necessarily the case, however; it may be that structured access mechanisms follow from deeper principles of cognition. In Chapter 5, I will suggest one such possibility. Abstraction in access mechanisms is one way that the parser might respond to functional pressure to minimize memory interference in a noisy architecture. If this line of argumentation is correct, then the claim of

structured access might be true even if there is no architectural commitment to modularity or domain-specificity. Instead, the parser's optimal strategy may be to pay attention to the most distinctive, minimal set of information to ease online processing, which would result in structured access for the dependencies considered here.

### *Outline of the dissertation*

The thesis has two main parts. Chapter 2 introduces the argument from interference, one of the most well-practiced arguments for cue-based parsing models and content-addressable memory architectures. I investigate the empirical interference profile of two long-distance dependencies in English: subject-verb agreement and antecedent-reflexive dependencies in English. An off-line judgment task shows that in offline measures, agreement and reflexive feature mismatches are similarly anomalous for English speakers. However across three eye-tracking experiments, a minimal comparison of English agreement and reflexive dependencies reveals qualitatively different profiles with respect to online interference patterns. Agreement dependencies are reliably susceptible to interference online, showing a widely observed interference profile that suggests incorrect access to structurally inaccessible features (Clifton et al 1999; Pearlmuter et al 1999; Wagers et al 2009). In clear contrast, reflexive dependencies do not show any reliable interference from structurally inaccessible antecedents. This indicates that the processes used to resolve these two grammatical dependencies are distinct, opening up the question of how it is that reflexive dependencies are immune to interference online. I argue that reflexives

engage structured access to recover their antecedent. In other words, the difference between agreement and reflexives obtains because the reflexive dependency is initially constructed only with reference to the structural relation between the anaphor and its antecedent; the morphological features are inert in the construction of the antecedent-anaphor chain, in line with much typological and theoretical work on the representation of the binding dependency (i.e. Lidz & Idsardi 1999; Büring 2005; Hornstein 2007; Kratzer 2009). The crucial conclusion from these experiments is that the syntactic dependencies licensed by reflexives are grammatically accurate, in the sense that the earliest stages of dependency formation appear to only access positions occupied by the local subject.

Chapter 3 takes up a discussion of the implications of the results in Chapter 2 for models of parsing. I critically assess the wider range of evidence from interference in sentence processing, and clarify the predictions of the models under consideration using a simple mathematical model of memory access. Using this formalization to model the agreement and reflexive dependencies in Chapter 2, I provide simulation evidence that further supports the claim that structured access, rather than passive memory dynamics, is the source of the empirical difference between agreement and reflexives observed in Chapter 2.

In addition to supporting the hypothesis of structured access in parsing, Chapter 3 provides an analysis of the diverse range of experimental results that have been labeled as ‘interference effects’. It is important to critically assess the range of these results, as they have been very influential in constructing models of the parser’s memory architecture. I argue that the range of constructions where the parser is

demonstrably ‘grammatically fallible’ in the early stages of processing (Wagers 2008; Phillips et al 2009) is narrower than generally assumed. The behavioral effects that risk being conflated under a single notion of an ‘interference effect’ actually correspond to two distinct underlying phenomena that are usefully dissociated in interpreting experimental data that investigates interference patterns in sentence processing. Once this distinction is made, it can be shown that for many phenomena, conclusions about grammatical infidelity rest on assumptions that may not be generally tenable. This critical assessment provides important support to the main claim of structured access by directly addressing potential counterexamples.

Chapter 4 builds upon the claim for structured access that was developed in Chapters 2 and 3 by examining another case of anaphor-antecedent dependency building, the case of long-distance reflexives in Mandarin Chinese. I provide an alternative argument for structured access by investigating the time course of reactivating antecedent noun phrases in completing local and long-distance binding relations. Time-course evidence in the form of speed-accuracy tradeoff (SAT) functions will suggest a role for structured access mechanisms in the resolution of long-distance reflexive dependencies, further supporting the conclusion that morphological and semantic features are not used to gate memory access for reflexive dependencies. The SAT evidence is supported by electrophysiological evidence about the processing of Mandarin long-distance reflexives. The ERP evidence suggests that familiar syntactic reanalysis processes do not drive the effect we observe in SAT, but that instead the difficulty observed in long-distance binding configurations shows a functionally similar profile to that observed in processing long-distance movement

dependencies and other memory-intensive tasks in sentence processing. Taken together, the investigation of the processing of *ziji* in Chapter 4 provides an alternative argument in support of structured access. Chapters 2 and 3 made the case that feature-matching but structurally inaccessible antecedents do not impact early memory access for reflexives. Chapter 4 demonstrates the converse: feature-mismatching structurally accessible antecedents are considered in the construction of a long-distance reflexive dependency, even in the presence of other feature-matched, accessible antecedents. This satisfies a clear prediction of structured access mechanisms: in both English and Chinese, structural relation to the anaphor, rather than feature content of potential antecedents, is the main determinant of online access.

Chapter 5 synthesizes the results presented in Chapters 3 and 4 with extant results in the literature, articulating and defending a revised theory of the relation between grammatical dependencies and memory access. Results from both English and Chinese point to the conclusion that reflexive dependencies access memory using purely structural cues. However, there remains the question of the generality of structured access in comprehension, and in Chapter 5 I begin to address this. To this end, I present data from two self-paced reading (SPR) tasks to provide a comparison of the processing of the syntactic anaphor *ziji* with the processing of the intensified pronominal *ta-ziji*. The SPR data provide another piece of converging evidence for the claim that *ziji* initially accesses only commanding antecedent positions. In contrast, *ta-ziji* appears to access all possible antecedent positions in parallel, suggesting that structured access is not pursued for *ta-ziji*. The distinction in behavior between *ziji* and *ta-ziji* provides confirmation of the result of the computational



modeling in Chapter 3, which shows that the apparent structural sensitivity of the reflexive anaphors presented in this thesis is not an artifact of linear position. More generally, they serve to sharpen the hypothesis of structured access in parsing: the contrast between *ziji* and *ta-ziji* demonstrates that it is not simply interpretive content that drives the use of a structured access mechanism. Not all pronominal dependencies access their antecedents in a structured fashion. I suggest that the parser deploys a structured access mechanism for all long-distance structural dependencies that are crucial to interpretation. I furthermore argue that structured access is a reflection of functional pressure to minimize memory interference: in order to guard against memory interference and misinterpretation for structurally constrained long-distance dependencies, more minimal and abstract retrieval procedures are to be preferred. In light of this claim, structured access can be seen as the optimal strategy for recovering the correct interpretation from a long-distance dependency.

Chapter 6 concludes.

## **Chapter 2: The argument from interference: English agreement and reflexives**

In this chapter I present evidence that syntactically inaccessible, but feature-matched linguistic material is not considered while processing reflexive dependencies. This evidence is built on a direct contrast between reflexive antecedent-anaphor agreement and subject-verb agreement in English, which shows reliable interference from feature-matched, non-subject noun phrases in comprehension (Pearlmutter et al 1999; Wagers et al 2009). This contrast is of theoretical importance for investigating the manner in which linguistic memory is accessed, because subject-verb agreement and reflexive-antecedent dependencies are superficially very similar: to a first approximation, they both require feature concord with the local subject. If feature-based access is the single option for building agreement dependencies with the subject, the two dependencies should behave similarly with respect to interference effects. If reflexives engage a structured access mechanism, however, they should show qualitatively different patterns, with reflexives being insensitive to the feature content of structurally inaccessible antecedents at the point of memory access.

The result of this comparison is that reflexives are found to be immune from interference in environments where agreement reliably demonstrates interference effects. The agreement interference is predicted on a straightforward feature-based access model, but the reflexive data are not. This indicates that the argument from interference—one of two primary arguments for feature-based memory access mechanisms in a content-addressable memory architecture—does not straightforwardly extend to reflexive dependencies. As reflexives do not appear to consider structurally inaccessible antecedents, this provides evidence that structural information, rather than morphological or semantic feature information, gates memory access during construction of these antecedent-anaphor dependencies. This provides the first of the two main empirical arguments I provide for structured access in comprehension: we do not see interference as widely as expected if feature content always controls access to stored information.

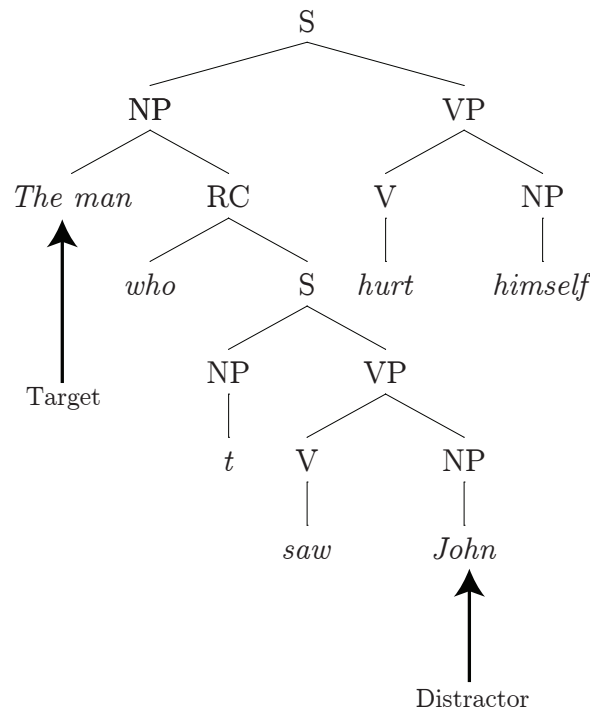
I first briefly outline the argument from interference that has been made in support of content-addressable memory architectures, and then review some of the experimental evidence that has been used to make this argument for subject-verb agreement. Note that there is a great diversity of results that have been attributed to memory interference, and I reserve a fuller discussion of this evidence until Chapter 3. In addition to a brief survey of relevant subject-verb agreement findings, I present a brief review of existing evidence on the processing of reflexives. Although existing literature appears to suggest that reflexives pattern very differently from agreement with respect to interference effects, this conclusion is hampered by the fact that past studies have varied widely with respect to their experimental materials and methods.

I present three eye-tracking experiments and one offline judgment task that directly compare the interference profiles of agreement and reflexives. The results of this comparison establish that, unlike subject-verb agreement, English reflexive dependencies reliably resist intrusion from feature-matched, but structurally inaccessible antecedents. This comparison supports the hypothesis of structured access in comprehension, in that syntactic position, rather than morphological feature content, is what gates memory access in the construction of reflexive dependencies. This is a crucial first step in establishing that feature-based access is not a general feature of linguistic comprehension: it appears that in some cases, a structured access mechanism is engaged that privileges syntactic position over feature match when accessing memory. In the discussion I take up a more in-depth consideration of the relation of the current results to prior findings.

### *The argument from interference*

One of the most well practiced arguments for content-addressable architectures in sentence processing is the argument from interference. Suppose that the parser needs to retrieve a certain constituent for some processing (the *target*). Given this task, the argument from interference has the following general form: if feature-matching, non-target memories disrupt processing or memory access, then the parser is employing an access procedure that uses this feature content to guide retrieval, rather than location in the memory store. Consider Figure 2.1. Given a hypothetical memory retrieval for a reflexive, the local subject memory is the target: on the assumption that

the string is grammatical, the desired antecedent will be found in this position. The embedded noun phrase *John* is a *distractor* memory: it may have the correct morphological feature makeup, but it is not the target of the reflexive's antecedent retrieval. The target memory is structurally accessible given the retrieval: it occupies the correct syntactic position. For English reflexives, this means that it occupies a c-commanding place in the syntactic hierarchy (Reinhart 1976; Chomsky 1981). In contrast, the distractor NP is structurally inaccessible. No matter how well feature-matched the distractor NP is, it is not an acceptable antecedent by virtue of its structural position.

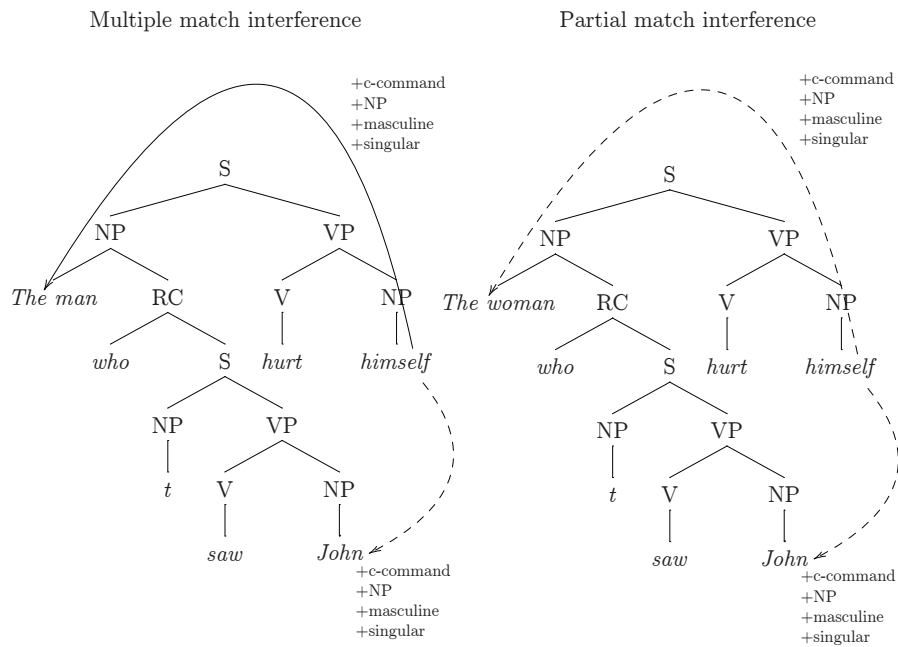


**Figure 2.1:** Target and distractor NPs when processing a reflexive pronoun in English.

The argument from interference follows directly from the definition of a content-addressable access mechanism: the desired feature content (the *cue*) is directly matched against the contents of all stored elements in short-term working memory. Stored memory representations resonate with the search cues when the cues are identical to the content in the memory image, in the same way that a tuning fork spontaneously begins to resonate in the presence of a pitch that matches its own characteristic pitch. By hypothesis, this process of resonating with the search cue renders a given memory available for further processing. When a memory is made accessible for further processing, it is said to have been retrieved.

In a content-addressable memory system, there is no formal requirement that memories be entirely distinct in terms of content, despite the fact that this content is used as an index for later reaccess. The prediction of interference derives from this fact. Without a unique manner of indexing any given memory, interference effects are bound to occur (Kohonen 1980). This stands in contrast to a register-based architecture, where each memory may be assigned a unique storage location. This feature of content-addressable architectures means that there may arise situations where the search cues do not uniquely identify the target memory. There are two relevant situations in which this might occur, and it is useful to distinguish them. One such situation is where many elements in memory contain some or all of the search cues, leading to a situation where multiple matching candidates resonate with the search cues. In this *multiple match* situation, the memories that resonate may be understood to compete with each other for selection. The alternative situation is when the best-matching memory only contains a subset of the desired cues, a *partial match*

situation. These are diagrammed in Figure 2.2. Multiple match is observed in the grammatical sentence in the left panel, because multiple NPs bear singular and masculine features. Partial match is observed in the ungrammatical counterpart of this sentence; the memory that is retrieved will only carry a subset of the features used to access memory, but the potential candidates are nonetheless distinct in feature content.



**Figure 2.2:** Multiple match and partial match interference configurations. For multiple match interference, the target memory is a perfect match to the search cues, but the distractor overlaps in some feature content with the target. In the partial match situation, neither target nor distractor is a perfect match to the search cues.

Neither situation occurs in a register-based memory. In these architectures, unique addressing would eliminate the possibility of multiple matches being returned in response to a query, and failure to find information in a specified address would result in a retrieval failure, rather than a partially matched memory being retrieved. Both the multiple match and the partial match situations are a sort of retrieval

interference, a prediction of content-addressable memory architectures. The distinction between these two types of retrieval interference is important, but I delay an in-depth characterization of the behavioral predictions of each until Chapter 3, where I argue that conflation of this distinction can lead to incorrect generalizations about what interference effects can tell us with respect to underlying mechanism.

The present studies focus exclusively on partial match interference, situations in which there is incorrect access of memories that only contain a subset of the necessary retrieval cues. This is because of the two cases of interference, only partial match interference unambiguously signals that the parser has retrieved structurally inaccessible material, and because the behavioral signature of partial match interference is well understood in the context of subject-verb agreement. In contrast, multiple match interference may be attributed to a number of different underlying mechanisms, as I detail in Chapter 3. For these reasons, multiple match interference thus does not provide the strongest evidence for incorrect access during parsing. Partial match interference, on the other hand, does not appear to be amenable to such alternative explanations, and so it provides the most stringent empirical test of incorrect access during sentence comprehension. Note that partial matching effects have also been referred to as *illusions of grammaticality* (Phillips et al 2010), but I avoid this term in order to not prejudge the underlying source of the effect. Although this is a very natural interpretation of these effects, partial-matching interference effects need not be linked to processes of ungrammaticality detection in any direct way; this is in fact true of the model to be presented in Chapter 3, as well as the ACT-R model of Vasishth et al (2008).



### *Partial-match interference in subject-verb agreement*

To date, two long-distance dependencies have been shown to be susceptible to partial-match interference in comprehension: negative polarity items (NPIs) and subject-verb agreement. The source of the NPI interference effect is a matter of debate (contrast claims in Vasishth et al 2008 and Xiang et al 2009), and I will put off discussion of these effects for the moment. Instead I focus on subject-verb agreement, where the phenomenon of interest is the well-known *agreement attraction* effect that has been noted by a wide range of researchers (e.g., Kimball & Aissen 1971; Kayne 1989; Bock & Miller 1991; den Dikken 2001). Agreement attraction occurs when the morphological features of a noun other than the local subject appear to control verbal agreement. For example, Bock and Miller (1991) presented subjects with sentence fragments as in (2.1):

- (2.1)        a.        The key to the cabinet...  
              b.        The key to the cabinets...

Bock and Miller observed that when the inaccessible noun *cabinet* was plural, there was a marked increase in the probability that participants would produce a plural verb form (i.e. *were*), compared to when *cabinet* did not bear plural features. The agreement attraction effect has been shown to obtain in a range of environments and across a range of different types of interfering noun (see Eberhard, Cutting & Bock 2005 for a review), and appears to be sensitive to the hierarchical, rather than

linear distance between the interfering noun position and the target verb (Franck, Vigliocco & Nicol 2002).

As with NPI dependencies, there remains some debate on the nature of subject-verb agreement interference effect, although much of the debate centers on the source of agreement attraction effects in production (Bock & Miller 1991; Vigliocco & Nicol 1998; Franck et al, 2002; Eberhard et al 2005; Staub 2010; Gillespie & Pearlmutter 2011). There is a sizeable body of research that has documented agreement attraction in production, including cross-linguistic documentation of similar effects in Spanish (Anton-Mendez, Nicol & Garrett 2002), German (Hartsuiker, Schrifers, Bock & Kikstra 2003), Dutch (Hartsuiker et al 2003), French (Franck et al 2002), Slovenian (Badecker & Kuminiak 2007), and Italian (Vigliocco & Franck 2001), among others. Importantly for current purposes, similar effects are also readily observed in comprehension measures (Clifton et al 1999; Pearlmutter et al 1999; Häussler & Bader 2009; Wagers et al 2009).

A number of researchers have argued that the subject-verb agreement attraction effects reflect interference effects in a content addressable architecture, in both production and comprehension (Badecker & Kuminiak 2007; Wagers 2008; Häussler & Bader 2009; Wagers et al 2009). If true, an important prediction of feature-based access in a content-addressable architecture is borne out. In comprehension, agreement attraction effects clearly constitute an argument from interference: structurally inaccessible, but feature-matching material is seen to exert a disruptive influence in the computation of subject-verb agreement. Furthermore, in comprehension this is a clear example of partial-match interference, occurring when

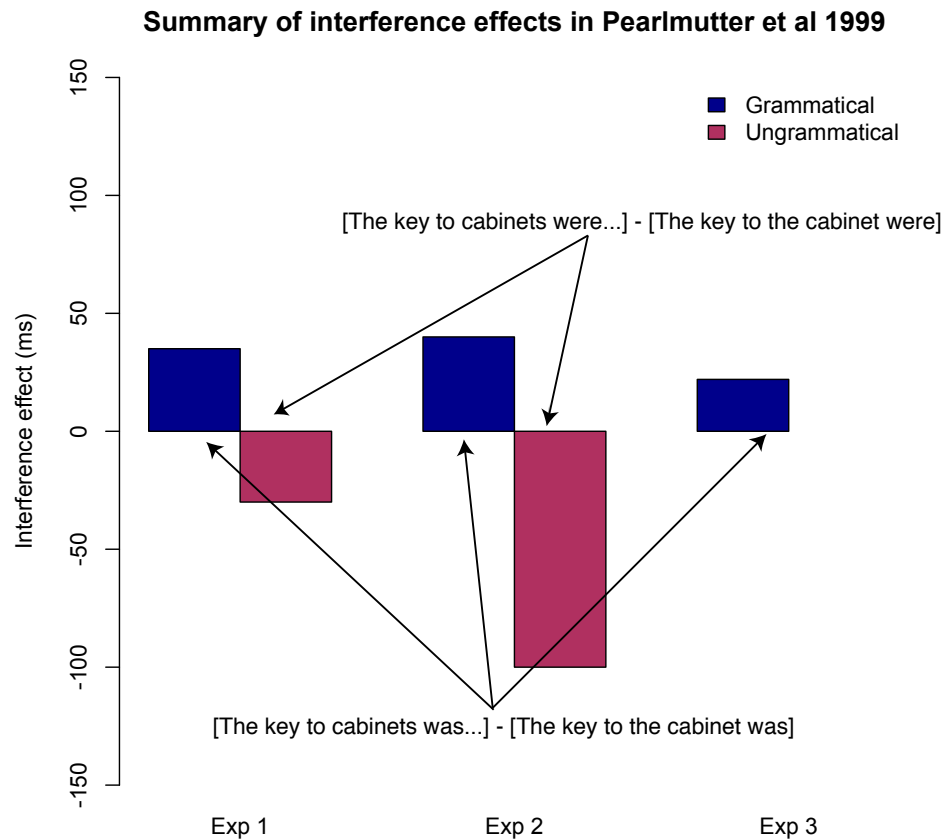
the target noun (the local subject) does not match in the desired features. In these situations this interference arises when the distractor memory (i.e. *cabinets* in (2.1)) is retrieved as a function of its morphological features. If a structured access mechanism that privileges structural information over feature content during memory access is employed to resolve agreement online, we should not observe these effects. This is because from the point of view of structured access, the subject (*the key*) is uniquely identifiable if only structural cues are used to access memory.

Across a number of studies, clear generalizations about the comprehension profile of this interference effect in agreement have emerged. For example, in one investigation of the interference profile of English subject-verb agreement in comprehension, Pearlmutter, Garnsey, and Bock (1999) employed both self-paced reading and eye-tracking to ask about the processing of grammatical and ungrammatical agreement dependencies. They examined sentences of the form in (2.2):

- (2.2) a. The key to the cabinet was rusty from years of disuse.  
b. The key to the cabinets was rusty from years of disuse.  
c. \*The key to the cabinet were rusty from years of disuse.  
d. \*The key to the cabinets were rusty from years of disuse.

In Experiment 1, they investigated these sentences using self-paced reading, and in Experiment 2, they employed eye-tracking methodology. In both replications, the findings were consistent. In the grammatical conditions, (2.2b) was read more slowly than (2.2a) from the interfering noun onwards in eye-tracking, and from the critical verb region onwards in self-paced reading. In the ungrammatical conditions, (2.2c)

was read more slowly than (2.2d), an effect that had a similar time course to the slowdown observed in the grammatical conditions. In eye-tracking measures, it was second-pass measures (i.e. re-reading times) that showed the effect. In Experiment 3, they only investigated the processing of the two grammatical variants of (2.2) above, in addition to variants with a plural head noun. A summary of the observed effects across the three experiments is presented in Figure 2.3. This summary plot shows the magnitude of the ‘interference effect’, which is obtained by subtracting the mean reading time for conditions that had singular intervening nouns (*cabinet*) from those that had a plural intervening noun (*cabinets*), for both grammatical and ungrammatical conditions.



**Figure 2.3:** Interference effects in Experiments 1-3 in Pearlmutter et al (1999). The interference effect is the difference in RTs at the critical region that is due to manipulating the feature content of the distractor, as shown.

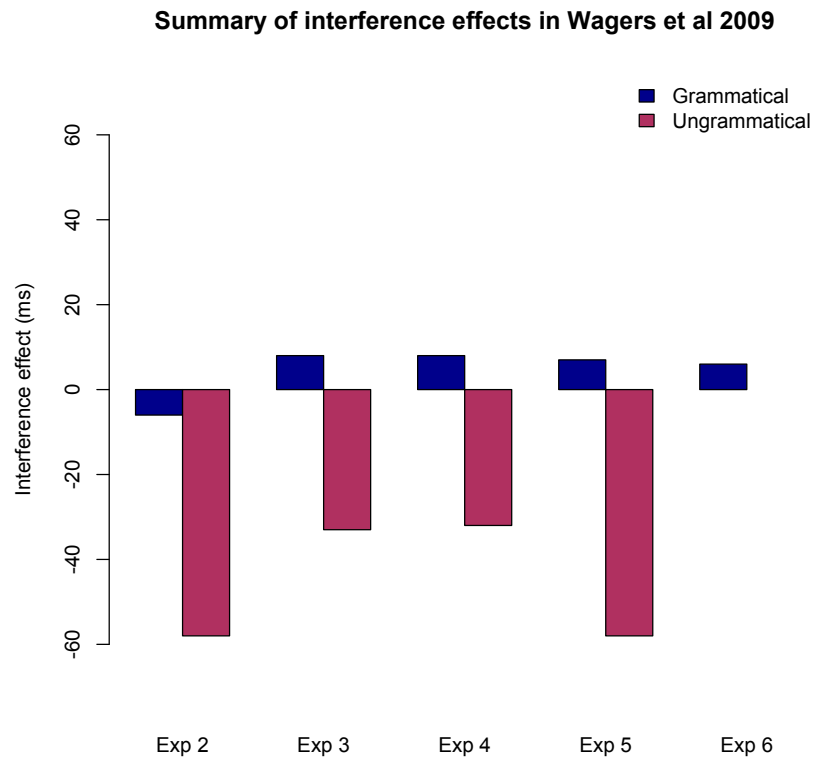
It can be seen from Figure 2.3 that across methodologies and replications the effect of the interfering noun position on the computation of subject-verb agreement was stable. The effect of a feature matched noun phrase in both ungrammatical and grammatical conditions was facilitatory. For the ungrammatical sentences, one way of understanding this effect is as an illusion of ungrammaticality (Phillips et al to appear), wherein the feature match with the inaccessible noun phrase leads to spurious acceptability, and hence eased processing, on some portion of trials in the experiment. This is reflected in the negative interference effects shown in Figure 2.3.

Pearlmutter and colleagues also observed a similar processing advantage when the interfering noun was singular in grammatical conditions: when the interfering noun was singular, matching the singular verb form, faster reading times (eased processing) were observed. This effect is not obviously attributable to an illusion of grammaticality, as the two conditions are equally grammatical, but the generalization in both grammatical and ungrammatical cases is straightforward. When the agreeing verb's features match the inaccessible noun's features, processing is facilitated, though this facilitation is much greater in the case of unacceptable sentences.

Wagers, Lau and Phillips (2009) presented another set of studies that investigated interference in agreement computation in comprehension. Like Pearlmutter et al (1999), they consistently found partial-match interference effects for agreement comprehension. They also gave evidence that interference occurred even when the interfering noun position was not in a linearly or hierarchically intervening position, as in (2.3a-d). However, unlike Pearlmutter et al (1999), they found that this effect was limited to ungrammatical sentences, for both PP and non-intervening relative clause environments. This finding was replicated across 6 experiments, wherein Wagers et al (2009) consistently found that in grammatical sentences (2.3a-b, e-f), the interfering noun's number marking did not exert any measurable influence on the relevant behavioral measure (reading times in self-paced reading, and accuracy measures in a speeded judgment task). Instead they found that across six experiments, and in two very different structural configurations, ungrammatical conditions (2.3c-d, g-h) consistently show an effect of partial-matching interference—that is, greater apparent acceptability and eased measures of processing. In Experiments 2-3, they

showed that this was the case for non-intervening relative clause environments (2.3a-d), and in experiments 4-7 they showed this in the same prepositional modifier environments (2.3e-h) that Pearlmutter et al (1999) investigated. Figure 2.4 provides an overview of their findings across all of the experiments.

- (2.3) a. The musician who the reviewer praises so highly...  
 b. The musicians who the reviewer praises so highly...  
 c. \*The musician who the reviewer praise so highly...  
 d. \*The musicians who the reviewer praise so highly...  
 e. The key to the cell unsurprisingly was rusty...  
 f. The key to the cells unsurprisingly was rusty...  
 g. \*The key to the cell unsurprisingly were rusty...  
 h. \*The key to the cells unsurprisingly were rusty...



**Figure 2.4:** Interference effects in Wagers et al (2009).

Wagers and colleagues did not find any evidence that the distractor noun impacted the processing of grammatical subject-verb agreement, and their Experiment 6 showed that even with a relatively high-powered study, the effect was not reliable. Instead, they suggested that the effects observed by Pearlmutter et al (1999) for grammatical sentences were due to an effect of plural complexity that arose at the previous region, which they addressed by including adverbs that intervened between the attracting noun and the critical verb, as well as performing mixed-effects regression analyses that factored out the effect of plural complexity. However, despite the fact that the effect in grammatical sentences was minimal and unreliable, there was consistency in the findings for ungrammatical sentences: inaccessible feature-matched nouns resulted in facilitated processing. The partial feature match led to a consistent interference effect.

Wagers and colleagues argued that cue-based retrieval interference was responsible for attraction in the comprehension experiments. In other words, the partial match that occurred in ungrammatical sentences led to incorrect retrievals of the distractor NP, indicated by eased processing profiles when there was an inaccessible feature match. The most compelling piece of evidence to this end, according to these authors, is the grammatical asymmetry prediction of the retrieval account: they did not observe an illusion of ungrammaticality, which is what would be expected if singular subjects with embedded plural nouns caused faulty encoding of the target subject's number feature (as in other models of the agreement attraction effect, Eberhard et al 2005). Though their account posited a mixture of forward- and backward-looking processes to derive this prediction, this grammatical asymmetry



prediction holds even if one assumes a purely retrospective agreement computation process (a point noted in Wagers 2008). The crucial difference controlling these predictions is the fact that in the grammatical sentences, there is multiple-match interference: in the grammatical interfering condition, there are multiple singular nouns, but the target head noun nonetheless is a perfect fit to the search cues. In contrast, for the ungrammatical sentences, there is a situation of partial-match interference: no noun fully licenses the agreement morphology on the verb, and so comprehension proceeds by considering either a structurally accessible feature mismatch, or a structurally inaccessible feature match.

Because of the clear and reliable behavioral patterns associated with partial-match interference in subject-verb agreement in English, this dependency provides an important benchmark in assessing the impact of interference effects in the parser more generally. Across production, judgment, and comprehension tasks, there is ample evidence that the agreement attraction effects in comprehension arise from an erroneous consideration of an inaccessible noun's morphological features. Across studies we observe that in grammatical sentences, no reliable effects are observed, and in ungrammatical sentences, eased processing is observed when there is a partial feature match in an inaccessible position. This interference pattern shown in subject-verb agreement is a clear and reliable behavioral signature of grammatically fallible, feature-based access in comprehension. Furthermore, these predictions follow from the adoption of a content-addressable model of memory access (as I demonstrate in Chapter 3). An important prediction of these models is that wherever feature content is used to access memory, agreement-style interference should be observed. Thus, in

order to evaluate the role of feature-based over structured access in comprehension, it is necessary to determine how widely observed these sorts of partial-matching interference effects are.

### *Lack of interference in reflexive dependencies*

Reflexive dependencies provide a natural point of comparison with subject-verb agreement in English. This is because at a very superficial level, they are subject to similar constraints. In English, they both require a feature-matched local subject in most cases. If the retrieval mechanism pools all the relevant linguistic constraints for a dependency to access memory, then for both dependencies the same mixture of morphological and structural features will be used to access memory. If this is so, then both agreement and reflexive dependencies should show similar interference profiles. However, in studies that have examined reflexive processing to date, there has been no reliable evidence for partial-matching interference as seen in agreement dependencies. A number of authors have argued from a wide variety of experimental results that syntactic constraints provide a hard constraint on the nouns that are considered for participation in a reflexive dependency (Nicol 1988; Nicol & Swinney 1989; Clifton et al 1999; Sturt 2003a,b; Xiang et al 2009). If the conclusion of these authors is correct, then there is a compelling case to be made for structured access in comprehension: reflexives access memory using primarily structural information.

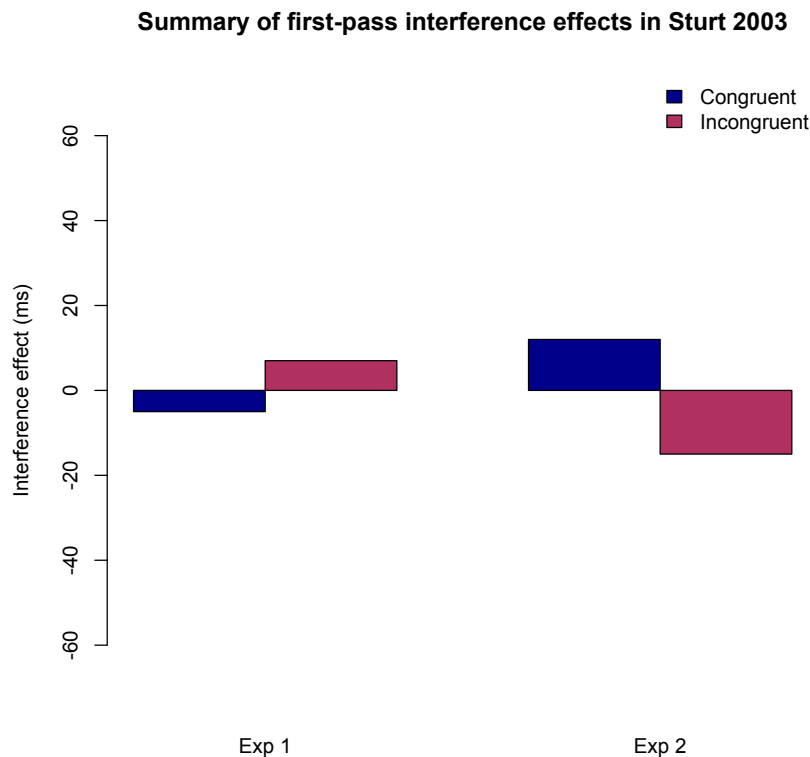
Sturt (2003a) used eye-tracking to determine whether or not a morphological feature match with structurally inaccessible antecedents exerted an influence on the

construction of the binding dependency. He examined small discourses of the form in (2.4) (Experiment 1) and (2.5) (Experiment 2). Note that unlike the agreement studies reviewed above, Sturt did not manipulate the feature content on the agreeing reflexive, instead manipulating the gender of the two noun positions in the sentence. In addition, rather than using number features, he manipulated the stereotypical gender of the noun phrase in the accessible position, and the actual gender of the noun phrase in the interfering position. Thus, none of the sentences in his experiments were globally ungrammatical, although because gender-biased nouns cause comprehenders to commit to the gender of a noun phrase, a temporary ungrammaticality arises in when the local subject's stereotypical gender mismatches the reflexive's morphological features. Previous ERP work demonstrates that this temporary percept of ungrammaticality due to violation of gender stereotype is reliable (Osterhout, Bersick, & Laughlin 1997). I label cases where the local subject does not agree with the reflexive in perceived gender as *incongruent* conditions and mark them with #.

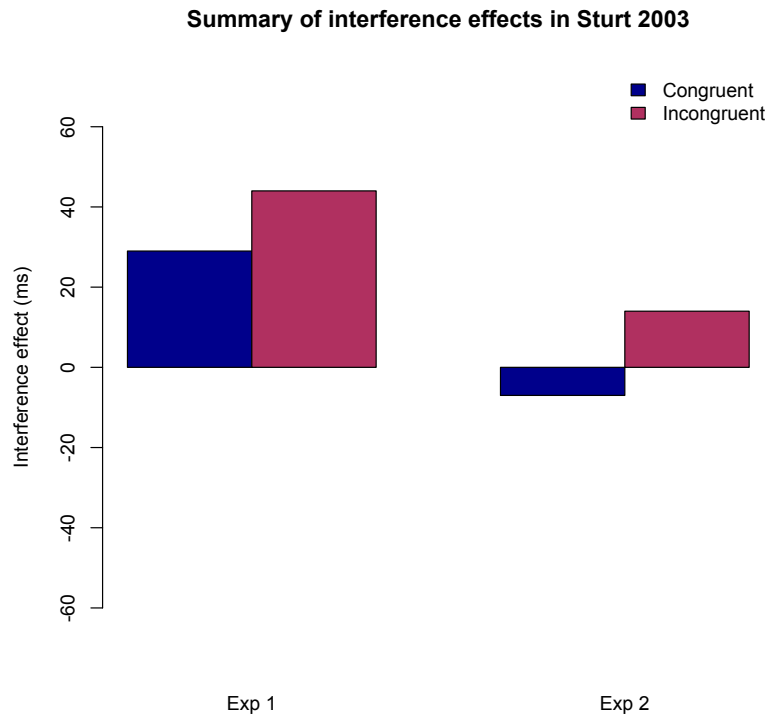
- (2.4) Jonathan was pretty worried at the City Hospital.
  - a. He remembered that the surgeon pricked himself with a needle.
  - b. #He remembered that the surgeon picked herself with a needle.
 Jennifer was pretty worried at the City Hospital.
  - c. She remembered that the surgeon pricked himself with a needle.
  - d. #She remembered that the surgeon pricked herself with a needle.
- (2.5) Jonathan was pretty worried at the City Hospital.
  - a. The surgeon who treated Jonathan had pricked himself with a needle.
  - b. #The surgeon who treated Jonathan had pricked herself with a needle.
 Jennifer was pretty worried at the City hospital.
  - c. The surgeon who treated Jennifer had pricked himself with a needle.
  - d. #The surgeon who treated Jennifer had pricked herself with a needle.

In order to facilitate comparison with the agreement studies presented in the

previous section, I present the key data from Sturt's Experiments 1 and 2 in a form that mirrors the design of the Pearlmutter et al (1999) and Wagers et al (2009) studies. In Figures 2.4 and 2.5, the interference effects for congruent and incongruent conditions in Sturt (2003a) are shown. The interpretation of the direction of the effects is identical to that in the previous section. It can be seen that the reliable behavioral pattern that signaled interference in subject-verb agreement comprehension is not found in either of these studies.



**Figure 2.5:** Interference effects for early (first-pass) measures in Sturt (2003).



**Figure 2.6:** Interference effects for late (total time) measures in Sturt (2003).

In neither experiment did Sturt find consistent evidence for a partial-match interference effect as seen in the agreement studies above; in fact, there were no reliable interference effects at all across measures. In Experiment 1, there were no reliable effects in first-pass measures, however an effect was observed in re-reading times in Experiment 1. For congruent conditions only, a feature-matched inaccessible antecedent caused faster reading times, and this difference reached statistical significance in pairwise and omnibus ANOVA analyses. This effect is consistent with partial-match interference, but this effect did not replicate in Experiment 2, and in fact, the direction of the numerical trend was reversed.

Sturt concluded that the syntactic binding constraints act as a hard constraint on

the early stages of constructing the binding dependency. To the extent that any effect of the inaccessible antecedent is observed in his experiments, it is limited in its scope and is not consistently obtained. Although the primary picture suggested by Sturt's study is one of structural fidelity, it is important to note that offline measures that Sturt examined in a follow-up to Experiment 1 showed that participants clearly do make more errors in conditions that contain a feature match in the inaccessible antecedent position. By asking a question such as "Who was pricked with a needle?" after as the sentences used in Experiment 1, and prompting participants with both characters in the story, Sturt found that the number of ungrammatical interpretations sharply rose in response to a feature-matched, inaccessible antecedent, for both unmatched and matched accessible antecedent sentences.

The finding that reflexives resist interference was subsequently replicated by Xiang, Dillon and Phillips (2009) using event-related potentials (ERPs). They employed materials that were based on Sturt (2003a), and contrasted the three conditions in (3.8):

- (2.6) a. The tough soldier that Fred treated in the military hospital introduced himself to all the nurses.
- b. #The tough soldier that Katie treated in the military hospital introduced herself to all the nurses.
- c. #The tough soldier that Fred treated in the military hospital introduced herself to all the nurses.

They contrasted these materials with a parallel set of conditions that examined the impact of inaccessible, feature-matched material on the processing of NPI dependencies, which are known to be susceptible to partial-matching interference

effects:

- (2.7) a. No restaurants that the newspapers have recommended in their  
dining reviews have ever...  
b. \*The restaurants that no newspapers have recommended in their  
dining reviews have ever...  
c. \*Most restaurants that the newspapers have recommended in their  
dining reviews have ever...

Their logic was as follows. If partial-matching interference obtains in either dependency, then the condition that has feature-matched, inaccessible material (either a gender-matched NP or an embedded negative quantifier) should show reduced processing difficulty on whatever ERP component indexes the difficulty associated with encountering an ungrammatical dependency. For both dependencies, grammaticality detection was reflected in the P600 component, a positive-going ERP component that occurs approximately 600ms post-stimulus, and has a primarily posterior distribution. This component is often observed in ungrammatical environments, though the exact relation of this response to ungrammaticality detection is unclear (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993). For NPI dependencies, the pattern in the P600 replicated earlier findings of partial-matching interference in NPIs (Drenhaus, Saddy & Frisch 2005). Specifically, the P600 response to the interfering condition (3.9b) was smaller than that associated with the fully ungrammatical condition (3.9c), again showing partial-matching interference. In fact, the response to the NPI interfering condition was statistically indistinguishable from the waveform evoked by the fully grammatical condition (3.9a).

An altogether different pattern was observed for reflexives. In a P600 response that began at approximately 450ms after the presentation of the critical reflexive, both gender-incongruent conditions (3.8b-c) showed an identical positive deflection relative to the baseline condition (3.8a). Pairwise analyses revealed no significant differences between the interfering condition and the entirely incongruent condition. In later time windows, however, there was a trend towards a greater P600 effect in (3.8b) relative to (3.8c). This did not reach significance, though it is important to note that this is exactly the opposite pattern to that found in the NPI conditions. It does not reflect the eased processing that is expected in partial-match interference situations. This mirrors the non-significant trends in parallel conditions in Sturt (2003a), where the effect of an intrusive, feature-matched subject was a numerical slowdown in reading times.

There was no trend consistent with partial-matching interference for reflexives in Sturt (2003a) and Xiang et al (2009) is suggestive, and matches well with suggestions of structural fidelity for reflexives that have been observed in other experiments (Nicol 1988; Nicol & Swinney 1989; Clifton et al 1999). Of all of the reflexive studies that have been published to date, there is no convincing demonstration of anything resembling the clear partial-matching interference effects that have repeatedly been observed for subject-verb agreement dependencies. There is no reliable indication that reflexives use their morphological features in retrieving their antecedents, which appears to support the hypothesis of structured access in comprehension.



### *Experiment 1: Direct comparison of agreement and reflexives*

The brief survey of agreement and reflexive studies given above suggests a contrast in the interference profiles between the two phenomena. As noted at the outset, whether or not this contrast is real is an important theoretical question when considering the hypothesis of structured access in comprehension. From the point of view of feature-based access mechanisms, subject-verb agreement and reflexive dependencies form a close minimal pair: they both require a local subject that agrees in morphological features. If they show differential sensitivity to feature-matching, but non c-commanding linguistic material, then there is a case to be made that the mechanisms used to access linguistic memory are qualitatively different for the two dependencies. In surveying previous results, I noted that partial-matching interference effects are reliably observed in agreement dependencies, and that this interference profile is the crucial behavioral signature of feature-based memory access. For reflexives, no result that is consistent with this interference profile has been observed to date, opening up the possibility that reflexive dependencies engage a different strategy for memory access. Since morphological features do not appear to gate memory access in behavioral measures, it is tempting to conclude that reflexives engage in structured access.

However, it is difficult to make a strong case for structured access with current studies, due to large differences in the environment where interference has been found for agreement and reflexives. All reflexive studies to date have investigated interference from nouns that are outside the local clause that contains the

reflexive. For example, in Sturt's (2003a) Experiment 2 and the experiment reported in Xiang et al (2009), the inaccessible noun was embedded inside an object relative clause that modified the subject noun. In these environments, it is well known that the magnitude of the agreement attraction effect is lessened in production, perhaps because of the scope of planning in production (Bock & Miller 1991) or because of the extra phrase-structural distance between the interferer and the head noun (Franck et al 2002; Eberhard et al 2005). Furthermore, agreement studies in English necessarily look at interference from number features, whereas reflexive studies have investigated the effect of (stereotypical) gender features on the computation of the reflexive dependency. It is possible that the role of these two features differs in comprehension, driving the observed differences between reflexives and agreement.

In order to first establish the contrast in interference profiles, and confirm whether or not reflexives are constructed in a fundamentally structure-sensitive manner, Experiment 1 presents a within-subjects comparison of subject-verb agreement and reflexives in English using eye-tracking. For both dependencies, the question is whether or not inaccessible, feature-matched noun phrases exert an influence on the computation of the dependency. The crucial test is whether or not the partial-match interference profile commonly observed for agreement obtains for reflexives, which would unambiguously indicate a feature-based access strategy. Structured access predicts that the feature match with the inaccessible antecedent should not affect the early stages of processing the reflexive dependency. In order to facilitate the comparison, the syntactic position and feature content of the interfering noun was held constant across dependencies. The prediction is that if structured

access is used for reflexives, Experiment 1 should replicate the partial-match interference pattern observed in Pearlmutter et al (1999) and Wagers et al (2009) only for agreement conditions; no observable such interference pattern should be seen for reflexives.

### Participants

40 members of the University of Maryland community participated in Experiment 1 (24 females, mean age 21.9). Participants gave informed consent, and were either paid \$10 for their participation or received course credit. The experimental session, including set-up and calibration, lasted approximately one hour.

### Stimuli

Forty-eight item sets of the form given in Tables 2.1 and 2.2 below were constructed. For both agreement and reflexive dependencies, the subject noun was the same. In all cases, the subject head noun (NP1) was modified by a subject relative clause that contained the intrusive noun (NP2). In order to ensure that the overt gender marking in singular reflexives did not provide extra cues to antecedent identity above and beyond number, both were chosen to have similar gender bias, based on the norms in Kennison & Trofe (2003). 24 of the item sets contained a pair of male-biased nouns, and the remaining 24 contained a pair of female-biased nouns. The verb inside the relative clause never overtly expressed agreement, and neither did the main

clause verb used in the reflexive conditions. For all conditions, the subject was followed by an adverbial that signaled the end of the relative clause.

### *Agreement conditions for Experiment 1*

---

Grammatical, no interference

1 The new executive/ who oversaw/ the middle manager/ apparently/ was dishonest / about the company's profits.

Grammatical, interference

2 The new executive/ who oversaw/ the middle managers/ apparently/ was dishonest / about the company's profits.

Ungrammatical, no interference

3 The new executive/ who oversaw/ the middle manager/ apparently/ were dishonest / about the company's profits.

Ungrammatical, interference

4 The new executive/ who oversaw/ the middle managers/ apparently/ were dishonest / about the company's profits.

**Table 2.1:** Summary of agreement conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined.

### *Reflexive conditions for Experiment 1*

---

Grammatical, no interference

1 The new executive/ who oversaw/ the middle manager/ apparently doubted/ himself on/ most major decisions.

Grammatical, interference

2 The new executive/ who oversaw/ the middle managers/ apparently doubted/ himself on/ most major decisions.

Ungrammatical, no interference

3 The new executive/ who oversaw/ the middle manager/ apparently doubted/ themselves on/ most major decisions.

Ungrammatical, interference

4 The new executive/ who oversaw/ the middle managers/ apparently doubted/ themselves on/ most major decisions.

**Table 2.2:** Summary of reflexive conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined.

For agreement conditions, the main verb was always a present tense agreeing form of *be* (*was* or *were*), followed by a predicative adjective and a four-word spillover region. For reflexives the main verb was always a non-agreeing, past tense

verb that was followed immediately by a direct object reflexive. When the reflexive was singular, it agreed in gender with the gender bias of the two nouns in the sentence: thus, 24 items contained *himself*, and the remaining 24 items contained *herself*. As in the agreement conditions, the reflexive was followed by a four-word spillover region.

The 48 target items were mixed with 152 fillers, for a total of 200 sentences. In addition to the 24 ungrammatical target items, there were 12 unrelated ungrammatical fillers (containing an illicit NPI dependency) for a grammatical-to-ungrammatical ratio of 4.6:1. Half of the target items and half of the fillers were followed by a comprehension question. Across items, comprehension questions addressed various parts of the sentence; this was done in order to prevent participants from adopting superficial reading strategies that extracted the information needed to answer comprehension questions without fully comprehending the sentence. The entire set of experimental stimuli for experiments 1-3 can be found at <http://people.umass.edu/bwdillon>.

### Offline judgments

One concern with investigating the effect of number mismatch on reflexive dependencies in English is the fact that plural reflexives for singular antecedents are tolerated in situations where the speaker does not wish to commit to a particular gender for the antecedent. Intuition suggests that sentences such as *the student hurt themselves during lunch break* are acceptable in a colloquial register, but this effect

appears to be subject to significant dialectal variation. However, for nouns that overtly signal the referent's gender, this option is degraded, as in *\*the girl hurt themselves during lunch break*. In order to test whether or not the number mismatch in the present materials was reliably rejected, an offline judgment study was conducted with the 48 items from Experiment 1, equally balanced for male and female biased nouns. These 48 target items were mixed with 100 fillers, and the materials were balanced so that across the experiment, half of the sentences were ungrammatical. The anomalies in the fillers comprised a variety of different grammatical errors, including unlicensed NPIs and unlicensed verbal morphology (e.g. *\*will eating*). 12 participants were asked to judge the acceptability of the sentences they read on a 7-point scale, where 7 was completely acceptable and 1 was completely unacceptable. Participants were instructed to judge the sentences with regard to whether or not they were acceptable in colloquial speech. The results are presented in Table 2.3.

	[+gram,-intr]	[+gram,+intr]	[-gram,-intr]	[-gram,+intr]
<i>Agreement</i>	5.36 ( $\pm 0.29$ )	5.08 ( $\pm 0.29$ )	3.00 ( $\pm 0.41$ )	3.32 ( $\pm 0.43$ )
<i>Reflexives</i>	5.60 ( $\pm 0.25$ )	5.68 ( $\pm 0.22$ )	3.08 ( $\pm 0.37$ )	3.24 ( $\pm 0.38$ )

**Table 2.3:** Mean judgments and standard error by subjects for Experiment 1 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable.

A three-way repeated measures ANOVA by participants revealed a significant main effect of grammaticality ( $F(1,11) = 34.0, p < 0.001$ ), as well as a significant interaction of grammaticality with interfering noun number ( $F(1,11) = 5.3, p < 0.05$ ). Resolving this interaction revealed that there was a significant interaction of

interfering noun number and grammaticality only for agreement conditions ( $F(1,11) = 10.4, p < 0.01$ ); the interaction did not reach significance for reflexives ( $F(1,11) = 0.1, p > 0.7$ ). Resolving this interaction further using paired  $t$ -tests revealed a marginal effect of interfering number for grammatical agreement conditions ( $t(11) = 1.8, p < 0.1$ ), and no effect in ungrammatical conditions ( $t(11) = -1.5, p < 0.2$ ). Thus there was no reliable evidence that the interfering number had an effect on offline judgments for either dependency.

Importantly, the effect of grammaticality was highly significant for reflexives ( $F(1,11) = 31.7, p < 0.001$ ), and the size of the grammaticality effect differed across dependencies. In fact, there was a slight trend towards a larger penalty for ungrammaticality in reflexives than there was in agreement (agreement:  $\Delta\mu = 2.1 \pm 0.37$ , reflexives:  $\Delta\mu = 2.5 \pm 0.44$ ).

The results confirm that in offline judgments, participants treat the reflexive and agreement anomalies in a qualitatively similar fashion. Importantly for the present purposes, there was no indication that the plural reflexive *themselves* was accepted to any degree with the gender-biased singular nouns in the experimental materials.

### Procedure

The 48 target item sets were distributed into 8 Latin Square lists, and five participants were assigned to each list. Each list was randomized along with the filler

sentences subject to the constraint that no two experimental sentences were presented next to each other.

The maximum number of characters allowed on a single line on the visual display was 142 characters, and all sentences in the experiment fit on one line. All sentences were presented in a 12-point fixed-width font (Courier), and all characters were 9 x 16 pixels on the display. The resolution of the visual display was 1280 x 720 pixels on an LCD screen.

Eye movements were recorded using an Eyelink 1000 tower-mount eye-tracker. Participants had binocular vision while movements were measured, but only the gaze of the right eye was tracked. The tower was 32 inches from the visual display, giving participants approximately 5 characters per degree of visual angle. The eye-tracker sampled eye movements at 1000Hz.

Prior to beginning the experiment, participants were familiarized with the apparatus and given four practice trials. While seated, participants' heads were immobilized using a chin rest and forehead restraint that was adjusted for comfort. Before the experiment, and whenever necessary throughout the experiment, the experimenter calibrated the eye-tracker with a 9-point display to ensure an accurate record of eye-movements across the screen. Participants began each experimental trial by fixating on a marker at the beginning of the sentence. Once the fixation in the target region was recognized by the experimental software, the trial sentence was displayed all at once. Participants ended the presentation of the trial sentence by indicating they had finished using a response pad. On trials with a question, the question was presented immediately after the trial sentence, and participants indicated



their response on the same response pad. Participants were allowed to take breaks at their discretion throughout the experiment. At a minimum, the experimenters asked the participants to take one short rest during the course of the experiment. After each break, participants were recalibrated to ensure accurate measurement of the eye movements.

### Data Analysis

Sentences in both reflexive and agreement conditions were divided into six regions of interest, as indicated in Tables 2.2 and 2.3. For all conditions, the complex subject was divided into three regions: the head noun, its determiner and adjective (NP1), the relative clause complementizer and the embedded verb, and the embedded noun along with its determiner and adjective (NP2). The remainder of each sentence was divided into a pre-critical region, a critical region, and a spillover region. For agreement conditions, the pre-critical region consisted of the main clause adverbial, the critical region consisted of the agreeing form of *be* and the predicative adjective, and the spill-over region consisted of the remaining four words. For reflexive conditions, the pre-critical region consisted of the main clause adverbial and the main clause verb, the critical region consisted of the reflexive and the following preposition, and the spill-over region was the remaining three words in the sentence. The larger analysis window for the agreement conditions was adopted because of a high rate of skipping of the inflected auxiliary. A larger window was adopted for the reflexive conditions to attempt to maximize the similarity of the critical reflexive

region to the critical agreement region. It should be noted, however, that similar patterns of results obtain when word-by-word regioning of the critical areas is used. Analysis for reflexive and agreement conditions proceeded separately. I present the data from four regions of interest in both agreement and reflexive conditions: NP1, NP2, the critical agreement/reflexive region, and the spillover region.

I report three measures for each of our regions of interest. Early measures reported here are the *first pass reading time* and the *probability of regression*. First pass reading time (FPT) is calculated by summing all fixations in a region of interest after participants first enter the region until the first saccade out of that region (either to the right or the left). The probability of regression (PR) corresponds to the probability that a regression is initiated from a particular region before exiting that region to the right. I also report a late measure, *total time* (TT), which is the total sum of all fixations in a particular region of interest, including first pass reading time and any time spent re-reading the region.

For all measures, statistical analysis was performed using mixed-effect linear regressions to assess the magnitude, direction, and reliability of the experimental factors on reading times and probability of regression. There are a number of advantages to this analysis. Mixed effects models have the advantage of allowing for simultaneous modeling of by-subject and by-item effects. Additionally, unlike repeated-measures ANOVAs, they generalize readily to data sets that have missing values (see Baayen, Davidson & Bates 2008). This is useful in the context of eye-tracking: for any given trial, there is a chance that participants will simply not fixate in a region of interest, leading to missing data. In adopting a mixed-effect models

analysis, missing values were left out, rather than adding zeros to trials where participants had no fixations for a particular measure.

The experimental fixed effects in the models were the factors GRAM (whether or not the sentence was grammatical), INTR (whether or not the embedded NP was plural), and their interaction. The fixed effects were coded using simple difference sum coding (grammatical conditions were coded as -.5, ungrammatical as .5; no interference conditions as -.5, interference as .5). Thus, all reported coefficients reflect the magnitude of the difference between levels of a given factor in milliseconds. Additionally, a fixed effect for the *order* of a trial in the experiment was considered. In addition to these fixed effects, I additionally considered random intercepts for participants and items, as well as random slopes for the experimental fixed effects by subjects and by items. In all cases, the significance of non-experimental fixed effects and random effects was assessed, and I report in all cases the best-fit model (following Baayen et al 2008; Jaeger 2008). I leave the experimental fixed effect structure constant across all models, because these effects were theoretically motivated by the design of the study. For most analyses, in addition to the experimental fixed effects, the best model included only random intercepts for subjects and items, as well as a fixed effect for trial order. The raw mean fixation times are presented in Tables 2.4 and 2.5 below. For the probability of regression measure, logistic mixed effects models were used (see Jaeger 2008), as the dependent measure was categorical (i.e. presence or absence of a first-pass regression for a given region of interest). A fixed effect was considered significant if its absolute *t*-value was greater than 2, which indicates that a given effect's 95% confidence

interval does not include 0 (Gelman & Hill 2005); thus all reported coefficients with a  $t$ -value whose absolute value is greater than 2 are significant at the  $p < 0.05$  level. Although I do not report the significance of non-experimental fixed effects (i.e., *order*), in most instances it did have a statistically significant coefficient in the final model.

### Results: Agreement

The by-region reading times for first-pass and total time measures for the agreement conditions, as well as first-pass regression probabilities, are presented in Table 2.4. Prior to the critical region, no significant effects for any of the experimental factors were observed in any measure.

<b>Agreement</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, no int</i>	582 (30)	627 (37)	373 (18)	795 (42)
<i>Gram, int</i>	588 (32)	612 (34)	389 (17)	837 (48)
<i>Ungram, no int</i>	583 (36)	593 (31)	445 (23)	805 (51)
<i>Ungram, int</i>	594 (39)	639 (33)	439 (21)	847 (44)
<b>Total Time</b>				
<i>Gram, no int</i>	867 (60)	955 (63)	579 (34)	1018 (46)
<i>Gram, int</i>	835 (51)	959 (50)	622 (40)	1059 (52)
<i>Ungram, no int</i>	891 (63)	971 (51)	811 (57)	1140 (70)
<i>Ungram, int</i>	881 (55)	1002 (70)	693 (35)	1074 (51)
<b>Pr(Regression)</b>				
<i>Gram, no int</i>	-	.18 (.03)	.13 (.03)	.65 (.04)
<i>Gram, int</i>	-	.20 (.03)	.17 (.02)	.67 (.04)
<i>Ungram, no int</i>	-	.18 (.03)	.25 (.04)	.71 (.04)
<i>Ungram, int</i>	-	.13 (.02)	.20 (.03)	.70 (.04)

**Table 2.4:** Table of means (in ms where applicable) for agreement conditions for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

The model results at the critical agreeing verb region (*was dishonest*) are summarized in Table 2.6. In the early first-pass measures, there was a significant effect of grammaticality. This was due to a significant slowdown in ungrammatical conditions of approximately 64ms ( $SE = 14.7$ ,  $t = 4.25$ ). In the spillover region, there were no significant fixed effects in first-pass reading times. In the total time measure, there was a significant effect of grammaticality, due to a slowdown of approximately 157ms ( $SE = 28.0$ ,  $t = 5.60$ ) for ungrammatical conditions relative to grammatical conditions. Additionally, there was an interaction of interference and grammaticality ( $\beta = -148\text{ms}$ ,  $SE = 56.4$ ,  $t = -2.64$ ). To resolve this interaction a second model was fit to evaluate planned pairwise comparisons, directly comparing differences due to interference within grammatical and ungrammatical sentences. This model revealed that the interaction was driven by a significant difference due to interference for ungrammatical sentences ( $\beta = -117.2$ ,  $SE = 39.6$ ,  $t = -2.96$ ); no such difference obtained in grammatical environments. This pattern replicates the interference asymmetry noted in Wagers et al (2009). In the spill-over region for agreement conditions, modeling revealed a significant effect of grammaticality ( $\beta = 78.7$ ,  $SE = 33.7$ ,  $t = 2.33$ ). No other fixed effects reached significance in this region.

In first-pass regression probabilities, modeling revealed a significantly greater probability of regressing in the ungrammatical conditions relative to grammatical conditions at the critical agreeing verb ( $\beta = .587$ ,  $SE = .180$ , Wald  $z = 3.26$ ,  $p < .005$ ). There was only a marginal interaction of grammaticality with interference ( $\beta = -.681$ ,  $SE = .360$ , Wald  $z = -.183$ ,  $p < .06$ ), although this is consistent with the pattern seen in total time measures, and in the predicted direction. However, a model containing

planned comparisons to resolve this interaction showed no significant differences due to interference within either grammatical or ungrammatical conditions. In the spillover region, there was only a marginal effect of grammaticality on the probability of regression ( $\beta = .273$ ,  $SE = .152$ ,  $Wald\ z = 1.79$ ,  $p < .08$ ).

<b>Reflexives</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, no int</i>	583 (31)	619 (30)	299 (12)	650 (49)
<i>Gram, int</i>	579 (32)	645 (35)	295 (13)	659 (35)
<i>Ungram, no int</i>	570 (36)	621 (29)	351 (17)	628 (34)
<i>Ungram, int</i>	567 (32)	651 (29)	342 (16)	631 (27)
<b>Total Time</b>				
<i>Gram, no int</i>	841 (49)	951 (55)	481 (20)	851 (48)
<i>Gram, int</i>	819 (47)	907 (48)	471 (24)	846 (42)
<i>Ungram, no int</i>	863 (56)	1023 (59)	588 (30)	918 (62)
<i>Ungram, int</i>	890 (59)	1040 (54)	580 (30)	882 (47)
<b>Pr(Regression)</b>				
<i>Gram, no int</i>	-	.19 (.03)	.14 (.03)	.73 (.04)
<i>Gram, int</i>	-	.18 (.03)	.16 (.03)	.77 (.04)
<i>Ungram, no int</i>	-	.16 (.03)	.09 (.02)	.77 (.03)
<i>Ungram, int</i>	-	.15 (.03)	.10 (.02)	.77 (.04)

**Table 2.5:** Table of means (in ms where applicable) for reflexive conditions for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

### Results: Reflexives

The by-region reading times and regression probabilities for the reflexives conditions are presented in Table 2.5. The results of models fit at the critical reflexive region are shown in Table 2.7. In first-pass measures, no significant effects were observed prior to the critical region. At the critical reflexive region, there was a significant effect of grammaticality, indicating a slowdown for ungrammatical conditions ( $\beta = 50.7$ ,  $SE = 10.5$ ,  $t = 4.81$ ). This slowdown in ungrammatical conditions was also observed in total time measures at the critical reflexive region ( $\beta = 110.1$ ,  $SE = 21.8$ ,  $t = 5.05$ ). In neither measure was there an interaction of grammaticality and interference, nor was there a consistent trend for the direction of this effect across measures. Although the interaction of interference and grammaticality did not reach significance, to maximize the chance of showing a difference due to interference I performed the same planned comparisons that we applied to the total reading times at the critical agreement region. This model failed to show an effect of interference for either ungrammatical or grammatical sentences. Surprisingly, the effect of ungrammaticality on the probability of regression for reflexive measures was a decreased probability of regression ( $\beta = .494$ ,  $SE = .207$ , Wald  $z = -2.38$ ,  $p < 0.05$ ). There was again no significant effect of either an interfering plural, nor was there an interaction of interference and grammaticality on the probability of regression from the critical reflexive region.

<i>Agreement</i>	$\beta$	SE	$t / z$
<i>First pass</i>			
GRAM	62.4	14.7	4.25*

INTR	3.8	12.5	0.31
GRAM×INTR	-21.8	25.0	-0.87
<i>Total Time</i>			
GRAM	157.2	28.0	5.60*
INTR	-43.0	28.1	-1.53
GRAM×INTR	-148.5	56.4	-2.64*
<i>Pr(Regression)</i>			
GRAM	.587	.180	3.26*
INTR	.004	.180	0.02
GRAM×INTR	-.681	.360	-1.83†

**Table 2.6:** Summary of fixed effects for best-fit models on agreement conditions at the critical agreeing verb region, including *t*-values (*z*-values for first-pass regression probability data). An asterisk (\*) indicates significance at  $\alpha = 0.05$ , while a cross (†) indicates significance at  $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds.

At the spillover region, there were no significant effects of the experimental manipulations in any of the measures. Likewise, there were no significant effects on the total reading at NP1. At the interfering NP position (NP2), however, modeling revealed a significant effect of grammaticality on total reading times for reflexive conditions. The NP2 position had significantly higher reading times ( $\beta = 100.0$ ,  $SE = 32.0$ ,  $t = 3.13$ ) when the reflexive was ungrammatical. There were no significant effects for interference or the interaction of interference and grammaticality at this region. There were no significant effects on probability of regression at this region.



<i>Reflexives</i>	$\beta$	SE	$t / z$
<i>First pass</i>			
GRAM	50.7	10.5	4.81*
INTR	-3.6	10.6	-0.34
GRAM×INTR	-6.3	21.1	-0.30
<i>Total Time</i>			
GRAM	110.1	21.8	5.05*
INTR	-2.6	19.3	-0.14
GRAM×INTR	1.25	38.6	0.03
<i>Pr(Regression)</i>			
GRAM	-.494	.207	-2.38*
INTR	-.055	.207	-0.27
GRAM×INTR	-.288	.415	-0.69

**Table 2.7:** Summary of fixed effects for best-fit models on reflexive conditions at the critical reflexive region, including  $t$ -values ( $z$ -values for first-pass regression probability data). An asterisk (\*) indicates significance at  $\alpha = 0.05$ , while a cross (†) indicates significance at  $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds.

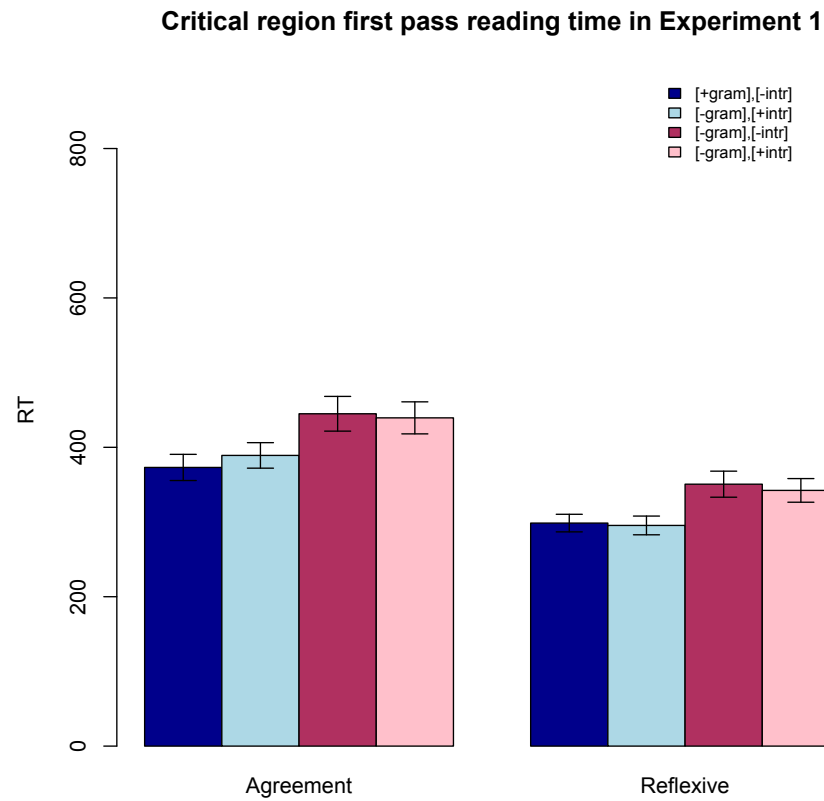
#### Direct comparison of interference effect

Results showed a reliable effect of the number of an interfering noun for agreement, and no corresponding effect for reflexives. In order to provide a direct measure of the interference effect across dependencies, however, a direct statistical comparison of the size of the effect across the two dependencies is called for. Within ungrammatical and grammatical sentences, I compared the size of the *interference effect*, a derived measure that was calculated by subtracting the reading time for [-intr] conditions from [+intr] conditions for total reading times in the critical region. A direct test of the size of the interference effect across agreement and reflexives revealed that there was a significantly larger interference effect for agreement in the ungrammatical conditions ( $t(39) = -2.5$ ,  $p < 0.02$ ; agreement  $\mu = -118.6\text{ms} \pm 42.8$ ,

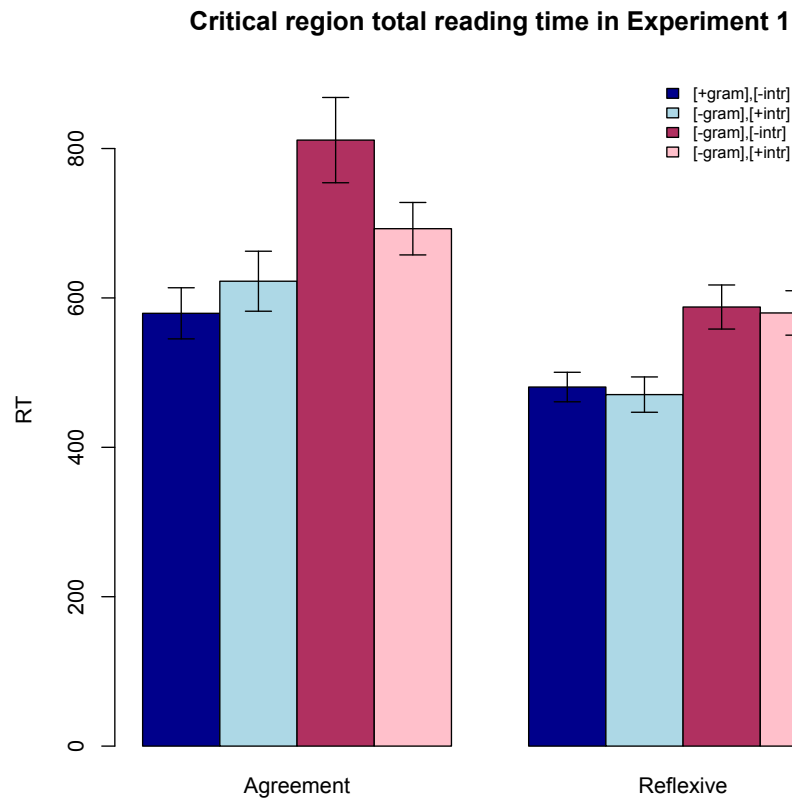
reflexive  $\mu = -7.8\text{ms} \pm 29.6$ ). There was no significant difference between agreement and reflexives in the grammatical conditions ( $t(39) = 1.4$ ,  $p < 0.3$ ; agreement  $\mu = 43.0\text{ms} \pm 31.6$ , reflexive  $\mu = -10.1\text{ms} \pm 28.8$ ).

### Discussion of Experiment 1

A summary of the first-pass and total reading times in the critical region, for agreement and reflexives, is summarized in Figures 2.7 and 2.8. For reflexive regions, across all three measures reported, and all other regions of interest (notably the interfering noun position NP2), the only significant fixed effect was the effect of grammaticality (GRAM). Ungrammatical conditions reliably lead to longer reading times, although surprisingly they led to fewer backwards regressions from the ungrammatical reflexive. In contrast, there were no significant effects of interference or the interaction of grammaticality with interference at any region or measure. Potentially as important as the lack of reliability in the interaction across regions and measures is the relative instability of the direction of the interaction effect across measures (Gelman & Hill 2005), showing that there was no consistent trend for interference in the reflexive conditions. There was thus no evidence, in any region or measure, for partial-matching interference in the reflexive conditions. This extends the findings of Sturt (2003) and Xiang et al (2009) in showing the relative robustness of reflexive dependencies to interference in ungrammatical conditions with number as well as gender.



**Figure 2.7:** Mean first-pass reading time at the critical region in Experiment 1. Error bars show standard error by participants.



**Figure 2.8:** Mean total reading time at the critical region in Experiment 1. Error bars show standard error by participants.

The processing of reflexive conditions clearly contrasts with the processing of agreement conditions, which show a qualitatively different pattern. Although ungrammaticality reliably slowed down the processing of both agreement and reflexives, only agreement showed reliable interaction effects (a significant effect at the critical agreement region for total times, and a marginal effect on probability of regression). Planned comparisons confirmed that this interaction was driven by interference in ungrammatical conditions. Interference led to shorter reading times in ungrammatical sentences, replicating earlier findings (Pearlmutter et al 1999; Wagers et al 2009) of clear partial-matching interference effects for agreement in

comprehension. It is worth noting that across all region and measure pairings that the effect of interference had the same direction: for ungrammatical sentences, an interfering partial-match noun phrase leads to speeded processing or fewer regressions, relative to sentences in which there is no partial match.

This contrast is telling, but it is always difficult to mount an argument based on null effects (the lack of an interaction in the reflexive conditions and the lack of any pairwise differences due to interference). This concern is mitigated by the fact that a direct comparison between the size of the interference in agreement and reflexive dependencies nonetheless indicated that the observed interference effects were reliably larger for agreement. In addition, the lack of reliability of the direction of the critical interaction effect, coupled with the very small effect sizes seen in the reflexive conditions, suggest a lack of interference effect when compared to the agreement conditions. Across a wide range of measures and regions, agreement reliably showed partial-matching interference effects. If the size and variance of the predicted interference effect should be similar for both agreement and reflexives, the finding that our experimental manipulation showed the effect for agreement suggests that the experiment had sufficient power to detect the reflexive effect. This argument rests on an assumption of equally sized effects that might not in fact be true, and I return to the question of the size of the predicted effect in Chapter 3.

In Figure 2.7 it can be seen that in first-pass measures, agreement and reflexives appear to show an identical profile. This may be due to an early processing stage in which they are in fact processed in the same manner. However, it may also reflect more ‘low-level’ effects. For both agreement and reflexives, the length of the

critical region (as well as the identity of the critical words) covaried with the factor of grammaticality. This makes it difficult to know if pattern in Figure 2.7 is due to effects such as word length, or if they index an early, identical ungrammaticality detection stage for agreement and reflexives. Experiments 2 and 3 aimed to shed some light on this early grammaticality effects in the first pass measures reported here. Experiments 2 and 3 employ designs with a full crossing of subject NP number and the number of the agreeing verb or reflexive, in order to further tease apart the processes involved in the earliest stages of processing reflected in the eye-movements here. If low-level factors drive the identical patterns seen in Figure 2.7, then this pattern should be insensitive to the head noun number. If instead the first-pass reading times in Figure 2.7 reflect an early grammaticality detection stage, then the pattern should be reversed for plural subjects. That is, lower reading times should occur for ungrammatical verb forms, rather than longer words.

An important additional goal of Experiments 2 and 2 is to provide additional evidence for the qualitatively different processing profiles exhibited by agreement and reflexives. To build an even stronger case for distinct interference profiles, Experiments 2 and 3 attempt to replicate the interference profiles observed in the agreement and reflexive conditions, respectively.

### *Experiment 2: Agreement revisited*

The results of Experiment 1 suggested a qualitatively different use of morphological features in comprehending reflexive and agreement dependencies. For

agreement features, a feature-matching but inaccessible NP gave rise to clear partial-matching inference effects across regions and measures, suggesting that the morphological features were used to access the subject in memory. No such interference was observed for reflexives. However, both agreement and reflexives patterned alike in early first pass measures; both dependencies showed a main effect of ungrammaticality. To determine whether or not this reflected an early stage of grammaticality detection free of interference, or was instead due to length differences in the critical word across the factor of grammaticality (*was* versus *were*), Experiment 2 expanded the agreement manipulation of Experiment 1 by adding the factor of head noun number. If the early effect of grammaticality is due to differences in word length, then increased reading times should be seen for all instances of *were* as opposed to *was*, regardless of head noun number. On the other hand, if the early effect genuinely reflects detection of grammaticality, then increased reading times should be observed whenever the verb does not match in features with the head noun. An additional goal of Experiment 2 was to replicate and strengthen the finding of partial-matching interference for agreement that was observed in Experiment 1.

### Participants

32 members of the University of Maryland community participated in Experiment 2 (24 females, mean age 21.9). Participants gave informed consent, and were either paid \$10 or received course credit for their participation. The experimental session, including set-up and calibration, lasted approximately one hour.

## Materials

The materials and fillers were identical to those in Experiment 1, with the exception that the four reflexive conditions were removed. Instead, an extra factor of head noun number was introduced into the agreement conditions, leading to a 2×2×2 factorial design that crossed head noun number, interfering noun number, and verbal number. The full set of experiment conditions in Experiment 2 is shown in Table 2.8.

### Agreement conditions for Experiment 2

---

Singular head, singular interferer, grammatical

1 The new executive/ who oversaw/ the middle manager/ apparently/ was dishonest / about the company's profits.

Singular head, plural interferer, grammatical

2 The new executive/ who oversaw/ the middle managers/ apparently/ was dishonest / about the company's profits.

Singular head, singular interferer, ungrammatical

3 The new executive/ who oversaw/ the middle manager/ apparently/ were dishonest / about the company's profits.

Singular head, plural interferer, ungrammatical

4 The new executive/ who oversaw/ the middle managers/ apparently/ were dishonest / about the company's profits.

Plural head, singular interferer, ungrammatical

5 The new executives/ who oversaw/ the middle manager/ apparently/ was dishonest / about the company's profits.

Plural head, plural interferer, ungrammatical

6 The new executives/ who oversaw/ the middle managers/ apparently/ was dishonest / about the company's profits.

Ungrammatical, no interference

7 The new executive/ who oversaw/ the middle manager/ apparently/ were dishonest / about the company's profits.

Ungrammatical, interference

8 The new executive/ who oversaw/ the middle managers/ apparently/ were dishonest / about the company's profits.

**Table 2.8:** Summary of agreement conditions in Experiment 2. Regions included in the analysis are underlined.

## Procedure

The experimental set-up and procedure was the same as in Experiment 1.



## Data Analysis

Sentences in Experiment 2 were divided into identical regions as those in Experiment 1. As in Experiment 1, I report analyses on first-pass, total time, and probability of regression measures.

<b>Singular Head</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, sing</i>	638 (26)	708 (29)	414 (15)	923 (42)
<i>Gram, pl</i>	595 (21)	722 (29)	445 (17)	976 (48)
<i>Ungram, sing</i>	615 (23)	674 (37)	530 (21)	975 (45)
<i>Ungram, pl</i>	669 (28)	697 (31)	497 (22)	950 (45)
<b>Total Time</b>				
<i>Gram, sing</i>	898 (60)	1064 (76)	569 (27)	1170 (65)
<i>Gram, pl</i>	867 (56)	1129 (88)	656 (52)	1296 (109)
<i>Ungram, sing</i>	950 (53)	1085 (94)	875 (52)	1359 (107)
<i>Ungram, pl</i>	980 (59)	1139 (68)	784 (46)	1294 (80)
<b>Pr(Regression)</b>				
<i>Gram, sing</i>	-	.16 (.03)	.16 (.03)	.54 (.05)
<i>Gram, pl</i>	-	.19 (.03)	.14 (.03)	.55 (.05)
<i>Ungram, sing</i>	-	.22 (.03)	.26 (.04)	.64 (.04)
<i>Ungram, pl</i>	-	.19 (.03)	.18 (.02)	.58 (.05)

**Table 2.9:** Table of means (in ms where applicable) for Experiment 2, agreement conditions with a singular head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

As in Experiment 1, all statistical analysis was performed using mixed-effect linear regressions. In addition to grammaticality (GRAM; whether or not the sentence was grammatical), there were fixed effects for plurality (PLUR; whether or not the embedded NP was plural), and head number (HEAD) that indicated whether or not the head noun was plural or singular. PLUR is identical to the fixed effect for interference in Experiment 1. In the context of the full factorial design, calling this factor interference risks being misleading. Whether or not a particular number value

‘interferes’ depends on the number of the verb, and so presenting this factor simply as embedded noun number is more straightforward. The fixed effects were coded identically to Experiment 1, using simple difference coding (e.g., HEAD was coded as -0.5 for singular heads, 0.5 for plural heads). The model-fitting procedure and reporting of the results was identical to Experiment 1. Significant interactions with head number were resolved by assessing the interaction of grammaticality and plurality within plural and singular head noun conditions. Any further interactions were subjected to the same planned pairwise comparisons that were used in Experiment 1 to test the effect of interference within the levels of the factor of grammaticality.

<b>Plural Head</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, sing</i>	638 (29)	718 (35)	507 (24)	941 (57)
<i>Gram, pl</i>	691 (27)	701 (24)	456 (18)	988 (42)
<i>Ungram, sing</i>	665 (28)	692 (34)	473 (22)	936 (52)
<i>Ungram, pl</i>	659 (29)	721 (29)	473 (16)	919 (46)
<b>Total Time</b>				
<i>Gram, sing</i>	981 (62)	1099 (67)	784 (44)	1256 (86)
<i>Gram, pl</i>	1012 (62)	1100 (68)	670 (37)	1277 (79)
<i>Ungram, sing</i>	989 (61)	1057 (76)	742 (47)	1247 (70)
<i>Ungram, pl</i>	997 (66)	1146 (69)	752 (36)	1287 (77)
<b>Pr(Regression)</b>				
<i>Gram, sing</i>	-	.17 (.03)	.20 (.03)	.61 (.04)
<i>Gram, pl</i>	-	.18 (.03)	.15 (.03)	.56 (.05)
<i>Ungram, sing</i>	-	.17 (.03)	.16 (.02)	.60 (.05)
<i>Ungram, pl</i>	-	.21 (.04)	.27 (.04)	.60 (.05)

**Table 2.10:** Table of means (in ms where applicable) for Experiment 2, agreement conditions with a plural head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

## Results

In the pre-critical NP1 region, models of both first-pass and total reading times suggested a slowdown for plural nouns relative to singular nouns (first pass:  $\beta = 32.6$ ,  $SE = 14.7$ ,  $t = 2.21$ ; total time:  $\beta = 65.2$ ,  $SE = 26.6$ ,  $t = 2.46$ ). Unexpectedly, the model for the first-pass times also included a significant term for a three-way interaction of HEAD, PLUR, and GRAM ( $\beta = -161.1$ ,  $SE = 59.0$ ,  $t = -2.73$ ), which was unexpected due to the fact that participants had not yet seen either the embedded noun or the matrix verb, and these were far outside of plausible parafoveal viewing regions. Resolving this interaction showed a significant interaction of PLUR and GRAM for singular head nouns ( $\beta = 97.2$ ,  $SE = 41.7$ ,  $t = 2.33$ ), though resolving this interaction showed no significant differences in planned pairwise comparisons, suggesting that the apparent interaction was spurious and not driven by any true differences between conditions. Additionally, in the same pre-critical NP1 region, no such interactions were observed in total reading times.

At the pre-critical NP2 region, no significant experimental effects were observed in first-pass, total time, or probability of regression measures.

The model fits to first-pass, total time and regression probability at the critical agreeing verb region are summarized in Table 2.11. Across all measures, a significant effect of grammaticality is observed, as are the interactions of GRAM, PLUR, and HEAD, and of HEAD and GRAM, though this later interaction did not reach significance in probability of regression.

Resolving the interaction of HEAD $\times$ GRAM for singular and plural head nouns showed that in both first-pass and total time reading measures, there was a significant effect of GRAM in singular head noun conditions (first pass:  $\beta = 83.2$ ,  $SE = 16.2$ ,  $t = -$

5.15; total time:  $\beta = 209.9$ ,  $SE = 28.2$ ,  $t = -7.43$ ); no significant effects were observed in plural head noun conditions. Replicating the findings of Experiment 1, a significant interaction of GRAM $\times$ PLUR was observed in singular head noun conditions for total reading times ( $\beta = -179.7$ ,  $SE = 56.5$ ,  $t = -3.19$ ), but this just barely reached significance in the first-pass reading times ( $\beta = -64.4$ ,  $SE = 32.3$ ,  $t = -2.00$ ).

	$\beta$	SE	$t / z$
<i>First pass</i>			
HEAD	4.6	11.4	0.41
PLUR	-13.5	11.4	-1.18
GRAM	37.2	11.4	3.26*
HEAD $\times$ PLUR	-22.5	22.9	-0.98
HEAD $\times$ GRAM	-92.1	22.9	-4.03*
GRAM $\times$ PLUR	-7.5	22.8	-0.33
HEAD $\times$ GRAM $\times$ PLUR	113.8	45.7	2.49*
<i>Total Time</i>			
HEAD	10.3	20.0	0.52
PLUR	-30.5	20.0	-1.53
GRAM	116.2	20.0	5.82*
HEAD $\times$ PLUR	-36.9	40.0	-0.93
HEAD $\times$ GRAM	-187.5	39.9	-4.69*
GRAM $\times$ PLUR	-34.2	39.9	-0.86
HEAD $\times$ GRAM $\times$ PLUR	291.0	79.8	3.65*
<i>Pr(Regression)</i>			
HEAD	0.073	0.140	0.53
PLUR	-0.114	0.140	-0.82
GRAM	0.401	0.140	2.87*
HEAD $\times$ PLUR	0.525	0.280	1.88†
HEAD $\times$ GRAM	-0.206	0.280	-0.74
GRAM $\times$ PLUR	0.387	0.280	1.38
HEAD $\times$ GRAM $\times$ PLUR	1.424	0.559	2.55*

**Table 2.11:** Summary of fixed effects for best-fit models at the critical agreeing verb region in Experiment 2, including  $t$ -values ( $z$ -values for first-pass regression probability data). An asterisk (\*) indicates significance at  $\alpha = 0.05$ ; a cross (†) indicates significance at  $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds.

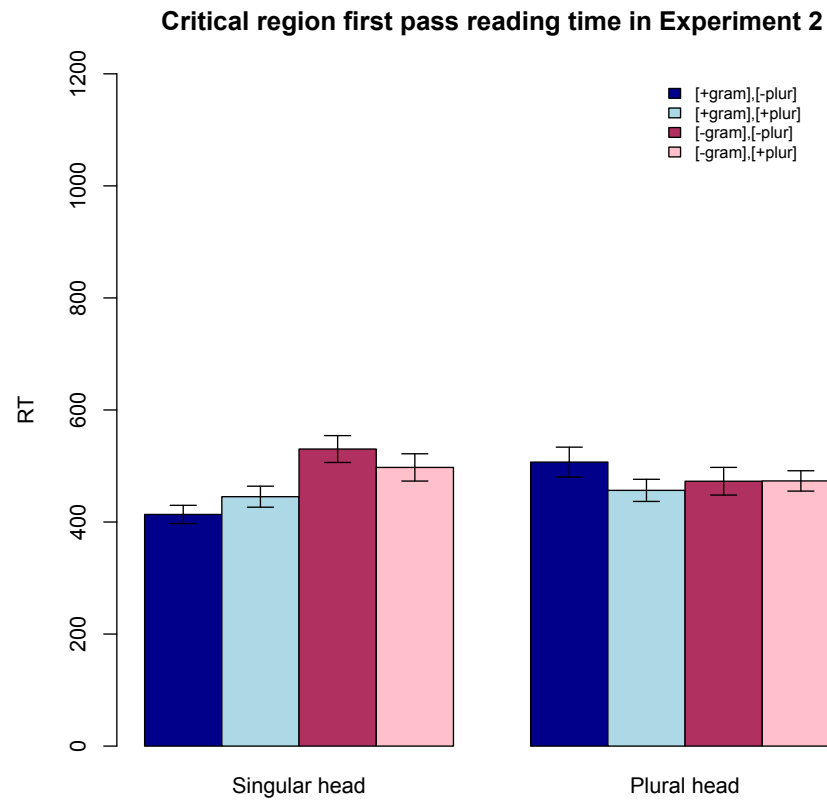
Further resolving this interaction for first-pass times showed no differences due to interference for either ungrammatical or grammatical conditions with singular head nouns, though surprisingly there was a significant difference due to interference for grammatical sentences in the plural head noun conditions ( $\beta = -49.4$ ,  $SE = 22.8$ ,  $t = -2.16$ ). In total reading times, faster reading times were observed when the embedded noun was plural in ungrammatical environments with singular head nouns ( $\beta = -101.9$ ,  $SE = 39.9$ ,  $t = -2.56$ ), replicating the illusion of grammaticality observed in Experiment 1. In addition, the interference effect in grammatical, plural head noun conditions was observed ( $\beta = -104.6$ ,  $SE = 39.9$ ,  $t = -2.62$ ).

A somewhat different pattern was suggested in the probability of regression measure. Resolving the interaction suggested that, as with the other measures, the difficulty due to ungrammatical conditions was observed only for singular head noun conditions ( $\beta = -0.504$ ,  $SE = 0.200$ , Wald  $z = 2.52$ ,  $p < 0.05$ ). No interaction of GRAM $\times$ PLUR was observed in singular head noun conditions, though it did reach significance in plural head noun conditions ( $\beta = 1.099$ ,  $SE = 0.391$ , Wald  $z = 2.81$ ,  $p < 0.01$ ). Resolving this interaction, showed an interference effect in ungrammatical conditions both for plural head nouns ( $\beta = -0.698$ ,  $SE = 0.266$ , Wald  $z = -2.62$ ,  $p < 0.01$ ) and singular head nouns ( $\beta = -0.540$ ,  $SE = 0.287$ , Wald  $z = -2.06$ ,  $p < 0.05$ ).

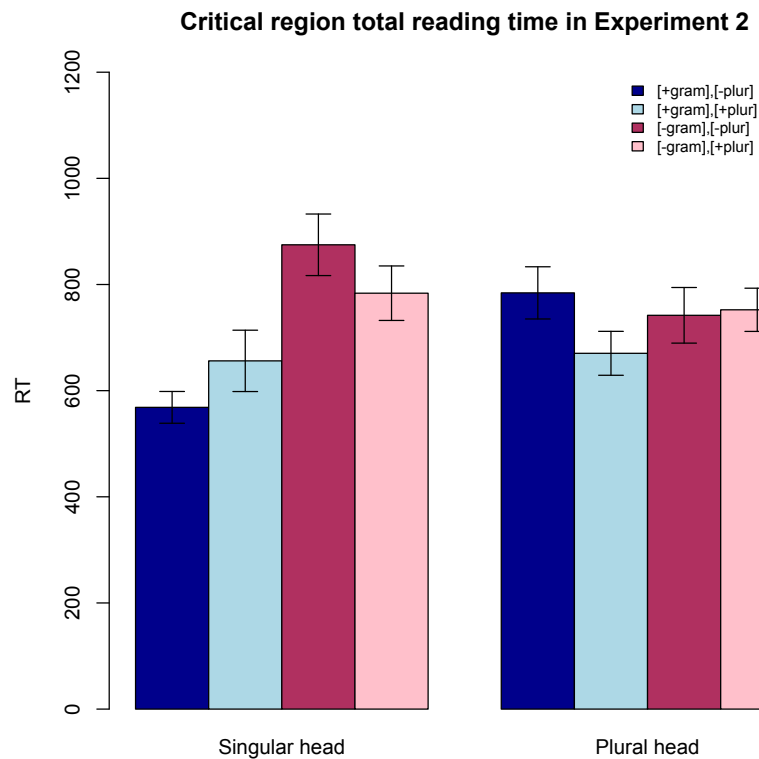
In the spillover region, no significant experimental fixed effects were observed in first-pass or total reading times. In the probability of regression measure, only a marginal effect of grammaticality was observed ( $\beta = 0.218$ ,  $SE = 0.121$ , Wald  $z = 1.81$ ,  $p < 0.08$ ).

## Discussion

The central finding from Experiment 2 is the replication of the interference effect in agreement for the singular head noun conditions. The first-pass and total reading times at the critical verb are shown in Figure 2.9 and 2.10. As in Experiment 1, ungrammatical sentences were read more quickly in the critical verb region when the interfering noun position had plural features that matched the verb. Somewhat surprisingly, there was only limited evidence for processing difficulty related to ungrammaticality in the plural head noun conditions. This contrasts with a finding by Pearlmutter et al 1999, who found that participants were sensitive to ungrammaticality in reading time measures with a plural head noun. The apparent lack of grammaticality effect for plural head nouns was supported by an interaction of head number and grammaticality that was found across measures and subsequent pairwise comparisons.



**Figure 2.9:** Mean first-pass reading time at the critical region in Experiment 2. Error bars show standard error by participants.



**Figure 2.10:** Mean total reading time at the critical region in Experiment 2. Error bars show standard error by participants.

In addition to replicating the crucial illusion of grammaticality effect, an important goal of Experiment 2 was to investigate the nature of the first-pass grammaticality effect observed in Experiment 1's agreement conditions (see Figure 2.7 above). In Experiment 1, both agreement and reflexives patterned alike in first-pass measures, with neither showing any interference effect. This led to the possibility that there was an early stage of processing in which agreement and reflexives patterned alike. However, this conclusion is not supported by the results of Experiment 2. Since the pattern of difficulty observed in first-pass measures in Experiment 2 were not identical for all instances of *was* vs. *were*, this effect is not



fully driven by visual or lexical properties of the agreeing verb. This finding in itself is consistent with there being an early stage of processing that is identical for agreement and reflexives. However, the first-pass pattern for agreement in Experiment 1 did not replicate in Experiment 2. Instead, even in first-pass measures there was a marginal interaction of NP2 plurality and grammaticality that suggested partial-matching interference. Thus, unlike Experiment 1, there does not appear to be a qualitative difference in the pattern observed in early and late measures, making it difficult to conclude that the pattern shown in Figure 2.7 suggests a distinct, interference-free stage of processing for agreement.

Although the attraction effect is usually limited to singular head noun environments in production (Bock and Miller 1991) and comprehension (Wagers et al 2009), in Experiment 2 it was evident for plural head nouns conditions in first-pass regression probability. That is, there were fewer regressions launched when the inaccessible NP position matched the number features of the ungrammatical singular verb *was*. Somewhat surprisingly, there was also a significant slowdown in grammatical plural head noun environments: when the interfering NP2 position was singular, reading times in the critical region were slower than when it was plural.

Lastly, there were also significant slowdowns at the head NP1 position when it was plural relative to when it was singular. This effect may be understood as a plural complexity effect (Wagers et al 2009), where the increased reading times correspond to extra processing engaged by plural nouns relative to singular nouns.

Experiment 2 thus replicates the crucial pattern of partial-matching interference in the present materials, and shows that this interference is evident from

the earliest stages of processing. The lack of any interference in early measures in Experiment 1 was not due to an early, interference-free stage of processing. Instead, it is more likely to reflect a lack of power in early measures.

### *Experiment 3: Reflexives revisited*

Together, Experiments 1 and 2 demonstrate that the computation of subject-verb agreement is reliably susceptible to partial-matching interference: across experiments, the effect of inaccessible, but feature-matching noun phrases was evident. In Experiment 1 no corresponding interference effect was observed for reflexives. Experiment 3's main goal was to replicate this result for reflexives, which would build a stronger case for structured access in the case of reflexives. As in Experiment 2, extra conditions that examine the effect of having a plural head noun were included, making a fully crossed  $2 \times 2 \times 2$  factorial design parallel to the design for agreement in Experiment 2.

### *Participants*

32 members of the University of Maryland community participated in Experiment 2 (24 females, mean age 21.9). Participants gave informed consent, and were either paid \$10 for their participation or received course credit. The experimental session, including set-up and calibration, lasted approximately one hour.

## Materials

The materials and fillers were identical to those used in Experiment 1, with the exception that the four reflexive conditions were removed. Instead, an extra factor of head noun number was introduced, leading to a  $2 \times 2 \times 2$  factorial design that crossed head noun number, interfering noun number, and verbal number. The full set of experiment conditions in Experiment 2 is found in Table 2.10.

### Reflexive conditions for Experiment 3

---

Singular head, singular interferer, grammatical

1 The new executive/ who oversaw/ the middle manager/ apparently doubted/ himself on/ most major decisions.

Singular head, plural interferer, grammatical

2 The new executive/ who oversaw/ the middle managers/ apparently doubted/ himself on/ most major decisions.

Singular head, singular interferer, ungrammatical

3 The new executive/ who oversaw/ the middle manager/ apparently doubted/ themselves on/ most major decisions.

Singular head, plural interferer, ungrammatical

4 The new executive/ who oversaw/ the middle managers/ apparently doubted/ themselves on/ most major decisions.

Plural head, singular interferer, ungrammatical

5 The new executives/ who oversaw/ the middle manager/ apparently doubted/ himself on/ most major decisions.

Plural head, plural interferer, ungrammatical

6 The new executives/ who oversaw/ the middle managers/ apparently doubted/ himself on/ most major decisions.

Ungrammatical, no interference

7 The new executive/ who oversaw/ the middle manager/ apparently doubted/ themselves on/ most major decisions.

Ungrammatical, interference

8 The new executive/ who oversaw/ the middle managers/ apparently doubted/ themselves on/ most major decisions.

**Table 2.12:** Summary of reflexive conditions in Experiment 3. Regions included in the analysis are underlined.

### Data Analysis

Data analysis was identical to Experiment 2.

## Results

In first-pass and total-time measures, modeling revealed a significant effect of head noun number at NP1, reflecting a slow-down for plurals (first-pass:  $\beta = 84.3$ ,  $SE = 16.6$ ,  $t = 5.01$ ; total time:  $\beta = 108.0$ ,  $SE = 26.1$ ,  $t = 4.12$ ). Likewise, there was a significant effect of noun number in first pass measures at NP2, also reflecting a plural complexity effect (first-pass:  $\beta = 40.3$ ,  $SE = 15.7$ ,  $t = 2.57$ ; total time:  $\beta = 79.1$ ,  $SE = 35.7$ ,  $t = 2.22$ ). At NP2, first pass regression probabilities revealed a significant three-way interaction of HEAD, PLUR, and GRAM ( $\beta = -1.23$ ,  $SE = 0.538$ ,  $t = -2.29$ ), but resolving this interaction led to no significant differences. Grammaticality effects were also observed in total time measures at both NP1 and NP2 regions (NP1:  $\beta = 87.8$ ,  $SE = 26.1$ ,  $t = 3.36$ ; NP2:  $\beta = 88.8$ ,  $SE = 25.3$ ,  $t = 3.50$ ).

<b>Singular Head</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, sing</i>	586 (28)	611 (25)	292 (11)	676 (44)
<i>Gram, pl</i>	585 (29)	678 (32)	302 (13)	708 (42)
<i>Ungram, sing</i>	604 (30)	656 (31)	327 (12)	664 (44)
<i>Ungram, pl</i>	604 (28)	670 (32)	345 (16)	696 (38)
<b>Total Time</b>				
<i>Gram, sing</i>	857 (63)	957 (50)	452 (24)	882 (61)
<i>Gram, pl</i>	872 (74)	1029 (56)	444 (20)	886 (56)
<i>Ungram, sing</i>	905 (57)	1050 (59)	574 (41)	952 (71)
<i>Ungram, pl</i>	956 (65)	1151 (64)	605 (32)	968 (54)
<b>Pr(Regression)</b>				
<i>Gram, sing</i>	-	.24 (.03)	.13 (.03)	.57 (.04)
<i>Gram, pl</i>	-	.19 (.03)	.09 (.02)	.47 (.05)
<i>Ungram, sing</i>	-	.19 (.02)	.15 (.03)	.64 (.05)
<i>Ungram, pl</i>	-	.21 (.03)	.12 (.02)	.67 (.04)

**Table 2.13:** Table of means (in ms where applicable) for Experiment 3, reflexive conditions with a singular head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

A summary of the fixed effects at the critical reflexive region is presented in Table 2.12. Both first-pass and total times revealed a similar pattern. Modeling of

both measures revealed a main effect of grammaticality (first pass:  $\beta = 17.5$ ,  $SE = 8.4$ ,  $t = 2.07$ ; total time:  $\beta = 75.5$ ,  $SE = 21.3$ ,  $t = 3.55$ ), as well as an interaction of head number and grammaticality (first pass:  $\beta = -44.4$ ,  $SE = 15.1$ ,  $t = -2.92$ ; total time:  $\beta = -146.1$ ,  $SE = 30.1$ ,  $t = -4.85$ ). Resolving this interaction revealed identical patterns across first-pass and total times. Within singular head noun conditions, there was a significant effect of GRAM (first pass:  $\beta = 39.6$ ,  $SE = 11.3$ ,  $t = 3.49$ ; total time:  $\beta = 148.5$ ,  $SE = 26.1$ ,  $t = 5.70$ ), but no significant effects of GRAM were obtained when the head noun was plural. There were no significant effects on the probability of regression at the critical region. Importantly were no significant effects of the interfering noun's number, nor were there interactions of interfering noun number with any of the other fixed effects. In addition, a significant effect of grammaticality was observed in the spillover region ( $\beta = 57.7$ ,  $SE = 22.7$ ,  $t = 2.55$ ), with no interaction of grammaticality and embedded noun plurality.

<b>Plural Head</b>	NP1	NP2	Critical	Spillover
<b>First Pass</b>				
<i>Gram, sing</i>	681 (46)	651 (27)	311 (11)	742 (57)
<i>Gram, pl</i>	688 (38)	655 (35)	306 (11)	686 (42)
<i>Ungram, sing</i>	683 (39)	621 (24)	297 (10)	684 (52)
<i>Ungram, pl</i>	665 (35)	695 (25)	309 (15)	707 (46)
<b>Total Time</b>				
<i>Gram, sing</i>	933 (62)	991 (46)	486 (21)	930 (86)
<i>Gram, pl</i>	967 (65)	1051 (66)	508 (26)	909 (79)
<i>Ungram, sing</i>	1046 (70)	1053 (56)	513 (26)	971 (70)
<i>Ungram, pl</i>	1077 (78)	1125 (57)	494 (27)	943 (77)
<b>Pr(Regression)</b>				
<i>Gram, sing</i>	-	.17 (.03)	.13 (.02)	.52 (.04)
<i>Gram, pl</i>	-	.22 (.03)	.15 (.02)	.58 (.05)
<i>Ungram, sing</i>	-	.24 (.03)	.15 (.03)	.59 (.05)
<i>Ungram, pl</i>	-	.18 (.04)	.18 (.04)	.61 (.05)

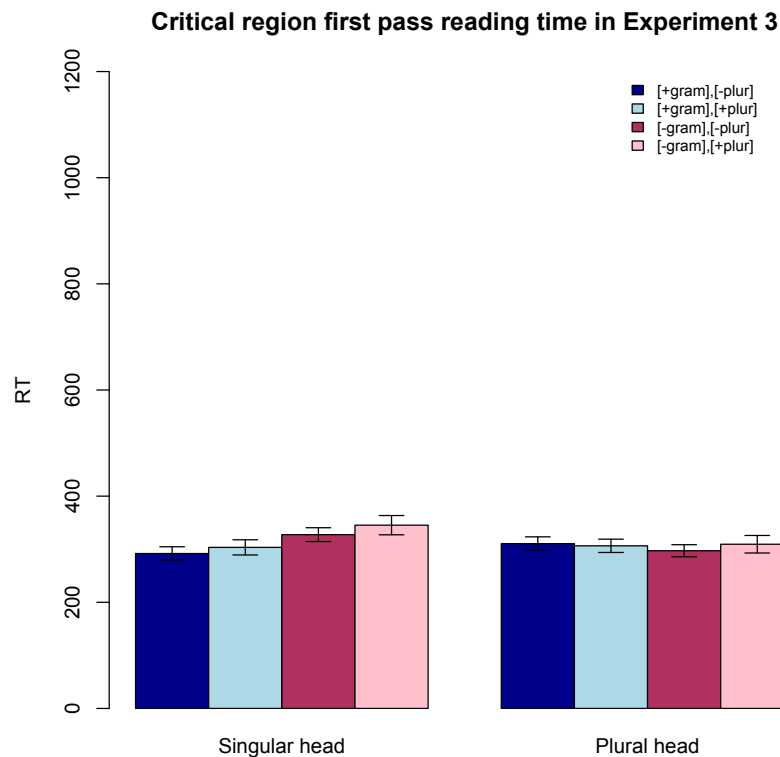
**Table 2.14:** Table of means (in ms where applicable) for Experiment 3, reflexive conditions with a plural head noun, for first pass, total time, and probability of regression. Standard error by participant is shown in parentheses.

	$\beta$	SE	$t / z$
<i>First pass</i>			
HEAD	-11.7	7.6	-1.55
PLUR	9.6	7.6	1.28
GRAM	17.5	8.4	2.08*
HEAD×PLUR	-9.0	15.1	-0.60
HEAD×GRAM	-44.1	15.1	-2.92*
GRAM×PLUR	12.1	15.1	0.80
HEAD×GRAM×PLUR	5.9	30.2	0.19
<i>Total Time</i>			
HEAD	-20.8	15.0	-1.38
PLUR	8.0	17.3	0.46
GRAM	75.5	21.3	3.55*
HEAD×PLUR	-2.9	30.1	-0.10
HEAD×GRAM	-146.1	30.1	-4.85*
GRAM×PLUR	-2.9	37.1	-0.01
HEAD×GRAM×PLUR	-90.7	60.1	-1.5
<i>Pr(Regression)</i>			
HEAD	0.242	0.158	1.54
PLUR	-0.025	0.158	-0.16
GRAM	0.256	0.158	1.62
HEAD×PLUR	0.452	0.316	1.43
HEAD×GRAM	-0.188	0.316	-0.59
GRAM×PLUR	0.080	0.316	0.25
HEAD×GRAM×PLUR	-0.114	0.631	-0.18

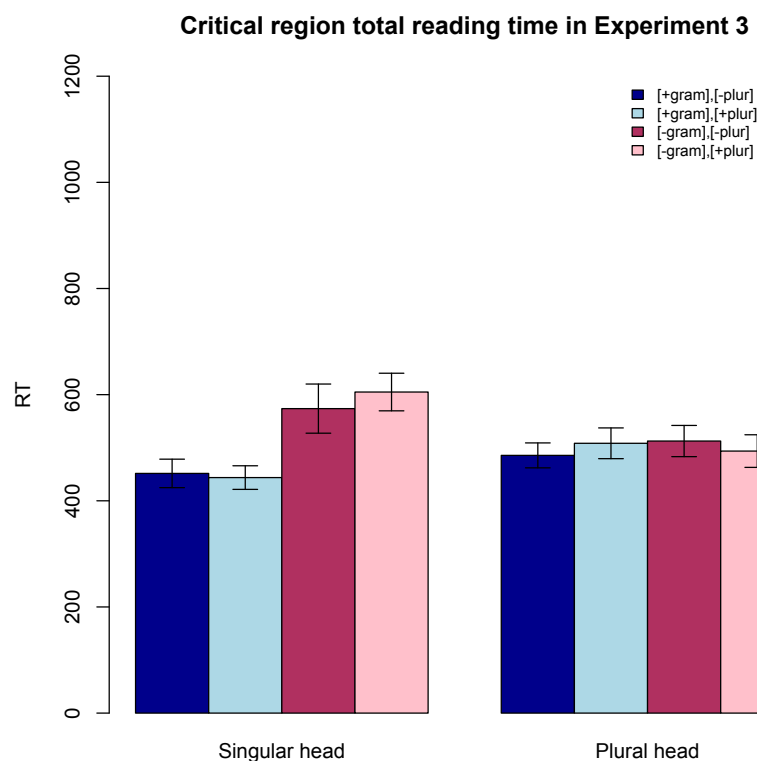
**Table 2.15:** Summary of fixed effects for best-fit models at the critical reflexive region in Experiment 3, including  $t$ -values ( $z$ -values for first-pass regression data). An asterisk (\*) indicates significance at  $\alpha = 0.05$ ; a cross (†) indicates significance at  $\alpha = 0.10$ . First-pass and total time coefficients are in milliseconds.

## Discussion

The most important finding from Experiment 3 is the replication of the processing profile for reflexives that was observed in Experiment 1. This can be seen in Figures 2.11 and 2.12, which show mean first-pass and total reading times, respectively. In no measure or region was there a measurable effect of interfering noun number, nor did it interact with any of the other factors, thus providing additional support for the claim that reflexives are initially processed in a way that is blind to the feature content of an structurally inaccessible NP.



**Figure 2.11:** Mean first-pass reading time at the critical region in Experiment 3. Error bars show standard error by participants.



**Figure 2.12:** Mean total reading time at the critical region in Experiment 3. Error bars show standard error by participants.

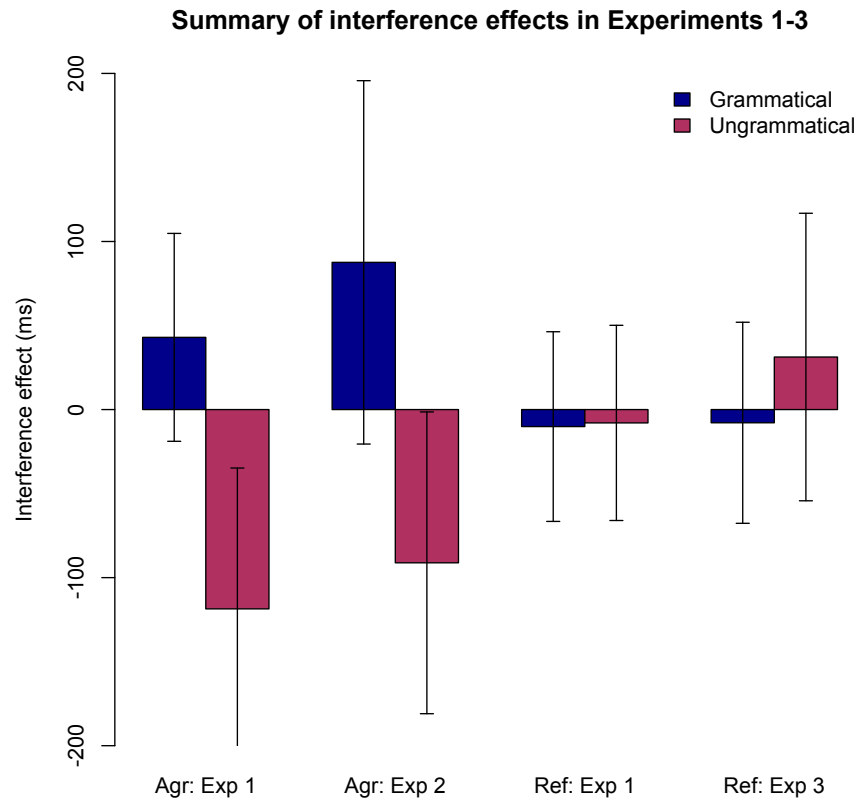
As in Experiment 2, at the critical reflexive region comprehenders appeared to be less sensitive to the feature match between antecedent and reflexive when the head noun was plural. This is surprising in light of the fact that there was no comparable asymmetry in the ungrammaticality response in any other region that showed an effect of grammaticality. These data could suggest that the process of recognizing feature mismatch is sensitive to the markedness of the features involved, but this is entirely speculative at this point. The empirical case for this is unclear, as the asymmetric grammaticality detection pattern was only seen in one region. In regions other than the critical region, a main effect of grammaticality was observed for both singular and plural head nouns alike. It is an interesting possibility that deserves



further study, but this result be treated with caution. This result needs to be replicated and extended before any conclusions about its import for theories of sentence processing can be drawn.

### *Overview of Experiments 1-3*

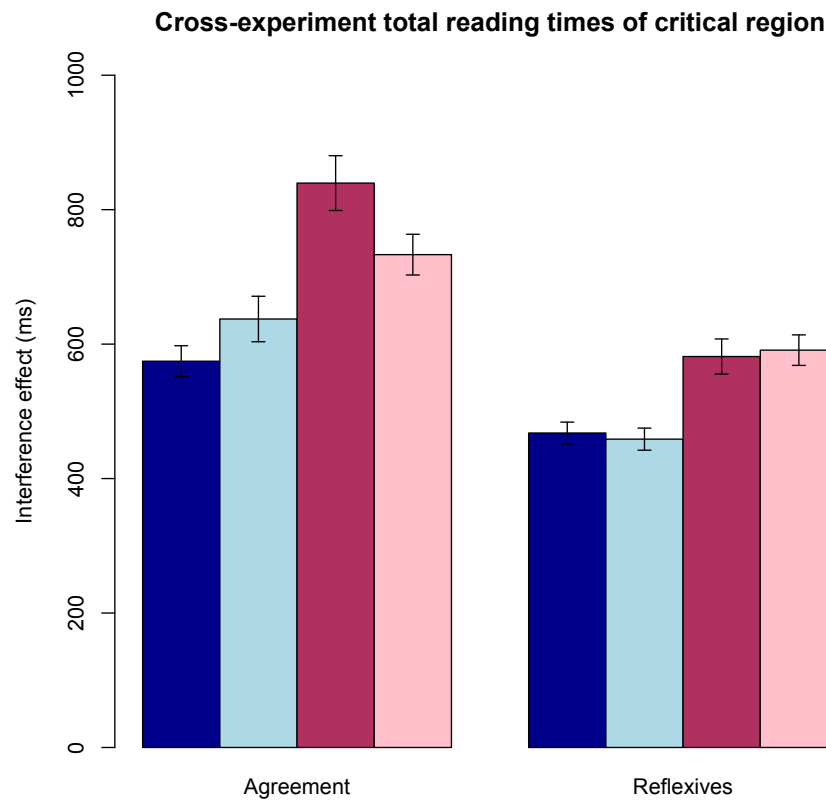
Across Experiments 1-3, a clear picture of the interference profiles for subject-verb agreement and subject-reflexive binding emerges. Focusing on the four conditions that were repeated across all experiments, i.e., the singular head noun conditions, it can be seen that agreement reliably shows the predicted partial-matching interference effects, and that reflexives consistently fail to show sensitivity to the number of the interfering NP. Figure 2.13 presents a cross-experiment summary of the observed interference effects in total time measures at the critical region, demonstrating the clear and reliable contrast between agreement and reflexives. Figure 2.14 presents a cross-experiment analysis, combining the conditions across experiments to highlight the processing profile evident in total reading times at the critical region for agreement and reflexive dependencies. The predictions of a structured access account for reflexives are clearly met: there is no observable interference for reflexives, whereas it is repeatedly observed for agreement. If a feature-based access strategy was pursued for reflexive binding, then interference should have been observed as it was in agreement. This extends prior findings on reflexives (Nicol 1988; Clifton et al 1999; Sturt 2003a; Xiang et al 2009), providing further empirical evidence in favor of structured access in comprehension.



**Figure 2.13:** Interference effects (in ms) observed for in total time measures across Experiments 1-3. Error bars reflect 95% CI by participants.

Additionally, in both Experiments 2 and 3 no interference was observed when the head noun was plural. The processing difficulty associated with anomaly detection for agreement and reflexive appeared to be smaller when the head noun is plural, though this was not a general finding across measures and regions. This has not been observed previously (see, e.g., Pearlmutter et al 1999; Wagers et al 2009), and further study is needed to investigate the source and reliability of these effects. This finding may be orthogonal to the main goal of the studies, however. It has been noted that in production and comprehension, agreement interference is not found in

plural head noun environments, as singular nouns do not generally cause interference (Bock & Miller 1991; Pearlmuter et al 1999; Eberhard et al 2005; Wagers et al 2009). For this reason, the agreement-reflexive contrast observed in the processing of singular head noun conditions provides the crucial contrast.



**Figure 2.14:** Total reading times (ms) at critical region, combining similar conditions across Experiments 1-3. Error bars are standard error by participants.

The central conclusion licensed by these studies is that there is a reliable difference between the interference profile observed for agreement and reflexives. The overall patterns of interference are replicable for agreement and reflexives, and the result of the direct comparison in Experiment showed that agreement reliably shows more interference than reflexives. This is predicted if structured access is

employed in resolving the reflexive dependency, as no interference from the structurally inaccessible NP should occur in this situation.

One possible concern with this finding is that it rests on null-effects logic: for reflexive dependencies, we did not find any trace of interference, therefore there is none. The direct comparison between the interference effects in Experiment 1 shows that this concern is not well-founded, as there is a reliable difference between agreement and reflexives. A counter-argument to this is that reflexives and agreement do differ in the size of the interference effect, but that this does not reflect a qualitative difference between the processing of agreement and reflexives. Instead, reflexives may simply have a smaller interference effect in these materials, one that we simply failed to observe given the power of the current studies. Chapter 3 takes up this question in further detail, using an explicit computational model to derive predictions about the size of the interference effects. To preview the results, modeling results support the claim that the empirical contrast between agreement and reflexives presented here does indeed reflect a qualitative difference in the strategies used to access linguistic short-term memory in the two types of dependency.

### *Discussion*

The finding that reflexives are processed in a manner consistent with structured access mechanisms is compatible with a range of studies that have previously been presented. Both Sturt (2003a) and Xiang et al (2009) suggested that initial processing of reflexives uses exclusively structural information to index and

retrieve its antecedent. To the extent that any influence of interfering nouns was observed, it was unreliable, appeared in later measures, and showed the opposite of the predicted direction for partial-matching interference effects. That is, rather than leading to faster processing, a feature-matched but inaccessible antecedent had the effect of making processing more difficult. The contention that reflexives are processed in a structured manner thus receives wide empirical support from a range of studies.

In line with the conclusion of structured access for reflexives, there are a number of other studies that have argued that the initial stages of binding an argument reflexive are not impacted by structurally inaccessible antecedents. One of the earliest and most widely cited set of results is the cross-modal lexical priming evidence offered by Nicol (1988; also discussed in Nicol & Swinney 1989). In her experiments, she presented sentences as in (3.10) to participants auditorily:

(3.10) The boxer told the skier that the doctor for the team would blame  
himself / him \* for the recent injury.

At the time point in the sentence indicated by the asterisk, participants were presented with a visual word about which they were asked to register a lexical decision. The reaction time for the lexical decision task was measured. Nicol reasoned that if an antecedent was reactivated (retrieved) at the point of processing the anaphor, then semantic associates of that antecedent should show priming relative to a neutral baseline. Thus, in the example given above, if *doctor* is retrieved at a given probe

point, then its semantic associate *nurse* should show faster lexical decision than an unrelated noun (e.g., *paper*). Nicol's findings were straightforward. When the sentence contained a reflexive (e.g., *himself*), only semantic associates of the local, structurally accessible noun showed priming in the lexical decision task: the size of the priming effect for associates of *doctor* was approximately 104ms, compared to -1ms and 11ms to *boxer* and *skier*, respectively. Interestingly, when the anaphor was the pronoun *him* instead of the reflexive, a complementary pattern of priming effects was seen. That is, priming to associates of *boxer* and *skier* was observed (43ms and 58ms, respectively). No such effect was seen for associates of the local noun *doctor* (-21ms) in the pronoun condition. Nicol suggested that this was because the familiar structural binding conditions of Chomsky (1981) were applied immediately online as hard constraints on antecedent consideration. Thus only antecedents that complied with Principles A and B were considered at the point of the pronominal element.

However, some authors have argued that the initial processing of reflexives is in fact prone to feature-based interference. One prominent counterexample to the generalization of structural fidelity that I have presented is presented by Badecker and Straub (2002). These authors criticized the results presented in Nicol (1988), arguing that if all antecedents were considered in a parallel, interactive fashion along the lines suggested in Gernsbacher (1989) or MacDonald and MacWhinney (1990), the cross-modal lexical priming task would be a poor test of access to structurally inappropriate antecedents. Badecker and Straub argued that if antecedents are selected according to both inhibitory and excitatory activation modulation (as in a parallel, neural-network based system), then a lexical decision task that probes only for heightened activation

should not be able to determine whether incorrect antecedents are actively inhibited as well (Badecker & Straub 2002, p 750). If an antecedent receives some amount of activation because of a feature match with a reflexive, but that activation is negated by inhibition due to an inaccessible structural position, they argued, then the cross-modal lexical priming paradigm would show no difference from baseline. This should lead to the incorrect conclusion that this position had not been accessed during processing the reflexive, whereas in fact it had been accessed, receiving equal and opposite excitation and inhibition. Badecker and Straub thus asked if interference from inaccessible antecedents would be observed in a more sensitive measure, such as self-paced reading. In the five experiments they presented, they carried out four comparisons that involved reflexives or reciprocals, similar to the following conditions (drawn from their Experiments 3-6):

- (2.8) a. Jane thought that Bill owed himself a second chance.  
b. John thought that Bill owed himself a second chance.
- (2.9) a. The attorney thought that the judges were telling each other  
which defendants had appeared as witnesses before.  
b. The attorneys thought that the judges were telling each other  
which defendants had appeared as witnesses before.
- (2.10) a. Jane thought that Beth's brother owed himself a second chance.  
b. Jane thought that Bill's brother owed himself a second chance.
- (2.11) a. It appeared to Jane that Bill owed himself a second chance.  
b. It appeared to John that Bill owed himself a second chance.

In Experiments 3 and 4, Badecker and Straub found evidence for what they called the *multiple-match effect*. That is, in (2.8b) and (2.9b), a slowdown was observed in the self-paced reading record relative to (2.8a) and (2.9a), respectively. In

Experiment 3, this was observed two words downstream from the critical reflexive, and in Experiment 4, they observed this effect if they pooled reading times at the four regions following the critical reciprocal. They did not observe any difference between the conditions shown in (2.10-2.11) (Experiments 5 and 6). Badecker and Straub concluded that for reflexives, the binding conditions are not used as an absolute filter on candidate generation for reflexive dependencies. They argued that the extra reading times observed at the reflexive reflected an extra process of candidate suppression when there was an extra, feature-matching (but structurally inaccessible) antecedent present in the string. They suggested that although it might seem that this is an effect of structural position—only the inaccessible subjects in (2.8) and (2.9) caused interference, whereas the structurally less prominent antecedents in (2.10) and (2.11) did not—their preferred interpretation for this finding was that the subject nouns impacted the resolution process through the discourse prominence normally associated with the subject position (Badecker & Straub, 2002, p 764).

It is unclear, however that these results provide a counterexample to the claims of structured access that I have argued for in presenting Experiments 1-3. There are two problems with these results. First, the extent to which interference is observed for reflexives in Badecker and Straub's studies is quite limited, only appearing in one out of four experiments. A clear demonstration of the effect is only observed in Experiment 3 (2.8a-b), causing a reading time slowdown when there is feature-matched but inaccessible antecedents in (2.8b). However, exactly the opposite pattern of results was obtained by Sturt (2003a) with structurally identical materials (Sturt's Experiment 1). Sturt found that sentences like (2.8b) were actually read faster



than (2.8a) in later reading time measures, an effect that he did not replicate in his Experiment 2. Thus, apparent exceptions to the generalization of structural fidelity for reflexives do not appear to be reliable across experiments; the bulk of evidence suggests that there is no interference from inaccessible antecedents during the initial processing of reflexive dependencies.

It is also worth noting that in a similar set of studies using self-paced reading, Clifton, Frazier and Deevy (1999) failed to find evidence for multiple match effects with related materials. In their Experiment 2, they found evidence for a general effect of having complex antecedents (i.e. *the son of the fireman* versus *the fireman*) on reading times after the verb, regardless of whether or not there was a reflexive or a referential DP in that position. In their Experiment 3, they used materials very similar to Badecker and Straub's Experiment 3, including the reflexive comparison in (2.8). They did not find a significant multiple-match effect, though at the critical reflexive region there was a non-significant trend in the direction of a multiple-match effect.

A second problem with the Badecker and Straub's effect concerns the direction of the effect, which is important when interpreting the underlying access mechanisms. In particular, the agreement interference results that have been provided here and elsewhere demonstrate that the clear behavioral signature of incorrectly considering an inaccessible antecedent is processing facilitation due to a feature-matched antecedent. As I demonstrate in Chapter 3, this is a direct prediction of a rational model of memory access. The effect observed by Badecker and Straub instead shows processing inhibition when there are multiple nouns that overlap in feature content, and so is better understood as multiple-match interference. The

opposite direction does not clearly signal that the incorrect antecedent was considered, and current models of memory offer many possible interpretations of such an effect. Importantly, models of this effect do not entail that the antecedent in the incorrect position was incorrectly retrieved during reflexive resolution. One possibility is that memory interference or ‘forgetting’ occurs when memories overlap feature content (Oberauer & Kliegl 2006), possibly due to a process of feature-overwriting (Nairne 1988, 1990). If this kind of interference deteriorates the quality of the memory representations, then the reading time slowdown in (2.8b) might be understood as impeded access to a degraded or unreliable memory trace. This could occur without any consideration of the inaccessible noun phrase at the point of retrieval. The crucial evidence for consideration of spurious, feature-matched antecedents is a partial-matching interference effect as seen in interfering agreement configurations, and to date this has not been reliably observed in the processing of reflexives.

### *Reflexive interpretation*

Although online comprehension evidence is nearly unanimous that reflexives are processed in a structured manner, there are interesting divergences with results from tasks that tap interpretation, as well as production tasks.

In a follow up to his Experiment 1, Sturt (2003a) presented a task where participants were asked to determine who was the agent of the reflexive action in the conditions presented in his Experiment 1 (repeated here as 2.12). After reading small

passages such as these, participants were asked about the patient of the action described in the second sentence, and given a choice between the two participants in the sentence (i.e. *Jonathan* or *the surgeon*).

- (2.12) Jonathan was pretty worried at City Hospital.  
a. He remembered that the surgeon pricked himself with a needle.  
b. #He remembered that the surgeon picked herself with a needle.  
Jennifer was pretty worried at City Hospital.  
c. She remembered that the surgeon pricked himself with a needle.  
d. #She remembered that the surgeon pricked herself with a needle.

Sturt found that the percentage of incorrect interpretations of the reflexive increased when the inaccessible antecedent (which corresponded to the topic of the mini-discourse) matched the features of the reflexive. Thus, in (2.12a) participants responded that Jonathan pricked himself with the needle in 17% of trials, and in (2.12d) they said that Jennifer pricked herself with a needle 31% of the time. When there was no feature-match between the stereotypical gender of the inaccessible NP and the reflexive, as in (2.12b-c), incorrect interpretations occurred at much lower rates (6% and 9% of trials, respectively). In a follow-up article (Sturt 2003b), Sturt suggested that although the binding principles are applied faithfully in a manner predicted by structured access models of memory access, this initial grammatically accurate parse does not bleed later application of discourse-level considerations. He considers the possibility that the ungrammatical offline interpretations that he found for the sentences in Experiment 1 reflect reanalysis in favor of the discourse-central antecedent, as might be expected if some discourse-level factors were brought to bear on the dependency construction process at a later stage (in line, possibly, with

theoretical proposals by Reinhart & Reuland 1993, Pollard & Sag 1992). This contention was supported by the fact that this effect was not observed in Experiment 2, where the interferer no longer held any prominent discourse status.

A two-stage account makes testable predictions about the time-course of the influence of structurally inaccessible antecedents on processing reflexive dependencies. In particular, one should be able to probe reflexive comprehension at early and late time points and observe that the influence of inaccessible antecedents occurs later than the influence of structurally accessible antecedents. One way to accomplish this would be to employ the speed-accuracy tradeoff task (SAT) to examine participants' accuracy on an interpretation task at various time lags from the presentation of the reflexive. If two processes, structural and discourse-level, are involved in producing the interpretation judgment, then the resulting SAT function should display non-monotonic growth. Such a non-monotonic function occurs whenever processing involves a mixture of processes with differing asymptotic accuracies (Ratcliff 1980; McElree & Doshier 1989). If such a two-stage process is correct, accuracy should show a decrease once discourse-level considerations influence the parsing process, indicating a late influence of inaccessible antecedents on comprehension.

In addition, if it is truly due to discourse-level considerations, as Sturt (2003b) suggests, then manipulating the discourse function of the interfering NP should modulate the availability of this effect. Reflexives are known to be sensitive to point-of-view effects (Kuno, 1987; Pollard & Sag 1992; for experimental evidence, see Runner, Sussman & Tanenhaus 2003; Kaiser 2006). For example, Kratzer (2009)

notes that the point of view of an embedded environment modulates the availability of bound readings of indexical pronouns in embedded environments. Borrowing this insight, one might test whether or not it was simply the point-of-view of the embedding environment that led participants to entertain incorrect interpretations in Sturt's Experiment 1b. Thus, *think* in (2.13a) asserts that the embedded clause takes the point-of-view of the thinker; the subject of *think* is the logophoric center of the sentence (a *pivot* or holder of the viewpoint according to Sells (1987)). No such point-of-view shift is required by *be unaware* in (2.13b), and there is no sense in which the subject of *be unaware* must be the pivot or perspective-holder of the embedded clause. If a late influence of the logophoric center drives these interpretive results then (2.13a) should show interpretive errors, but (2.13b) should not show errors:

- (2.13)      a. Jennifer thinks that the surgeon pricked herself with a needle.  
               b. Jennifer is unaware that the surgeon pricked herself with a needle.

These possibilities merit further study, but for present purposes I note that a mixture model of reflexive comprehension of the sort described here would be consistent with both the offline judgment results and the claim of structured access in comprehension.

#### *Attraction in reflexive production*

The primary empirical claim of this chapter is that agreement and reflexive

dependencies are constructed in fundamentally distinct manners in comprehension, implicating structured access in comprehension. Whereas comprehension evidence is near unanimous on this point, this claim is surprising from the point of view of a number of studies that have investigated these dependencies from the point of view of production. The results in this domain suggest a tighter link between the processes involved in agreement and reflexive feature expression than that implied by the comprehension evidence. Most importantly, in addition to interpretive errors, there is some evidence from production tasks that reflexives are equally (or more) susceptible to attraction effects than agreement. This evidence is important, and requires careful consideration in light of the claims advanced by the experiments presented here.

Experimental evidence to this effect is presented by Bock, Nicol, and Cutting (1999). These authors aimed to further explore the phenomenon of agreement attraction in production (e.g. Bock & Miller, 1991), and in particular, they asked whether or not the feature transmission processes involved in subject-verb agreement and pronominal agreement reflected identical underlying mechanisms. Their answer was both ‘yes’ and ‘no’, because of two major findings that suggested similarities and differences in the way agreement and pronominal elements (reflexives and tag pronouns in their study) were licensed.

The experimental paradigm they used was the same employed by Bock and Miller (1991) to induce agreement attraction errors in the lab. In this task, participants listen over headphones to a preamble to a sentence that contains a target subject noun phrase. The participants are then instructed to repeat this preamble and complete the sentence in an appropriate fashion. Responses are scored with respect to whether or

not the correct agreement morphology is produced. In their experiments on pronouns and agreement, Bock and colleagues manipulated two features of the subject noun phrase that they cued participants with. All preambles were of the form in (2.14) below, where the subject NP is modified by a PP that contains an interfering NP position. The first factor they manipulated was the number value of the head noun, which was singular, plural, or collective (i.e. notionally plural but grammatically singular). They crossed this with the number of the interfering noun position, which was either singular or plural. In addition to this within-subjects manipulation, they manipulated the type of dependency that participants were asked to produce between subjects. For subjects that were in the agreement conditions, the task was identical to Bock & Miller (1991). Participants in the reflexive or tag pronoun conditions were asked to produce a continuation with an appropriate pronominal form. In order to ensure that participants would produce the target pronominals, for these conditions participants received explicit instructions that described what a reflexive or tag pronoun was. In addition, the verb that participants were to produce was given in the preamble. In other words, the task for subjects in reflexive or tag pronoun lists was simply to produce the either a plural or singular pronominal form after the preamble was played. The verbs were all transitive for the reflexive lists, and intransitive for the tag pronoun lists.

(2.14) The gang leader / gang leaders / gang with the dangerous rival /  
rivals (armed / vanished) ...

The first major finding that Bock and colleagues report is that anaphors are

significantly more sensitive to the notional number of the subject noun than verbal agreement. Thus, for collective head nouns such as *choir*, participants opted for the plural form in verb agreement in approximately 35% of trials, but for reflexives and tag pronouns in the same environment, the plural form was chosen 75% of the time. The authors conclude that this is a major point of variation in the source of agreement features for reflexives and agreement.

However, with respect to the attraction effect (the production analog of the interference effects in comprehension), the authors found similar effects for both anaphoric elements and verbal agreement. The magnitude of the effect sizes (i.e. the number of inappropriate plurals that were produced following preambles like *the gang leader with the dangerous rivals...*) was similar across tag pronouns, reflexives, and agreement. For agreement dependencies, a singular head noun with a singular interfering noun elicited plural agreement in 2% of trials, but when the interfering noun was plural, plural forms were produced in 10% of trials. In the non-interfering, singular-singular conditions, reflexives and tag pronouns were pronounced in a plural form in 4% and 2% of trials, respectively. But as with agreement, the proportion of trials with plural forms increased in response to an interfering plural noun; the observed rate of plural production was 18% and 17% of trials for reflexives and tag pronouns respectively. Furthermore, the attraction effect apparently obtained in the same manner across all three dependencies. Namely, as has been noted for subject-verb agreement (Eberhard et al 2005), it seemed to be primarily a morphological effect, occurring above and beyond any effect of notional number engendered by the head noun's number value. The authors did note that the apparent rate of attraction



for pronominal forms used in the study was approximately twice as much as for agreement. They suggest that this difference, if reliable, may reflect a ‘two-source’ account of attraction for anaphora, whereby incorrect coreference and agreement attraction both contribute to the incorrect number marking on the anaphors. Even so, the authors hypothesized that the source of agreement attraction is at its core identical for both agreement and anaphoric dependencies, and reflects primarily morphosyntactic agreement relations. Any residual differences that are observed (namely, the sensitivity to the notional number of the head noun) are accounted for as extra processes that are engaged in the case of anaphora.

Bock, Eberhard, and Cutting (2004), as well as Bock, Butterfield, Cutler, Cutting, Eberhard & Humphreys (2006) replicated these the two basic findings on pronominal agreement—increased sensitivity to notional number and canonical agreement attraction—with respect to tag pronouns and reflexives, respectively. As in Bock et al (1999), pronouns were significantly more sensitive to the notional number of the antecedent than was verb agreement. These studies again found similar rates of attraction for both pronouns and verb agreement. An important further contribution of these studies was the finding that pronominal agreement attraction is primarily determined by the morphological number, rather than notional number on the interfering noun. That is, in environments with a collective noun in the interfering position (i.e. *the gang leader with the dangerous group*), no increase in attraction was observed for anaphoric forms. This finding is surprising in light of the marked increase in sensitivity to the notional number of their antecedents that is observed for anaphoric forms in general. This was viewed as strong support for the authors’

contention that agreement attraction, in both pronominal and agreement forms, reflects a primarily morphosyntactic phenomenon. Even for dependencies that are demonstrably more dependent on notional number, attraction remains unaffected by the notional number of the interfering noun.

These results were summarized in Eberhard, Cutting, and Bock (2005), and formalized using a simple linear model that the authors constructed to predict rates of plural form production in the context of their experiments. The model states that the probability of producing a plural on any given agreeing form is a function of a linear combination of the morphological and notional number on all of the nouns contained in the head noun, a value that they refer to as the SAP value (*Singular and Plural*):

$$(2.15) \quad S(r) = S(n) + \sum_j w_j S(m_j)$$

$S(r)$  refers to a linear predictor that, when transformed using an inverse logistic function, predicts the probability that a plural form will be used for a dependent element in any given context.  $S(n)$  refers to a contribution from notional number, a continuously valued term that varies from unambiguously plural to unambiguously singular. The second term refers to the contribution from morphological number marking  $S(m)$  on each of the  $j$  nominal elements contained in the subject noun phrase, each of which has its own weight  $w_j$  that is inversely correlated to the depth of embedding in the noun phrase.

The authors interpret the SAP predictor as feature transmission processes that are present on all trials, providing a real-valued number for the subject noun phrase.

The impressive fits of their mathematical model itself do not directly confirm this claim, however. The model itself is entirely compatible with alternative views of the agreement attraction process, such as the interference-based view adopted here and elsewhere (Badecker & Kuminiak 2007; Wagers et al 2009). The successful model fits simply reflect the fact that notional and morphological number of the subject NP's constituent elements predict, in a linear fashion, the observed amount of plural marking on dependent elements. This view is more or less shared by a variety of different accounts, though the authors do note that their model differs crucially from that offered by Vigliocco and colleagues (Vigliocco & Frank 1998, 2001; Vigliocco & Hartsuiker 2002), who argued for a greater and more direct influence of notional number on the agreement resolution process.

The mathematical formalization provided by Eberhard et al (2005) clarifies where agreement and pronouns part ways in production. Recall that in discussing the agreement-pronoun contrast, Bock and colleagues (1999) argued that agreement and pronoun attraction reflect the same underlying process, and that to the extent that differences between agreement and pronouns are observed, they are due to extra processes that are engaged for pronouns. In this model, the exact difference between reflexives and agreement is clear. The crucial difference, the authors argue, is the fact that pronouns are lexical items in their own right, with their own set of features that influence the calculation of the SAP values. Thus, the calculation of the SAP linear predictor for the pronominal elements is modified as follows:

$$(2.16) \quad S(r_{pro}) = 2S(n) + \sum_j w_j S(m_j) + w_{\square} S(m_{pro})$$

The formulation of the linear predictor for pronouns  $S(r_{pro})$  shown above represents a rearrangement of the terms present in the formula for the SAP value for pronouns given in Eberhard et al (2005, pp. 545), intended to highlight the unity in the process that the authors envision between agreement processes in reflexives and in pronouns. Pronouns reflect the SAP value of their antecedent in a manner that is fundamentally identical to subject-verb agreement with i) a hard constraint that there is a twice as large an influence of notional number for pronouns, compared to agreement and ii) an extra, weighted boost from the morphological features of the pronominal ( $S(m_{pro})$ ). In concrete terms, the authors assume that this additional effect reflects the fact that in production, the notional number of a referent (at the message level) feeds the selection of a lexical item for the pronoun, which leads to a greater role for notional number in the marking process. Taken together, this SAP value affects later reconciliation processes that may alter the morphological features of the pronominal that was selected. It is this later step of reconciliation in which agreement attraction occurs for pronominal elements, as in agreement. Thus in the final formulation of the model, attraction or interference effects have an identical source for both agreement and pronominal elements, but the flow of information in production for pronominals differs in a way that increases their sensitivity to the notional number of the correct antecedent only.

Eberhard and colleagues refer to the differences between agreement and reflexives by referring to them with different terms: since verb agreement arises in the morphological marking stage, under the influence of syntax, it is *structural control* (p.

22). On the other hand, since the lexicon is assumed to bear more of the burden in determining the observed number marking in anaphora, they call the task of selecting lexical items to reflect similar message-level meanings in coreferential elements *lexical concord*. Though these reflect constraints on the expression of number marking at distinct levels of production, the authors retain the use of the superordinate term agreement to refer to both types of processes.

Importantly, even in production, it cannot be the case that agreement on verbs and pronominal elements are controlled by the same SAP value. Although reflexives and agreement appear to pattern similarly with respect to attraction in production, the differential sensitivity of anaphora to notional number suggests an additional process or source of number marking specific to pronominal elements. Bock and colleagues have suggested that this is due to an effect of lexical retrieval of the pronoun, as well as the effect of message level meaning on this selection. However, the fact still remains that comprehension and production show apparently very different interference profiles for reflexives. In comprehension, the sensitivity to attraction varies between the two, and in production, the sensitivity to notional number varies between the two. An account of the differences between agreement and reflexives should be able to account for these two basic differences.

#### *The difference between agreement and reflexives*

In comprehension, reflexive and agreement diverge sharply with respect to interference effects, and in production, notional number plays a greater role in the

determination of reflexive number. Nonetheless the empirical results do point to a limited role for shared processes across the two phenomena: both susceptible to superficial agreement attraction in production. A fully specified processing model of agreement and reflexive dependencies must account for the fundamental differences in the two dependencies (the reliance on notional number), as well as the apparent similarities (attraction effects in production).

One hypothesis for the difference has been proposed by Bock and colleagues (Bock et al 1999; Eberhard et al 2005), who maintain that two features distinguish the processing of reflexives from the processing of agreement: first, reflexives are lexical items in their own right and second, that their role as referring elements causes their antecedent's semantic representation to play a greater role in the dependency construction process. This accounts for the greater impact of notional number on agreement, but not the attraction effect for agreement and reflexives in comprehension. On this model, the attraction effects for both dependencies stem from a single, post-syntactic reconciliation process in the Eberhard et al (2005) model. However, an alternative hypothesis is that the difference between reflexives and agreement is that there are two dissociable processes engaged by reflexive elements: a process of binding chain construction and a process of feature concord. The separation of feature concord and the core dependency formation, combined with differences in information flow in production and comprehension, may be sufficient to understand the production-comprehension asymmetry.

The major claim advanced by Eberhard et al (2005) is that the process difference between agreement and anaphora is that for anaphoric elements, the

message-level representation of the antecedent directly guides selection of the pronominal from the lexicon and its notional meaning, which in turn affects the calculation of the SAP value. Although intuitively appealing, there are a number of concerns with this account of the differences. Most importantly, it is unlikely that the presence or absence of a lexical item to be selected is the crucial factor that divides pronouns and agreement. An agreeing form's sensitivity to notional number and grammatical number does not reflect whether or not an item has lexical features of its own to contribute. Corbett notes this in his discussion of the agreement hierarchy, an implicational hierarchy that describes the relative likelihood of semantic (notional) versus syntactic (grammatical) agreement in any given language (Corbett, 2006). Thus although in British English nouns that routinely control plural agreement due to their notional number (*The committee are going to decide*) do not command plural agreement on demonstrative determiners (*This committee*; cf. *\*These committee*, Corbett 2006). If one makes the plausible assumption that determiners are selected from the lexicon, then these facts are unexpected. That is, the Eberhard et al (2005) model would predict that subject-verb agreement should show less reliance on notional number, whereas there should be greater reliance on notional number for determiners; this is exactly the opposite of what is observed.

In addition, it is unclear that bound pronouns reflect pre-structural lexical retrieval more than subject-verb agreement. First, lexical retrieval is very likely involved for suppletive verb forms such as *was* and *were*, and many of the results in comprehension involved exactly these verbs. Second, it is unclear that reflexive pronouns are in fact retrieved from the lexicon prior to construction of syntactic

structure. The idea that syntactic control determines a bound pronominal's morphological shape is at the heart of many approaches to anaphora (Lees & Klima 1963; Lidz & Idsardi 1999; Kratzer 2009), and from the point of view of production, this assumption seems natural. The model advocated by Bock and colleagues suggests that the lexical boost for reflexives comes from a production stage where the anaphor has been retrieved from the lexicon, but structure has not been created. However, the selection of the correct referring expression depends on the structural relation it bears to its antecedent, and so it would seem that lexically retrieving the pronoun prior to structure generation in production is an inappropriate sequence of operations. On this model, there is no guarantee that the retrieved pronoun would be the correct form given its eventual structural relation to its antecedent. If instead the correct pronominal form is retrieved after the creation of structure, producers could more easily select the appropriate anaphor. This is essentially the model advocated by Bock and colleagues for subject-verb agreement in English, and it is not obvious that anaphor selection is any different.

As noted above, the lexical differences approach captures the greater effect of notional number on reflexives in production. However, in Eberhard and colleagues' model (2005) the same post-structural reconciliation process is invoked in both agreement and reflexives to account for the attraction findings. Given that the inherent lexical differences model addresses the more direct influence of notional number of pronominal agreement, It is not clear that the different comprehension and production with respect to interference can be made to follow from this model. However, if the relevant difference between agreement and reflexives is that the



binding dependency between a reflexive and its antecedents and the feature concord between the two forms must be modeled as distinct processes, then the production-comprehension asymmetry may be a function of differences in the flow of information for these two modes of language use.

A common feature of contemporary theoretical accounts of reflexive pronouns is a separation between the process of building the binding dependency and the process of ensuring feature concord between antecedent and anaphor are distinct (Pollard & Sag 1992; Reinhart & Reuland 1993; Lidz & Idsardi 1999; Buring 2005; Hornstein 2007; Kratzer 2009). This is a natural theoretical move, as the semantic representation of reflexives may in fact be devoid of feature content. The intuition is that the core dependency to be accounted for by a proper theory of reflexive anaphors is the mapping from a surface form such as (2.19a) to an interpretation such as (2.19b) (Ross 1967; Sag 1976; Hankamer & Sag 1976; Sag & Hankamer 1984; Fiengo & May 1994).

- (2.19)      a. Only John hurt himself.  
               b.  $\llbracket \text{Only John} \rrbracket \lambda x . x \text{ likes } x$

The representation of the dependency as a featureless bound variable ensures the proper interpretation (that the two arguments of the predicate obligatorily covary in reference), and accounts for the semantically inert nature of the morphological features. Theoretical accounts that separate the process of feature concord from the process of anaphoric interpretation are well positioned to account for the difference between agreement and reflexives in both production and comprehension.

Once feature concord and binding relations are distinguished, the production-comprehension difference may simply reflect differences in the flow of information with respect to morphological features: in comprehension, the reflexive's features need to be checked, and in production, those features need to be assigned. The attraction in production reported by Bock et al (1999) may reflect an error in a feature-assignment stage, which is entirely absent in comprehension. The post-syntactic reconciliation process in the Eberhard et al (2005) model suggests exactly this possibility. Furthermore, if the construction of the binding dependency precedes the selection of the morphological features in the feature-assignment stage, then the increased sensitivity to notional number may reflect the semantic value of the binding dependency. In contrast, there is no comparable feature-assignment stage that occurs in comprehension; instead, once the binding dependency is constructed in comprehension, the features on the incoming reflexive need to be verified. Thus it is possible that in both production and comprehension, the binding process occurs prior to the feature concord process. The different nature of the feature concord process in comprehension and production could in fact reduce to a difference in the fidelity of feature assignment versus feature checking.

The role of the reflexive's morphological features in comprehension on this account is unclear. Although they are apparently not used in memory access, it is clear that the antecedent-anaphor feature mismatch is evident to comprehenders early (modulo the results on processing plural subjects in Experiment 3). Comprehenders evidently assess the feature concord rapidly online, but the time course of this feature check relative to the construction of the binding dependency is unclear. Likewise, the

main explanatory aim of the lexical differences model in Eberhard et al (2005) was the increased sensitivity to notional number for pronouns. This does not automatically follow from the separation of binding and feature concord processes. It is possible that the process of creating a referential dependency between the anaphor and its antecedent prior to the feature assignment stage in production drives the increased sensitivity to notional number (perhaps by temporarily reactivating the referent's interpretive content), but there is no direct evidence for this conjecture.

Determining which account is a better model of the difference between reflexives and agreement is an important goal for future research. While I have provided a sketch of how attraction might differ for reflexives in comprehension and production, the source of the increased sensitivity to notional number is not obvious. Although important questions about the production-comprehension link remain, the current data demonstrate that reflexives engage in structured access in comprehension. Even in comprehension, however, important questions remain unanswered. In particular, the scope of structured access in comprehension remains unknown. The conditions under which structured access is deployed are taken up in detail in Chapter 5.

### *Conclusion*

The primary goal of this chapter was to determine whether or not subject-verb agreement and reflexive dependencies showed identical interference profiles in English. It was shown that in a tightly controlled comparison, there are systematic

differences in the role that morphological features play in the agreement and reflexive dependencies.

For agreement, the current data replicated and extended the partial-matching interference effects previously observed in comprehension (Pearlmutter et al 1999; Wagers et al 2009). This finding supported the predictions of a content-addressable architecture for sentence comprehension, showing that the morphological feature content of noun phrases in a parse is used as a cue to gate memory access in constructing the agreement dependency online.

For reflexives, however, an altogether different profile was observed. There was no evidence in any experiment that the morphological features of the reflexive were used in memory access. This conclusion rests on the repeated failure to find the partial-matching interference effect that was observed with agreement dependencies, which provides clear evidence of feature-based memory access. Instead, the available evidence suggested that reflexives are uniquely sensitive to the feature match with the structurally accessible antecedent, implicating a memory access procedure that leverages structural position, rather than feature match, in building the binding dependency.

I suggested that the process of structured access in comprehension tracks the construction of a binding dependency, rather than the fact that reflexives have interpretive or lexical content (as in the Eberhard et al 2005 model). On this view, structural binding is the process that gives rise to the profile of structured access in comprehension.

Before a full evaluation of this hypothesis, its predictions, and its fit with the current data, it is important to provide a firmer basis for the claim that reflexives access their antecedents in a fundamentally structured fashion in comprehension. One remaining empirical concern with the data presented in Experiments 1-3 is the possibility that the qualitatively interference profiles for reflexives and agreement actually reflect underlying quantitative differences in what is actually the same memory access procedure across the two dependencies. If this objection were true, it would undermine the case for structured access from the current data. Chapter 3 will focus on formalizing a model of memory access in order to critically evaluate this possibility, and clarify the relation of the present findings to previous research on interference in sentence processing.

### **Chapter 3: Revisiting the interference argument: optimal information retrieval**

In Chapter 2, I argued that the interference profile observed for reflexives implicated a structured access mechanism: when comprehenders retrieve the reflexive's antecedent from memory, it is primarily syntactic information that controls memory access. This argument is to some degree independent of specific architectural commitments: it is built on the observation that the behavioral signature that characterizes feature-based access for subject-verb agreement is consistently absent for reflexives. Instead, reflexives show sensitivity only to the feature content of structurally accessible antecedents. This suggests that structural position, rather than feature match, gates memory access for reflexives, a conclusion that is compatible with a wide range of architectural assumptions. The primary goal of this chapter is to provide an alternative computational argument in favor of structured access. By providing a simple but explicit model of structured memory access in reflexive dependencies, I show that structured access models capture the data better than feature-based models.

Although the claim of structured access is in principle independent of particular architectural commitments, in the present chapter I present an explicit model of structured memory access that builds on existing models of memory architecture in sentence comprehension (Lewis et al 2006). Structured access can be straightforwardly implemented in a content-addressable architecture, if the retrieval cues are sufficiently rich to selectively access the local subject position. I demonstrate this with a simple structured access model based on the ACT-R architecture (Anderson 1989; Anderson & Milson 1989; Lewis & Vasishth 2005; Lewis et al 2006). Providing an explicit model serves two goals. First, it provides a framework for understanding the relation of current results to previous work on interference effects in comprehension. This is important as it clarifies what conclusions are justified from previous work on interference effects, and makes concrete predictions about what data are necessary to falsify the claim of structured access that I have made. In addition, the explicit behavioral predictions derived from the model address some residual empirical concerns with Experiments 1-3. The results of this modeling solidify the case for structured access in comprehension by showing that the reflexive non-interference effect cannot be reduced to superficial differences between agreement and reflexives, such as linear or hierarchical position relative to the antecedent.

I start by presenting an abstract characterization of the information retrieval problem in sentence comprehension that draws on the seminal work by John Anderson (Anderson & Milson, 1989). Casting the problem as one of *rational memory access* allows the application of very general principles of optimal behavior

and rational analysis in reasoning about memory access. This abstract characterization is at the heart of the cognitive architecture ACT-R (Anderson 1990), which characterizes cognition, in part, as an optimal statistical adaption to the structure of the environment (Anderson 1990; Anderson & Lebiere 1998). I recruit this architecture as an explicit model of memory access in sentence comprehension (following Lewis & Vasishth 2005). With the principles of rational memory access and an explicit model of memory access in comprehension, I present a model of the agreement and reflexive data in Experiments 1-3. Using the model, I directly compare the predictions of feature-based and structured access models of reflexive dependency formation. Results show that the structured access model predicts the observed data significantly better than feature-based access models, providing an alternative argument in favor of structured access in comprehension.

Lastly, I contrast findings that have been argued to reflect interference effects with the predictions of the memory model assumed here and elsewhere (Lewis & Vasishth 2005, Lewis et al 2006). Patterns of inhibition and patterns of facilitation have both been attributed to interference effects (compare claims in Van Dyke & McElree 2006 with those in Vasishth et al 2008), with both taken to be indicative of online access to incorrect items in memory. However, formal analysis and simulation evidence show that inhibition and facilitation have distinct mechanistic sources. This analysis suggests that only patterns of facilitation, as seen in cases of partial-matching interference, provide evidence in favor of incorrect access online. This finding suggests that structurally incorrect access may occur less frequently in comprehension



than is generally assumed, and makes a clear prediction about the data that is needed to falsify claims of structured access for reflexives.

### *Rational memory access*

Memory access procedures play an important role in parsing long-distance dependencies. For this reason, it is important to give careful consideration to how memory access might proceed in the ideal case. Note that this is likely true even if there is a substantial role for prospective structure-building processes (Lau, 2009). The architectural assumption that ensures a substantial role for memory retrieval is a limited focus of attention (McElree 2000; McElree, Foraker, & Dyer, 2003; Lewis & Vasishth 2005; Lewis et al 2006, Wagers 2008). In adopting a limited focus of attention, we make the claim that not all information in memory can be maintained in an active, accessible state in parallel; when information is not in an active state, it needs to be retrieved or reactivated in order to participate in further processing. This assumption allows the parsing process to be cast, in part, as a series of retrievals that shunt information into and out of an active processing state that is referred to as *focal attention* (McElree et al 2003; Lewis & Vasishth 2005; Lewis et al 2006; Wagers 2008).

On this view, memory retrieval is an important informational bottleneck in sentence processing, and so a formal consideration of the properties of that access procedure provides potentially useful insight into the operation of the parser. In the context of models such as the ACT-R model of sentence processing (Lewis & Vasishth 2005), it is arguably the memory access mechanism that controls the most

interesting and novel predictions of the architecture (see also Van Dyke 2007). For this reason, in the discussion that follows I abstract away from model details that are arguably orthogonal to the issues at hand, such as the role of lexical access or the possibility of specialized buffers for dealing with particular parsing processes (a number of these issues are discussed in Lewis & Vasishth 2005).

An important milestone in understanding the properties of memory access is Anderson and Milson's (1989) observation that the behavior that characterizes memory access might be understood as rational, in the sense of performing optimally given the constraints on the system (see also Anderson 1990). Anderson and Milson cast the central problem of memory modeling as one of information retrieval, which has well-understood methods of optimization. This abstract approach allowed them to formalize how the memory access should behave in the best case, without making any commitments to the actual mechanisms that implement rational memory access. To borrow terminology familiar from linguistic theorizing, this might be understood as a 'competence'-level characterization of the access procedure, abstracting away from the algorithms that implement memory access in performance in order to gain insight into how the system behaves. This is an important abstraction that makes conclusions based on this reasoning very general; this analysis is compatible with any number of mechanistic implementations that might achieve the same extensional results.

At this abstract level we might first consider one statement of optimal behavior in the task of information retrieval under formal decision theory. If the task is to retrieve the target memory given some evidence (e.g. a set of retrieval cues), then the optimal decision rule is simple. Under these conditions, the Bayes optimal

decision—the behavior that will produce the least error in the limit—is to choose the memory  $m$  that has the highest probability given some evidence  $E$ :

$$(3.1) \quad \operatorname{argmax}_m p(m|E)$$

This reflects a posterior belief over the probability of a given memory being the target, given some retrieval cues. Intuitively, this statement says that for any given retrieval cues, the optimal behavior for a memory system is to retrieve the memory best fits those cues. Put differently, it is not a good idea to retrieve memories that are a poor fit to your retrieval cues. This formalization is useful, as we can decompose this statement of optimal behavior to further gain insight into the system. This posterior belief about the best memory given some cues can be broken down using Bayes' rule into likelihood function of the evidence (i.e., the probability that a particular set of retrieval cues would be used if  $m$  was the target memory), and a prior belief about the likelihood of a memory  $m$ . Note that the denominator, the probability of the evidence,  $p(E)$ , is constant over  $m$  and so not relevant for finding the most likely memory:

$$(3.2) \quad p(m|E) \propto p(E|m)p(m)$$

The assumption that memory access should be rational provides the starting point for Anderson's theorizing about memory: if memory access is rational, then behavioral indices of memory retrieval should be monotonically related to  $p(m|E)$

(Anderson & Milson 1989; Anderson 1990). Anderson and Milson (1989) derive an explicit linking between this high-level characterization of Bayesian optimal behavior to an information retrieval characterization of memory access, and they show that this simple characterization of the problem predicts the existence of frequency effects, recency effects, fan effects, and a number of other observed working memory phenomena. Bayesian optimal performance—fastest access or recall latencies, highest task performance—is achieved by always retrieving the memory with the highest posterior probability, though this abstract analysis of the system’s behavior is silent with respect to the actual mechanisms that implement this behavior. On the assumption that behavioral measures are directly related to  $p(m|E)$ , these widely documented memory access effects may be understood as reflecting optimal adaptation to the statistical structure of the environment. In other words, people optimize their information retrieval routines to do the best they can given cognitive constraints, and the variety of empirical phenomena observed in working memory tasks reflect this simple fact. An important feature of this characterization is that it is independent of implementation, as it only makes the assumption that the goal of memory access is to get the ‘right’ memory for some search query. As such, the above characterization is compatible with a wide range of assumptions about the fine structure of memory.

For example, consider the basic fan effect in list memory (Anderson 1974). The fan effect refers to the phenomenon whereby the more facts are studied in relation to a concept, the slower participants are to retrieve any one fact about that concept. If one adopts a rational model of memory access, there is an easy

explanation for this effect. Informally, this is because the more memories a given cue is associated with, the less diagnostic that cue is: the more red objects are represented in memory, the less helpful a query such as “retrieve the red object” is. The less diagnostic a cue is, the less likely it is to be used given a desire to access some memory, lowering  $p(E|m)$  above. This translates into slower reaction times for items that have greater fan, given the crucial linking assumption that time to access a memory is inversely related to the posterior probability of that memory given some retrieval cues.

This abstract characterization of the memory access problem is extremely general, and it provides a powerful tool for reasoning about how access should proceed at a high level. Using Marr’s terminology, the description of memory access as Bayesian optimal is an example of a *computational level* characterization of a cognitive problem (Marr, 1982). That is, it provides a general statement of the goals of how cognition should proceed, but it does not in itself specify the algorithms or processing routines that guarantee optimal function in memory access. This is an important point, as the comparisons that the analyst performs in determining what behavior would be ‘rational’ given a set of constraints and a goal appear to have an algorithmic character: we calculate the probability of a set of memories for some set of retrieval cues, and then select the most probable memory. However it is important to bear in mind that this comparison is an analytic tool rather than a claim about actual access algorithms. There are, however, explicit algorithmic-level descriptions of optimal memory access that have been developed.

### *Implementing rational retrieval: the model*

The computational characterization of the memory access problem described above forms the basis of the ACT-R cognitive architecture (Anderson & Lebiere 1998). The memory retrieval process in ACT-R reflects one particular implementation of a rational model of memory access (Anderson 1990; Anderson and Schooler 1991; Lewis & Vasishth 2005). The essential idea of this model is that each memory image contained in the memory store (a *chunk*) is associated with a hypothesized activation level  $A$  that indexes its availability. This activation is evaluated at the point of retrieval, and is intended to track the log probability that a given chunk is going to be needed. It is a function of a given chunk's retrieval history and the current context (i.e. the current set of retrieval cues). In other words, activation is a function of a posterior probability (activation prior to retrieval) and a likelihood (fit with retrieval cues): thus the activation  $A$  for a given memory  $m$  is proportional to  $\log(p(m|E))$ , the log of the posterior probability of the memory given the search cues

With this understanding, the link between the abstract, formal statement of the optimal model of memory access and the ACT-R implementation is easier to see. The final activation of a chunk in ACT-R is calculated by summing terms that correspond to the prior ( $p(m)$ ) and the likelihood ( $p(E|m)$ ) in the statement of optimal memory access, as shown in (3.3). The history of retrievals reflects the prior probability, as this indexes the past usefulness of a given memory. Two terms are related to the likelihood of the evidence given a memory: an associative boost for retrieval cues that match the information contained in the memory, and a penalty for retrieval cues that

are not matched by information in the chunk. Together, these implement the likelihood  $p(E|m)$ .

$$(3.3) \quad A_i = B_i + \sum_j W_j S_{ji} + \sum_k PM_{ki} + \epsilon$$

The term that corresponds to the prior,  $B$ , reflects the history of prior retrievals. It is taken by summing the time elapsed since the  $m^{\text{th}}$  retrieval of chunk  $i$ , over all  $n$  retrievals of that chunk. The time elapsed since a prior retrieval is weighted by a decay factor  $d$  (3.4). This implements the intuition that the more times a memory has been accessed, the more likely it is to be accessed again in the future.

$$(3.4) \quad B_i = \ln[\sum_m t_m^{-d}]$$

In addition to base activation, the activation level for a given memory at the point of a retrieval operation is a function of the degree to which it conforms to the retrieval cues. Intuitively, this is because the better the retrieval cues match any given memory, the more likely that memory is to be the target. There are two terms that accomplish this, and they correspond to the usefulness of the cues used (more distinctive is better) and the mismatch between features in the cues and features in the target (all other things being equal, it is better to avoid using search cues that mismatch the target's features). The first determines the strength of association of the retrieval cues to the information contained in chunk  $i$ . This is done by multiplying each of the  $J$  retrieval cue's weights ( $W_j$ ) by the strength of association for that cue on

chunk  $i$  ( $S_{ji}$ ).  $S_{ji}$  is reduced by the *fan* of a cue. As noted above, *fan* is related to  $p(E|m)$  (Anderson & Reder 1999). In other words, the more chunks in memory contain that cue, the greater the fan, and the lower the value of  $S_{ji}$ . Intuitively, this is because a cue that has a high fan does a poor job of discriminating the target memory chunk from distractors, and thus should be down-weighted in trying to access memory in the most effective manner possible. It is calculated in the following manner:

$$(3.5) \quad S_{ji} = S - \ln (fan_j)$$

Where  $S$  is a free parameter of the model (the maximum strength of a cue).  $Fan_j$  is simply the number of items in memory that contain a given cue. In addition to the associative strength between the retrieval cues and the information contained in a memory chunk, there is also a term that penalizes the activation of a given chunk if any of the retrieval cues are not matched by information on the chunk.  $P$  is a negatively-valued free parameter in the model, known as the mismatch penalty.  $M_{ki}$  is a logical value that is 1 for each of the retrieval cues that is not matched by information in the chunk, and 0 otherwise. Thus the total mismatch penalty is  $P$  times the number of cues that are not matched by the information in chunk  $i$ .

In addition, there is assumed to be a certain amount of stochastic noise in the system,  $\epsilon$ , which is distributed according to a logarithmic distribution with a mean of 0 and a variance that is a function of the ‘noise’ parameter  $s$ :

$$(3.6) \quad \sigma^2 = \frac{\pi^2}{3} s^2$$



At the point of a retrieval, ACT-R implements rational memory access in the following fashion. The activation value for every memory chunk is calculated with respect to the search cues being used, as well as each memory chunk's baseline activation, which tracks each chunk's previous retrieval history. The memory chunk that has the greatest activation (equivalent to the greatest posterior probability of being the target memory) is the memory that is retrieved. This follows from assuming an optimal decision metric, and can be understood as a 'winner-take-all' situation. In this way, memory retrieval operates in a manner similar to that endorsed by race models of syntactic comprehension, which posit that in syntactic disambiguation, the fastest available syntactic analysis is the one that is adopted (McRoy & Hirst 1990; Traxler, Pickering & Clifton 1998; Van Gompel, Traxler & Pickering 2001). This is because the most active chunk has the shortest retrieval latency, which is calculated by exponentiating the activation value, and multiplying by a latency scaling factor  $F$ :

$$(3.7) \quad T_i = F e^{-A_i}$$

From (3.7) it can be seen that greater activations translate to faster access. The direct relation between activation (probability of retrieval in context) and retrieval latency ensures that the fastest memory to be retrieved is guaranteed to be the most likely on this model. The race aspect of the ACT-R model is important in making sense of the model predictions, and I return to the relationship between race models

and retrieval latencies in ACT-R below (the link is also noted in Lewis & Vasishth 2005 p 399).

By adopting an explicit model of memory retrieval, it is possible to compare quantitative model predictions to the results from Experiments 1-3. By linking model predictions to experimental results, firmer conclusions about the underlying system may be drawn. Before presenting the results of this modeling, however, it is important to reiterate the crucial linking assumption between retrieval latencies and online reaction time measures made here and in other applications of this model to sentence processing (Lewis & Vasishth 2005; Vasishth et al 2008). The crucial linking hypothesis is that the behavioral measures that are thought to index retrieval operations in experimental settings are inversely and monotonically related to  $p(m|E)$ . In other words, faster retrieval should result in faster reading times, and slower retrieval should result in slower reading times. Although this is a useful starting assumption, it need not be the case: there are any number of processes that might obscure the relationship between retrieval time and observed reading time, especially in the context of sentence processing. When predicted retrieval times are used to model behavioral measures such as reaction time, this implicitly assumes that no interpretive processes, reanalysis procedures, or error signals are indexed in the behavioral measure. This seems unlikely, as retrieving a memory in sentence comprehension is presumably carried out in the service of enabling further processing. It is almost certain that these processes are also reflected in reading time measures. However, for purposes of the current discussion, I assume with others (Lewis & Vasishth 2005; Lewis et al 2006; Vasishth et al 2008) that whatever extra

processes are engaged as the result of a retrieval, they do not disrupt the monotonic relation between memory retrieval times and online reading time measures.

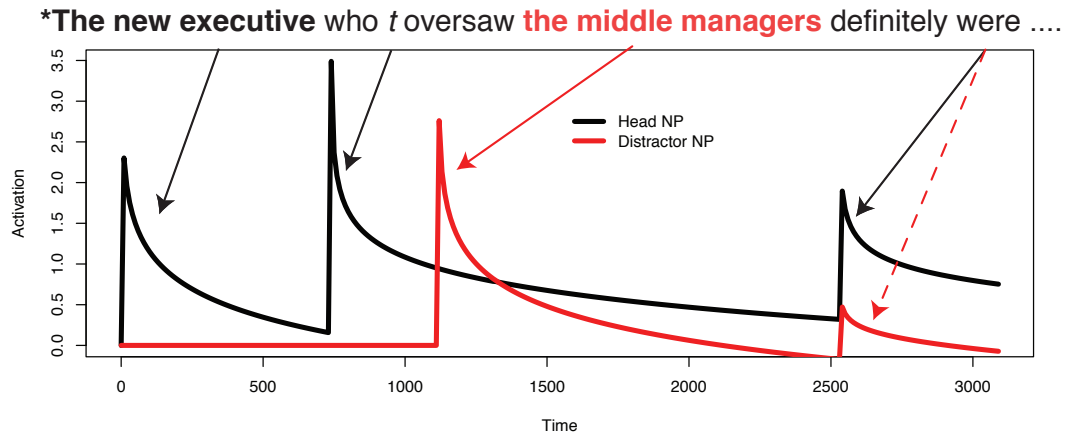
### *Modeling reflexive and agreement dependencies*

#### *The relation between model predictions and experimental findings*

The ACT-R model described above can be used to provide quantitative predictions about the hypothesized retrieval processes that occur during the processing of long-distance dependencies. The explicit predictions derived from the model above have the potential to be informative in understanding the model results presented in Chapter 2. One possible concern with the results presented in Chapter 2 is the assumption that the interference effect sizes for agreement and reflexives should be equal. This assumption is implicitly made in contrasting the presence versus absence of interference effects in the within-subjects comparison of agreement and reflexives in Experiment 1. The logic of this comparison is the following: we observed interference in agreement, so the experiment has the power to detect the interference effect. Therefore, the failure to observe it for reflexives means it is not there. This logic is reasonable if the interference effects for both dependencies are of comparable magnitude.

However, it is not clear that the assumption of equal-sized effects is valid, which risks undermining this logic and weakening the conclusion. In particular, there are unavoidable differences between agreement and reflexives with respect to their

linear position in the string, and the parsing processes that precede the critical subject retrieval process for agreement and reflexives. For example, consider Figures 3.1 and 3.2, which are derived from an ACT-R model of the retrievals involved in parsing agreement or reflexive dependencies, respectively. These show an average trace of the activation of the target and distractor NPs across the parse, leading up to the critical retrieval. One crucial difference between agreement and reflexives is evident: for reflexives, but not agreement, the target memory (i.e. the local subject) is reactivated just prior to the critical retrieval, giving the local subject a large activation advantage over the interfering NP at the point of the critical reflexive retrieval. This fact leads to a plausible alternative explanation of the experimental results in Chapter 2: the lack of interference effects for reflexives may simply reflect differences in passive memory dynamics that are unrelated to the construction of the antecedent-anaphor dependency (the *baseline differences hypothesis*). If there is a baseline bias in favor of the local noun immediately after processing the matrix verb, this may selectively eliminate interference effects for reflexive dependencies. If the baseline differences hypothesis is true, then the results presented in Chapter 2 may not in fact provide evidence for structured access.



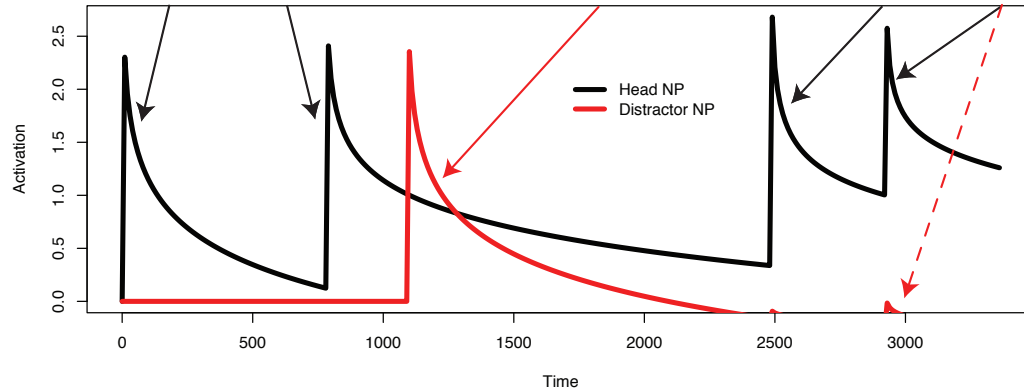
**Figure 3.1:** Average activation for target (black) and distractor (red) NPs for a sentence that shows partial-match interference at the agreeing verb. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural verb *were*.

It is difficult to assess the impact of that this baseline bias should exert on any reflexive interference effects without an explicit model, and for this reason the quantitative fits obtained by modeling provide useful insight on the experimental results. There are two key questions that the models may be used to answer. The first is whether or not the predicted size of the interference effect for reflexives is demonstrably smaller than that for agreement. If this is true, then a second, more critical question is raised. Given the observed interference effects in Chapter 2, is more likely that they reflect a feature-based model or a structured access model of reflexives? This second question is critical for the main claim of this thesis. If a

feature-based model of reflexive antecedent access can adequately capture the experimental data, then the claim for structured access based on the data in Chapter 2 is weakened. If, on the other hand, structured access reflexive models better fit the observed data, then the models provide an additional argument for structured access in comprehension.

Each of the two questions was investigated in a separate computational experiment. Experiment 4 models the size of the predicted interference effect for reflexives if they retrieved their antecedent using a full set of morphological and structural cues. This experiment directly compares a feature-based access model of agreement to a feature-based access model of reflexives, to assess whether or not passive memory dynamics do in fact minimize any interference effect that might occur for reflexives. To preview the findings, the modeling results in Experiment 4 suggest that reflexives should in fact show a reduced interference effect relative to agreement, even if they retrieve the subject in the exact same fashion. This finding makes it important to determine the answer to the second question: do structured access or feature-based access models better capture the experimental data in Chapter 2? Experiment 5 directly compares the predictions of a feature-based model of reflexives to a structured access model of reflexives to attempt to answer this question.

\*The new executive who *t* oversaw the middle managers apparently doubted themselves



**Figure 3.2:** Average activation for target (black) and distractor (red) NPs for a sentence that shows partial-match interference at the reflexive. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural reflexive *themselves*.

### Modeling feature-based and structured access

The crucial contrast that is drawn here is between feature-based and structured models of memory access. Feature-based models use a mixture of structural and feature-based cues to access memory, as is maintained by interference models of the agreement attraction effect (Badecker & Kuminiak 2007; Wagers 2008; Wagers et al 2009), as well as models of NPI interference effects (Vasishth et al 2008). In contrast, a structured access model means that morphological feature information is not used to access memory; instead, only structural cues are used in retrieval. In the context of

the ACT-R model, this means that the search cues are selectively restricted to only structural information. This has the effect of rendering access effectively blind to the morphological feature content of the interfering position, and so there should be no effect of the morphological feature content of the distractor noun on the retrieval process.

Admittedly, the notion of ‘structural cue’ in the context of a content-addressable architecture is unclear, and there are a number of important issues that arise when considering the plausibility of representing structured linguistic representations on individual memories (as noted by Vasishth et al 2008; Wagers 2008). The primary difficulty is that a relational notion like c-command is defined over pairs of constituents, which makes it qualitatively different from the lexical content of constituents (such as morphological features). Inherent content can straightforward be used as a retrieval cue in a content-addressable architecture, but relational notions like [+c-command] cannot be encoded on any single chunk, making it difficult to use it as a retrieval cue. However, the issue is somewhat orthogonal to the present concerns. Whether or not the relational notions that characterize linguistic structure can be adequately encoded in retrieval cues is a question that speaks to the feasibility of content-addressable models of sentence processing, rather than the possibility of structured access. For the current purposes I follow Vasishth et al (2008) in hypothesizing the existence of a cue that is functionally equivalent to [+c-commanding], even though significant extra architectural commitments may be required to implement such a cue. Even if [+c-commanding] does not prove to be a plausible retrieval cue due to its inherently relational nature, the fact remains that in



order for a content-addressable parser of the sort presented here to function at a basic level of accuracy, there must be some cue that discriminates structurally accessible content from inaccessible content (a similar approach was taken by Lewis & Vasishth 2005). Although I adopt the term [+c-commanding], this is merely a notational convenience. No matter what label it is assigned, the role of this structural cue in the present model is formally equivalent.

### The model

In Experiment 4, all 8 conditions from Experiment 1 were modeled, and in Experiment 5, only the reflexive conditions were modeled. The structure of the crucial experimental sentences is repeated in (3.8). In Experiments 1-3, the structure of the subject noun was held constant across dependencies. The subject noun was always a singular, gender-biased noun that was modified by a subject relative clause. As before, the main clause subject noun is the *target* memory due to its structurally accessible position (i.e. it is the c-commanding local subject), regardless of its feature content. The *distractor* noun was the object NP that was embedded inside the relative clause; since it does not c-command the verb or reflexive, the distractor was always structurally inaccessible. The *probe* refers to the memory retrieval of theoretical interest. For agreement sentences, the probe retrieval was engaged by the agreeing verb form, and reflected the need to find and attach the subject of the VP. For reflexive dependencies, the probe retrieval was initiated by the reflexive element, and reflected the processes involved in retrieving the antecedent.

- (3.8) a. [The new executive]<sub>target</sub> who oversaw [the middle manager]<sub>distractor</sub> apparently [**was**]<sub>probe</sub> dishonest about the company's profits.  
b. [The new executive]<sub>target</sub> who oversaw [the middle manager]<sub>distractor</sub> apparently doubted [**himself**]<sub>probe</sub> on most major decisions.

It is important to consider the role of the free parameters in a model such as ACT-R, as there are a number of model parameters that the modeler is free to manipulate: these include the variance on the noise distribution ( $s$ ), the magnitude of the penalty for mismatching features ( $M$ ), the amount of maximum activation for each chunk ( $G$ ), the decay rate ( $d$ ), the base associative strength for a cue ( $S$ ), and the scaling parameter ( $F$ ). There are in general default settings for many parameters (Anderson & Lebiere 1998; Lewis & Vasishth 2005), and model fits in previous work have adopted a single parameter setting to assess the quantitative fit provided by the model. I adopt a different approach here. The focus of the modeling reported here is not to determine whether or not the model can accurately fit a given set of experimental results, as has been the focus in previous applications of ACT-R to sentence processing results (Lewis & Vasishth 2005; Vasishth et al 2008). Instead, I am comparing the fit of two separate model structures to the data, and asking which model is better suited to capturing the observed data. Thus rather than focusing on the fit of a single set of parameters, I consider the effects of a range of parameter settings that span values assumed in previous work. In all reported results, 324 different model combinations are reported. The most important advantage of this approach is that it allows an assessment of the robustness of a predicted effect: if an effect survives under various parameterizations of the model, it can be regarded as a reliable

prediction of the model. Effects that are present only under certain parameter settings require a greater number of assumptions to motivate; they require independent motivation for adopting that parameter setting, beyond noting that it predicts a given set of data. The sole exception to this approach is the scaling parameter  $F$ , which was chosen to ensure that predictions of the model were on an appropriate scale (in all simulations,  $F = 2.0$ ).

The model of the crucial memory retrieval engaged by the probe provides two dependent measures of interest. The first measure is the proportion of retrievals that result in recalling the target memory position. Recall that the memory that is retrieved is the one that has the greatest activation at the point of retrieval, given the search cues; in other words, retrieval selects the most probable memory given a retrieval query. The percentage of accurate retrievals thus refers to the percentage of trials in which the target memory was the most probable memory at the point of retrieval. The second measure provided by the model is a measure of the *interference effect*, which is a measure that is derived from predicated retrieval latencies that the model generates. As in Experiment 1, the interference effect refers to the effect on predicted retrieval times that occurs when the distractor NP contains interfering feature content. It is calculated by subtracting the predicted retrieval time for [+intr] conditions from [-intr] conditions, and is a measure of the predicted impact of interfering material on retrieval times.

The Lewis and Vasishth (2005) ACT-R sentence processing model is a fully specified parser, including an implementation of a left-corner parser and a lexical access module. In order to focus on the predictions of the memory access component

of the parser, the model presented here abstracts away from these and simply stipulates the schedule of retrievals required to parse the sentences in (3.8). Thus the model presented here is a minimal implementation of the crucial retrieval process. The schedule of non-critical retrievals, which controls the prior history component of the activation process, is given in Appendix A; this schedule is derived from empirical estimates of processing times from Experiments 1-3, as well as a consideration of the hypothesized parsing processes that engage both the target and distractor NPs.

*Experiment 4: comparison of interference effect for agreement and reflexives*

Experiment 4 asked whether or not there is a substantial difference in the predicted size of the interference effect for reflexives relative to agreement, on the assumption that both types of items retrieve their antecedents in exactly the same manner. Thus, the model for reflexives and agreement accesses the local subject in the exact same manner, using a feature-based search that uses a combination of morphological and structural cues. As described above, the two dependencies differ on a number of dimensions (such as position relative to the verb and linear distance from the subject). Thus it is worthwhile to assess the nature of any baseline differences that exist. Both agreement and reflexive retrievals are modeled with the same feature-based access model: the subject is retrieved using the cues *local subject* and *number*; in addition, singular reflexives probe memory with the cue *gender*. Due to syncretism in the preterite verb form in English, the retrieval of the subject at the

verb that directly precedes the reflexive (see Figure 3.2) does not include *number*; thus there is a targeted re-activation of the local subject at this point.

As described above, a range of model parameterizations were considered, totaling 324 distinct combinations in all. For each combination, 5000 Monte Carlo simulations were run. Each simulation included the full series of hypothesized retrievals, including random noise, and provided a prediction for i) the retrieval latency of the most probable memory and ii) the identity of the most probable memory (target or distractor). Within each parameterization, the average retrieval latency of the most probable memory in each individual trial was calculated, providing an estimate of the average retrieval time that occurs for that retrieval. The average retrieval time does not distinguish between target and distractor retrievals; the predicted retrieval latency does not contain error signals generated by retrieving the distractor NP. In addition to the average retrieval latency, within each model parameterization the proportion of correct retrievals was calculated by observing on how many trials the target NP was correctly retrieved. All eight conditions from Experiment 1 were modeled, and they are repeated in a simplified form here as (3.9) and (3.10) for convenience:

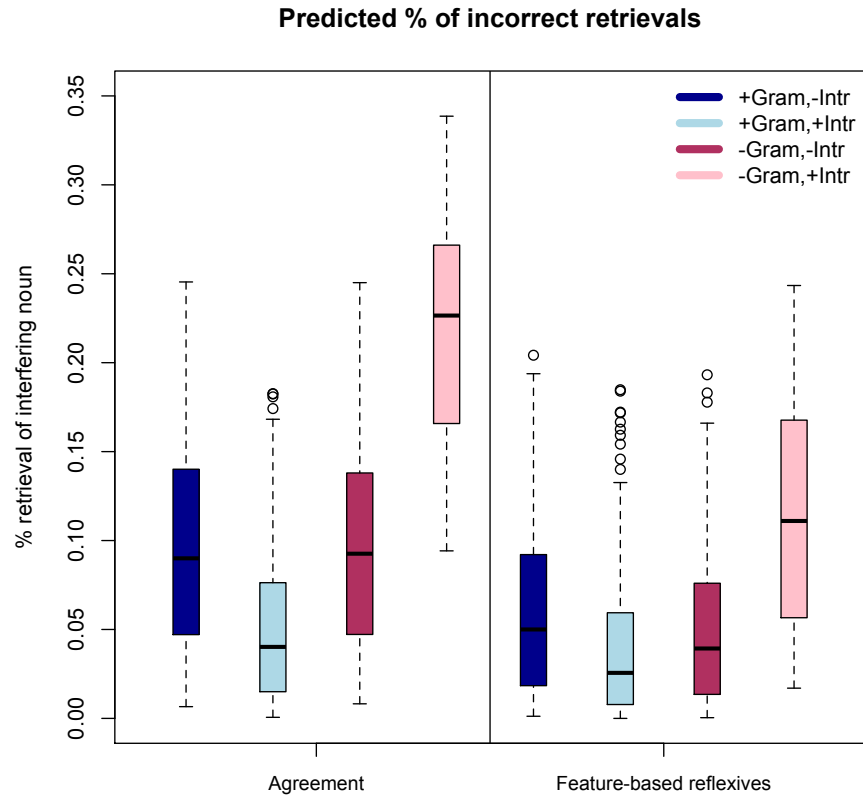
- (3.9) a. The executive who oversaw the manager apparently was  
dishonest about the company's profits. [+gram,-  
intr]  
b. The executive who oversaw the managers apparently was  
dishonest about the company's profits. [+gram,  
+intr]  
c. The executive who oversaw the manager apparently were  
dishonest about the company's profits. [-gram,-  
intr]

- d. The executive who oversaw the managers apparently were dishonest about the company's profits. [-gram, +intr]
- (3.10) a. The executive who oversaw the manager apparently doubted himself on most major decisions. [+gram, -intr]  
 b. The executive who oversaw the managers apparently doubted himself on most major decisions. [+gram, +intr]  
 c. The executive who oversaw the manager apparently doubted themselves on most major decisions. [-gram, -intr]  
 d. The executive who oversaw the managers apparently doubted themselves on most major decisions. [-gram, +intr]

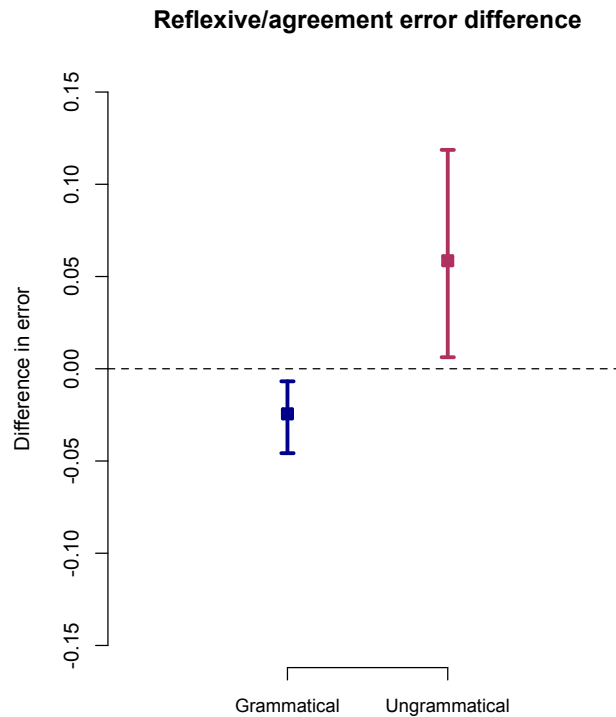
The predictions for each condition were determined in independent runs. The interference effect was determined by subtracting the retrieval latency for [-intr] conditions from [+intr] separately for grammatical and ungrammatical sentences.

The rate of incorrect retrievals for each condition is plotted in Figure 3.3. It can be seen that there is a sizeable decrease in the number of incorrect retrievals in interfering environments for reflexives compared to agreement. To determine whether or not this held within each model parameterization, the difference in error rate as a result of interference was calculated for both grammatical and ungrammatical sentences alike. These values were then directly compared across reflexive and agreement models, for each parameter setting. The distribution of these differences is shown in Figure 3.4, which indicates that for almost every single model setting, the percentage of incorrect retrievals due to interference was lower for reflexives than for agreement dependencies. This indicates that the baseline activation differences that

are evident in comparing Figures 3.1 and 3.2 have a clear impact on the amount of interference that is expected to occur for reflexives.



**Figure 3.3:** Percentage of retrieval of distractor NP, for all parameterizations ( $n=324$ ), at critical probe position for agreement and feature-based reflexive models.

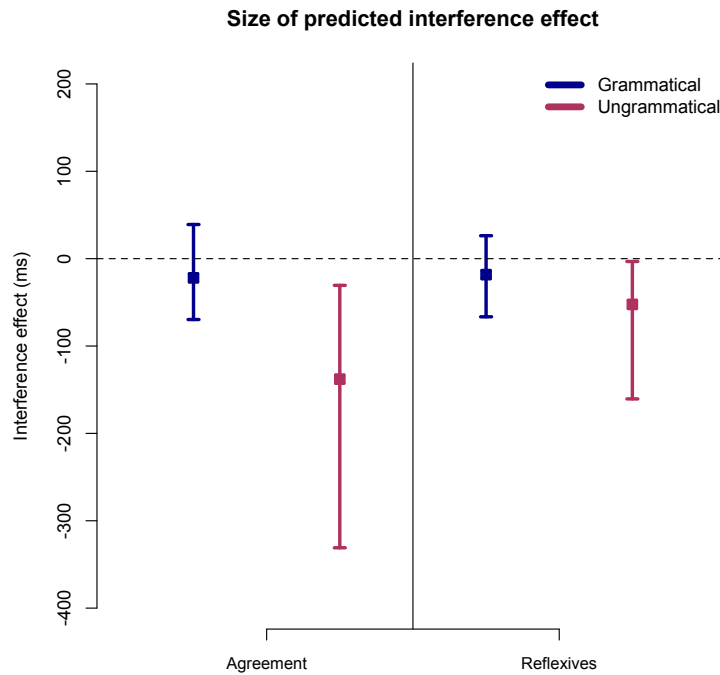


**Figure 3.4:** Difference in interference error between agreement and feature-based reflexive models (n=324). Error bars indicate 95% confidence intervals.

The predicted interference effects for agreement and reflexives are summarized in Figure 3.5. The values in Figure 3.5 reflect the average retrieval time for [+intr] conditions minus the retrieval time for [-intr] conditions, for grammatical and ungrammatical sentences alike. Note that the average retrieval time for any given condition is a mixture of retrieval latencies for incorrect and correct retrievals (as in Vasishth et al 2008). There are two main findings from these results. First, the direction of the model's predicted interference effect for grammatical conditions is not stable across parameterizations; the model's predicted interference effect for grammatical sentences is instead centered closely around zero. This is an important observation that I return to below. The second clear finding is that the predicted size

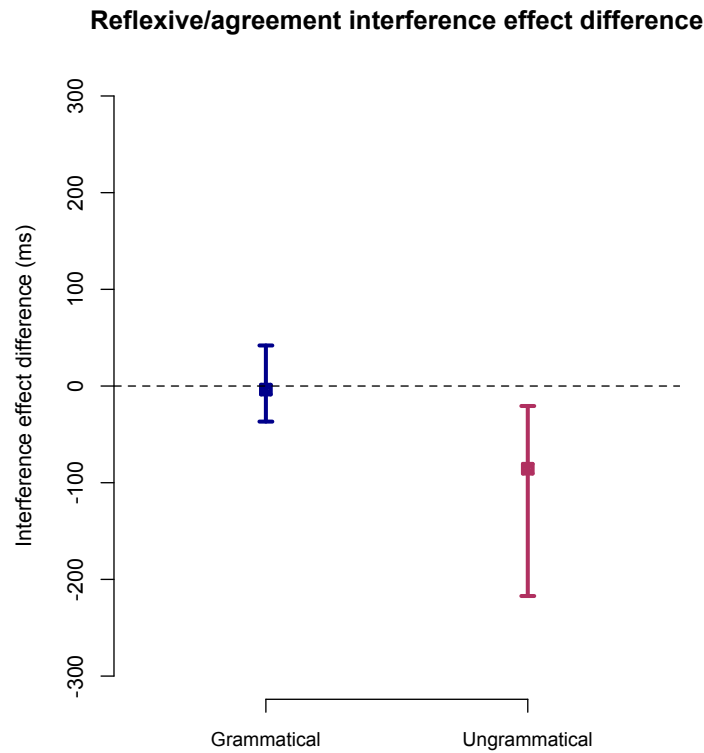


of the interference effect is smaller for reflexives than it is for agreement. In ungrammatical sentences, the mean interference effect for agreement is approximately -137ms, whereas for reflexives, it is less than half that (-52ms). Thus reflexives are expected to show approximately half as much interference as agreement, an observation that holds independent of the magnitude of the interference effect. Figure 3.6 shows a direct comparison of the agreement and reflexive interference effect for each parameterization, calculated by subtracting the agreement interference effect from the reflexive interference effect. It can be seen that across parameterizations, there is no reliable difference between agreement and reflexives for grammatical conditions. However, in ungrammatical conditions, almost every parameterization predicts a smaller interference effect for reflexives than for agreement.



**Figure 3.5:** Predicted interference effect ([+intr]-[-intr] conditions) for agreement and feature-based reflexive models ( $n=324$ ). Error bars indicate 95% confidence intervals.

Thus main finding in Experiment 4 is that the susceptibility of reflexives to interference is predicted to be significantly less than that for agreement simply due to baseline activation differences between the two dependencies. The local subject NP is reactivated at the verb immediately prior to the reflexive, and this provides an activation boost that lowers error rates and minimizes the interference effect. This generalization held true of almost every single parameterization tested.



**Figure 3.6:** Difference in predicted interference effect (agreement-reflexive conditions) for all models ( $n=324$ ). Error bars indicate 95% confidence intervals.

The second important finding from Experiment 4 is that there are no clear model predictions for grammatical sentences. This is an important result that I will return to in the general discussion below, but the intuition for this fact is straightforward. It stems from the fact that there are opposing interference effects at work in grammatical sentences: inhibition from similarity-based interference and facilitation from retrieval error. These two effects both exert an effect on grammatical sentences, when the interfering feature (an embedded [+singular] distractor NP when the probe searches for [+singular]) is both more similar to other elements in the parse (i.e. the target subject noun) and causes increased retrieval error. In ungrammatical

sentences, on the other hand, these two effects are dissociated: greater retrieval error occurs when the two nouns are more dissimilar. Understanding the source of the equivocal predictions for grammatical sentences is important in interpreting previous literature, and I return to a fuller discussion of this after Experiment 5.

The main finding of Experiment 4—reduced interference for reflexives even if they are processed in the same manner as agreement—corroborates the concern about Experiments 1-3 raised above. There is a smaller predicted interference effect for reflexives compared to agreement. This raises the possibility that the observed difference between the interference profile for agreement and reflexives reflects not a qualitative difference in memory access strategy, but rather an underlying quantitative difference stemming from baseline differences between the two. Experiment 5 directly compares the predictions of the two access strategies against the empirical data to ask which access strategy for reflexives is a better fit to the observed empirical data from Experiments 1-3.

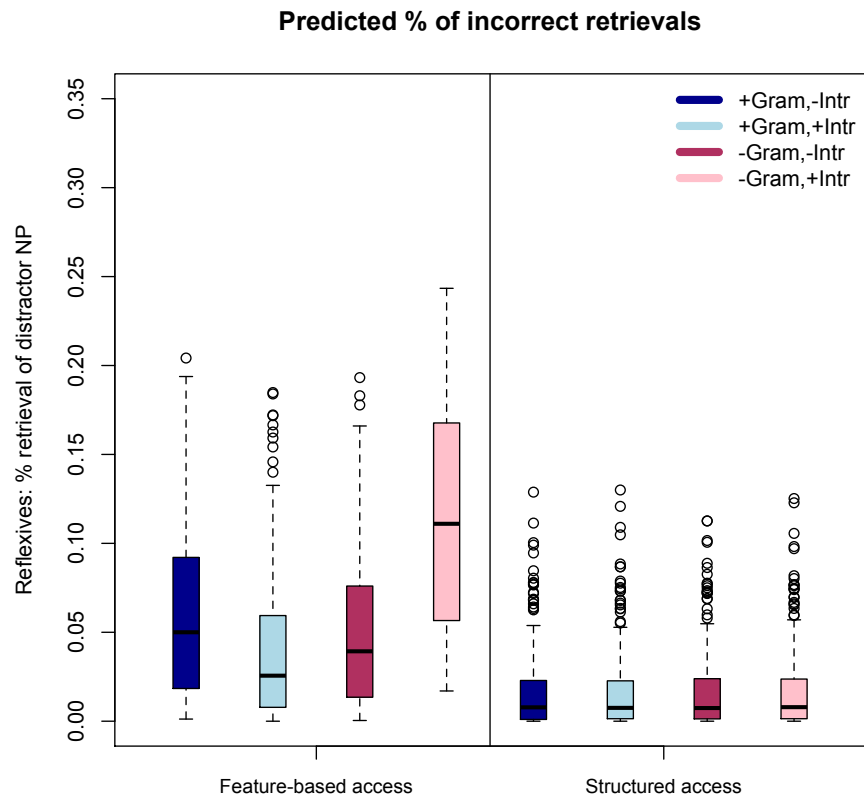
#### *Experiment 5: comparison of access strategies for reflexives*

In order to assess whether a feature-based or a structured access account for reflexives better captures the results of Experiments 1-3, Experiment 5 directly compared the predictions of the two modes of access against the observed empirical data. Feature-based models reflexive dependencies retrieve the subject using a mixture of structural (i.e., local subject) and morphological cues (number for *themselves*, number and gender for *himself/herself*). Structured access models retrieve

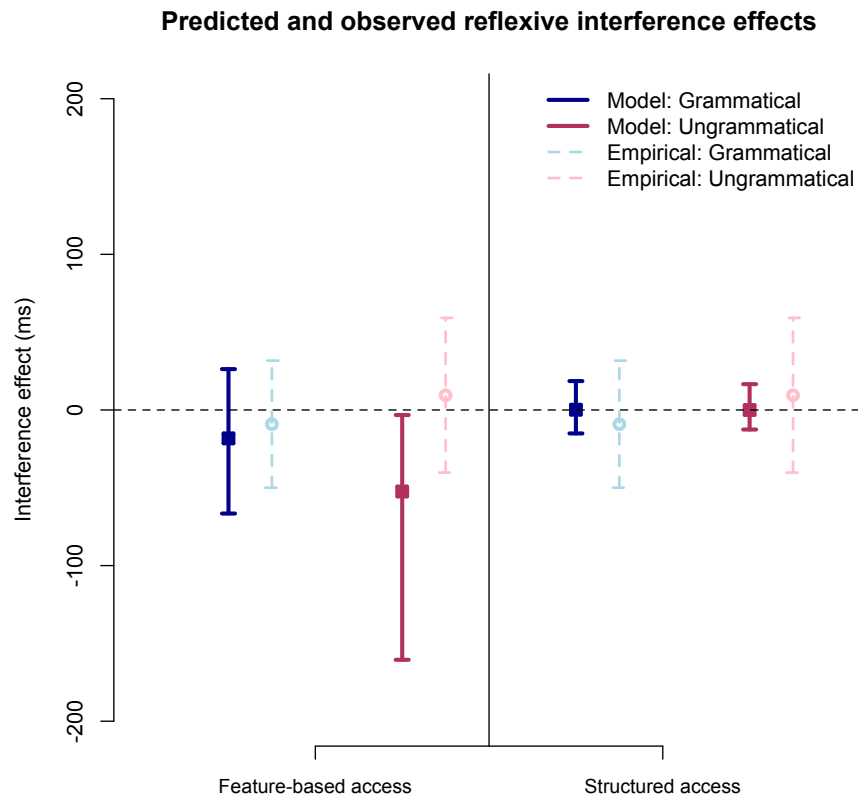
the reflexive's antecedent based only on structural cues. In the context of the current model, this effectively implements a targeted search for the local subject. It was seen in Experiment 4 that the predicted interference effect for feature-based reflexives was smaller than that for agreement, but there was still a clear prediction for a facilitatory interference effect. As no interference effect was observed in the experiments in Chapter 2, an important question is whether or not this finding is more likely to have been generated by a feature-based or structured access model for reflexives.

The empirical data of interest are the interference effects for reflexives observed in Experiments 1 and 3, as reflected in total reading times at the critical region. Total times were chosen because they provided the largest interference effect for agreement, and so provided the best opportunity for an interference effect to be observed for reflexives, if there was indeed any interference. For each participant ( $n=40$  from Experiment 1,  $n=32$  from Experiment 3), the average total reading time in each condition was obtained, and an interference effect score was derived.

The predicted error rates for both reflexive models are shown in Figure 3.7. It can be seen that as in Experiment 4, feature-based models of reflexive antecedent access are prone to greater error in the presence of feature-matched, but inaccessible constituents. This is not true of structured access models, which show similar rates of error across all four conditions. This is due to the fact that in structured access models, the feature manipulation across conditions is uncorrelated with the structural search cues; thus, error rate and feature manipulation do not covary. The result is that there is no differential erroneous retrieval across conditions.



**Figure 3.7:** Percentage of incorrect retrievals for reflexive conditions for feature-based and structured access models ( $n = 324$ ).



**Figure 3.8:** Comparison of predicted interference effects (solid) for reflexive conditions ( $n = 324$ ) and observed reflexive interference effects from Experiments 1 and 3 (by participants,  $n = 72$ ). Error bars indicate 95% CI.

Figure 3.8 presents a comparison of the predicted interference effects and the observed interference effects in total reading times estimated from a joint consideration of Experiments 1 and 3. As seen in the predicted error rates, the lack of correlation between the search cues and experimental manipulation in structural access models leads to no consistent differences in retrieval time across conditions. For this reason, the predicted interference effects are tightly centered around zero milliseconds. Feature based models, on the other hand, predict consistent direct interference effects (processing facilitation) for ungrammatical conditions. Again, as

in Experiment 4, there is no clear prediction for the direction of the interference effect in grammatical sentences, due to the tradeoff between inhibitory and facilitatory interference effects. Instead, the average prediction is for no interference effect.

The distribution of the empirical interference effect in Figure 3.8 appears to better accord with the predictions of the structured access model. In order to assess which model better predicts the observed data, Bayes Factor model comparison (Gallistel 2009) can be used to determine which of the two models, feature-based or structured access, better predicts the observed data. One way of measuring this is with likelihood of any one model given the observed empirical evidence, which can be derived from Bayes' rule as in (3.11):

$$(3.11) \quad L(M|D) = \int_{\theta} L(\theta|D)p(\theta|M) d\theta$$

In (3.11), the predictions of the model serve as the prior  $p(\theta|M)$ , which is combined with the likelihood of the size of the interference effect given the data. The greater the degree to which the model predictions and empirical data converge, the greater the likelihood of the model that generated those predictions. By taking the ratio of the likelihood of the structured model given the data over the feature-based access model, we derive the odds in favor of one model over another.

Results indicate that the data is better fit by the predictions of the structured access models. For ungrammatical sentences, the odds are 3.2:1 in favor of structured access models over feature-based access models. This odds ratio can be interpreted in an intuitive manner, as the odds that the structured model is correct over the feature-



based model. Generally, values of greater than 3:1 are regarded as ‘substantial’ evidence (Jeffreys 1961) in favor of a hypothesis. For grammatical sentences as well, there was some evidence in favor of structured access model, with the odds favoring them over feature-based access by 2.2:1. Although this value is generally considered only weak or ‘anecdotal’ evidence in favor of a hypothesis, it is notable that the tight predictions of the structured access model of reflexives better fits the observed data for both comparisons.

The results of Experiment 5 clearly support the hypothesis of structured access in comprehension. Crucially, for ungrammatical sentences where feature-based and structured access predictions differed, an evaluation of the observed interference effects showed that there is substantial evidence in favor of the structured access model over the feature-based access model. There was additionally weak evidence in favor of structured access models for grammatical sentences. Although feature-based and structured access models did not make clearly different predictions for the size of the interference effects, the predictions of structured access models were more narrowly centered around zero, which gave it a slightly better fit to the observed empirical data. This analysis gives important support to the hypothesis of structured access for reflexives. In particular, the modeling results help to alleviate one concern over the null-effects logic of the reflexive manipulations of Experiments 1 and 3. In matching specific model predictions to empirical data, Experiment 5 formulated the ‘null’ as a specific range of predicted values derived from the structured access model, and provided positive evidence that the predictions of the structured access

model provide a better description of the empirical data than the alternative hypothesis of feature-based access.

### *Interim conclusions*

The models constructed in Experiments 4 and 5 clarify the conclusions that are licensed from Experiments 1-3. Experiment 4 demonstrated that reflexives are predicted to show less interference than agreement simply due to their linear position after the matrix verb. This position confers an activation boost to the local subject, and reduces the predicted interference effects observed at the immediate post-verbal position. Error rates and interference effects for feature-based models of reflexives were consistently smaller than those predicted for agreement in similar models. This led to the possibility that the different interference profiles in Experiments 1-3 were due not to qualitatively different access strategies, but rather simply a reflection of an underlying quantitative difference in the size of the interference effect due to baseline activation differences in the local subject NP between agreement and reflexive dependencies at the point of retrieval. Experiment 5 showed that this alternative hypothesis did not adequately capture the empirical data, and provided further support for structured access models of reflexives.

Together, the modeling experiments evaluate and reject an alternative explanation of the data presented in Chapter 2. The empirical observation that reflexives appear to access structurally commanding NPs preferentially is in fact indicative of an underlying structured access strategy. This confirms the first

prediction of a structured access model of memory access for reflexives: the memory access procedure does not retrieve feature-matched, but structurally inaccessible NPs. Instead, the main determinant of whether or not an antecedent is considered is its structural position, as expected if reflexives engage an access mechanism that effectively targets the local subject.

### *The predictions of rational memory access models*

The clearest result of the models presented here is the facilitation effect observed for interference in ungrammatical sentences. In this section I examine the source of this effect and show that an analysis of the model's behavior supports the claim that facilitatory interference is the evidence that is necessary to show that structurally inaccessible positions are retrieved during memory access. An additional interesting finding from Experiments 4 and 5 is that in grammatical sentences, the model provides no consistent prediction for the direction of the interference effect. An examination of these two results clarifies the architectural conclusions are licensed when interference effects are observed in online processing.

There is a fairly straightforward characterization of the reason that the model predicts no interference for grammatical sentences. To see this, consider the summary of the critical conditions that is presented in Table 3.1. Here it can be seen that the experimental design manipulates two dimensions along which interference has been claimed to occur: similarity between the target and distractor NPs in memory (Gordon et al 2001; Gordon, Hendrick & Johnson 2004) and the match between the retrieval

probe and the distractor NP (Van Dyke & McElree 2006; Van Dyke 2007). Similarity between NPs in memory occurs when the two nouns in the sentence share the same value for a given feature; within the current experimental manipulation, this occurs in conditions where both target and distractor NPs are singular. On the other hand, the conditions also vary with respect to whether or not the distractor NP matches the features required by the agreeing verb (the probe-distractor match). These are different for the grammatical and ungrammatical sentences, reflecting the fact that the probing verb requires [+singular] in the grammatical case, and [+plural] in the ungrammatical case.

	<b>Target-Distractor match</b>	<b>Distractor-Probe match</b>
<b>The executive</b> <sub>+SG</sub> that oversaw <b>the manager</b> <sub>+SG</sub> definitely <b>was</b> <sub>+SG</sub> ..	✓	✓
<b>The executive</b> <sub>+SG</sub> that oversaw <b>the managers</b> <sub>+PL</sub> definitely <b>was</b> <sub>+SG</sub> ..	✗	✗
<b>The executive</b> <sub>+SG</sub> that oversaw <b>the manager</b> <sub>+SG</sub> definitely <b>were</b> <sub>+PL</sub> ..	✓	✗
<b>The executive</b> <sub>+SG</sub> that oversaw <b>the managers</b> <sub>+PL</sub> definitely <b>were</b> <sub>+PL</sub> ..	✗	✓

**Table 3.1:** Distribution of NP feature match and match of the inaccessible NP to retrieval cues across experimental conditions.

Viewed in this way, the reason for the no interference prediction for the grammatical conditions is fairly straightforward. In short, interference due to the target-distractor NP match has an inhibitory effect on retrieval times, and interference due to probe-distractor match has a facilitatory effect. The pair of grammatical conditions covary on these two distinct dimensions of interference, and their opposing effects cancel each other out in aggregate. These two types of interference are not simultaneously present for any of the ungrammatical sentences: NP match and inaccessible match interference occur in different conditions. When deriving

predictions for interference in grammatical conditions, the relative size of the inhibition and facilitation effects are controlled by separate model parameters. Thus, the relative weighting of parameter over the other determines the model's predictions for any one parameterization, but on average, the prediction is for no difference between these conditions. In ungrammatical sentences, on the other hand, the configuration that has an NP match is not the same as the configuration that causes increased retrieval error. Importantly, the underlying source of inhibition for matching features across the two NPs is distinct from the source of facilitation due to the inaccessible match. Of the two sources of interference, inhibition and facilitation, it is only facilitation that provides evidence that the incorrect position was retrieved.

The source of the inhibition effect due to the NP match is straightforward, and it may be understood in a number of ways. In the context of the ACT-R model used here, it is one instance of the well-known fan effect from general working memory tasks (Anderson 1974; Anderson & Reder 1999). The fan effect refers to the fact that the more memories a retrieval cue is associated with, the longer retrieval takes. Above it was noted that this follows directly from a rational model of memory access. When there are multiple items with the same feature specification, that cue is overall less diagnostic. This in turn leads to a less robust retrieval process, causing slower access time. On this model, the inhibition comes from using a retrieval cue that is less effective, slowing retrieval times.

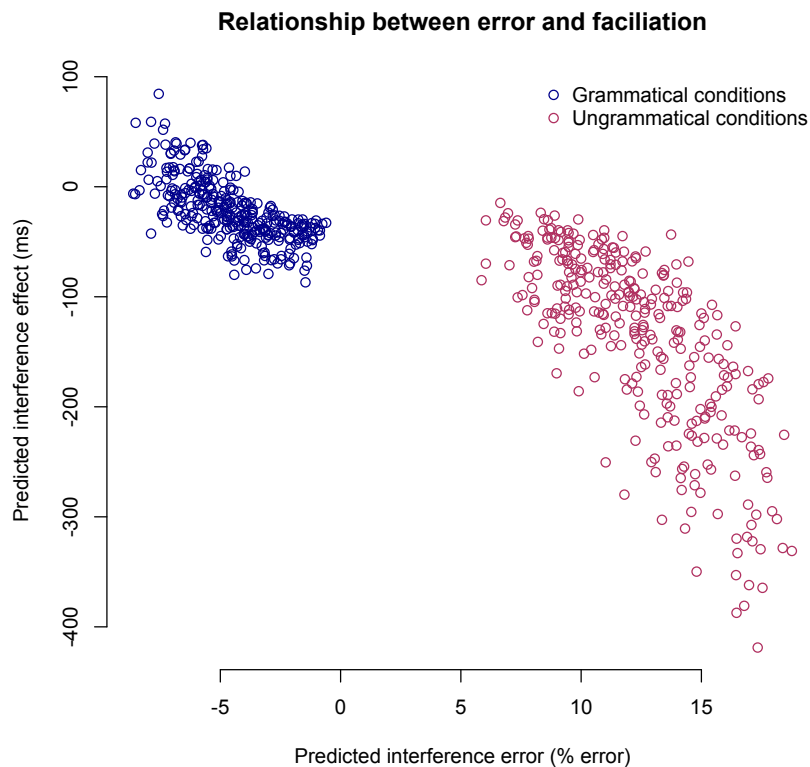
However, other models of this inhibitory effect do not require that the inhibiting feature actually be used as a retrieval cue. These models predict similar effects whenever two memories share feature content (Nairne 1988, 1990; Oberauer

& Kliegl 2006). In these models, feature similarity is the primary cause of forgetting. One way to understand this is through feature-overwriting (Nairne 1988, 1990). Feature-overwriting can occur whenever two memories overlap in a given feature value, leading to that feature getting ‘erased’ from the representation. This leads to less robust representations for the memories involved, and slows subsequent recall (Nairne 1988). These models accord well with interference accounts of information loss in short-term memory, which are considered to capture most known facts about short-term forgetting (Ricker, AuBouchon & Cowan 2010).

In addition to the large body of work that suggests that such feature-overwriting occurs in general working memory, there are a number of results in sentence processing that appear to require us to posit such a process. In particular, the results of a number of studies by Peter Gordon and colleagues (Gordon et al 2001, 2004, 2006) suggest that it is NP similarity, rather than probe-to-item similarity, that causes inhibition in sentence processing. Gordon and colleagues observed inhibition when two nouns in a sentence were similar along dimensions that are unlikely to be used as retrieval cues, such as whether or not both NPs are proper names. Because of this, inhibitory effects due to NP similarity in memory do not straightforwardly license conclusions about the features used in retrieval, and more importantly for the present purposes, they do not entail that the interfering memory was actually retrieved and considered. The multiple-match effect arises in a more indirect fashion.

However, facilitation due to probe-item match does license the conclusion that illicit material is retrieved online. The assertion that retrieval error leads to faster retrieval times is not as immediately intuitive as the similarity-based inhibition effects

that obtain in NP match environments. The facilitation that stems from retrieval error has been described as an illusion of grammaticity effect (Wagers 2008; Wagers et al 2009; Phillips et al 2010), reflecting the intuition that when retrieval error happens in ungrammatical sentences, the sentence is perceived as grammatical, and the average reading time is lowered due to those trials in which no error signal was registered. However, the source of this effect is more general than this: retrieval error leads to faster retrieval latencies in grammatical and ungrammatical sentences alike, as shown in Figure 3.9.



**Figure 3.9:** Relationship between interference error and interference effect on average retrieval latency. Blue points indicate comparisons between grammatical conditions, and red points are ungrammatical comparisons, for each of 324 parameterizations.

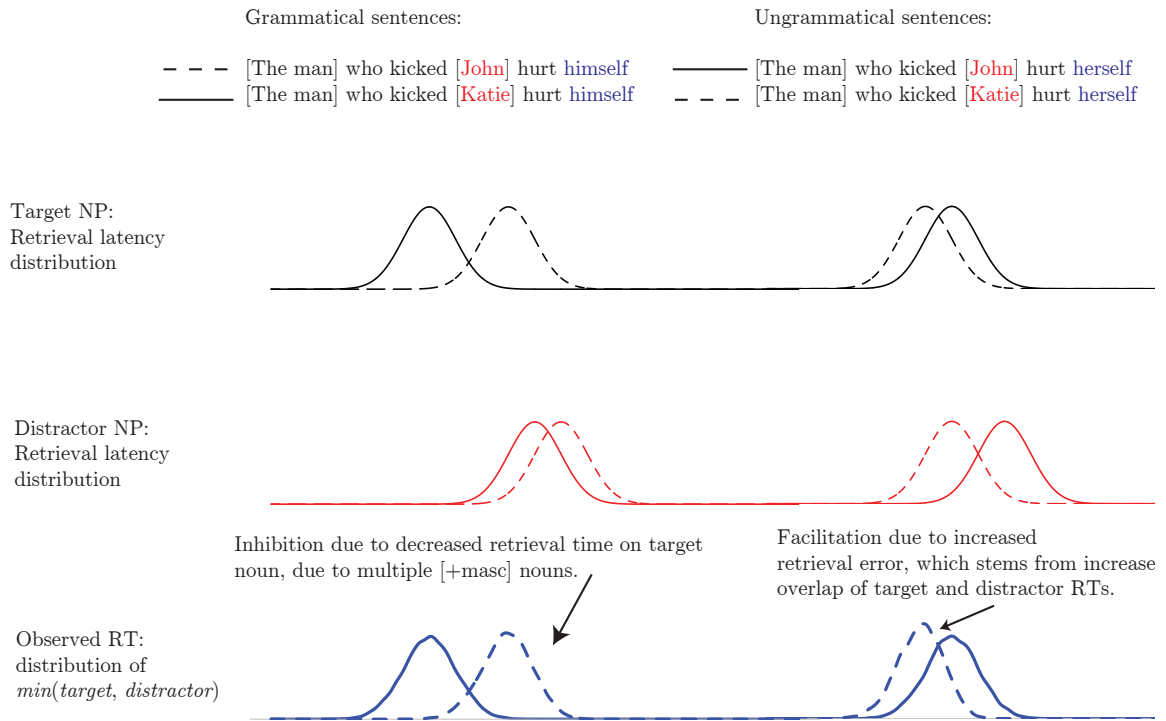
The race model aspect of the retrieval process drives this relationship between retrieval error and facilitation. The crucial property is the optimal decision rule assumed by the model: the memory that has the highest activation in light of its past history and a set of retrieval cues is the memory that is retrieved. Given the relationship between activation and retrieval latency, this means that the fastest memory to respond to the search cues is always the winner, whether or not it is the ‘target’ memory. This makes the retrieval process in ACT-R in important ways similar to race models of sentence processing (Traxler, Pickering & Clifton 1998; Van Gompel, Pickering & Traxler 2001; see also Frazier 1979), which maintain that the parser adopts the fastest syntactic analysis to become available in cases of ambiguity. If the parser incorrectly retrieves ungrammatical material in cases such as *\*the key to the cabinets were...*, then it effectively creates a syntactic ambiguity at the point of retrieval, even if the resulting structure is not well-formed. This spurious ambiguity in the retrieval process leads to processing facilitation in the exact same way that ambiguity speeds processing in race models: there are two possible retrievals, and the faster one is adopted. This has a facilitatory effect because the minimum of two random variables is never greater than the minimum of either random variable on its own:  $\min(x,y)$  is always less than or equal to  $\min(x)$  or  $\min(y)$ . If incorrect retrievals occur, then the distractor and target NPs have comparable retrieval latency; in the aggregate, choosing the quicker option will skew the reaction distribution negatively (see Figure 3.10).

‘Incorrect retrieval’ thus simply means that quicker relation between the probe and the distractor can be built more rapidly than between the probe and the target; this



provides the processing facilitation in the context of an illicit retrieval. On the assumption that our processing measures are monotonically related to the latency of retrieval, then due to the race aspect of the model, retrieval errors should facilitate reading times on average. On the further assumption that retrieval drives structure generation, this claim can be seen to be equivalent to previous claims of ‘race’ type phenomena in sentence comprehension (Traxler et al 1999; Van Gompel et al 2001). For this reason, this observation is more general than the ACT-R implementation. There is a deeper way of understanding this feature, however: facilitatory interference follows from a rational model of memory access, as I will show below.

Figure 3.10 summarizes the multiple-match and partial-match interference scenarios. It can be seen that in the multiple match interference environments that occur in the grammatical sentences, the interfering case (*the man who kicked John hurt himself*) is slowed due to feature overlap between *man* and *John*. This results in an inhibitory interference effect in the observed RT distributions. Any facilitation due to retrieval error is obscured by the larger inhibitory effect of similarity. For this reason, it is not possible to conclude from a behavioral slowdown that incorrect structure has in fact been generated. In contrast, in ungrammatical sentences, there is no feature overlap in the interfering case (*\*the man who kicked Katie hurt herself*). The partial match causes the distribution of the retrieval times for the target and distractor to overlap to a greater degree, causing greater retrieval error. There is no corresponding inhibition due to feature overlap. Because of the race aspect of the model, this leads to greater retrieval errors and an overall facilitatory interference effect.



**Figure 3.10:** Effect of interference in multiple match and partial match comparisons.

The inhibition in multiple match interference is driven by decreased retrieval latencies on the target noun, due to feature overlap. The facilitation in partial match interference is an increased overlap in the distribution of target and distractor retrieval distributions. The race aspect of the retrieval process leads to an overall facilitation effect, which unambiguously indicates that incorrect access has occurred online.

As mentioned, the link between behavioral facilitation and incorrect structure generation is more general than the ACT-R model. The fact that retrieval error leads to facilitation can be shown to follow directly from the rational structure of the model. Since this holds of the description of the system at an abstract, computational level, it holds across implementations of the rational access model. This provides another way of understanding race models of syntactic ambiguity (Traxler et al 1998; Van Gompel et al 2001, 2005), which might be said to operate in the same ‘rational’

fashion as memory access here: in all cases, the parser is doing what it thinks is correct in the fastest manner possible.

To see how this follows from a high level computational description of the retrieval process, suppose that we have two conditions, one where a distractor NP matches to some degree a search cue  $E$  and one where it does not match at all. Let these be  $m_{i+}$  and  $m_{i-}$ , respectively. We may express the advantage of the matching interfering memory over the non-matching interfering memory as the following odds ratio:

(3.12)

$$\frac{p(m_{i+}|E)}{p(m_{i-}|E)} \propto \frac{p(E|m_{i+})p(m_{i+})}{p(E|m_{i-})p(m_{i-})}$$

On the assumption that prior terms reflect only the history of usage (Anderson & Milson 1989), and that usage is determined by grammatical role in the sentence rather than feature content, then this odds ratio can further be reduced to the following:

(3.13)

$$\frac{p(m_{i+}|E)}{p(m_{i-}|E)} \propto \frac{p(E|m_{i+})}{p(E|m_{i-})}$$

From here, an advantage for  $m_{i+}$  is evident across a wide variety of assumptions about the likelihood term here: the probability of using some search term  $E$  for a memory should invariably be greater when the memory contains some of the

search cues ( $m_{i+}$ ) than when it does not ( $m_{i-}$ ). There is, of course, one exception: degenerate cases where no partial cue overlap is allowed, which would arise if the search cues had to perfectly match the target memory. In this case, as neither memory is a perfect match (mismatching on structural cues both), the likelihood of both interfering memories is zero. In this degenerate case, both likelihoods are equal, and the posterior should be equal across these (i.e. zero for both). Thus, in all situations the likelihood ratio of the two posterior probabilities for the two interfering memories is equal to or greater than one:

(3.14)

$$\frac{p(m_{i+}|E)}{p(m_{i-}|E)} \geq 1$$

In the context of a given set of retrieval cues, the probability of the interfering memory is always equal to or greater than that of the non-interfering memory, a fact that is mirrored in other models (see, e.g., Van Dyke & McElree 2006). Crucially, however, the increased probability of the interfering memory actually has a facilitatory effect given the current linking assumptions between access times and memory probability. Recall that the retrieval latency of the winner ( $m_{win}$ ) in a Bayesian optimal system always is the memory with the highest posterior probability:

(3.15)

$$p(m_{win}|E) = \max_m p(m|E)$$

It follows from the result above that the posterior of the winner in the context of an interfering element should always be equal to or greater than the posterior of the winner in the context of a non-interfering element:

(3.16)

$$\frac{\max_{m \in \{m_{i+}, \dots\}} p(m|E)}{\max_{m \in \{m_{i-}, \dots\}} p(m|E)} \geq 1$$

This last result guarantees that when errors occur, the average probability of the winning memory is greater than when they do not. Because this probability is inversely related to retrieval latency, then it must be the case that incorrect retrieval of inaccessible material must generate equal or eased retrieval times relative to non-interfering material.

The observation that ambiguity actually has a facilitatory effect in a rational model has been noted previously (Traxler et al 1998; Van Gompel et al 2001, 2005; Levy 2008). The result presented here shows that the incorrect retrievals in a content-addressable architecture should give rise to the same facilitation, for essentially the same reason: incorrect retrievals create a spurious ambiguity where there is none permitted by the grammar. Like cases of grammatically licensed ambiguity, the parser is considering which of two structures to build.

Thus interference effects in sentence comprehension fall into two classes, each with distinct mechanistic sources. Similarity among items in a sentence inhibits recalls of those items, whether it is due to retrieval processes (Lewis & Vasishth

2005) or feature-overwriting processes (Nairne 1988, 1990; Gordon et al 2001, 2004; Oberauer & Kliegl 2006). In contrast, retrieval error leads to processing facilitation. This distinction is important to consider, as it clarifies what data can be used as evidence for incorrect retrieval online. In previous literature both inhibitory (Van Dyke & McElree 2006) and facilitatory effects (e.g. Vasishth et al 2008; Wagers et al 2009) have been attributed to incorrect retrievals; however, the analysis presented here suggests that only facilitatory effects can be taken as evidence of incorrect retrieval online.

#### *Relation to previous work*

In light of the somewhat counterintuitive finding that incorrect access should always result in faster retrieval, in both ungrammatical and grammatical sentences, it is worth revisiting prior work on interference effects and the conclusions that have been drawn from this work.

There have been a range of effects that have been labeled as interference effects. For example, Gordon, Hendrick & Johnson (2001) used self-paced reading to examine sentences like the following, containing object relative clauses with either pronouns or common nouns as subjects:

- (3.17)           a. The banker that the barber praised climbed the mountain.  
                  b. The banker that Ben praised climbed the mountain.

At the point of processing the embedded verb *praised* and at the immediately following word, they observed processing difficulty reflected in longer reading times in (1a) relative to (1b). Gordon and colleagues interpreted this as an interference effect that obtains at the point of retrieving the subject, which is then subsequently attached to the verb. The feature overlap between the two subject NPs (i.e. they are both definite noun phrases) in (1a) makes it relatively difficult for comprehenders to recover the target NP, presumably due to the content-addressable property of memory access. They found a number of similar effects in Gordon et al (2002), as well as in Gordon, Hendrick & Johnson (2004, 2006). In these experiments, they found that the degree of feature overlap between the two NP positions correlated with observed processing difficulty. Thus, nouns that were more similar on some dimension (for example, both being proper names) inhibited each other at the point of retrieval. Not all dimensions of similarity caused such interference: definiteness and specificity did not (Gordon et al 2004). They interpreted this finding as an interference effect of the sort that is predicted by a content-addressable architecture.

Importantly, in this case similarity-based interference causes inhibition and slower reading times. On the predictions of the model discussed above, this is not due to incorrect access but rather due to interference degrading the memory representations of similar items. This is the account of these effects that Gordon and colleagues adopt. For these experiments in particular, it is difficult to imagine that the match between the retrieval cues and the nouns is the source of difficulty. For example, it is difficult to imagine that retrieval cues that select proper names are used to find the subject for the verbs in the experiment: it is unlikely that *praised* in (3.17)

retrieves its subject with a feature like [+proper name]. Rather than being due to a degraded cue (a fan effect), this inhibition reflects the match of the noun phrases in memory. This would appear to support feature-overwriting accounts of similarity-based interference (Oberauer & Kliegl 2006; Logacev & Vasishth to appear), and makes it more difficult to retrieve memories that have been degraded in this manner. These findings suggest that interference at the point of retrieval can arise because of mutual inhibition among elements in a parse, and so inhibitory interference in the context of multiple match interference cannot license conclusions about incorrect access, or the content of the retrieval cues.

However, Van Dyke and McElree (2006) challenged the characterization of these effects as the result of a disrupted encoding. Their primary question was whether or not they could find interference effects that stemmed entirely from the match of the retrieval cues to the items in memory. To do this, they compared two clefted sentences as in (3.18a-b). In (3.18a), the clefted noun *boat* is perhaps the only plausible object for *sailed*. In (3.18b), the verb *fixed* is instead plausible with a wider range of object nouns. Van Dyke and McElree reasoned that the semantic cues at *sailed* should resonate more strongly with *boat* than the cues at *fixed*. In this way their manipulation addressed the content of the retrieval cues, rather than feature matches amongst items in memory.

In addition, Van Dyke and McElree crossed the retrieval cue manipulation with a memory load manipulation. For memory load conditions, participants had to maintain a list of words in short-term memory while they read sentences. Importantly, the memory list contained plausible completions for *fixed*, but not *sailed*. After



memorizing the list and reading a sentence, a list recall task and a comprehension question were asked, in that order. Van Dyke and McElree asked whether or not the content of the remembered list would interfere with the retrieval of a clefted noun at the gap site. If the memory list contained semantically plausible candidates for the gap site, they reasoned, then interference effects should be observed at the gap site. In contrast, sentences where the gap site was not a plausible host for any of the nouns in the memorized list should be relatively easy to process at the retrieval site. The crucial comparison was between (3.18a) and (3.18b):

- (3.18)      a. List: *TABLE-SINK-TRUCK*  
                 It was the boat that the guy who lived by the sea sailed *e* in two days.  
                 b. List: *TABLE-SINK-TRUCK*  
                 It was the boat that the guy who lived by the sea fixed *e* in two days.

Upon reaching the critical verb *fixed* in (3.18b), reading times were longer than at *sailed* in (3.18a) by approximately 38ms, and longer than reading times for the same sentences when there was no associated memory load. Van Dyke and McElree's conclusion was that it was the fit of the retrieval cues of *fixed* to the content of the interfering list that caused the disruption, rather than the similarity in items in memory, as this was held constant across memory load conditions.

Van Dyke and McElree attributed this to the greater likelihood of recalling an incorrect item from the memory load list, due to its match with the retrieval cues. However, the result in the previous section calls this interpretation into doubt. In the absence of further assumptions about the relationship between retrieval error and processing difficulty, this is exactly the opposite pattern expected on the current

model. If the interfering memory list items are incorrectly retrieved, then facilitated processing should be observed. The inhibition that Van Dyke and McElree observe is instead consistent with a fan effect: the semantic cues to thematic integration at the verb are less distinctive for *fixed* than for *sailed* in the context of the memory load, which would lead to inhibition without incorrect retrieval of the nouns from the memory load list. An account of these results in terms of incorrect retrieval is possible: for instance, it may be that an incorrect retrieval of material that is not integrated into a linguistic parse leads to a sort of type mismatch that specifically inhibits further processing in the case of retrieval error. However, in the absence of these extra mechanisms, these results suggest similarity-based inhibition, rather than incorrect access as the source of difficulty. Unlike the results from Gordon and colleagues (2001, 2002, 2004, 2006), it seems that these results stem from a fan effect.

In another set of studies, Van Dyke and Lewis (2003) and Van Dyke (2006) looked at interference effects from additional syntactic material. They asked if a greater number of subject positions interfered with the ability of a verb to retrieve the target subject NP from memory. For example, Van Dyke and Lewis (2003) contrasted configurations such as the following (subject NPs are bolded to highlight the interfering positions):

- (3.19) a.     **The worker** was surprised that **the resident** who *t* was living  
                   near the dangerous warehouse was complaining about the  
                   investigation.  
           b.     **The worker** was surprised that **the resident** who *t* said that **the**  
                   **warehouse** was dangerous was complaining about the investigation.

In (3.19a), at the point of processing the critical verb phrase *was complaining*, Van Dyke and Lewis hypothesized that the subject noun phrase (i.e. *the resident who...*), would have to be retrieved from memory and that furthermore, this retrieval would be subject to interference effects. Thus, the presence of an additional subject position in (3.19b) (*the warehouse*), relative to (3.19a), should impact processing measures at the critical region. Their hypothesis was supported by the data: increased reading times were observed for (3.19b) relative to (3.19a). The locus of the interference effect was at the point of retrieval, rather than upon encountering the second interfering subject. They argued that the source of these effects was interference at a retrieval stage, rather than interference during encoding. Although the time course of the observed effect locates the source of the interference in retrieval, rather than encoding, these results do not distinguish between similarity-based inhibition due to feature-overwriting and interference due to less distinct retrieval cues. On the current assumptions, in neither case is the online processing profile consistent with incorrect access. In these studies, while multiple subject-like elements cause similarity-based inhibition, there is no direct evidence that incorrect elements are retrieved in early parsing.

In a later study, Van Dyke (2007) showed that semantic similarity further contributed to the interference effects observed here. She constructed additional conditions like those seen in (3.20), which replace *warehouse* in (3.19) with *neighbor*:

- (3.20) a.       **The worker** was surprised that **the resident** who *t* was living near the dangerous neighbor was complaining about the investigation.

- b. **The worker** was surprised that **the resident** who *t* said that **the neighbor** was dangerous was complaining about the investigation.

Neighbors, but not warehouses, match the selectional requirements (e.g. animacy) for the retrieving verb phrase *was complaining*. Van Dyke argued that if both semantic and syntactic cues were used to retrieve the subject verb phrase, then both manipulations should result in interference effects. Increasing interference by increasing the number of subject NPs as well as increasing interference by increasing the number of NPs that match the critical verb's selectional restrictions both caused slowdowns in an eye-tracking task, across early and late measures. Again, this sort of processing slowdown does not implicate incorrect access online. It is compatible with a range of mechanisms that predict inhibited access to items in memory as a function of their similarity to other items in memory.

However, a cloze-format offline task probed participants' comprehension of the critical VP by asking them to decide which of the three NPs present in the sentence truthfully completed sentence fragments of the sort "\_\_\_ *was complaining about the investigation.*" Both semantic and syntactic sources of interference decreased participants' accuracy on this offline task; they were more likely to select the interfering noun in all interfering contexts in a way that suggested additive effects of semantic and syntactic interference. However, Wagers (2008) suggested that an independent source of complexity in the Van Dyke & Lewis (2003) materials complicates their interpretation; namely, that at the point of retrieval, syntactic interfering conditions had two clause boundaries where non-interfering conditions had only one. After controlling for this by contrasting lexical versus expletive

subjects, he replicated their offline interpretation findings, but did not replicate their finding of increased reading times in response to an additional syntactic subject position.

One possible objection to the reasoning advanced here is that it is possible that incorrect retrievals selectively inhibit processing, and so results like those in Van Dyke's study do in fact indicate incorrect access online. It is unclear what these processes might be, however. It seems unlikely that they are due to reinterpretation or ungrammaticality detection, as the offline measures suggest that neither reanalysis nor detection of ungrammaticality occurs in these situations (Sturt 2003a; Van Dyke 2007; Wagers 2008; Wagers et al 2009; Phillips et al 2010). Since no selective reanalysis or ungrammaticality detection appears to be engaged in case of an incorrect retrieval, it may be that the parser does not make a distinction between retrieving a target or distractor memory, treating both situations alike. If this argument is correct, then there seems to be no basis for concluding that more difficult processing occurs when a non-target memory is retrieved online.

Interference effects have been obtained in a number of other studies. For example, Fedorenko, Babyonyshev and Gibson (2004) found processing slowdowns at the verb in Russian when there were multiple nouns that share identical case, although interestingly these effects only obtained when both abstract case and phonological realization were identical across the interfering elements. Suckow, Vasishth, & Lewis (2005) found that when there were multiple animate NPs in German sentences, inhibition was observed due to similarity-based interference. As with the results reviewed above, increasing either the similarity of the material to

other constituents in the sentence or to the retrieval cues causes inhibition. These results are all architecturally interesting, in that they clearly indicate that the memory architecture of the sentence allows items in memory to interfere with each other in some fashion. Again, however, they do not demonstrate the facilitatory interference profile that indicates incorrect retrieval, and so it is unclear what to infer about incorrect access online from results that suggest inhibited processing in the face of interference. All of these results are compatible with a parser that can index and correctly access the desired items in memory in accord with structured access, if one assumes that items in memory can directly interfere with each other through feature-overwriting or conflicting bindings (Logacev & Vasishth to appear).

In contrast to the many examples of inhibitory interference effects, there is a somewhat sparser body of work that demonstrates facilitatory interference effects. To date there has been no clear demonstration of facilitatory interference for reflexives, with studies either showing no detectable interference (the present work, along with Nicol 1988; Clifton et al 1999; Sturt 2003a; Xiang et al 2009) or inhibitory effects (Badecker & Straub 2002; Sturt 2003a). As noted in Chapter 2, the two dependencies that reliably exhibit facilitatory interference effects are subject-verb agreement and NPI processing (Clifton et al 1999; Pearlmutter et al 1999; Drenhaus et al 2005; Wagers 2008; Vasishth et al 2009; Xiang et al 2009). Across these studies, facilitation is seen in ungrammatical sentences that have inaccessible feature-matched elements. There is a wide range of results that have been attributed to interference effects, but the conclusion that the parser routinely constructs ungrammatical parses (see, e.g. Van Dyke 2007) is not obviously warranted. Only a narrow subset of dependencies—

agreement and NPI processing—appear to support this conclusion, and they may be the exception rather than the rule when considering parsing more generally, a topic that I return to in Chapter 5.

### *Conclusion*

The theoretical and computational analysis presented here strengthen the case for structured access in comprehension on a number of fronts. First, the explicit computational models of the experimental results on the processing of agreement and reflexives support the conclusion that reflexives access the subject using structured access mechanisms. The narrow predictions of the structured access model provide a superior fit to the experimental data than did a feature-based access model of reflexive antecedent access, making a stronger case for structured access in comprehension.

Second, an analysis of the predictions of a rational access memory showed that facilitatory, rather than inhibitory interference is the behavioral signature of incorrect access. It has long been noted that in certain situations, structural ambiguity causes facilitated processing (Van Gompel et al 2001, 2005; Levy 2008). Facilitatory interference is a qualitatively similar phenomenon: if the parser incorrectly retrieves and attaches structurally licit and illicit material alike, it creates a situation of (temporary) spurious ambiguity. This apparent ambiguity leads to facilitated processing, as observed in cases of true syntactic ambiguity (Van Gompel et al 2001, 2005). However, this facilitatory interference is only observed for a limited range of

syntactic dependencies: the partial-matching interference effects observed in NPI and agreement dependencies are the only robust demonstrations of structurally incorrect access in online parsing. Instead, a large portion of work on interference effects in sentence processing focuses on inhibitory effects due to increased similarity, but these effects do not provide evidence that the parser has generated ungrammatical parses by retrieving inaccessible linguistic material.

These two findings build further support for structured access mechanisms in comprehension, by strengthening the conclusion of structured access reached in Chapter 2, and demonstrating that there are few counterexamples to structured access in the existing literature. The wider range of interference results are compatible with structured access in that they indicate similarity-based inhibition, rather than spurious retrieval of non-target memories, in online parsing processes.



## **Chapter 4: Processing long-distance reflexives in Mandarin Chinese**

The preceding chapters built the case for structured access by investigating reflexive dependencies in English. It was shown that the interference profile for reflexives is qualitatively different than that for agreement. The observed pattern of interference for agreement suggests that morphological features are used to retrieve memory representations in constructing English subject-verb agreement dependencies. English reflexives, on the other hand, were reliably immune to such interference, indicating that feature-based access is not used to retrieve the local subject to resolve a reflexive's reference. This conclusion was supported by computational evidence presented in Chapter 3, which showed that a structured access mechanism better predicted the observed set of experimental results than did a feature-based model.

This set of findings constitutes one type of argument for structured access: there was no behavioral evidence that comprehenders considered inaccessible

antecedents. In the current chapter, I provide a complementary argument for structured access. I demonstrate that comprehenders retrieve and consider feature-mismatched antecedents if they are structurally accessible, even if they are not the target of retrieval; in order to make this argument, I examine the Mandarin Chinese long-distance reflexive *ziji*. Long-distance reflexives are a good test case in investigating structured access because of a number of unique properties. Like English reflexives, they must be structurally bound, but unlike English *himself*, their antecedent can potentially be found in any clause that dominates *ziji* (Huang & Liu 2001; Huang, Li, & Li 2009). Here an investigation of the processing of *ziji* reveals that comprehenders check the local subject position before a more distant (but licit) subject position, regardless of the semantic features of the two subjects. The observed contrast between local and long-distance binding provides a useful diagnostic of the role of syntactic structure in memory access. Rather than investigating the impact of an illicit noun phrase on reflexive resolution, the studies reported here more directly test of the role of syntactic structure in finding an antecedent. Evidence that unacceptable antecedents are considered based on their syntactic position, even in the presence of a better candidate antecedent, would provide a strong argument for structured access.

A comparison of long-distance and local binding configurations for *ziji* provides evidence that this is indeed the case. Across the two experiments presented in this chapter, syntactic position appears to provide the primary means of accessing linguistic memory when comprehenders try to construct a *ziji*-antecedent dependency. In particular, comprehenders appear to preferentially access a feature-mismatching

local subject position before accessing a feature-matched, but more structurally distant NP. This is expected if uniquely syntactic information is used to access *ziji*'s antecedent and consideration of antecedents occurs serially.

In order to make this argument, I rely on time course evidence using the speed-accuracy tradeoff paradigm (SAT; Experiment 6), as well as event-related potentials evidence (ERP; Experiment 7). In previous chapters, the argument for structured access was built on previous arguments about memory access from the presence or absence of interference effects. In contrast, the argument in this chapter is centered on the argument from the time course of processing. I first present a brief overview of the argument from time course for content-addressable memory architectures, which introduces the SAT technique and the logic of investigating the time course of decision-making in processing. I then present SAT time course evidence that shows that comprehenders are slower to accept grammatical long-distance *ziji* interpretations than local interpretations of the anaphor. This is expected on a structured access account, where comprehenders consult the local subject position before more distant antecedent positions when processing *ziji*; this set of findings is not compatible with an account in which semantic feature cues are directly used in the search for an antecedent. The ERP evidence replicates the locality bias with an alternative methodology, and shows that canonical syntactic reanalysis is not engaged when shifting from local to long-distance bindings. Instead, the difficulty is indexed by an ERP component that has been associated with working memory difficulty: the left anterior negativity (LAN; Kluender & Kutas 1993).

### *Linking structured access and structured search*

Processing evidence on the time course of memory access has primarily been used to address questions of memory architecture, contrasting content-addressable versus serial-register memory systems. This architectural contrast does not directly address the structured access hypothesis, which instead makes a claim about the content of the information used to index and access memory during parsing. Nonetheless, a consideration of the argument from time course is useful in the present context because the structured access hypothesis makes a prediction for the time course of memory access. On the assumption that retrieval is serial (i.e. multiple memories cannot be retrieved and processed in parallel), then in situations where the cues used in retrieval underdetermine the target of retrieval (i.e. when there are numerous acceptable syntactic targets), then the need to consider multiple targets will result in slowed processing. This delayed processing is a consequence of pursuing a narrowly syntactic access strategy, because a more restricted access strategy has the possibility of creating a fleeting ambiguity that needs to be ruled out by checking multiple positions. Consider a sentence like *John gave Mary a book about himself*. If the arguments in the preceding chapters are correct, then *himself* retrieves antecedents using only structural information. Thus in trying to fix the reference of the anaphor, there is a fleeting ambiguity from the point of view of memory access: the parser could in principle access either of the local c-commanding NP positions as an antecedent. Although the formal constraint supplied by the gender features makes the sentence unambiguous, resolving this ambiguity may require checking both noun

positions to remove the unacceptable antecedent *Mary* from consideration. In the English studies in Chapter 2, I argued that the initial candidate set is smaller than predicted on a feature-based access account. In contrast, for sentences like *John gave Mary a book about himself*, the situation is reversed: structured access predicts that all structurally accessible (local, c-commanding) antecedents are in the initial candidate set (*John* and *Mary*), whereas featured access limits this set only to those that match in formal gender features (*John*).

In this way, structured access suggests the possibility of *structured search*: in situations where the retrieval cues underdetermine the target of retrieval, the parser might need to search through the syntactically licit positions, considering each in turn. This requires the assumption that is implicit in the memory models considered thus far: only one memory is retrieved for processing at a time (perhaps due to limitations on the scope of focal attention), forcing a sequential retrieval process. If this model is correct, then as the number of positions to be considered grows, so does average processing speed. On the further assumption that some positions are more likely to be considered prior to others, then there should be processing speed differences associated with accessing syntactic positions that are distant from the probe site compared to local syntactic positions.

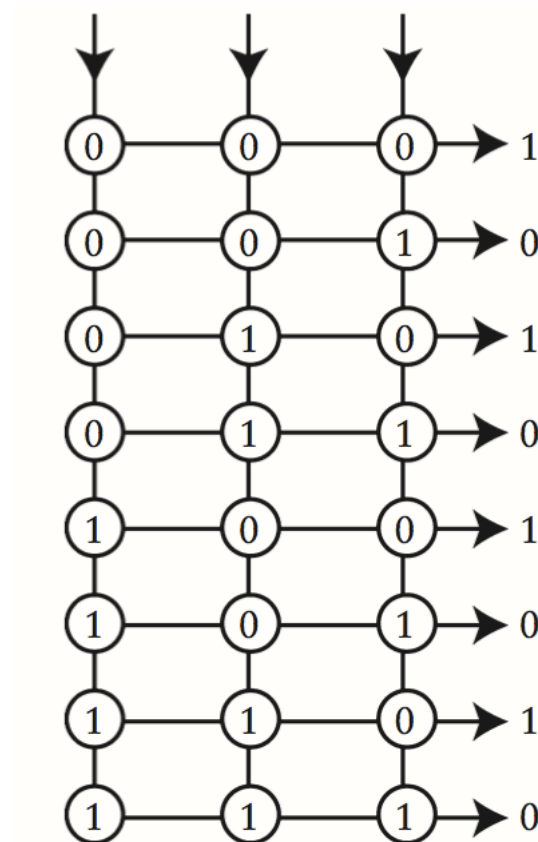
Thus, a structured search account maintains that the parser is not allowed to entirely skip positions that could in principle be relevant in memory access, given a set of retrieval cues. This claim differs in important ways from the claims supported by existing time course evidence, as I argue below. The bulk of time course evidence has been taken to show that memory retrievals in sentence comprehension have the

property of direct access: accessing a target memory occurs in roughly constant time with respect to the size of the search path, due to the fact that direct access allows the parser to ‘skip’ irrelevant memory positions in retrieving linguistic material. Because the focus has been on whether or not the parser considered irrelevant memories in memory access, this evidence does not directly speak to the prediction of a structured search account, which maintains that the parser cannot skip memory positions that could potentially be relevant but for some reason do not contain the target memory.

### *The argument from time course*

Arguments from the time course of memory access have been a strong source of support for content-addressable memory architectures (McElree 2000; McElree et al. 2003; Foraker & McElree 2007; Martin & McElree 2008, 2009). The logic of the argument from time course is as follows: if a serial search process is used in the access procedure for querying linguistic memory, then the time required to retrieve information should grow as a function of the size of the search path. If, on the other hand, memory access is mediated by a content-addressable memory architecture, then the size of the search path should not affect retrieval times. This is because of the direct access property of content-addressable memory architectures. Direct access refers to the fact that when memory is queried, the target memory directly resonates with the retrieval cues, without the need to consult memories that do not match the retrieval cues. Direct access may be thought of as a parallel search for content of all memories in a given store. The architecture of a content-addressable architecture is

schematized in Figure 4.1 (from Gallistel & King 2009). In this representation, the memories consist of three bits, and during a memory retrieval, each bit is probed with the retrieval cues. Those memories that match are returned. All memories are evaluated with respect to their fit to the retrieval cues, and then a decision about which memory is passed to further processing is made. The architecture requires that all memories be checked and for this reason the time to query memory is constant over the size of the search path.



**Figure 4.1:** Architecture of a content-addressable memory, from Gallistel & King (2009). Memories consist of three bits, and each bit is probed in parallel for a match. In the present case, all memories that contain a 0 in third position are returned in response to a retrieval query.

Importantly, the constant-time property of content-addressable architectures holds only over dimensions such as the size of the search space or ‘physical’ distance to the target memory. They do not generally have the constant-time access property on other dimensions, however. For example, in the content-addressable ACT-R model presented in Chapter 3, access time is constant across the number of memories, but varies as a function of how well the cues match the target memory. Many content-addressable systems have this property; consider another well-known implementation of a content-addressable memory system, the Hopfield Net (Hopfield, 1982). Hopfield nets are a type of neural network architecture that stores ‘memories’ in the connections between binary units in a fully-connected, recurrent architecture. Once a network has been trained on a series of memory images, querying the network with a noisy version of the memory causes the network to ‘retrieve’ the desired memory by settling into a stable state that corresponds to the learned pattern that most closely matches the input. Crucially, however, the time course of retrieval (i.e. the number of iterations needed to reach convergence) is a function of the degree to which the retrieval cue matches the target memory. The closer the match between the probe and the target memory, the fewer iterations are necessary for the network to settle into a stable memory state. Presumably, if this were a model of an actual memory retrieval process, a greater number of iterations would correspond to a greater retrieval latency on some behavioral measure. This provides an important qualification of the direct access property of content-addressable architectures: while location and the size of the search path do not affect retrieval times, the degree of match to a search cue and noise



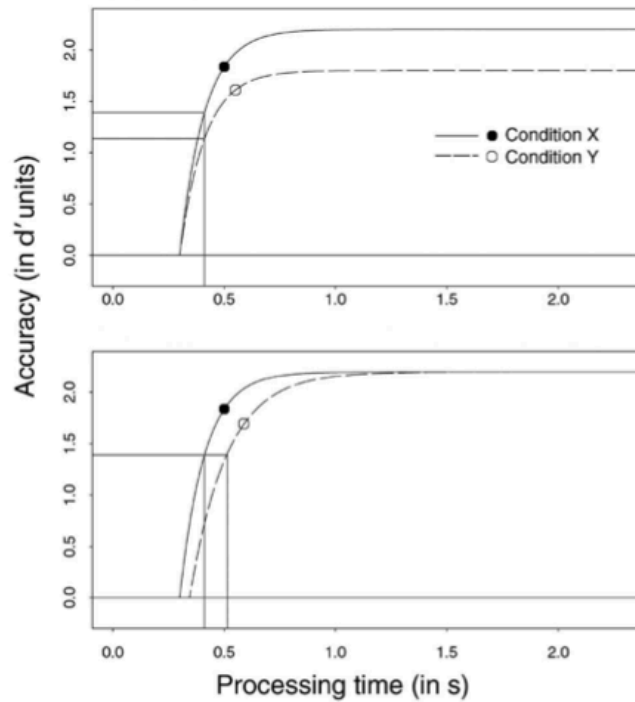
in the system can (and on many models do) affect retrieval times, as in Hopfield Nets or the ACT-R model (Anderson & Lebiere 1998; Lewis & Vasishth 2005).

It is also important to note that the property of constant-time access is not unique to content-addressable architectures. Standard random-access memories (RAM) used in digital computers are also constant-time architectures. Memories are retrieved by loading a query onto a series of bus-lines that are probed in parallel against the contents of all memory stores, looking for the location that matches the query and returning it in a time that is functionally constant across the size of the search path and location of the target memory. This stands in contrast to disk media such as CD-ROMs or hard-drives that need to be physically traversed to reach the desired address. As noted by Gallistel and King (2009), RAM architectures are in essence architecturally identical to content-addressable architectures, and can trivially implement a content-addressable look-up table if desired. This is because they share the time-course properties of content-addressable architectures and storage capabilities. The primary difference between content-addressable architectures and RAMs is what is allowed as a retrieval cue. In RAM architectures, retrieval “cues” correspond to the location of a given memory, and parallel search proceeds over all possible memory locations. The contents of any one location are not considered at access, however. Thus constant-time access is not a unique to content-addressable architectures. Rather than provide definitive evidence in favor of a content-addressable system, then, the argument from time-course provides evidence that for eliminating serial search architectures from consideration.

The first argument for content-addressable architectures of comprehension, the argument from interference, has been made with a number of experimental techniques (behavioral reading measures, electrophysiological measures, off-line and on-line patterns of judgments). In contrast, the argument from time-course has typically been made exclusively with the speed-accuracy tradeoff (SAT) paradigm (also known as the response-signal paradigm; see Wickelgren 1977). The SAT procedure probes participants for responses at cued intervals, producing a curve that describes subjects' accuracy as a function of time. As the decision processes tapped by the SAT task are by hypothesis parasitic on the retrieval of the information needed to support them, the time course of the decision is taken to be proportional to the time needed to access the information in memory.

The SAT task has been applied to language processing by Brian McElree and colleagues to ask about the role that structural or linear distance plays in the speed or the accuracy of linguistic processing. As mentioned, studying the speed of processing typically involves SAT measures, which license inferences about the time course of retrieval mechanisms employed in sentence processing by providing a direct measure of the time course of information accrual. Traditional RT paradigms are limited in how informative they are about the dynamics of memory processes. This is because participants can trade speed and accuracy in a task (Wickelgren, 1977), opting to spend more time processing for greater accuracy on a task. For this reason, estimating a single mean reaction time in a given experimental condition (or even a single RT/accuracy pair) can obscure differences between the success of a process, on the one hand, and true processing speed, on the other. By providing information about

task accuracy as a function of time, the resulting SAT functions allow researchers to build a picture of the dynamics of task completion that dissociates the speed of completing the task (measured in the *dynamics* parameters, see below) and the accuracy of task completion (referred to as the *asymptotic accuracy* parameter). For example, consider Figure 4.2 (from Öztekin & McElree 2010), which shows hypothetical SAT functions for two processes that differ in the speed of processing (4.2, bottom panel), as well as processes that differ only in their asymptotic accuracy (4.2, top panel). The circled points on each curve represent hypothetical reaction times that could have been sampled from these processes in a simple RT task. Note that the single RT/accuracy pairs do not reveal if the observed difference is due to an underlying difference in processing speed or processing accuracy. For this reason, time-course measures such as SAT are essential to making an argument about the time course of processing. SAT functions allow the experimenter to quantify both the speed and the probability of successfully completing a given process, and provide dissociable measurements of both.



**Figure 4.2:** Hypothetical SAT curves showing a) two processes that differ in asymptotic accuracy only (top panel) and b) two processes that differ in processing speed only (bottom panel) (figure from Öztekin & McElree 2010). Vertical and horizontal lines indicate that point at which each curve is at 50% of asymptotic accuracy.

The relatively direct measure of processing speed that is measured with SAT curves can be used to more straightforwardly test predictions about the time course of processing. In the domain of memory access, these predictions are clear. The prediction for a parallel-access retrieval mechanism is that all relevant representations should be accessed with similar temporal dynamics, as all memories are probed in the same processing step. That single processing step is constant across the location of the target in the memory store, and so should result in constant access speeds across the growth part of the curve. Memory access that has any serial component, by contrast, opens up the possibility that some representations are contacted before others. These extra processing steps should be reflected in differences in temporal

dynamics of the SAT function. For example, in a fully serial search, these changes would be observed in response to changes in the structure of the search space (e.g., more intervening material or greater hierarchical distance to the retrieval target). In much work that has made architectural arguments from the time course of memory access, the primary contrast of interest is between content-addressable architectures and serial-register architectures, as the different commitments that each architecture makes to the time course of memory access are clear.

Some of the earliest applications of this argument focused on unstructured lists in memory (McElree & Doshier 1989). McElree and Doshier (1989) found that the time course of recognition for words in an unstructured list was constant across serial position, with the exception of the most recent item, which was assumed to be in focal attention. They concluded from this finding that retrieval of items in the list is mediated by a content-addressable, direct-access mechanism. In contrast, the speed of a search process should be modulated by factors like linear position or size of the search cohort (Sternberg 1969). This architectural conclusion is licensed in part because of the unstructured nature of the word list. That is, in this context, it seems that the identity or content of the probe word is the only information that is available to be used in accessing the memorized list at the point where participants are required to make a recognition judgment. Thus is linear position or size of the list do not impact speed, the word identity itself—its content—must be guiding access. However, an important difference between unstructured memory lists and sentences concerns the availability of structure to guide memory access. As argued above, the nature of the cues used to access memory control how retrieval events proceed in

sentence comprehension. The rich linguistic structure in the form of possible morphological, syntactic and semantic cues makes it more difficult to directly draw architectural conclusions from this argument in the domain of syntactic processing. For this reason, although the debate often contrasts content-addressable and search-based memory architectures, in the domain of sentence comprehension it is difficult to address this question without also considering the nature of the information used to index and access memory.

In order to start addressing the question of memory architecture in the domain of sentence processing, McElree and colleagues have investigated the time course of dependency completion for a number of different linguistic dependencies. These include filler-gap dependencies (McElree, 2000; McElree et al., 2003), subject-verb thematic dependencies (McElree et al., 2003), pronoun antecedent resolution (Foraker & McElree, 2007), and verb phrase ellipsis (Martin & McElree, 2008, 2009). By and large, the results from these studies point in the same direction. They have generally indicated that the retrieval of information during sentence processing occurs at a constant rate regardless of the distance from the probe point to the target, suggesting a general content-addressable (or parallel-access) architecture during sentence processing.

One influential example of the argument of time course was presented by McElree, Foraker and Dyer (2003). They constructed experimental sentences as in (4.1). The topicalized NP *the scandal* needs to be interpreted as the object of the most deeply embedded verb in all examples. McElree and colleagues interpolated varying amounts of clausal material between the fronted NP and the gap position, ranging

from no intervening material in (4.1a) to two intervening clauses in (4.1c). In doing so they increased both the linear and the hierarchical distance to the gap site. They reasoned that if the difficulty associated with longer filler-gap distances was due to an increase in search path in conjunction with a serial access memory architecture, then the difficulty associated with longer filler-gap distances should affect the speed (*dynamics*) portion of the resulting SAT functions. On the other hand, a content-addressable architecture should show no effects of distance on retrieval dynamics. Instead, only the terminal accuracy (*asymptote*; probability of correct retrieval) should be affected.

- (4.1) a. It was **the scandal** that the celebrity relished *e*.  
 b. It was **the scandal** that the model believed that the celebrity relished *e*.  
 c. It was **the scandal** that the model believed that the journalist reported that the celebrity relished *e*.

McElree and colleagues found that access time (as indexed by the SAT curve's dynamics parameters) for the filler was constant across all three examples in (4). The distance between the filler and gap was only evident in the pattern of terminal accuracies: the longer the distance, the less likely participants were to retrieve the correct filler to interpret at the gap site. This finding was compatible with a content-addressable memory access mechanism, in that it displayed constant-time access and loss of memory fidelity with increasing temporal decay or interference. McElree and colleagues found similar results across a number of other constructions including subject-verb integration (*the book ripped* versus *the book that the editor admired ripped*) and double gap topicalization constructions (*this is the album that*

*the stamps were difficult to mount e in e versus this is the album that the stamps which obviously angered the fussy collector were difficult to mount e in e).*

One possible objection to these results, noted by McElree et al (2003), Martin & McElree (2008) and Wagers (2008) is that these results rest on data from the processing of dependencies that are often thought to involve predictive resolution strategies. The most well-known of these hypotheses is the ‘active-filler’ strategy for processing *wh*-movement dependencies, which posits that the parser actively tries to find a gap for a filler held in memory (Wanner & Maratsos 1978; Frazier 1987; Frazier & Flores d’Arcais 1989). If this is the case, then the interpretation of these data is unclear. If the filler is held in a special store or engages special parsing routines (as in Wanner & Maratsos 1978), then it is not clear that the filler actually is more ‘distant’ in (4.1c) than (4.1a), making it difficult to argue that the constant access dynamics implicate a content-addressable architecture. Furthermore, there are independently motivated reasons for thinking that the intermediate gap sites (the [spec,CP] positions along the path between *wh*-element and its gap site) might contain some representation of the moved constituent, with support for this coming both from theories of grammatical organization (Chomsky 1973, 1977; McCloskey 2001) and processing considerations (Frazier & Clifton 1989; Gibson & Warren 2004). If this line of reasoning is correct, then it is not clear that there is any difference in the distance between the filler and gap in the examples in (4.1).

To address this concern and provide support for content-addressable architectures, it is necessary to investigate dependencies that are not amenable to anticipatory or predictive processing strategies. In order to make a stronger case for



content-addressable architectures in comprehension, Martin and McElree (2008, 2009) examined the processing of verb phrase ellipsis (VPE) constructions like those in (4.2).

- (4.2) a. The editor [<sub>VP</sub> admired the author's writing], but the critics did not *e*.  
b. The editor [<sub>VP</sub> admired the author's writing], but everyone at the publishing house was surprised to hear that the critics did not *e*.

Unlike the filler-gap dependencies tested in earlier SAT studies (McElree 2000; McElree et al 2003), these VPE examples do not generally allow comprehenders to adopt a predictive strategy for resolving the antecedent of the ellipsis: upon completing a VP, it is not explicitly marked as participating in an upcoming ellipsis dependency. Despite the entirely retrospective nature of the VPE dependency, Martin and McElree again found that the antecedent VP was retrieved in constant time, no matter how distant it was from the ellipsis site. Thus the findings of these studies confirm and extend those of prior studies: extra hierarchically intervening material between the antecedent and the ellipsis site did not produce any differences in the dynamics of memory retrieval. Instead of impacting processing speed, distance between the ellipsis site and the antecedent VP instead led to a lower probability of retrieval, as reflected in lower accuracy (a lower asymptotic portion of the curve).

Similar findings have been found for filler-gap dependencies (McElree & Griffith 1998; McElree 2000; McElree et al 2003), VPE (Martin & McElree 2008, 2009), and coreferential antecedent-pronoun relations (Foraker & McElree 2007),

among others. The repeated failure to find effects of the size of the search space in the SAT paradigm constitutes the argument from time-course against serial search memory architectures. For the range of dependencies considered, McElree and colleagues conclude that a content-addressable architecture best captures the observed time course evidence. This conclusion aligns nicely with the argument from interference, which provides a qualitatively different source of evidence that points in the same direction: during sentence comprehension, access to memory (the contents of the parse) is gated by a procedure that is constant-time (with respect to search path) and prone to interference, two defining features of content-addressable memory systems (see also Lewis et al 2006).

#### *Chinese long-distance anaphors*

On the current evidence, it has been argued that long-distance dependencies in natural language do not engage a serial memory search (McElree et al 2003; Martin & McElree 2008). In all long-distance dependencies that have been considered, the speed of processing is constant across distance and the size of the search path. However there is one significant point of variation between the dependencies that were tested that could be crucial to understanding the pattern of results across studies. As mentioned, dependencies such as *wh*-movement and clefting that have been studied using the SAT paradigm are *prospective* dependencies. In prospective dependencies, the left edge of the dependency clearly signals information that will be relevant farther downstream in parsing. In *wh*-movement, for example, the fronted

*wh*-word signals to the comprehender that a gap will occur at a later point in the sentence, and thus potentially allows for preprocessing strategies that may obviate the need for memory retrieval (as already noted in McElree et al., 2003). Not all dependencies that have been studied using SAT methods have this property: in particular, verb-phrase ellipsis is a fully retrospective dependency that cannot be reliably anticipated ahead of the integration site (a point noted by Martin & McElree, 2008). However, there are a number of independent reasons why a structured search might not be deployed in finding the antecedent VP for VPE. First, there are no structural constraints on the location of the antecedent verb phrase (Johnson 2001), and so there is no reason to suspect that structural information should be used to narrow the search space. Second, and perhaps more importantly, in the case of Martin & McElree's studies (2008, 2009), the experimental design only included one potential, complete VP antecedent. For this reason, although VPE findings do suggest a content-addressable architecture that does not need to serially consider irrelevant material, they do not rule out a serial consideration of syntactically licit candidate positions.

For these reasons, in order to probe for the existence of structured search, it is crucial to test a dependency that is retrospective, and for which multiple syntactic positions might contain the tail of the dependency. One such dependency is the relation between the Mandarin long-distance reflexive *ziji* and its antecedent. Upon reaching the reflexive element *ziji*, comprehenders need to initiate a search for an antecedent. This is an entirely retrospective process, as there are no cues prior to *ziji* that signal the presence of a dependency (with the exception of contextual bias and

inherently reflexive verbs in Chinese, see Jin 2003; Li & Zhou 2010). There are a number of key properties of *ziji*'s interpretation that make structured search a potentially useful option. These are i) long-distance binding possibilities; ii) syntactic constraints on possible linguistic antecedents; and iii) blocking effects. In order to satisfy these constraints, *ziji* must consider a potentially unbounded search space, and within that space, search must be limited to antecedents that occupy particular syntactic positions relative to both *ziji* and to other potential antecedents. This situation provides one context where a structured search would provide an efficient way to satisfy these constraints during the construction of a dependency between *ziji* and its antecedent.

The search space for *ziji* is unbounded because, unlike the English reflexives *himself/herself*, *ziji* does not require that its antecedent be in the same clause. *Ziji* is an example of the cross-linguistically well-attested class of *long-distance reflexives*. Long-distance reflexives are pronominal elements that have structural requirements on their linguistic antecedents, but unlike English reflexives, long-distance reflexives may be bound outside of their minimal clause (Büring 2005). Local and long-distance binding possibilities for *ziji* can be seen in (1) and (2).

- (4.1)        *Lisi    nongshang-le ziji*  
               Lisi    harm-PERF    self  
               “Lisi harmed herself”

- (4.2)        *Zhangsan<sub>i</sub>    shuo    Lisi<sub>j</sub>    nongshang-le ziji<sub>i/j</sub>*  
               Zhangsan    says    Lisi    harm-PERF    self  
               “Zhangsan says that Lisi harmed him / herself”

Like many long-distance reflexives, *ziji* imposes a number of constraints on potential linguistic antecedents (Büring 2005; Huang et al 2009). There are significant

syntactic constraints placed on antecedents: they must be subjects and they must be contained in the same clause or a higher clause than *ziji* (Huang & Liu, 2001). In addition to these syntactic constraints, there are a number of discourse-pragmatic constraints on the use of *ziji*. Antecedents must be animate and sentient, and must be prominent in the current discourse (Xue, Pollard & Sag 1994; Huang & Liu 2001). In the absence of an appropriate antecedent in the immediate sentential context, *ziji* may refer to the speaker, presumably as a reflex of the prominent discourse status that is automatically afforded to the speaker (Kuno 1987; Huang & Liu 2001). There is ongoing debate about the relative prominence of discourse-pragmatic and structural licensing conditions on *ziji* (Huang, Cole & Hermon, 2006; Huang et al 2009), but it is relatively uncontroversial that both sorts of constraints are operative, and that accurate resolution of the antecedent-anaphor dependency requires the comprehender to systematically exclude inaccessible referents from consideration.

In addition to limiting a potentially unbounded search space to a subset of syntactically or pragmatically accessible antecedents, correct resolution of the antecedent of *ziji* also requires verifying that potential antecedents stand in particular syntactic relations to each other. Consider the following sentences (from Huang & Liu 2001: p.2-3; example (5) modified from their (11), p.6):

(4.3)      *Zhangsan<sub>i</sub>      renwei Lisi<sub>j</sub>      hen      ziji<sub>i/j</sub>*  
               Zhangsan      think   Lisi      hate   *ziji*  
               “Zhangsan thinks that Lisi hates himself/him.”

(4.4)      *Zhangsan<sub>i</sub>      renwei ni<sub>j</sub>      hen      ziji<sub>\*i/j</sub>*  
               Zhangsan      think   you      hate   *ziji*  
               “Zhangsan thinks that you hate yourself/\*him.”

- (4.5)      *Ni<sub>i</sub>      renwei Zhangsan<sub>j</sub>      hen      ziji<sub>?i/j</sub>*  
               You   think   Zhangsan      hate   *ziji*  
               “You think that Zhangsan hates himself/?you.”

Although the matrix subject *Zhangsan* is generally considered an acceptable antecedent for *ziji* as in (3), when the local subject is a second-person pronoun *ni*, then this binding possibility is no longer considered acceptable (4). These examples show an instance of a person blocking effect, where the presence of a first- or second-person antecedent blocks binding from higher subjects (Huang & Liu 2001). Sentences (3) and (4) illustrate the blocking effect for *ziji*: in (4), assessing the feature content of the local subject is necessary to determine whether or not the matrix subject is an acceptable antecedent. Similarly, it has been reported that singular subjects block binding from higher plural subject positions (Tang 1989; Huang & Tang 1991). Thus, it is not only the presence of a pronoun with first- or second-person features that impacts the availability of other subjects, but also its relative position to those subjects.

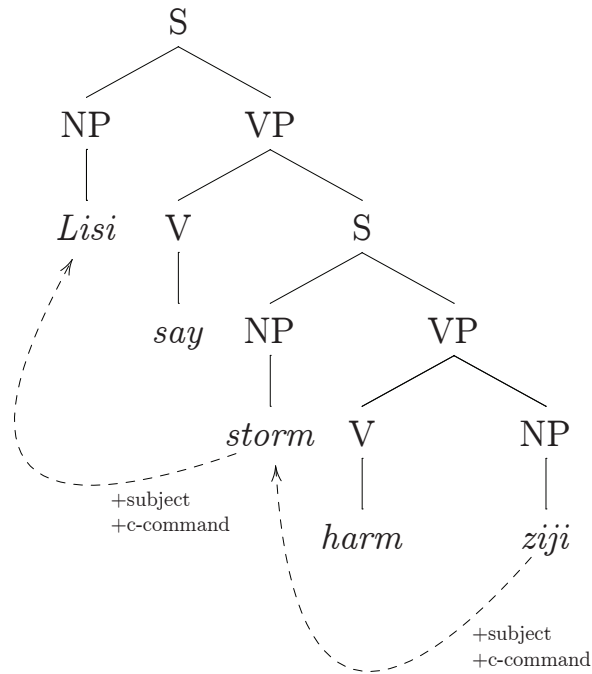
Experimental evidence to date suggests that there is a locality bias in processing *ziji*. For example, Li & Zhou (2010) provide ERP evidence that long-distance binding of *ziji* elicits a P300/600 component relative to local or ambiguous binding of *ziji*. I take up this finding in more detail in our discussion of Experiment 7. In addition to ERP results, a number of cross-modal priming studies have shown a bias towards reactivating the local antecedent upon reaching *ziji*. In one study, Gao, Liu & Huang (2005) showed that in sentences such as “the teacher asked the journalist to respect *ziji*”, reaction times for probes related to the object *the journalist* were faster than reaction times for probes related to the matrix subject *the teacher*.

Liu (2009) replicated and extended this result by varying the SOA of the probe word relative to *ziji*, probing for lexical decisions at 0ms, 160ms and 370ms after the offset of *ziji*. In a very suggestive finding, reaction times to probes related to the local subject were faster at the 0ms SOA, but this pattern reversed at the 160ms probe latency, suggesting greater reactivation for the long-distance antecedent. For these same two SOAs, the opposite pattern of facilitation was observed for the pronoun *ta*. At the longest SOA, both antecedents were equally activated for both pronominals. This result is suggestive of a structured search, with early consideration of the local antecedent and relatively delayed consideration of the long-distance antecedent only for *ziji*, and not for other pronominal elements. However, the reaction times for the lexical decision task do not provide a direct window on the time course of activating antecedents for *ziji*.

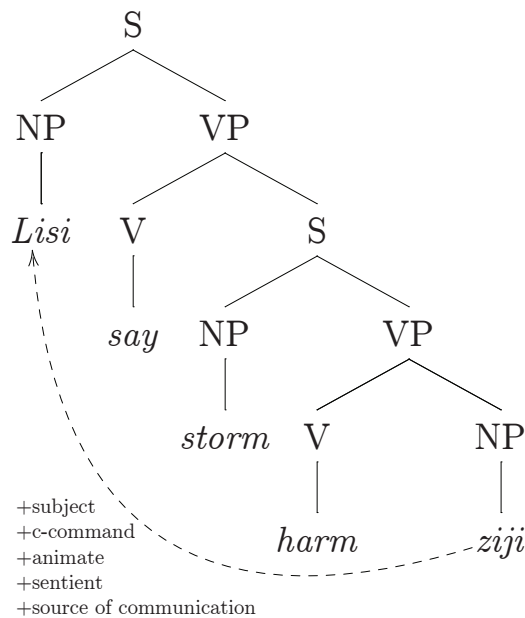
Taken together, the structural constraints and blocking configuration constraints on antecedents provide a good context to observe a structured search, if structured search is indeed an option for the parser to pursue. Structured search is here understood as a process in which potential antecedent positions are sequentially considered as potential antecedents, and this follows naturally from the structured access hypothesis and the serial retrieval of memories assumed by cue-based parsing models like ACT-R. This is because if only syntactic information is used to access memory, then multiple positions may need to be considered before an acceptable, feature-matched antecedent is found. Although the need for speed in language comprehension is an important factor that may favor parallel-access mechanisms (cf. Lewis et al 2006), the parser's ability to reliably recover the correct interpretation

may also be a factor in determining an optimal online strategy for constructing a linguistic dependency. By limiting consideration to specific structural positions, the task of finding an antecedent for *ziji* can arguably more straightforwardly respect structural and blocking constraints. This is because the structural fidelity is presumably enhanced if only structural cues are deployed initially, and the blocking constraints can be naturally implemented by terminating the search if the local subject contains a ‘blocking’ feature specification. A schematic comparison of the antecedent retrieval process is shown in Figures 4.3 and 4.4. If only structural cues are used to access memory, then when the correct feature-matched antecedent is distant, multiple operations are needed to recover the correct antecedent, as in Figure 4.3. If, however, all semantic and structural cues are used in parallel to retrieve the antecedent, then only a single retrieval step is needed to retrieve the antecedent. This is shown in Figure 4.4, where the enriched cue set allows the antecedent to be recovered in a single access step.





**Figure 4.3:** Example of a structured search process for finding *ziji*'s antecedent in the sentence *Lisi shuo fengbao hai-le ziji* "Lisi said the storm harmed him". The hypothetical structural cues do not allow comprehenders to rule out consideration of the local subject *fengbao* "storm". Thus comprehenders must evaluate multiple subject positions in the search for the correct antecedent *Lisi*. This structured access predicts that processing time should grow with the number of subject positions that need to be evaluated.



**Figure 4.4:** Example of feature-based access for finding *ziji*'s antecedent in the sentence *Lisi shuo fengbao hai-le ziji* "Lisi said the storm harmed him". The mixture of structural and semantic cues allow direct access to the correct antecedent *Lisi*. Feature-based access predicts that processing time should be constant with the number of subject positions that need to be evaluated.

If even the retrospective, structurally constrained *ziji*-antecedent dependency is resolved using a feature-based, parallel-access mechanism, then a strong case can be made for such a feature-based access mechanism across all levels of sentence processing. If, on the other hand, structured search is implicated in the processing of *ziji*, then it would suggest that the parser can engage in a structured search of linguistic memory for at least some linguistic long-distance dependencies, due to a selective limiting of the search cues to only structural cues.

#### *Experiment 6: SAT Evidence*

Experiment 6 employed the multiple-response speed-accuracy tradeoff (MR-SAT) procedure to provide time-course evidence about whether or not *ziji* accesses all potential antecedent positions in parallel, or whether or not certain positions are checked before others. Figure 4.3 presented a schematized comparison between a serial, structured access to potential antecedents, and Figure 4.4. shows the alternative mechanism, direct-access using semantic features. I follow the same logic employed in prior SAT studies (McElree 2000; McElree et al 2003; Foraker & McElree 2007; Martin & McElree 2008, 2009). If all possible antecedent positions are considered in parallel for *ziji*, then local and non-local antecedents should show similar processing speed (identical dynamics parameters in the SAT function). If, on the other hand, certain positions are checked prior to other positions, then we should observe different processing speeds for local and long-distance bindings. I also predict that due to the nature of the blocking effects described above, any speed advantage should favor local antecedents. The availability of the long-distance binder depends on the nature of the local binder, and so a reasonable strategy would be to check a local subject before a non-local subject, in order to determine whether the non-local subject position should even be considered. Furthermore, the linear recency of the local subject may itself confer an advantage over the long-distance subject.

### Participants

Twenty college students from Beijing Normal University participated in the experiment. Data from 2 participants were excluded due to atypical (sigmoidal) SAT

curves. The remaining 18 participants included 10 females, and had a mean age of 23.5 years. Each participant completed six 1-hour sessions spaced at least a day apart, in addition to a 1-hour practice session for familiarization with the multiple-response speed-accuracy tradeoff procedure (MR-SAT; see below). All participants were native Mandarin Chinese speakers and had normal or corrected-to-normal vision. They were paid 35 RMB per hour for their participation in the experiment.

### Materials

Experiment 6 investigated the processing of Mandarin sentences that contained a matrix attitude verb and an embedded transitive complement clause. Two experimental factors were manipulated. One was the position of an animate subject; it was either the subject of the matrix clause, the subject of the local clause, or not present. The second factor was the identity of the object, which was either *ziji*, an acceptable definite NP, or an unacceptable definite NP. Since *ziji* forms a long-distance binding dependency with its antecedent, the predictions of a structured search are that the position of the target antecedent should impact processing time. Control conditions with an acceptable definite NP (thus with no binding dependency) were included to ensure that any processing speed differences that are observed are not due to baseline differences due to the sentential context. Because a definite NP object does not form a binding dependency with its animate subject, no processing speed differences should obtain when this factor is manipulated for control conditions.

This design provided three critical reflexive conditions designed to investigate the processing of *ziji* (conditions 1-3 in Table 1 below). Based on the position of the animate subject, *ziji* either took a long-distance antecedent (*LD antecedent* condition; 1), a local antecedent (*local antecedent* condition; 2), or had no antecedent in the sentence (*no antecedent* condition; 3). In these conditions, the animacy of the subjects was the factor that controlled which antecedent *ziji* was forced to take. As noted above, an animate NP in either the main clause or embedded subject NP position can function as a grammatical antecedent for *ziji*. In addition to the critical *ziji* condition, two control sentences were constructed with definite descriptions instead of *ziji* in the embedded object position (*LD control*, *local control*, *no antecedent control*; conditions 4-6). The well-formed control conditions replaced *ziji* with a full NP that was a plausible object of the embedded verb (e.g., *the batsman* in conditions 4-6 below); whereas the unacceptable control conditions replaced *ziji* with a full NP that was an implausible object of the embedded verb (e.g., *glasses* in conditions 7-9 below), resulting in a semantic anomaly. The anomaly conditions (conditions 7-9) and the conditions with no animate subjects (conditions 3,6 and 9) were included for purposes of *d'* scaling, as described below, and so they were not directly analyzed. The primary experimental contrast was the effect of subject animacy on the speed and accuracy of processing sentences that contained *ziji*, and sentences that contained acceptable definite NP objects.

Therefore, each set of items consisted of three critical (conditions 1-3) and six control sentences (conditions 4-9). The control sentences additional helped to prevent participants from forming strategies based on the pattern of animate and inanimate

referents that preceded the critical object NP. Thus, because of the structure of the control conditions, neither the presence of *ziji* nor the acceptability of the continuation was predictable from the preamble. This was done to ensure that all processing measures reflect processes initiated at the object NP itself. All three conditions consisted of a main clause that contained a verb of reporting, and an embedded transitive clause. Additionally, to increase complexity and difficulty of the task, a temporal adverbial clause was interpolated between the embedded subject and the embedded verb. In all conditions, an animate NP was also used as the subject of the temporal adverbial phrase. However, since it occupied a structural position that does not c-command *ziji*, it was not a grammatical antecedent for *ziji*. In the local antecedent and control conditions, the main clause NP was always a ‘media’ noun (e.g. book, documentary, memo) to ensure compatibility with the meaning of the main clause verb. In the no antecedent condition, both structurally accessible NPs were inanimate, and therefore this condition did not contain a grammatical antecedent for *ziji*. None of the inanimates used in any position could be construed metonymically; metonymic interpretations of inanimates (i.e. *the newspaper* being used to refer to the employees of the newspaper) may be used as antecedents for *ziji*.

Forty sets of the 9 sentence types (5 acceptable and 4 unacceptable) were generated. The 360 sentences were equally distributed in 6 presentation lists, one for each of the 6 sessions, to minimize the repetition of content material within a session. Crucially, no two instances of *ziji* sentences (conditions 1-3) from the same set were included in a single presentation list. Within a session, each participant viewed 206 sentences, of which 60 were drawn from the current study. Since only one third of

target sentences contained *ziji*, the critical *ziji* conditions comprised around 10% of all sentences within and across sessions. The order of presentation within a session was randomized.

### Procedure

I employed the multiple-response SAT paradigm, following Wickelgren, Corbett & Doshier (1980). Stimulus presentation, timing, and response collection were all carried out on a personal computer using the Linger software by Doug Rohde (available at <http://tedlab.mit.edu/~dr/Linger/>). Each trial began with a 500 ms fixation cross presented in the center of the screen. Each word appeared in the center of the screen for 400 ms, followed by 200 ms of blank screen. All words were presented using simplified Chinese characters, and the last word of each sentence was marked with a period (。). Immediately prior to the onset of the final word, a series of 18 auditory response cues (50 ms, 1000 Hz tone) was initiated. The cues occurred every 350 ms, and the final word of the sentence remained on the screen. Participants were trained to initially respond by pressing both response keys simultaneously to indicate an undecided response. They were then trained to give a response after each tone, and to switch their response to either the ‘accept’ or ‘reject’ key as soon as they could. Importantly, they were also trained to modify their responses if their assessment changed. During the 1-hour practice session, participants were told that some of the sentences were complex, but nevertheless were meaningful sentences. Each

participant performed six 1-hour sessions, and in each they saw one of the lists of materials. The order of lists was randomized across participants.

# Condition	Example
1 LD antecedent	张教练 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 自己 Coach Zhang say [that report [when team not perform well time] underestimate ziji] “Coach Zhang says that that report underestimated <i>self</i> when the team was doing poorly.”
2 Local antecedent	回忆录 表明 张教练 [在 团队 未能 发挥 水准 的时候] 低估了 自己 Auto-biography say [coach Zhang [when team not perform well time] underestimate ziji] “The auto-biography says that coach Zhang underestimated <i>self</i> when the team was doing poorly.”
3 No antecedent	*回忆录 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 自己 *Auto-biography say [that report [when team not perform well time] underestimate ziji] * “The auto-biography says that that report underestimated <i>self</i> when the team was doing poorly.”
4 LD control	张教练 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 那位击球手 Coach Zhang say [that report [when team not perform well time] underestimate that batsman] “Coach Zhang says that that report underestimated <i>the batsman</i> when the team was doing poorly.”
5 Local control	回忆录 表明 张教练 [在 团队 未能 发挥 水准 的时候] 低估了 那位击球手 Auto-biography say [coach Zhang [when team not perform well time] underestimate that batsman] “The auto-biography says that coach Zhang underestimated <i>the batsman</i> when the team was doing poorly.”
6 No antecedent control	回忆录 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 那位击球手 Auto-biography say [that report [when team not perform well time] underestimate that batsman] “The auto-biography says that that report underestimated <i>the batsman</i> when the team was doing poorly.”
7 LD control anomaly	*张教练 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 眼镜 *Coach Zhang say [that report [when team not perform well time] underestimate glasses] *“Coach Zhang says that that report underestimated the glasses when the team was doing poorly.”
8 Local control anomaly	*回忆录 表明 张教练 [在 团队 未能 发挥 水准 的时候] 低估了 眼镜 *Auto-biography say [coach Zhang [when team not perform well time] underestimate glasses] * “The auto-biography says that coach Zhang underestimated the glasses when the team was doing poorly.”
9 No antecedent control anomaly	*回忆录 表明 那篇报导 [在 团队 未能 发挥 水准 的时候] 低估了 眼镜 *Auto-biography say [that report [when team not perform well time] underestimate glasses] * “The auto-biography says that that report underestimated the glasses when the team was doing poorly.”



**Table 4.1:** Summary of conditions in experiment: Critical *ziji* conditions. Critical conditions are 1-2 and 4-5; conditions 3 and 6-9 were included for purposes of  $d'$  scaling (see text).

### Data Analysis

To derive the full time-course information in SAT analysis,  $d'$  scores are calculated by comparing the judgments in an acceptable condition and a closely matched unacceptable condition at each of the response points, within participants. The resulting series of  $d'$  values at each time point  $t$  is fit using a shifted exponential function:

$$(4.6) \quad \begin{aligned} d' &= \lambda(1 - e^{-\beta(t-\delta)}) & , t > \delta, \\ d' &= 0 & , \text{otherwise} \end{aligned}$$

Here,  $d'$  is the standard measure of discrimination:  $d' = z(\text{hits}) - z(\text{false alarms})$  (Wickens, 2001). The shifted exponential in (4.6) describes the growth of accuracy over time  $t$  (in ms) with three parameters: asymptote ( $\lambda$ ), rate ( $\beta$ ), and intercept ( $\delta$ ). The current experiment used common scaling in its design: in order to derive the  $d'$  scores for LD and local antecedent conditions, their hit rates for the critical conditions (Conditions 1 and 2) were scaled against the pooled false alarm rate at the corresponding time lag from the three ungrammatical controls (Conditions 7-9), following McElree & Griffith (1998). Both critical *ziji* and control conditions were scaled using this pooled scaling, to allow a more straightforward comparison of any potential differences in the resulting time courses that may arise. All differences

between *ziji* and control conditions stem from differences in the hit rates between the two. One reason for adopting this measure, instead of scaling against the no-antecedent *ziji* condition was the somewhat high acceptability of the no antecedent condition (see below). If the no antecedent condition were used for scaling purposes, the false alarm rate used in the  $d'$  calculations would have been high, leading to low  $d'$  scores (less than 1 in most cases). For such low scores, small variations in hit rate would cause very large changes in the observed  $d'$ , leading to unreliable SAT function estimates. An additional benefit is that pooled scaling produced SAT curves for the *ziji* conditions and the control conditions with comparable asymptotic accuracy, ensuring that neither floor nor ceiling effects could be responsible for any observed effects. Thus, this scaling measure allows direct comparison between the critical *ziji* conditions and their non-*ziji* acceptable control counterparts (conditions 4 and 5), and we can straightforwardly estimate the dynamics of the acceptable judgment alone (i.e., the successful completion of the dependency).

In order to determine whether the SAT functions for these conditions differed in asymptote ( $\lambda$ ), rate ( $\beta$ ), or intercept ( $\delta$ ), the analysis proceeded in two steps: a *model selection analysis* and a *parameter estimation analysis* (Li & Smith 2009). In the model selection analysis, the best fit model was determined using the adjusted  $R^2$ -statistic using a hierarchical model-testing scheme over the averaged and individual data, an approach pursued in prior work on SAT in sentence comprehension (McElree et al 2000, 2003; Foraker & McElree 2007; Martin & McElree 2008, 2009). In the parameter estimation analysis, only the fully specified model was considered, and any differences between the critical conditions on the parameters of interest were assessed

using familiar hypothesis testing measures. In all analyses the SAT function given above was fit to the measured  $d'$  values at each time lag. In order to obtain parameter estimates, it is necessary to fit non-linear regressions of the SAT function (1) against the observed  $d'$  score. In the present study, I employed Gibbs sampling, a Monte Carlo method for approximating the posterior of a distribution of two or more random variables. Gibbs sampling is often used as a form of Bayesian statistical inference (Gelman & Hill 2005). For the current analysis, Gibbs sampling was used to estimate the posterior distributions over the three SAT parameters, and the best-fit model for a given set of data was chosen by selecting the median of the resulting posterior distributions. Note that this method of fitting SAT curves contrasts with the method often employed in the psycholinguistic SAT literature (STEPIT fits, McElree & Griffith 1998; McElree et al. 2000, 2003; Martin & McElree). In this approach, the SAT function is fit with an iterative hill-climbing algorithm (Reed, 1976; similar to STEPIT, Chandler, 1969), and fit quality was assessed by an adjusted  $R^2$ -statistic, which measures the variance accounted for by the fit, with a penalty for an increasing number of model parameters. In almost all cases the two methods provide identical results, and I primarily employ the Gibbs sampling method because in addition to providing information about the estimate of a value, it provides an estimate of the variance associated with that value. However, to ensure that the estimation technique did not change the qualitative pattern of results, I also report STEPIT fits in the parameter estimation section. A direct comparison of STEPIT and Bayesian fits, as well as code for implementing the models here, is available at <http://people.umass.edu/bwdillon/>.

For all Monte Carlo simulations, 4000 samples were used as a burn-in period, followed by 15,000 iterations of the sampler. 3 parallel sampling chains were run in each simulation, and convergence was checked by evaluating the potential scale reduction factor ('r-hat') statistic (Gelman & Hill 2005). For all reported parameter values, the potential scale reduction factor was effectively 1, indicating satisfactory convergence of the MCMC chains.

In addition to model-fitting on the  $d'$  data, analysis was performed on participant and item mean final acceptance rates (the *empirical accuracy*), which was obtained by taking the average rate of acceptance over the last response point.

Data from two participants were excluded due to unreliable dynamics estimates. The empirical  $d'$  scores from these participants appeared to be better fit by a sigmoidal rather than an exponential function, leading to unrealistically large and unreliable differences in the critical conditions in the crucial intercept and rate parameters when fit with the SAT function in (4.6).

#### Empirical Accuracy Analysis

The rate of acceptance of for the critical *ziji* conditions was 85% for the LD antecedent condition, 81% for the local antecedent condition, and 48% for the no antecedent condition. The rate of acceptance for the corresponding acceptable control conditions was 91%, 87% and 91%. A mixed-effects logistic linear regression was fit with orthogonal fixed-effect contrasts for *locality* (comparing long-distance versus local antecedent conditions) and *binding* (comparing both *ziji* conditions with

antecedents to the no antecedent condition), as well as random slopes for subject, item, and session. For critical *ziji* conditions, there was a significant effect of *binding* ( $\beta = -1.3$ ,  $z = -17.5$ ,  $p < .0001$ ) and a marginal effect of *locality* ( $\beta = .28$ ,  $z = 1.94$ ,  $p < .06$ ). For control conditions, the observed empirical accuracy was a significant effect of *locality* only ( $\beta = .70$ ,  $z = 3.97$ ,  $p < .0001$ ). For unacceptable controls, accuracy for long-distance, local, and no antecedent controls was 99%, with no differences between these conditions.

### Model Selection Analysis

For the model selection analysis, individual participants' data was fit with a series of nested models separately for control and *ziji* conditions. This analysis compared long-distance and local animate configurations. The model-fitting analysis pitted a series of nested models (including shared or separate parameters for the two conditions of interest for each of the intercept, rate and asymptote) against each other on adjusted  $R^2$ , following McElree et al (2003) and Li & Smith (2009). This was done by participants to determine if extra parameters led to a significantly better model fit, as measured by adjusted  $R^2$ -statistic. Adjusted  $R^2$  (4.7) gives an estimate of the variance accounted for by the (non-linear) regression against the SAT curve, weighted by the number of parameters used in constructing that curve ( $k$ ). In (4.7),  $d$  refers to the observed  $d'$  value,  $\hat{d}$  refers to the predicted  $d'$  value, and  $n$  refers to the number of data points.

$$(4.7) \quad R^2 = 1 - \frac{\sum_{i=1}^n (d_i - \hat{d}_i)^2 / (n - k)}{\sum_{i=1}^n (d_i - \bar{d})^2 / (n - 1)}$$

For the critical *ziji* conditions, adding separate asymptote parameters for LD and local antecedents led to a reliable increase in adjusted  $R^2$  across participants ( $\Delta\mu = 0.03 \pm 0.01$ ;  $t(17) = 3.9$ ,  $p < 0.01$ ). Compared to the two-asymptote baseline, the addition of an extra rate parameter led to a small but reliable increase in adjusted  $R^2$  ( $\Delta\mu = 0.002 \pm 0.0005$ ;  $t(17) = 4.3$ ,  $p < 0.001$ ). Any model that included an extra intercept parameter led to a significantly poorer fit on adjusted  $R^2$ . Thus the best fitting model across participants allotted separate asymptotes and separate rate parameters for local and long-distance antecedents for *ziji* ( $2\lambda-2\beta-1\delta$ ).

For control conditions, the addition of an extra asymptote for long-distance and local controls led to a marginally significant increase in adjusted  $R^2$  ( $\Delta\mu = 0.032 \pm 0.016$ ;  $t(17) = 2.1$ ,  $p < 0.055$ ). Neither additional rate parameters nor additional intercept parameters for LD and local control environments led to a reliable increase in adjusted  $R^2$ . Across participants, the best-fitting model for control conditions had shared dynamics parameters for LD and local conditions, but separate asymptotes ( $2\lambda-1\beta-1\delta$ ).

The model-fitting analysis suggests that the critical *ziji* conditions have different processing speeds, as allotting separate rate parameters  $\beta$  to LD and local antecedent conditions led to a reliable increase in adjusted  $R^2$ . However, with the model-fitting approach one might be concerned by the possibility of over-fitting, a concern magnified by the small differences in adjusted  $R^2$  among the best fit models

for any given participant's data (Liu & Smith 2009). For this reason it is important to ensure that the parameter estimates across participants display a consistent ordering (see, e.g., McElree et al 2003) by directly analyzing the estimated SAT parameters across participants.

### Parameter Estimation Analysis

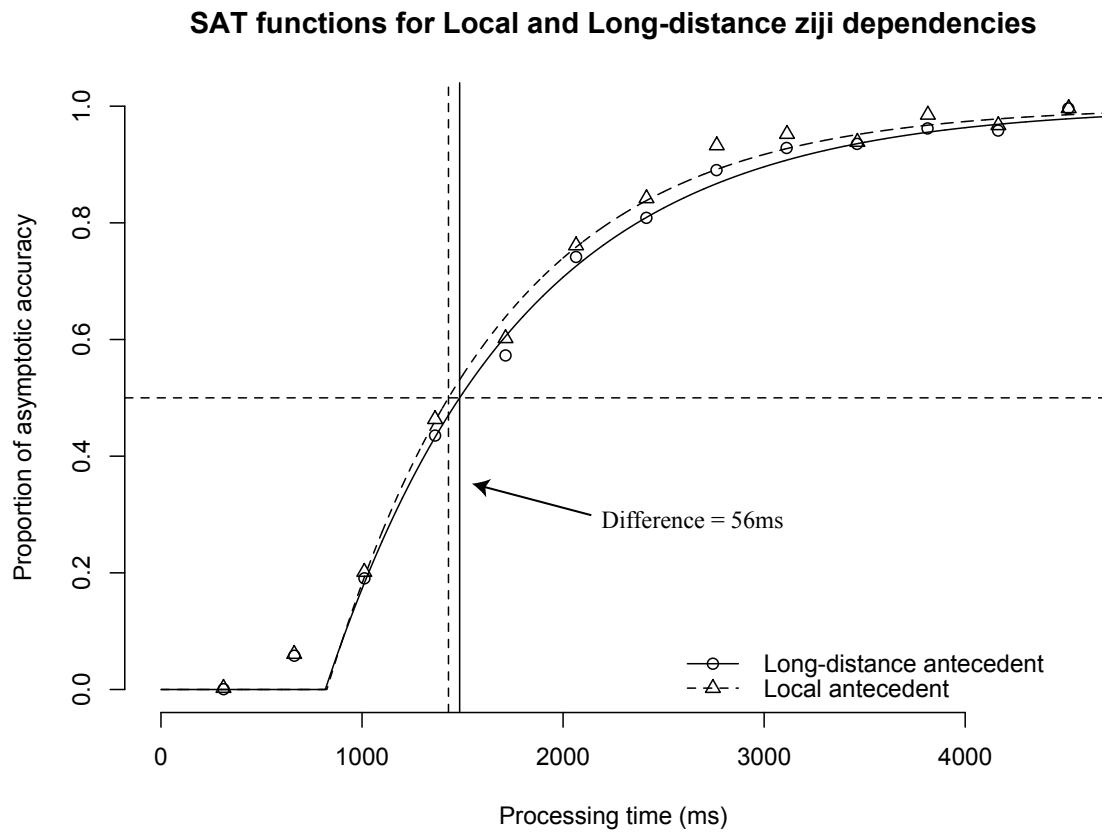
The focus of the parameter estimation analysis, following Liu & Smith (2009), was the magnitude of the difference between parameters for local and long-distance antecedent configurations. Here this was done by fitting fully saturated models ( $2\lambda-2\beta-2\delta$ ) for each participant and testing for consistent ordering of parameters across participants. In addition to presenting absolute parameter values for the processing of local and LD conditions, I also present the differences between the two values of a given parameter across these conditions. This can be understood as the advantage enjoyed by one condition over the other in speed (for differences in dynamics parameters) or accuracy (for differences in the asymptote).

Parameter estimates for each of the participants included in the analysis were estimated using the Bayesian method, although I also present fits from the hill-climbing algorithm employed in other SAT work (McElree 2000; McElree et al 2003; Martin & McElree 2008). For Monte Carlo fits, the median of the resulting posterior distribution for each participant was chosen as the estimate and submitted to further analysis. Figure 4.7 summarizes the differences in all dynamics parameters for both critical *ziji* and control conditions, for each individual participant. Values above 0 in

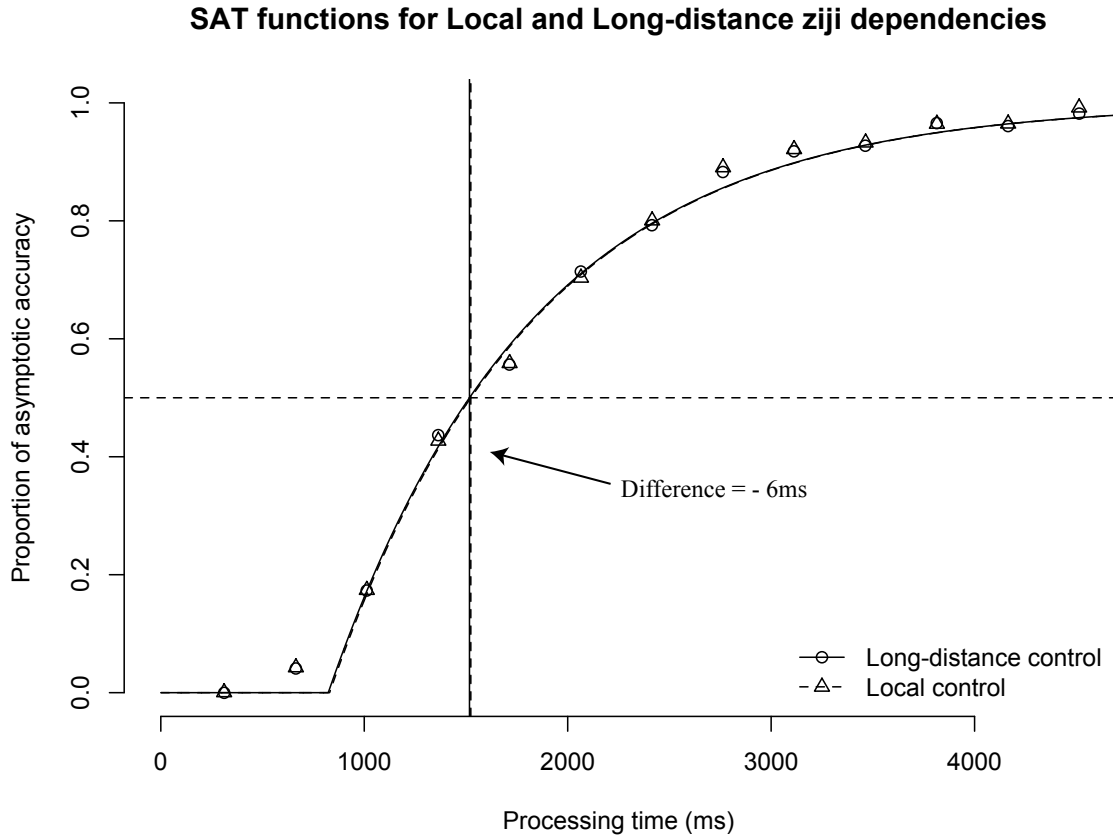
the figures indicate a dynamics advantage for local antecedent configurations. In addition to rate and intercept estimates, I report the composite processing measures speed ( $\beta^{-1} + \delta$ ), a measure that helps to guard against parameter tradeoffs in the dynamics estimates (Carrasco, Giordano & McElree, 2006). In all cases, I present the best estimate for the size of the difference between the parameters in local and long-distance configurations. For this comparison, we take the long-distance configuration to be the ‘basic’ configuration; thus, all parameter differences presented below reflect the dynamics advantage in local antecedent configurations.

The average advantage for local antecedents in the *ziji* and control conditions is summarized numerically in Table 4.2. Fully saturated model fits to average data are presented in Figures 4.5 and 4.6, scaled to proportion of asymptote in order to highlight the differences in the growth portion of the SAT curve. Figures 4.7-4.10 show the mean parameter values for accuracy, rate, intercept and speed for both *ziji* and control conditions. The results from both methods of parameter estimation are largely in agreement. However, the STEPIT fits appear in general to trade greater speed in the rate parameter for local configurations for later intercepts; no such trade-off occurs for Bayesian fits. However, for the compound speed measure, where these dynamics parameter tradeoffs are controlled, the estimates from both approaches are closely matched.





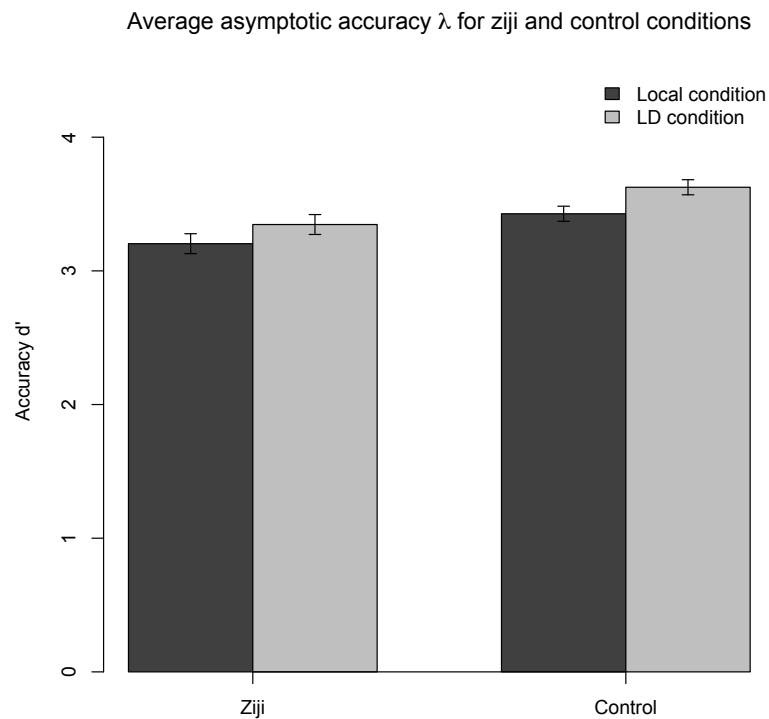
**Figure 4.5:** SAT functions for LD and local antecedent *ziji* conditions with fully saturated models ( $2\lambda-2\beta-2\delta$ ), over average data (not averaged parameters). Accuracy is scaled to show proportion of asymptote; vertical bars indicate time point at which 50% accuracy is reached.



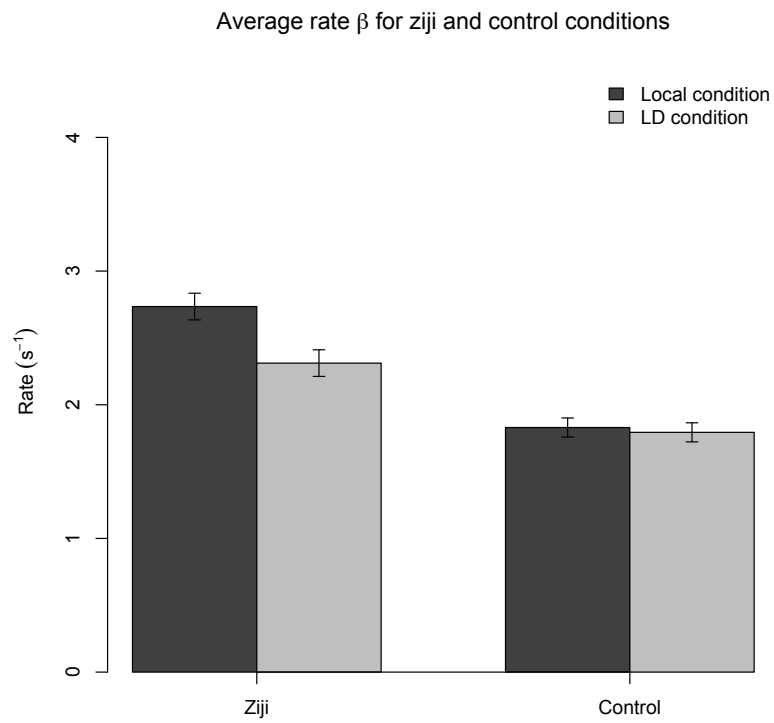
**Figure 4.6:** SAT functions for LD and local control conditions with fully saturated models ( $2\lambda-2\beta-2\delta$ ), over average data (not averaged parameters). Accuracy is scaled to show proportion of asymptote; vertical bars indicate time point at which 50% accuracy is reached.

For the critical *ziji* comparison, parameter estimates from STEPIT fits indicated no reliable difference in estimated asymptote ( $\lambda$ ) parameters between the two critical *ziji* conditions. There was a reliable difference between the *ziji* conditions for rate ( $\beta$ ) parameters, as well as for the compound speed measure ( $t(17) = 2.47, p < 0.05$  and  $t(17) = 2.52, p < 0.05$ , respectively). In addition, for the intercept ( $\delta$ ) parameter there was a marginal effect of condition ( $t(17) = -2.09, p < 0.06$ ). For Bayesian fits, the pattern of findings was qualitatively similar. There were no reliable differences for local versus long-distance configurations in either the

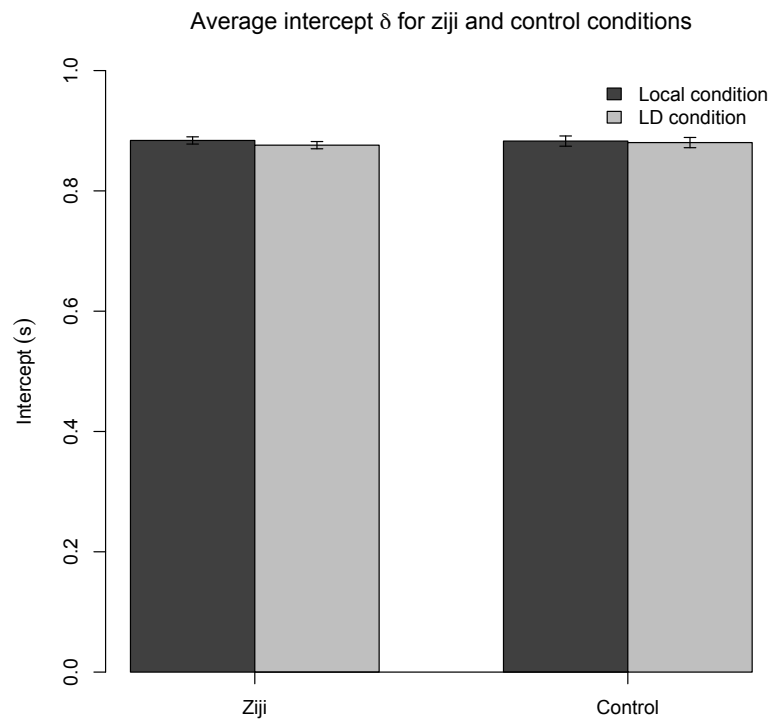
asymptote parameter or the intercept parameter. There were, however, reliable effects in both the rate parameter ( $t(17) = 2.21, p < 0.05$ ) and the compound speed measure ( $t(17) = 2.28, p < 0.05$ ).



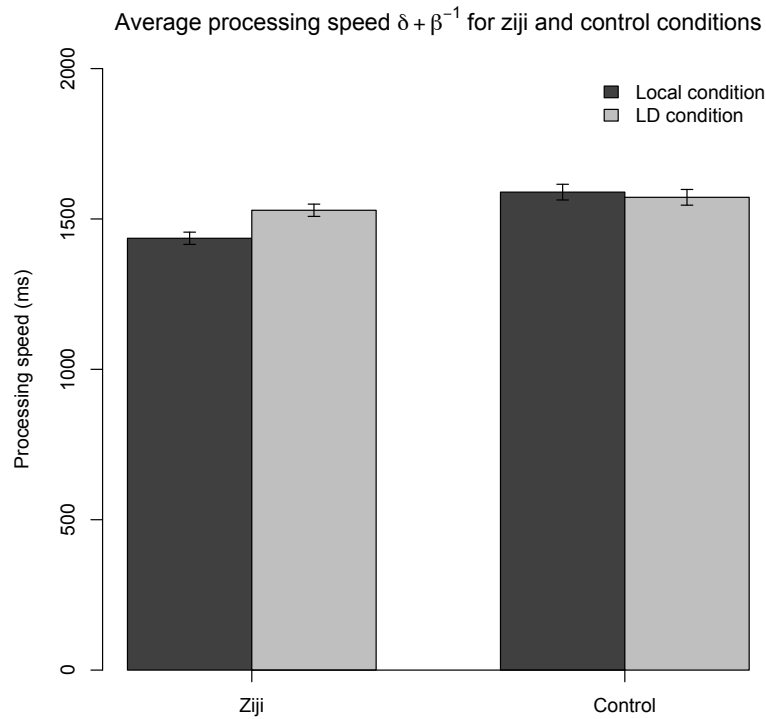
**Figure 4.7:** Average asymptotic accuracy ( $\lambda$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show  $\pm 1$  SE, corrected for between-participant variance.



**Figure 4.8:** Average rate ( $\beta$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show  $\pm 1$  SE, corrected for between-participant variance.



**Figure 4.9:** Average intercept ( $\delta$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show  $\pm 1$  SE, corrected for between-participant variance.



**Figure 4.8:** Average speed ( $\delta + \beta^{-1}$ ) across individual participant SAT function fits with Bayesian parameter estimation. Error bars show  $\pm 1$  SE, corrected for between-participant variance.

On control conditions, STEPIT fits revealed no reliable differences between the local and long-distance configurations for any parameter. Likewise, Bayesian estimates of the parameters revealed only a marginal effect in the asymptote ( $\lambda$ ) parameter ( $t(17) = -1.75, p < 0.1$ ), and no other reliable differences between the control conditions in any of the dynamics parameters.

It is unclear whether the processing speeds of the control conditions and the *ziji* conditions are directly comparable, as the processes involved in rendering a judgment in the two cases are distinct. Nonetheless, planned pairwise comparisons were performed to directly compare parameter differences for the *ziji* conditions to

those of the control conditions. For STEPIT fits, the locality advantage in *ziji* conditions over control conditions was significant for both rate and speed parameters ( $\beta$ :  $t(17) = 2.98, p = < 0.01$ ,  $\beta^{-1} + \delta$ :  $t(17) = 2.29, p < 0.05$ ). The same pattern was found for the Bayesian fits, although this effect was only marginally significant for the compound speed measure ( $\beta$ :  $t(17) = 2.69, p < 0.05$ ,  $\beta^{-1} + \delta$ :  $t(17) = 1.74, p < 0.1$ ).

One possible concern about the analysis presented here is that the fully saturated models presented here posit separate asymptote parameters for the critical *ziji* conditions. If the true model does not contain separate asymptotes for each condition, a potential worry is that non-significant trends in accuracy drive the observed dynamics effects. However, the accuracy effect for the critical *ziji* conditions was nearly reliable in the empirical accuracy analysis, and so it is unlikely that modeling a given participant's data with a single asymptote model is appropriate. Furthermore, since almost every participant presented data that was significantly better fit with a two-asymptote model, modeling the data with a single asymptote would have led to significant distortions of the estimates of the dynamics parameters. For participants who have greater asymptotic accuracy for LD antecedent conditions, a single-asymptote model inappropriately increases the processing speed estimates for long-distance configurations, and conversely for participants who show the opposite pattern of accuracy.

Parameter	<i>Ziji</i>		Control	
	Bayesian fit	STEPIT fit	Bayesian fit	STEPIT fit
Asymptote	-0.14 ( $\pm .15$ )	-0.15 ( $\pm .14$ )	-0.20 ( $\pm .11$ )	-0.17 ( $\pm .11$ )
Rate ( $s^{-1}$ )	0.42 ( $\pm .20$ )	0.86 ( $\pm .35$ )	0.04 ( $\pm .15$ )	-0.04 ( $\pm .13$ )
Delta ( $s$ )	-0.01 ( $\pm .01$ )	-0.32 ( $\pm .15$ )	0.00 ( $\pm .02$ )	0.02 ( $\pm .02$ )
Speed ( $s$ )	0.09 ( $\pm .04$ )	0.09 ( $\pm .04$ )	-0.02 ( $\pm .05$ )	0.03 ( $\pm .04$ )

**Table 4.2:** Difference in parameter estimates between local and long-distance configurations on the critical *ziji* and control conditions. Values greater than 0 indicate a processing advantage for local antecedents. Standard errors by subject are in parentheses.

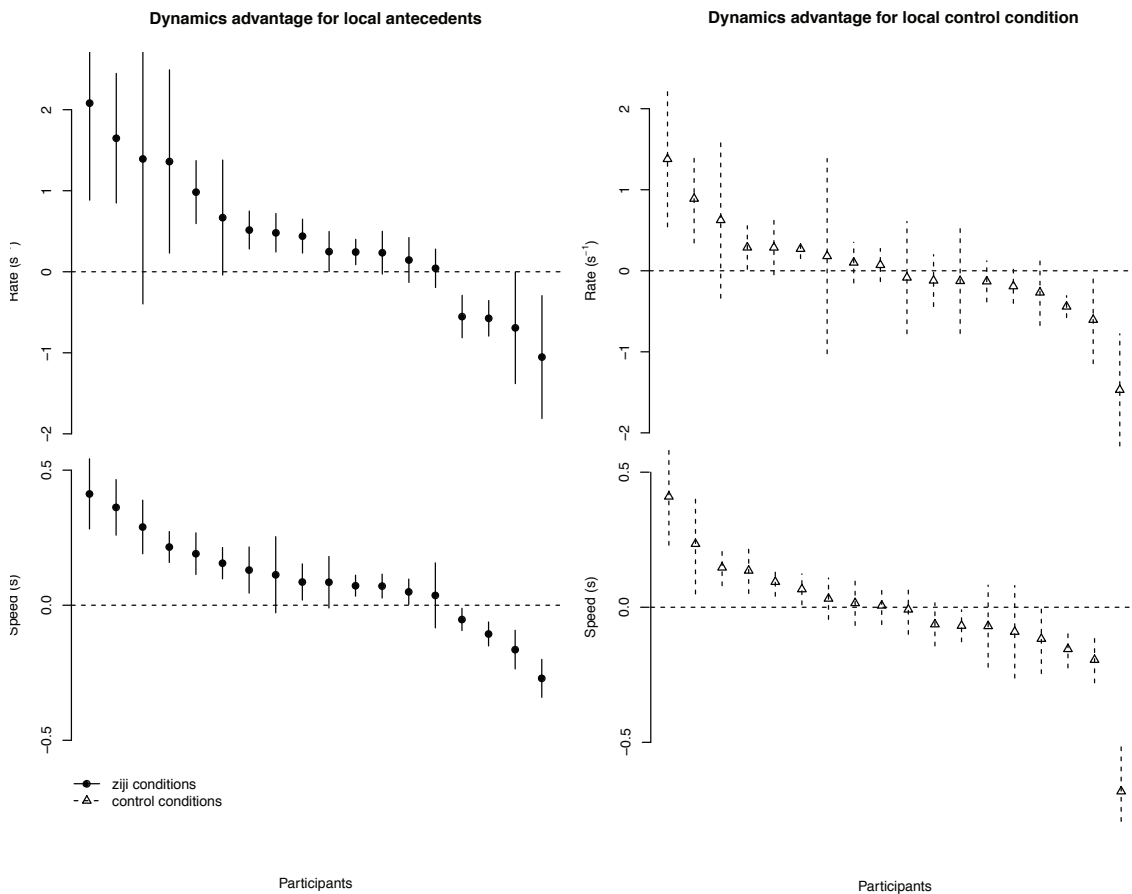
### Discussion

Both analyses of the SAT data indicate that the critical *ziji* conditions differed in processing speed, and that the control conditions did not. Model-fitting analyses showed that for the critical *ziji* comparison, the best fitting model attributed separate rate ( $\beta$ ) parameters and asymptote parameters ( $\lambda$ ) to local and long-distance antecedent conditions. This indicated that the two *ziji* conditions reliably differed in speed and accuracy across conditions. The control comparison, on the other hand, showed no such advantage for extra dynamics parameters; only separate asymptotic accuracy parameters for each control condition reliably improved model fit. This analysis was supplemented with an analysis of the resulting parameter estimates, in order to check for consistent ordering of parameters. This analysis showed that in both rate and speed measures of the *ziji* conditions, the local antecedent was processed reliably faster across participants than was the LD antecedent condition. No such difference was observed in the control conditions, although there was a marginal effect of asymptotic accuracy for these conditions.

Thus the two analyses of the SAT data show that long-distance and local *ziji* conditions differ reliably in processing speed. Furthermore, the analysis of the direction of the parameters indicates that when *ziji*'s antecedent is local, it is reaccessed faster than when it is contained in a dominating clause. The most direct



measure of this effect is the compound speed measure indicated in Figure 4.11. The speed measure combines both rate and intercept parameters, and provides a measure of average processing speed in seconds. It can be seen that on average, local *ziji* antecedent configurations have approximately a 90ms processing advantage over the long-distance configurations.



**Figure 4.11:** By-participants summary of dynamics advantage (in both rate and speed parameters) for local antecedent over LD antecedent conditions for *ziji* and control conditions. Participants are ranked by size of advantage; order is not identical across *ziji* and control conditions. Error bars represent 1 standard deviation of the posterior distribution of that parameter's estimate.

This finding supports the structured search hypothesis. Comprehenders appear to access the local antecedent before accessing the long-distance subject, a finding that is compatible with a number of architectural implementations. If *ziji* activates its antecedent with purely structural information, like English *herself/himself*, then this finding is expected; the sort of iterated structural access mechanism that would derive this processing profile is presented in Figure 4.3. It can be seen that, on the assumption that the local antecedent is more active, then recovery of the the long-distance antecedent requires multiple retrieval processes.

This finding stands in contrast to previous SAT work, which has generally found that memory access in sentence processing occurs with essentially constant time. I argued that the syntactically constrained, fully retrospective nature of *ziji* makes it an ideal dependency for testing the possibility of structured search, if such an access mechanism is ever engaged. The evidence presented here provides positive evidence that the parser does engage in structured search. Although they stand in contrast to prior SAT findings, this finding about the access profile of *ziji* converges nicely with prior findings on the processing of *ziji*, suggesting both serial access (Gao et al 2005; Liu 2009) and a locality preference (Liu 2009; Li & Zhou 2010).

One unexpected finding was the high acceptance rate of the no antecedent *ziji* condition, which participants accepted on 48% of trials. It is known that in the absence of linguistic binding, *ziji* can refer to the speaker (Huang & Liu 2001). In the context of this experimental task, it is unclear how participants calibrated the point of view of the sentences they observed. The perceived point of view of the sentence is presumably the main factor that controls the recoverability of this indexical

interpretation of *ziji*. Since this aspect was not controlled, however, it is unclear how participants perceived these unbound *ziji* sentences. The relationship of unbound *ziji* to the linguistically bound examples presented here is an important question, but I delay further discussion of it until after Experiment 7.

The SAT evidence presented here is indicative of a structured search strategy, but there remain some unaddressed questions. If structured search proceeds by checking and rejecting the local subject position before constructing a dependency with an acceptable, but distant subject position, as in Figure 4.3, then a process that functionally approximates reanalysis is engaged in recovering a long-distance interpretation. One important question is the relationship between structured search and reanalysis processes that are engaged by this type of structured search and those that are engaged in more familiar cases of ambiguity, such as garden-path sentences (Bever 1970; Kimball 1973; Frazier & Fodor 1978).

A related question concerns the automaticity of this process. Does structured search reflect a strategy that participants adopt in the context of a particular experimental manipulation, or is it indicative of a general structure-building process that is uniformly invoked when processing long-distance reflexives? This is a general concern about SAT experiments; repeated exposure to the experimental materials over multiple hour-long sessions increases the likelihood that participants detect the crucial experimental manipulation and adapt their processing strategies accordingly while performing the task. It is possible that comprehenders adopted a strategy to consider for rendering quick decisions about the acceptability on *ziji* by performing a superficial check of the two subject positions as if they were words in a list. This

would suffice to perform the task, but would presumably not reflect the actual processes engaged by building an anaphoric dependency. Although the control conditions were designed to disrupt this kind of anticipatory strategy, it is possible that comprehenders nonetheless engaged this sort of (relatively) conscious strategizing to perform the judgment task.

Experiment 7 investigated long-distance and local *ziji* bindings using event-related brain potentials (ERP) to investigate these questions. ERP is an attractive technique to complement the SAT evidence presented here. The SAT evidence provides crucial support to the contention that it is processing speed that differs between local and long-distance *ziji* binding, but the nature of the processes that slow computation for long-distance antecedents remains unclear. The ERP signal can be broken down into a number of distinct waveform components, each of which is thought to index different aspects of linguistic processing. Determining which ERP components index the locality advantage could shed light on the nature of the structured search process.

#### *Experiment 7: ERP Evidence*

The evidence in Experiment 6 shows that the process of reaccessing a long-distance antecedent for *ziji* occurs more slowly than it does for local antecedents. The explanation for this effect on a structured search account is that the local subject is first selectively reaccessed, and then rejected due to a poor fit with *ziji*. The extra processing steps involved in rejecting the local antecedent and accessing the long-

distance antecedent cause a slow-down in processing time when compared to the local antecedent. However, the SAT data do not indicate the exact processes that cause this slow-down. In Experiment 7, I investigate the processing of *ziji* using ERP in order to determine which stages of processing are impacted when a long-distance dependency needs to be constructed. An important feature of ERPs is that they provide a multidimensional view on language processing, with multiple components indexing arguably distinct cognitive processes that support linguistic processing. For this reason, in determining which ERP component indexes the locality advantage in the SAT materials, we may gain further insight into the nature of the processed speed advantage enjoyed by local antecedents relative to long-distance antecedents.

The exact functional significance of many ERP components is a matter of active research and debate, and there appears to be a many-to-many mapping between hypothesized linguistic processes and ERP components. To a first approximation, however, semantic and syntactic processing are characteristically associated with different ERP components. For instance, words that are anomalous with respect to morphological or syntactic features have long been recognized to generate the P600 response, a late posterior positivity that generally peaks around 600 ms post-stimulus (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), as well as an earlier anterior negativity termed the (E)LAN (Kluender & Kutas 1993; Coulson, King, & Kutas, 1998; Friederici et al., 1993; Hagoort, Wassenaar, & Brown, 2003; Lau, Stroud, Plesch, & Phillips, 2006; Neville, Nicol, Barss, Forster, & Garrett, 1991). Although these components are sensitive to similar factors (morphosyntactic wellformedness), they are widely regarded as

reflecting distinct processes in the computation of a syntactic representation (Friederici 1995; Hahne & Friederici 1999; Hagoort 2003; Bornkessel & Schlesewsky 2006). On the other hand, semantic anomalies in otherwise syntactically well-formed sentences typically elicit a central negativity around 400 ms known as the N400 (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008). These characterizations are far from exceptionless, however. For example, there appear to be instances of “semantic” error that engender P600 responses (Kolk, Chwilla, van Herten, & Oor, 2003; Kim & Osterhout, 2005; Kuperberg, 2007), as well as N400 effects that have been linked to processing case anomalies (Hopf, Bayer, Bader, & Meng, 1998) or syntactic reanalysis effects (Bornkessel, Schlesewsky & McElree, 2004).

Interestingly, there is support for the idea that early and late ERP components seem to characterize automatic and controlled processes, respectively (e.g. Hahne & Friederici 1999; Pulvermüller, Shytrov, Hasting & Carlyon 2008). For example, it has been noted that later components such as the P600 are readily modulated by the proportion of ungrammatical sentences in an experimental session (Coulson et al 1998, Hahne & Friederici 1999). This suggests that these responses reflect relatively controlled processes, and that experimental participants modulate the processing routines that underlie this component in order to adapt to the experimental environment. Earlier components, such as the LAN, are apparently robust to this manipulation (Hahne & Friederici 1999). As it is relatively invariant across task manipulation, these earlier responses have been argued to reflect fast, automatic

processes involved in structure-building (Friederici 1995; Hahne & Friederici 1999; Pulvermüller et al 2008).

Here these distinctions are potentially helpful in shedding light on the SAT data presented in Experiment 6. One question that remained from Experiment 6 was whether or not the observed speed delay for long-distance antecedents was related to strategic processing, or if it is best characterized as reflecting early, automatic processes necessary for constructing an anaphoric dependency. On the assumption that the early structure-building components (e.g. the (E)LAN) index more automatic structure-building processes, then we may gain some insight into this question. If we observe a LAN effect at *ziji* when its antecedent is long-distance, then there is a case to be made that the processes that contribute to the SAT slowdown reflect early, automatic processing routines associated with building a structural anaphoric dependency. However, if later components such as the P600 are observed, then this claim is not necessarily supported. A P600 opens the possibility that the shift from local to long-distance antecedents involves more controlled processes, potentially similar to those observed in other syntactically ambiguous structures such as garden-path sentences. A number of authors have noted that P600 effects obtain in situations of grammatical ambiguity (Osterhout, Holcomb & Swinney 1994; Hopf et al., 1998; Friederici, Hahne & Saddy 2002; Hagoort, 2003; Kaan & Swaab, 2003). If the reanalysis suggested by Figure 4.3 reflects reanalysis processes that are more widely engaged in sentence processing, or a conscious, task-specific strategy, then one expects to observe a P600 in long-distance antecedent situations.

Previous ERP work on the processing of *ziji* has demonstrated that long-distance antecedents elicit a posterior positivity (P300/600) relative to local antecedents, supporting the locality bias for *ziji* (Li & Zhou 2010). However, this study differed in important ways from the materials in the SAT study. In Li and Zhou's study, both the local and the long-distance antecedent positions contained animate NPs that were featurally compatible with *ziji*. In their manipulation, the semantics of the embedded verb served to disambiguate the antecedent of *ziji*, as the verb was either inherently reflexive or anti-reflexive (Li & Zhou 2010). Thus it is contextual or thematic knowledge that comprehenders need to draw upon to exclude the local subject from consideration, but the dependency between the local subject and the anaphor is well-formed. In contrast, the dependency between the inanimate local subject and the anaphor in the SAT study is not formally well-formed, which drives a further search for an acceptable antecedent as in Figure 4.3. It is possible that the use of contextual or semantic knowledge to exclude a licit binding dependency in case of a true ambiguity is an entirely different process from the structured search for an antecedent, which presumably occurs prior to the contextual evaluation of the antecedent-anaphor dependency. For this reason it is difficult to straightforwardly conclude that the P300/P600 observed by Li and Zhou is the electrophysiological index of structured search. If the structured access account is correct, then only when the local subject is unacceptable is a more extensive consideration of potential antecedents required. Thus in order to more directly compare the ERP reflex of *ziji*'s locality bias with the SAT results, I present an ERP study of the processing of long-



distance and local *ziji* in environments that were parallel to those tested in the SAT study.

### Participants

Twenty-four college students from Beijing Normal University participated in the experiment, including 13 females (mean age 22). The 24 participants (13 females) had a mean age of 22, were all healthy, native speakers of Mandarin Chinese with no history of neurological disorder, and all were strongly right-handed based on the Edinburgh handedness inventory (Oldfield, 1971). All participants gave informed consent and were paid 50 RMB/hour for their participation, which lasted around 2½ hours, including set-up time.

### Materials

The experimental materials consisted of three conditions designed to investigate the processing of *ziji* (conditions 1-3 in Table 4.3 below). These three conditions matched the three critical *ziji* conditions in Experiment 6 on a number of important dimensions. *Ziji* either took a long-distance antecedent (*LD antecedent* condition; 1), a local antecedent (*local antecedent* condition; 2), or had no antecedent in the sentence (*no antecedent* condition; 3). As noted above, an animate NP in either the main clause or embedded subject NP position can function as a grammatical antecedent for *ziji*. As in Experiment 6, all three conditions consisted of a main clause

that contained a verb of reporting, and an embedded clause, of which *ziji* was the object. Likewise, the binding possibilities for *ziji* were manipulated by manipulating the animacy of the subject NPs: LD antecedent conditions contained an animate matrix subject, and local antecedent contained an animate embedded subject. All other pre-critical NPs were inanimate, and in the no antecedent condition, both subject NPs were inanimate, and therefore this condition did not contain a grammatical antecedent for *ziji*. In order to ensure that the critical *ziji* was not in a sentence final position, all conditions followed *ziji* with a conjunction (e.g. *ye* ‘also’, *que* ‘but’) and second clause continuation. Although the processing profile of conjoined anaphors is known to be different from that of single argument anaphors (see, e.g. Harris, Wexler & Holcomb 2000; Burkhardt 2005), none of the conjunctions in the experimental materials can be used as NP-level conjunctions, ensuring that participants did not interpret *ziji* as part of a conjoined NP.

120 sets of the 3 sentence types were generated. The 360 sentences were equally distributed in 3 presentation lists and subjects were assigned to each list in a Latin Square fashion. Within each list, the 120 targets were interleaved with 240 unrelated fillers of similar complexity, and the list was divided into six blocks of 40 sentences. The order of the 6 blocks was randomized across subjects. Of the 360 sentences in each session, half were considered to be acceptable, half were considered to be unacceptable.

# Condition	Example
1 LD antecedent	厨师表示 [油锅 曾经烫伤了 自己], 所以 辞职了。 Chef say [deep-fryer scalded <i>ziji</i> ], so ( <i>pro</i> ) resigned. “The chef said the deep fryer scalded him, so he resigned.”

2 Local antecedent	医疗报告 表示[厨师 曾经烫伤了 自己], 所以 辞职了。 Medical report say [chef scalded <i>ziji</i> ], so ( <i>pro</i> ) resigned. “The medical report said the chef scalded himself, so he resigned.”
3 No antecedent	*/?医疗报告 表示 [油锅 曾经烫伤了 自己], 所以 辞职了。 */?Medical report say [deep-fryer scalded <i>ziji</i> ], so ( <i>pro</i> ) resigned. */? “The medical report says that the deep fryer scalded <i>self</i> , so he/she/I resigned.”

**Table 4.3:** Summary of conditions in Experiment 2.

### Procedure

Participants were comfortably seated in a dimly lit testing room around 100 cm in front of a computer monitor. Sentences were presented one word at a time in white letters on a black background in 30 pt simplified Chinese characters. Each sentence was preceded by a fixation cross. Participants pressed a button to initiate presentation of the sentence, which began 1000 ms later. Each word appeared on the screen for 400 ms, followed by 200 ms of blank screen. The last word of each sentence was marked with a period, and 1000 ms later a question mark prompt appeared on the screen. Participants were instructed to read the sentences carefully without blinking and to indicate with a button press whether the sentence was an acceptable Mandarin sentence. Each experimental session was preceded by a 12-trial practice session that included both grammatical and ungrammatical sentences. Participants received feedback and were able to ask clarification questions about the task during the practice session. The experimental session was divided into six blocks of 75 sentences each. Breaks were permitted after each block as necessary.

### EEG Recording

EEG was recorded from 32 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, O1/2. Recordings were referenced online to the right mastoid, and re-referenced offline using linked mastoids. An additional electrode was placed on the left and right outer canthus, and above and below the left eye to monitor eye movements and eye blinks. EEG and EOG recordings were amplified and sampled at 1000 Hz using a bandpass filter of 0.1-70 Hz. Impedances were kept below 5 k $\Omega$ .

### EEG Analysis

All analyses were conducted over single trial epochs, consisting of the 100 ms preceding and the 1000 ms following the critical presentation of *ziji*, normalized using a 100 ms pre-stimulus baseline. In order to exclude motion and ocular artifacts, normalized epochs with activity greater than  $\pm 50 \mu\text{V}$  were removed, as were trials that had a peak-to-peak voltage difference of greater than 100  $\mu\text{V}$  in the EOG (Luck 2005). The total rejection rate with these criteria was 19%, ranging between 18%-20% across critical conditions. Averaged waveforms were filtered offline using a 20 Hz low-pass filter for presentation purposes; however, all statistics were performed on unfiltered data. The latency intervals that were analyzed statistically were chosen based on previous conventions in the ERP sentence processing literature: 0-200 ms,

200-400 ms, 400-600 ms, 600-800 ms, and 800-1000 ms, Regions of interest were defined as follows: left anterior (FT7, F3, FC3), midline anterior (FZ, FCZ, CZ), right anterior (F4, FC4, FT8), left posterior (TP7, CP3, P3), midline posterior (CPZ, PZ, OZ), and right posterior (CP4, P4, TP8).

In order to assess the reliability of the effects elicited by the experimental manipulations, I employed linear mixed effects (LME) modeling (Pinheiro & Bates 2000). There are a number of advantages to this approach over traditional approaches. One important advantage in the current context is that this approach readily accommodates missing data (Gelman & Hill 2005; Baayen et al 2008). Since epoch rejection rates tend to vary a good deal across participants in ERP studies, LME models are an attractive analysis option.

Analysis proceeded separately for each time interval, with average values for each epoch within that time interval estimated for the electrodes included in the analysis. The best-fit model for each time interval was selected by hierarchical model comparison, by adding terms for fixed effects and checking for a significant increase in model likelihood using a  $\chi^2$ -test. The logic and interpretation of this approach is qualitatively similar to the hierarchical interpretation of ANOVA results that is often reported in ERP papers: fixed effects (factors) and their interactions are evaluated to ensure that they explain a significant amount of variance, and the resulting best-fit model is then evaluated to determine the nature of these effects. For the present experiment, there were two fixed effects of electrode position, *anteriority* (with the levels *posterior* and *anterior*) and *laterality* (with the levels *left*, *midline*, and *right*), as well as two orthogonal experimental contrasts for binding and locality. Main

effects and interactions for electrode position and order were entered into the models prior to the planned orthogonal contrasts using simple difference coding. The *binding* contrast assessed the effect of having a linguistic antecedent in the sentence, and compared the no antecedent condition to both LD and local antecedent conditions (coding: .5 for the antecedent conditions, -.25 for both LD and local antecedent conditions). The *locality* contrast assessed the effect of having a local antecedent, and directly compared local and LD antecedent conditions (coding: .5 for LD antecedents, -.5 for local antecedents), 0 for no antecedent). Random intercepts for trial and participants were included; random intercepts for items and random slopes for the experimental fixed effects did not significantly increase the likelihood for any model, and so were not included in the final model fits described here. This exclusion did not change the pattern of results found. For each time interval, we describe the best-fitting model and the fixed effect coefficients. All *p*-values for linear model coefficients were estimated using MCMC methods as implemented in the *LanguageR* R package (Baayen 2008), with  $n=10,000$  samples.

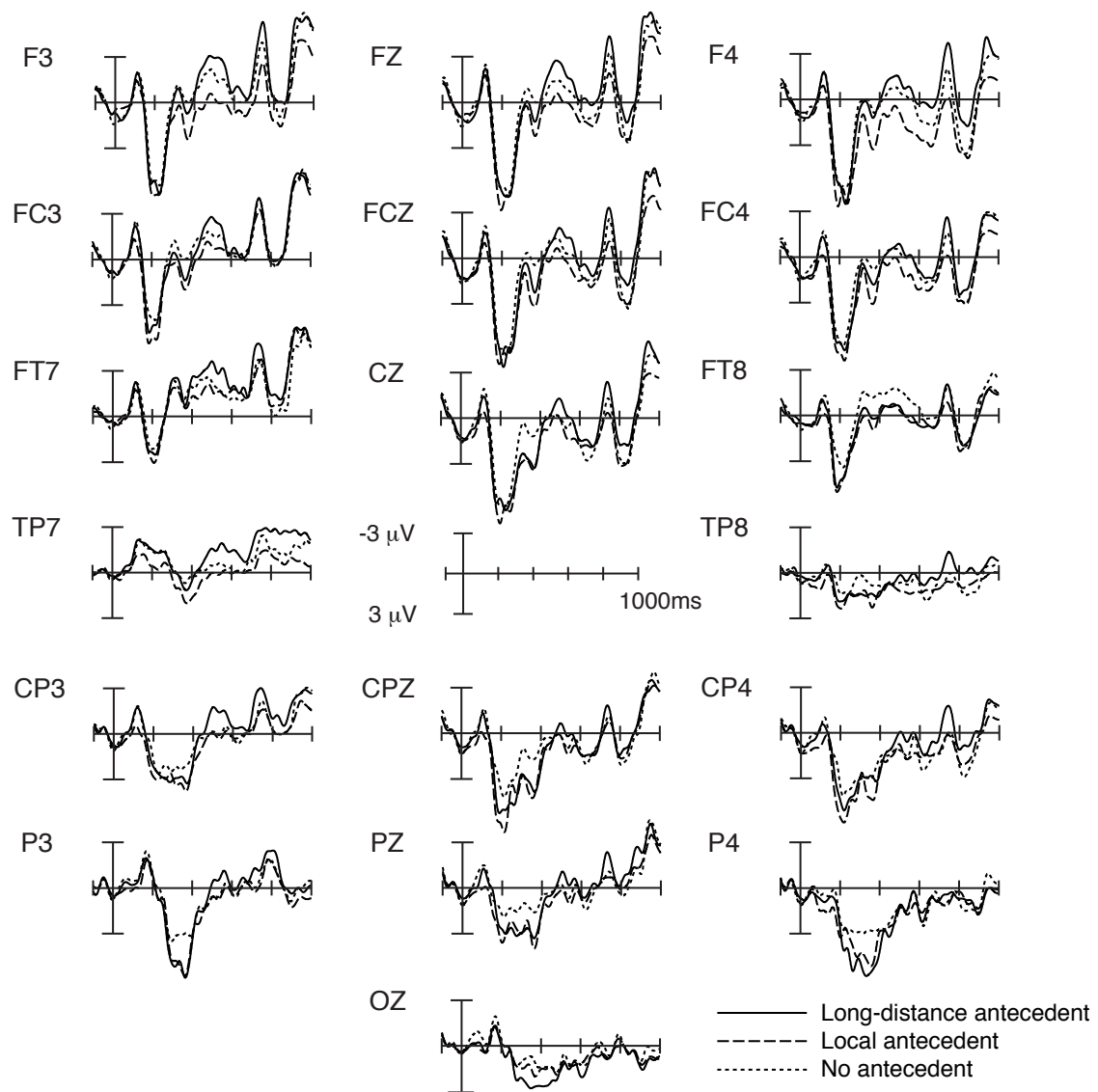
### Results: Behavioral Data

The average rate of acceptance was 81% for the long-distance binding condition, 73% for the local binding condition, and 51% for the no-binding condition. A logistic linear mixed effects model with crossed random intercepts for participants and items revealed significant effects of both the *binding* contrast ( $\beta = -.81$ ,  $z = -14.0$ ,

$p < .0001$ ) as well as the *locality* contrast ( $\beta = .47, z = 4.2, p < .0001$ ). An analysis of reaction times did not reveal any significant differences among the conditions.

#### Results: ERP Data

The grand average ERPs are presented in Figure 4.12. A summary of the fixed effects for *binding* and *locality* are presented in Table 4.4. In the 0-200ms time interval, there were no significant fixed effects for experimental contrasts. In the 200-400ms time interval, the best-fitting model included an interaction of *binding* with *anteriority* and *laterality*. Resolving this interaction revealed a broadly distributed negativity for unbound *ziji*, relative to the two bound *ziji* conditions. Numerically, this effect was largest over posterior regions. In the 400-600ms time window, analysis revealed a negativity for the LD antecedent relative to the local antecedent binding conditions, which reached significance in the mid and left anterior ROIs. A negativity was also observed in the 600-800ms time window, with a focus over the mid and right anterior ROI. No experimental fixed effects were observed between 800-1000ms.



**Figure 4.12:** Grand average ERPs for *ziji* conditions in Experiment 7, low-pass filtered for visual presentation. Only electrodes included in the ROI analyses are presented.



	0-200ms	200-400ms	400-600ms	600-800ms	800-1000ms
<b>Binding</b>					
<i>Left Ant.</i>	-	-.64 ± .44	-	-	-
<i>Mid Ant.</i>	-	-1.18 ± .44**	-	-	-
<i>Right Ant.</i>	-	-.91 ± .44*	-	-	-
<i>Left Post.</i>	-	-.79 ± .44†	-	-	-
<i>Mid Post.</i>	-	-1.22 ± .44**	-	-	-
<i>Right Post.</i>	-	-1.07 ± .44*	-	-	-
<b>Locality</b>					
<i>Left Ant</i>	-	-	-1.02 ± .43*	-.73 ± .47	-
<i>Mid Ant.</i>	-	-	-.91 ± .43*	-.90 ± .47†	-
<i>Right Ant.</i>	-	-	-.75 ± .43†	-.99 ± .47*	-
<i>Left Post.</i>	-	-	-.84 ± .43†	-.72 ± .47	-
<i>Mid Post.</i>	-	-	-.18 ± .43	-.13 ± .47	-
<i>Right Post.</i>	-	-	-.31 ± .43	-.48 ± .47	-

**Table 4.4:** Table of experimental fixed effects (coefficients in  $\mu\text{V}$ , with standard error). Experimental fixed effects only shown if the best-fit model included a significant interaction of experimental effect with *anteriority* and *laterality*. † =  $p < 0.1$ , \* =  $p < 0.05$ , \*\* =  $p < 0.01$ .

### Discussion

The ERP results revealed two distinct components associated with the processing of *ziji* in our materials. For the *binding* contrast, an early negativity associated with the processing of unbound *ziji* was observed in the 200-400ms time window. This negativity had a primarily central-posterior distribution. This is consistent with an N400 effect, though visual inspection of the ERPs suggests that the peak of this negativity is slightly earlier than the canonical N400 effect. For the *locality* contrast, a qualitatively different component was observed, differing both in time course and distribution. LD antecedents caused an increased negativity over the

anterior ROIs in the 400-600ms window, with some effects also being observed in the subsequent 600-800ms window. The distribution of this component is consistent with a (L)AN effect, although the time course is somewhat later than the canonical LAN effect.

One important insight from these results is that the processes involved in recovering a long-distance bound interpretation and revising to an indexical interpretation of *ziji* are at least partially distinct. The exact interpretation of this effect depends on assumptions about the identity of the component, and that component's functional significance. For example, if this component is best characterized as an N400, then it might be understood as indexing the difficulty in accessing information in the lexicon (e.g. Lau et al 2010). On this interpretation, this effect suggests that *ziji* is less predicted in the no antecedent conditions. Another important distinction between indexical and bound *ziji* is that in order to recover an indexical interpretation, participants must change the point-of-view of the sentences (Huang & Liu 2001; Anand 2006). If this is necessary when encountering *ziji* in the context of a discourse with no animate participants, then the negativity may reflect this process of context-shifting.

The difficulty associated with long-distance interpretations of *ziji* was observed in a (L)AN component at approximately 400ms post-stimulus.. Anterior negativities of this sort have been linked to a variety of morphosyntactic violations as well as working memory difficulty associated with forming long-distance dependencies (Neville et al 1991; Friederici et al., 1993; Kluender & Kutas, 1993; King & Kutas, 1995; Coulson et al, 1998; Hagoort et al 2003). If structured search

proceeds as in Figure 4.3, then the observed anterior negativity is compatible with either interpretation of the (L)AN component. On the assumption that the LAN indexes a morphological violation, then it might be caused by the mismatch between the animacy features of the local subject and *ziji* when the local subject is reaccessed and considered. Alternatively, if the observed negativity reflects working memory difficulty (as argued by Kluender & Kutas 1993), then it may be indexing the fact that more retrieval operations are necessary to recover the long-distance antecedent (as in Figure 4.3), leading to a greater anterior negativity. Both of these interpretations are compatible with a view of structured search as an early process engaged during the construction of anaphoric dependencies, but further work is necessary to distinguish between them.

Interestingly, recovering a long-distance interpretation did not cause ERP components that are commonly observed during syntactic or semantic reanalysis. Reanalysis that is associated with syntactic ambiguity or garden-path type sentences is often reflected in a posterior P600 component (Osterhout et al 1994; Friederici et al 2002; Hagoort, 2003; Hopf et al 2003; Kaan & Swaab, 2003), although it has been noted that in some cases the N400 may reflect this sort of reanalysis processes (Bornkessel et al 2006). Neither component was observed in the processing of long-distance antecedents. This supports the contention that the relative processing disadvantage shown by long-distance interpretations of *ziji* in Experiment 6 is not due to the same reanalysis processes that are engaged by processing garden-path type sentences. It is important to note that these results stand in contrast to those reported by Li & Zhou (2010), who did report that long-distance interpretations of *ziji* caused

both P300 and P600 effects. However, in their study, the local subjects were always animate and compatible with *ziji*; long-distance interpretations of *ziji* were forced by manipulating the verb semantics. In this case, if retrieval of the local subject yielded a feature-matched antecedent, it is unclear what processing would be necessary to reanalyze to a long-distance interpretation. If structured search terminates upon finding a compatible antecedent, then it is possible that entirely different reanalysis processes were required to recover the distant antecedent in that study, causing a P300/600 rather than an anterior negativity. If this reasoning is correct, then this account generates testable predictions for future work about when the different ERP components will be observed in processing *ziji*. In particular, the difficulty associated with the structured search necessary to find an initial, acceptable antecedent for *ziji* should be reflected in a LAN. On the other hand, difficulty associated with selecting among multiple antecedents based on contextual information should be associated with P600 effects.

If the distinction between early, automatic processes and later controlled processes is correct (Hahne & Friederici 1999), then an interesting claim about *ziji* can be made. The finding that structured search causes a LAN effect is compatible with the view that the (relatively difficult) process of recovering a long-distance antecedent for *ziji* is a fast, automatic process engaged by encountering a bound instance of *ziji*. This is consistent with the contention that the SAT dynamics effects seen in Experiment 6 were not due to task-specific strategies that participants adopted over the course of the SAT experiment. On the strongest interpretation of this result,

this pattern of results suggests that the structured search for an antecedent reflects the earliest structure-building procedures engaged upon encountering *ziji*.

Overall, the ERP effects provide an alternative source of evidence in support of structured search by confirming that fully grammatical (and preferred) long-distance interpretations of *ziji* are associated with extra processing difficulty. The time course of the difficulty associated the LD antecedent conditions is consistent with structured search reflecting an early, automatic process, rather than later, controlled processes that are sometimes associated with syntactic reanalysis (Osterhout et al 1994; Hahne & Friederici 1999; Friederici et al., 2002). The (L)AN associated with the long-distance interpretation of *ziji* may reflect either the feature mismatch with the local subject, or the extra memory retrievals necessary to reaccess the distant antecedent. Interestingly, the response that indexes the shift to an indexical interpretation of *ziji* temporally preceded the difficulty associated with long-distance interpretations. Furthermore, since no LAN was seen in this condition, it is possible that the decision to pursue an indexical interpretation of *ziji* obviated the need for the parser to check the local subject position at all. This may indicate that participants quickly decide to pursue a shift in point-of-view rather than attempting to bind *ziji* to an antecedent within the sentence. The nature of this decision remains unclear, but I tentatively suggest that the lack of any animate discourse entities may have biased readers in favor of a first-person interpretation of *ziji*. Thus there is a quick use of top-down discourse information about which parse to pursue, but once the decision to pursue a linguistically bound interpretation has been engaged, the parser engages a structured search. As in Experiment 6, however, participants varied a great deal in the

degree to which they accepted the no antecedent *ziji* condition. Further study that more carefully controls the perceived point-of-view of the sentences is necessary to get a clearer picture of the processing of unbound *ziji*.

### *General Discussion*

The current experiments investigated the processing of antecedent-anaphor dependencies involving the Mandarin Chinese long-distance reflexive *ziji* using both time-course analysis and electrophysiological measures. The SAT results showed that in constructing a *ziji*-antecedent dependency, local antecedents are accessed more rapidly than long-distance antecedents. This locality processing advantage was reflected in both the rate and speed measures in SAT, and this difference in dynamics was absent from control conditions without *ziji*. These results stand in contrast to previous SAT studies, which suggested that processing advantages due to locality only impacted asymptotic accuracy (i.e., the probability of successfully computing the dependency; McElree 2000; McElree et al 2003; Martin & McElree 2008).

The ERP results confirmed the locality advantage for *ziji* antecedents. Long-distance interpretations of *ziji* elicited a (L)AN component relative to local interpretations of *ziji* rather than a P600 component, a finding that suggests that the locality advantage seen in the SAT experiment is not due to strategies employed in the context of a long SAT experiment, but rather is due to automatic structure-building processes required to form an antecedent-anaphor dependency.

Both experiments support the hypothesis that comprehenders initially consider a feature-inappropriate, but structurally accessible antecedent when pursuing a bound interpretation of *ziji*. This is a direct prediction of structured access in sentence comprehension: if access is guided by structural position, rather than feature-content, then participants have no choice but to initially consider the local subject. On this model, there is no information available to the parser during the course of constructing the binding dependency to exclude this position from consideration. This informational bottleneck forces participants to engage a structured search when there are multiple structurally licit positions to consider for forming a dependency. If comprehenders only employ structural information to retrieve anaphoric antecedents, then they have a natural mechanism for limiting the dependency to structurally appropriate antecedents only. As discussed above, such a mechanism is agnostic with respect to certain architectural commitments. In particular, it is not obviously at odds with other time-course results in retrospective dependencies that suggest a role for content-addressable memory architectures in sentence processing (e.g. Martin & McElree, 2008). The structured search process may reflect iterated retrievals from an underlyingly content-addressable architecture (as in the model in Chapter 3), provided that the processor has access to cues that can effectively distinguish the local subject from other positions in the sentence. By iterating retrievals with varying sets of retrieval cues, a content-addressable architecture can implement this sort of structured search when the need to faithfully implement structural constraints outweighs the need for speed in dependency formation. The relevant difference between the *ziji* dependency and the VP-ellipsis dependencies examined by Martin

and McElree (2008, 2009), on this viewpoint, is that only for *ziji* is the parser willing to suffer a slower retrieval procedure in the service of structural accuracy.

### *Locality bias in ziji dependencies*

The time course evidence presented here is compatible with theories of dependency construction that invoke notions of structured search. Furthermore, it suggests that at the point of initiating the antecedent-anaphor dependency, comprehenders initially access and check the local subject position. This is not consistent with the view that all *ziji*-antecedent dependents are considered in parallel, which would predict constant SAT dynamics for all antecedent positions. An interesting finding is that participants appear to prefer to access the local subject position upon encountering *ziji*. This does not necessarily follow from the formulation of structured search given above, which only holds that multiple syntactically licit positions need to be checked in an effectively serial order. If comprehenders preferred to access the highest subject in a sentence, and search downwards, this would be entirely compatible with structured search as I have presented it. This leads to an interesting question of why the local subject is preferred in this case.

One possibility is that the locality bias reflects a general advantage for linguistic material contained within the local clause relative to material outside the local clause. In other words, the local subject may be available more quickly by virtue of its being contained within the local clause, which is still in the process of being parsed when *ziji* is encountered. This account is consistent with studies on sentence



recall that suggest that the local clause has a privileged role in online sentence processing (Jarvella, 1971; Jarvella & Pisoni, 1970). This account does not predict any necessary preferred order of access for subject positions outside of the local clause. This may also be equivalent to the claim that that some elements remain concurrently available in the focus of attention while others are displaced and must be later retrieved (McElree, 2006; Jonides, Lewis, Nee, Lustig, Berman & Moore, 2008). If the local subject remains in the focus of attention, while the long-distance subject requires retrieval, then the observed distinction in dynamics would be predicted. This interpretation seems less likely, in light of findings from list memory experiments that indicate that the focus of attention is extremely limited in size and scope, corresponding to just one task-relevant encoding (McElree & Doshier, 1989; McElree, 1998). If only one element occupies focal attention before *ziji* is processed, it is likely to be the verb, in anticipation of the upcoming object NP. However, the available data on the contents of the focus of attention is limited for connected linguistic representations, which have considerably richer structure than do word lists. It is known that full clauses are sufficient to displace information about their embedding environment (McElree et al., 2003; Wagers & McElree, 2009). In the context of the present experiment, the temporal adjunct clause that intervened between the subject and the verb seems likely to have pushed the local subject out of the focus of attention. However, it is presently unknown whether information about the subject is later restored and thus focally available during the initial processing of verb-phrase internal arguments.

Alternatively, the locality advantage may reflect a strict ordering of access, where subject positions are accessed in progressively dominating positions. This ordering could reflect a useful strategy for processing *ziji* in particular, due to the blocking constraints on *ziji*. Consider (4.4) and (4.5), repeated as (4.7) and (4.8) below:

- (4.7)      *Zhangsan<sub>i</sub>      renwei ni<sub>j</sub>      hen      ziji<sub>\*i/j</sub>*  
               Zhangsan      think    you      hate    *ziji*  
               “Zhangsan thinks that you hate yourself/\*him.”
- (4.8)      *Ni<sub>i</sub>      renwei Zhangsan<sub>j</sub>      hen      ziji<sub>?i/j</sub>*  
               You      think    Zhangsan      hate    *ziji*  
               “You think that Zhangsan hates himself/?you.”

These sentences display the blocking effect for *ziji*. The important generalization about these effects is that *ziji* cannot access subjects that dominate first- or second-person subjects (Huang & Liu 2001); similarly, it has been reported that *ziji* cannot be bound by singular subjects that dominate plural subjects (Tang 1989; Huang & Tang 1991). If comprehenders consider antecedent positions from the most local to the most distant in a strict order, then this constraint can be easily implemented by terminating search upon reaching an indexical or plural subject. This intuition is reflected in grammatical accounts of *ziji* that invoke cyclic movement of *ziji* to progressively higher subject positions in the derivation of an antecedent-anaphor chain (Pica 1986; Cole, Hermon & Sung 1990; Cole & Sung 1994; Cole, Hermon & Lee 2001). This account very naturally explains blocking effects. On these accounts, the feature match between *ziji* and a subject is evaluated at each subject position, from the most local to the most distant.

Both of the retrieval-based accounts presented above share a common feature in that they require the use of positional or structural information in retrieval, and thus support the existence of structured access mechanisms. Note that this requirement does not hold if the local subject is maintained in focal attention. In order to implement a search that serially samples subject positions, or that preferentially accesses information in the local clause, the positional information inherent in those two specifications needs to be available to guide memory access. It is possible, in principle, to empirically distinguish the two accounts. The fully serial, bottom-to-top search and the locality bias account reflects a make distinct predictions about the time course of activating long-distance subjects that are two versus three clauses distant from *ziji*, as in (4.9).

- (4.9)            *Lisi<sub>i</sub>   shuo   Xiaoming<sub>j</sub>   renwei fengbao   hai-le*  
                     *ziji<sub>i/j/k</sub>*  
                     Lisi   says   Xiaoming   thinks storm   hate   *ziji*  
                     “Lisi says Xiaoming thinks the storm harmed himself/her.”

On either account, the inappropriate *fengbao* should be considered before either *Lisi* or *Xiaoming*. If the locality bias observed here reflects a general advantage for the local clause, then participants should be equally likely to consider *Lisi* or *Xiaoming* in the second stage of access. In SAT, this would mean constant access dynamics for non-local position, reflecting the fact that on some trials *Lisi* is the first non-local antecedent considered, and on the rest, *Xiaoming* is considered. If the locality bias indicates a strategy whereby comprehenders progressively consider more distant subject positions, then SAT dynamics should show that the intermediate subject

position is reliably accessed before the highest subject position. Future work will examine the role of hierarchical distance beyond the local clause boundary in an attempt to tease apart these competing hypotheses.

### *Alternative accounts of the data*

An alternative account of the current results is that the difference in processing dynamics between dependencies with local and long-distance antecedents reflects a reanalysis from a local interpretation to the long-distance interpretation. This is consistent with some linguistic accounts of *ziji* that have suggested that *ziji* as a local anaphor is lexically distinct from *ziji* when its antecedent is distant, based on differences in meaning and pragmatics of usage in these different environments (e.g., Huang & Liu, 2001; Anand, 2008). Reanalysis has been noted to cause delays in SAT dynamics parameters (McElree, 1993; Bornkessel et al 2004), and thus if the first option that comprehenders attempt upon recognizing *ziji* is the local-antecedent interpretation, they should fail and require lexical reanalysis in the long-distance antecedent conditions.

Note, however, that the LAN we observed in the ERP record is not associated with general syntactic reanalysis. Instead, the ERP component that is commonly associated with reanalysis processes is the P600 (Friederici et al., 2002; Hagoort, 2003; Hopf et al., 2003; Kaan & Swaab, 2003), though it has been argued that certain subcases of reanalysis (specifically, reanalysis related to case marking) engender N400 effects (Hopf et al 1998; Bornkessel et al 2004). On the assumption that local

and long-distance interpretations of *ziji* required distinct lexical items and structural analyses, then it can be argued that P600 effects should be observed. Thus given relatively well-accepted assumptions about the functional significance of common language-related ERP components, it can be argued that the processing slowdown related to long-distance interpretations of *ziji* does not reflect general syntactic reanalysis mechanisms.

As mentioned above, however, it is unclear whether an account that invokes reanalysis is an entirely distinct alternative to accounts that invoke structured search. In particular, the notion of structured search that we have suggested posits that in order to recognize that a local antecedent for *ziji* is inappropriate in a sentence with a long-distance antecedent for *ziji*, the local subject position must first be retrieved and rejected due to its unacceptability as an antecedent for *ziji* (due to either morphosyntactic constraints, or possible discourse-level interpretive constraints). It is not unreasonable to call this intermediate step in a structured search procedure a step of ‘reanalysis.’ The main conceptual difference between a reanalysis account and a structured search account is that in the latter, the local subject is initially targeted for access using targeted cues that single out the local subject. This need not be so on a reanalysis-style account of these results, which instead would posit that the local subject is initially retrieved because more local phrasal material is more active in the parse. However, this local activation advantage would need to be so great as to cause access of the local subject during retrieval despite the fact that the long-distance subject has a number of semantic features (e.g., animacy, sentience, being a source of communication) that are known to invite anaphoric co-reference (Kaiser, Runner,

Sussman, & Tanenhaus, 2009). In contrast, the local subject contains no semantic cues to support retrieval. None of the nouns used in this position supported metonymic interpretations that could license *ziji*, such as *the corporation* being used to refer to the corporation's employees. Such an account seems unlikely in light of the fact that the local subject is not always reliably retrieved in other instances where it needs to be retrieved, such as English subject-verb agreement. In computing subject-verb agreement, an inaccessible feature-match causes incorrect access during online parsing, showing that the feature match can in fact overcome any bias for the local subject. The data presented here thus seem to favor structured access strategies over reanalysis caused by general memory dynamics and a preference for material contained in the local clause.

A related alternative account might be formulated in terms of the ACT-R model in Chapter 3, which posits retrieval time varies as a function of the fit to the search cues. It may be the case that it is simply more difficult to retrieve the more distant antecedent, but that the intermediate antecedent is not in fact considered in the course of recovering the distant antecedent. However, this alternative explanation requires the search cues to be limited to structural cues only. Otherwise, the semantic cues contributed by the distant antecedent would potentially make it a much better fit with the retrieval engaged by *ziji*. This would predict that the long-distance feature-matched antecedent of *ziji* is retrieved more quickly than the local feature-mismatched noun phrase. But if only structural search cues are used, it is unclear on what grounds comprehenders retrieve the distant subject over the local subject. It seems that on a structured access account with serial retrieval of memories, there is a

significant portion of trials for which the local subject will be retrieved, leading to a situation where structured search is necessary to recover the distant interpretation.

### *Linguistic and discourse antecedents for ziji*

One consistent finding from both Experiments 6 and 7 is that sentences that do not contain an explicit linguistic antecedent for *ziji* are not consistently rejected as unacceptable by speakers. This is not entirely surprising, as it is possible to interpret *ziji* as a first-person indexical that refers to the speaker of an utterance. It is not clear what relation the process of arriving at this egocentric interpretation bears to the process of finding a linguistic antecedent in memory. In Experiment 7, the no antecedent condition elicited an apparent N400 relative to both local and LD antecedent conditions. Interestingly, the no antecedent condition appeared to pattern in between LD and local antecedent binding conditions with regards to the anterior negativity observed. If the time course implied by the ERP components reflects the order of operations, then it suggests that the egocentric interpretation is not an ‘elsewhere’ interpretation that is adopted only after an exhaustive search of the parse fails to return a licit antecedent. Instead, it would suggest that participants engage a quick decision to pursue either a bound or indexical interpretation of *ziji* before any antecedent reactivation occurs. Interestingly, the offline judgment data suggests that this interpretation is not as readily accepted as interpretations with explicit linguistic antecedents, which may suggest some difficulty in recovering this interpretation. This apparent difficulty may be an artifact of the tasks used in the present set of studies: it

is difficult to know what perspective participants adopt when interpreting out-of-context sentences in a laboratory setting. Further work that explicitly manipulates the perceived perspective of the sentences may be useful in determining the source of this difficulty in the present tasks.

If it is the case that the egocentric interpretation is considered before an exhaustive search of the parse, then this raises the question of what information allows comprehenders to decide between linguistically binding *ziji* and adopting an egocentric interpretation. One possibility is that top-down information about the discourse model is applied to disambiguate early: in the case where there are no sentient entities in the discourse, the egocentric interpretation is the only licit option. In other cases of structural ambiguity, it has been argued that comprehenders use the number of discourse entities in a heuristic fashion to decide which parse to pursue (van Berkum, Brown & Hagoort 1999b). Thus, in a context with two girls, the ambiguous input *John told the girl that...*, comprehenders quickly pursue a relative clause interpretation of *that*. In a context with only one girl, however, *that* is preferentially parsed as introducing a sentential complement to the verb *told*. A similar discourse heuristic might be used to disambiguate which interpretation of *ziji* is to be pursued. If the discourse contains appropriate sentient entities that could potentially antecede *ziji*, then comprehenders deploy structured search to check that those discourse entities stand in an appropriate structural relation to the anaphor. A discourse-based heuristic for making a decision about how to parse *ziji* generates testable predictions. If the discourse contains a sentient entity that is linguistically introduced in a syntactically inaccessible position, then access to egocentric



interpretations of *ziji* should be inhibited relative to sentences that introduce no sentient actors to the discourse model. In the case of an inaccessible antecedent, comprehenders may be tempted into considering a bound interpretation for *ziji* that is not globally correct.

At this point this suggestion remains speculative, as the present study does not provide evidence to support this notion. The relationship between linguistically and non-linguistically bound *ziji* in online processing routines remains an important unresolved question, and I leave this to future research.

### *Conclusion*

In this chapter I examined the time-course of antecedent-anaphor dependency construction using the Mandarin Chinese long-distance anaphor *ziji*. It was found that local antecedents are accessed more rapidly than long-distance antecedents, suggesting that the information necessary to complete local antecedent dependencies is present before the information needed for long-distance antecedent dependencies. This finding was supported by converging evidence from ERPs, which showed that long-distance bindings of *ziji* elicited a LAN component relative to local antecedent bindings.

These findings are compatible with several implementations, but the crucial features that all accounts share is that the local subject is reactivated using primarily structural cues, without regard to the semantic fit of their content. This satisfies an important prediction of the hypothesis of structured access: as in English reflexives in

Chapters 2 and 3, it is structural position, rather than morphological or semantic compatibility with the anaphor, that guides memory access. But while the argument from interference showed that structured access provided a more restricted initial candidate set, the evidence presented here demonstrated the complementary prediction of structured access. For long-distance reflexives like Mandarin *ziji*, structured access actually leads to an inappropriately large initial candidate space relative to a feature-based access mechanism. In the SAT and ERP experiments presented here, I showed that this larger potential candidate space causes measurable delays in processing speed, suggesting that comprehenders serially check potential antecedent positions.

## **Chapter 5: Structured access across dependency types**

The preceding chapters have argued that it is syntactic position, rather than morphological or semantic feature content, that guides memory access in both English argument reflexives and Chinese long-distance reflexives. For both of these dependencies, the hypothesis of structured access appears to accurately describe the manner in which they are parsed. There remain a number of important questions about structured access in sentence comprehension, however. In particular, it does not appear to be the only manner in which information is retrieved from memory in online parsing: agreement dependencies have been repeatedly shown to use morphological features in a direct-access fashion (Clifton et al 1999; Pearlmutter et al 1999; Wagers et al 2009; Chapter 2 of the present thesis). The selective deployment of structured access mechanisms for constructing long-distance dependencies leads to an important theoretical question: what is the role of structured access in comprehension, and how generally is such a mechanism used in online processing?

This chapter presents a first attempt at addressing this question. There is one straightforward hypothesis that suggests itself based on the contrast between agreement and reflexives. Namely, it may simply be the case that anaphoric dependencies are constructed at a different level of representation than are agreement dependencies. For example, agreement might be computed over the syntactic parse, but anaphoric relations might instead be stated as interpretive rules over logical form representations (Jackendoff 1972; Wasow 1972; Chomsky & Lasnik 1993; Fiengo & May 1994). If structured access is simply a property of referential dependencies, then structured access should hold equally well for all anaphor-antecedent dependencies. In order to test this hypothesis, in this chapter I contrast the processing of *ziji* with the intensified pronoun *ta-ziji* that has related but distinct licensing properties from *ziji* (Pan 1998, 2000; Huang 2000; Bergeton 2007). Self-paced reading evidence confirms that *ziji* initially accesses the local subject position before considering other subject positions. However, in a closely matched experiment, *ta-ziji* shows a qualitatively different processing profile in identical contexts, suggesting immediate direct access to all licit antecedent positions in parallel.

The contrast in behavior between *ziji* and *ta-ziji* provides important insight into the role of structured access in comprehension. It demonstrates that structured search is not deployed for all anaphoric dependencies: *ziji* and *ta-ziji* both participate in anaphoric dependencies, and both are presumably of similar relevance to the task of computing the message intended by any given utterance. Nonetheless, they display qualitatively different access profiles, casting doubt on the hypothesis that all interpreted dependencies use structured access mechanisms. The processing profile

associated with structured memory access is not a general property of interpretive procedures (as opposed to morphological or syntactic dependencies).

These findings suggest that interpretive content is not a sufficient condition for the parser to engage structured search mechanisms. In the last part of this chapter, I consider the possibility that interpreted dependencies that are subject to syntactic licensing conditions are likely to engage structured search. If this is so, then structured search may be understood as an optimal strategy for satisfaction of grammatical constraints in online parsing processes. If the goal of parsing is to reconstruct the intended meaning of the input string as quickly and accurately as possible, then a more narrowly syntactic access mechanism may convey advantages in speed or accuracy when a) a dependency requires a constrained search of the parse space, and b) when failure to build the correct dependency would result in misunderstanding.

If this line of reasoning is correct, then structured access is an adaptive strategy to minimize interference and maximize interpretive fidelity in a noisy memory environment. By narrowing retrieval cues to the minimal distinctive set necessary to accomplish the task at hand, structured access minimizes interference. The selective deployment of this strategy reflects the fact that structured access can be costly in terms of processing time, as seen in Chapter 4, and so is only worth engaging when syntactically constrained dependencies contribute to meaning. Even this restricted characterization of the mechanism predicts a wide application of this mechanism, however: general thematic integration operations, most syntactic attachment decisions, and bound anaphor dependencies (among others) are all

syntactic operations that have direct interpretive reflexes as well as syntactic licensing conditions. Across the range of these dependencies, structured access should be apparent.

The ‘optimal strategy’ argument for abstract structural retrieval cues is similar to arguments in Mitchell and colleagues’ *Tuning Hypothesis* (Mitchell, Cuetos, Corley & Brysbaert 1995). On both accounts, the pressure to state parsing procedures at an abstract structural level is motivated by functional pressures, either the need to derive robust structural frequency estimates (Mitchell et al 1995) or to minimize interference from distracting elements in memory (on the present account).

#### *A puzzle for the hypothesis of structured access in comprehension*

In Chapters 2-4, it was seen that syntactic information provides the primary means of accessing memory for certain dependencies. This argument rests on the observation that English reflexives and Mandarin long-distance reflexives are able to initiate accurate retrieval of the local subject of their clause, apparently without interference from other feature-matched noun phrases in the parse tree, whether or not they occupy a licit antecedent position. The crucial conclusion these studies license is that the parser can in principle accurately target and retrieve particular syntactic positions, at least for the dependencies under consideration.

However, this conclusion poses a puzzling question in light of the results in Chapter 2: if it is the case that the parser can in principle accurately retrieve and check the local subject, why don’t comprehenders do this when checking subject-verb

agreement in English? If comprehenders can accurately index the local subject for retrieval during sentence processing, then agreement processing could plausibly proceed by accurately retrieving the local subject and then verifying that its features match the features of the verb. However, this does not appear to be the strategy that comprehenders pursue: subject-verb agreement in English reliably shows interference effects, which suggests that morphological features are used to access memory in a content-addressable fashion. Why is it that comprehenders engage different access strategies for the two dependencies?

One intuitive possibility is that since reflexives participate in a semantically interpreted dependency, they initiate a retrieval of the local subject in a qualitatively different manner. It may simply be the case that reflexive dependencies and agreement dependencies are constructed at different levels of representation, as would be expected if reflexive dependencies are best understood as constraints on interpretation, rather than structure-building processes (Jackendoff 1972; Wasow 1972; Chomsky & Lasnik 1993). If one makes a secondary assumption that the parser constructs somewhat shallower or less structured parses before shunting constituents off to more structured stores for interpretation (Frazier & Fodor 1978), then this prediction could be easily accommodated. Fallibility may simply reflect the memory access mechanisms that are deployed to retrieve information from uninterpreted, temporary syntactic stores. In contrast, structured access mechanisms may be the preferred manner of access for structure built at a more compact or global level of representation. A number of well-studied parsing models have an architecture in which material is associated in a less structured manner before being passed to a more

compact, global, or structured representation (see, e.g., the two-stage model in Frazier & Fodor 1978 or Townsend & Bever 2001). For any parsing model that posits separate levels of representation that differ on the degree to which linguistic information is structured in memory, then the distinction between agreement and reflexives might simply reduce to differences in the access strategies employed by different representational levels in the parse.

This account appears to capture a straightforward intuition behind the agreement and reflexive dependency contrast: reflexives mean something, and agreement generally does not, at least in English. To evaluate this intuition, consider the range of grammatical dependencies that have been shown to be susceptible to facilitatory interference, briefly summarized in Table 5.1. If the reasoning in Chapter 3 is correct, then those dependencies in 5.1 that do not show facilitatory interference are those that are constructed with structured access mechanisms. The range of dependencies that have been considered remains somewhat sparse; conclusions drawn from such a narrow sample should be approached with caution.

	<b>Facilitatory interference</b>
English subject-verb agreement	✗
Spanish subject-verb agreement	✗
English reflexives	✓
Mandarin Chinese reflexives	✓
German, English NPIs	✗

**Table 5.1:** Summary of interference properties of long-distance dependencies.



Subject-verb agreement in English provides perhaps the clearest example of a dependency that is clearly prone to grammatical illusions due to feature-matched, but inaccessible material. Similar comprehension results have been found in Spanish (Alcocer & Phillips 2009; Lago, Alcocer & Phillips 2011). Another example of a dependency that is prone to grammatical illusions is NPI dependencies, both in English (Xiang, Dillon & Phillips 2006, 2009), and in German (Drenhaus et al 2005; Vasishth et al 2008). This appears to run counter to the generalization that structured access holds of interpreted or semantic dependencies. However, it is not clear that this represents the same phenomenon as agreement interference. For instance, Xiang and colleagues (2009) argued that the NPI effect was not due to a feature overlap with an inaccessible negative element. Instead, they argued that the best account of these data was one that invoked the generation of spurious pragmatic inferences due to the interaction of embedded negation and the semantics of the restrictive relative clause. Additional support for this view comes from a number of subsequent studies. It has been shown that NPI interference is negatively correlated with an individual's pragmatic reasoning skills (as measured by the Autistic Quotient; Baron-Cohen, Wheelwright, Skinner, Martin & Clubley 2001), which supports the claim that 'overactive' pragmatics is the source of this effect. For individuals who have impaired pragmatic abilities, as indexed by large AQ scores, the NPI interference effect is diminished, while the agreement attraction effect is unaffected (Xiang, Grove & Giannakidou 2011). Additionally, self-paced reading evidence presented by Parker & Phillips (2011) suggests that negation embedded in complement clauses does not cause interference in the same manner as negation embedded inside restrictive

relative clauses, which has formed all of the evidence for NPI interference to date. If complement clauses change the pragmatics involved with these sentences without drastically changing the structure, then the overactive pragmatics account of NPIs is supported.

If it is true that the NPI illusion effect is generated by an altogether separate mechanism, then agreement provides the only clear and reliable case of a long-distance dependency that is structurally fallible online. It appears that the range of evidence for feature-based access is rather limited.

Although the evidence for feature-based, grammatically fallible memory access seems restricted, it must be acknowledged that the evidence for structured access is equally narrow. Just as agreement provides the only clear case of fallible, feature-based direct access dependency construction, reflexive dependencies in English and Chinese as investigated here provide the clearest examples of structured access in Table 5.1. Given the data that is currently available, the question over the role of structured access in comprehension apparently reduces to a comparison between verbal agreement and reflexives. Initial hypotheses about the role of structured access in comprehension must at a minimum account for this distinction, but there are many dimensions on which the two dependencies differ. As suggested above, perhaps the most salient difference between the two dependencies is that only reflexives involve a referential dependency. This intuition leads to a natural first hypothesis about the range of dependencies that engage structured access: if a dependency involves constructing an interpreted anaphoric dependency then the parser operates in a more structured, accurate fashion.

Experiments 8 and 9 test this view by comparing the Mandarin long-distance reflexive *ziji* to the related Mandarin anaphor *ta-ziji*. This comparison is interesting because both anaphors have very similar licensing environments, although they have been argued to have very different underlying structure (Pan 1998, 2000; Bergeton 2007). Both may be used as reflexive anaphors bound by the local subject (Huang, Li & Li 2009). However, unlike *ziji*, *ta-ziji* is more likely to function as a contrastive pronominal element in many environments, leading to the possibility that apparently syntactic constraints on *ta-ziji* are epiphenomenal (König & Siemund 1999; Bergeton 2007). If *ziji* and *ta-ziji* pattern alike with respect to the access of non-local elements, then the hypothesis that structured access is a general property of interpreted anaphoric dependencies is supported. If they pattern differently, however, then this would suggest that structured access is deployed for satisfaction of a narrower range of grammatical constraints.

### *Revisiting Chinese anaphors*

Recall that Mandarin Chinese *ziji* is a long-distance reflexive: it is subject to a number of structural constraints, requiring either a linguistic antecedent in a certain structural configuration or defaulting to an indexical interpretation. In a sentence like (5.1), *ziji* can be bound by either the local or the long-distance antecedent. There is broad agreement that linguistically bound *ziji* is essentially a syntactic anaphor, even if there are additional discourse-pragmatic constraints on potential antecedents

(Huang et al 2009). In accord with this characterization, I showed in Chapter 4 that *ziji* engages a structured search through the parse space to find its antecedent.

- (5.1)            *Zhangsan<sub>i</sub>    shuo    Lisi<sub>j</sub>    nongshang-le   ziji<sub>i/j</sub>*  
                  Zhangsan    says    Lisi    harm-PERF    self  
                  “Zhangsan says that Lisi harmed him / himself”

In addition to *ziji*, however, there is another Mandarin pronominal that may be used to indicate that two arguments of a verb corefer: *ta-ziji*. Many discussions of *ta-ziji* simply state that it is a Principle A anaphor with a distribution that more or less mirrors English *himself* (Chomsky 1981), without further discussion (see, e.g., Huang & Liu 2001; Huang et al 2009). This characterization appears to hold in (5.2), where unlike *ziji*, *ta-ziji* cannot be bound by the distant antecedent. *Ta-ziji* is a bimorphemic compound that consists of the pronoun *ta* affixed with the reflexive element *ziji*, meaning literally *him-self*. The characterization of *ta-ziji* as a local anaphor squares well with a well-known cross-linguistic generalization about the morphological complexity of long-distance versus local reflexives. It has been noted that morphological complexity correlates with locality restrictions (Pica 1986). Simplex anaphors more often tend to allow long-distance binding, whereas morphologically complex forms are generally subject to stricter locality conditions. This fact has been taken as support for a head-movement analysis of long-distance binding, which predicts the phrase/head asymmetry in locality restrictions (Cole & Sung 1994). For this reason, from a typological point of view the analysis of *ta-ziji* as a principle A anaphor makes a lot of sense.

- (5.2)            *Zhangsan<sub>i</sub>    shuo    Lisi<sub>j</sub>    nongshang-le   taziji<sub>\*i/j</sub>*  
                  Zhangsan    says    Lisi    harm-PERF    self  
                  “Zhangsan says that Lisi harmed himself/\*him”

However, this characterization of *ta-ziji* obscures more complex licensing conditions. The distribution of *ta-ziji* and English local reflexives diverge when a wider range of data is considered, weakening the case for *ta-ziji* as a Principle A anaphor. For instance, long-distance bindings of *ta-ziji* are readily obtained simply by altering the animacy of the local subject in (5.2):

- (5.3)            *Zhangsan<sub>i</sub>    shuo    naben shu<sub>j</sub>    nongshang-le   taziji<sub>i/\*j</sub>*  
                  Zhangsan    says    that-CL book    harm-PERF    self  
                  “Zhangsan says that book harmed him”

No such contrast is found for English reflexives (c.f. *\*John said that the book hurt himself*). A number of authors have suggested that in light of data like (5.3), the characterization of *ta-ziji* as a Principle A anaphor cannot be maintained (Pan 1998, 2000; Huang 2000; Bergeton 2007). The exact nature of *ta-ziji* remains contentious, but existing accounts share a common insight: *ta-ziji* preferentially takes ‘prominent’ antecedents, where prominence is defined as some mixture of lexical semantic properties (e.g. animacy), structural properties (locality, dominance) and discourse properties (the ability to be contrasted with a contextually relevant set of individuals). For example, Pan (1998, 2001) accounts for this data by positing that the relevant notion of prominence is determined by a graded scale of animacy, ranging from animate human entities to inanimates. Given this notion of prominence, Pan suggests that the binding domain for *ta-ziji* is computed relative to the highest, most prominent

NP. Bergeton (2007) suggests, instead, that *ta-ziji* is essentially an intensified pronominal identical to *he himself* in English (a view that was endorsed by Tang 1989 as well). On his account, intensification is licensed when the antecedent may be contrasted with a set of alternatives in the discourse. Interestingly, even for unambiguously pronominal elements in English (intensified pronouns such as *he himself*) a contrast similar to that in (5.2) obtains:

- (5.4)           a.       ?John thinks Mary said that he himself went to the store.  
                   b.       John thinks Mary said that she herself went to the store.

(5.4a) is anomalous, except on the marginal reading where the focus domain associated with the intensifier *himself* is the VP (i.e. *John went to the store himself*). The contrast in (5.4) supports Bergeton's claim that the discourse licensing conditions on intensification can mimic syntactic locality conditions (for similar observations, see Baker 1995; Zribi-Hertz 1995; König & Siemund 1999). This mimicry is apparent even for English subject pronominals.

On either style of account, the strict locality of the binding domain for *ta-ziji* suggested by examples like (5.2) is epiphenomenal: *ta-ziji* may well take long-distance antecedents if they are sufficiently prominent, either morphosyntactically or in a discourse context. Similar conclusions have been reached for Norwegian *sig selv* by Lødrup (2009). The distribution of *sig selv* and Lødrup's account for this distribution make *sig selv* appear quite similar to *ta-ziji*. The intensification conveyed by the extra *selv* or *ziji* element appears to mimic Principle A effects, although given

appropriate configurations of animate discourse entities, long-distance readings are readily obtained.

If *ta-ziji* is an intensified pronominal whose licensing conditions are stated primarily in terms of animacy (Pan 1998, 2000) or discourse (Bergeton 2007) prominence, then unlike *ziji*, there is no need to limit antecedent search to structurally licit positions. If the apparent structural constraints are epiphenomenal, then simply accessing antecedents that are prominent on the relevant dimensions would achieve the desired result. Prominence, rather than solely structural position, is the more direct cue to *ta-ziji*'s antecedent, and so it is reasonable to think that an effective parsing strategy for *ta-ziji* will make direct reference to these cues rather than (or in addition to) syntactic structure.

If this line of argumentation is correct, then *ziji* and *ta-ziji* make an interesting minimal pair to test the domain of application of structured access procedures. In many cases, the binding possibilities for these anaphors overlap, but there is reason to suspect that the locality restrictions on *ziji* and *ta-ziji* are of a qualitatively different nature. If structured access is a strategy that is applied to interpreted dependencies, regardless of their underlying grammatical constraints, then we expect *ziji* and *ta-ziji* to both show the characteristics of structured search: preferential access of local subject positions, and delayed access to distant or inaccessible positions. However, an alternative possibility is that structured search is more narrowly applied to dependencies that have particular structural requirements on them. If it is the nature of the underlying structural constraints that drives structured access in reflexive dependencies, then *ta-ziji* should not show the behavioral signature of structured

access. Instead, we expect to observe direct access to licit, prominent antecedents regardless of position in the structure, based solely on their feature-compatibility with the anaphor.

Experiments 8 and 9 test the behavioral profile of these two anaphors to determine whether structured access reflects a fact about interpreted dependencies more generally, or if it instead is an expression of the particular linguistic constraints on *ziji*. Of critical interest is whether or not distant antecedents for *ta-ziji* are accessed in a direct-access manner, or if they pattern with *ziji* showing structured access and impeded processing when the antecedent is structurally distant.

*The scope of structured access: contrasting ziji & ta-ziji*

In order to test the manner in which *ziji* and *ta-ziji* access their antecedents, I focus on the processing difficulty associated with accessing sub-commanding antecedents. Sub-command refers to structural configurations where non-c-commanding antecedents may bind an anaphor, provided that the antecedents meet certain structural conditions. Sub-command has been long noted in descriptions of the licensing conditions on *ziji* (Tang 1989). Consider the following example from Tang (1989, p 100):

- (5.5) [Zhangsan<sub>i</sub> tou dongxi de] shishi bei ziji<sub>i</sub> de laoban faxian-le.  
 [Zhangsan steal thing DE] fact by self DE boss discover-  
 PERF  
 “The fact that Zhangsan stole something was discovered by his boss.”

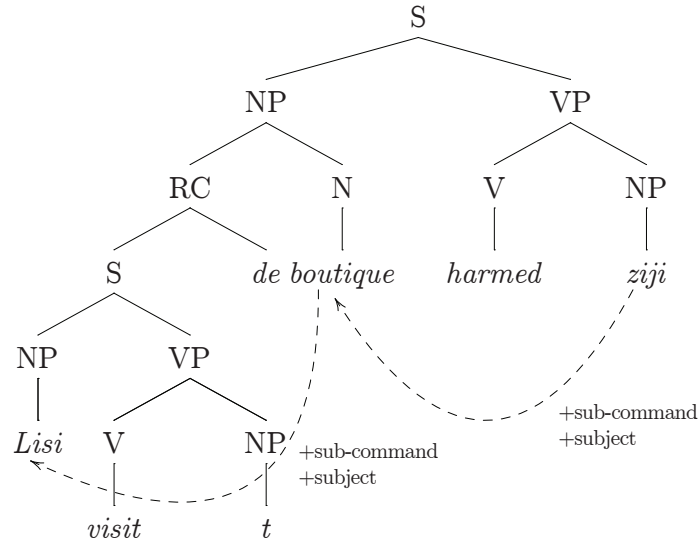


In this example, *Zhangsan* is able to bind *ziji*, despite the fact that it does not c-command the anaphor; it is not a sister to a node that dominates the anaphor at any level of representation. C-command has long been noted as an important condition on reflexive binding and quantifier-variable relationships (Reinhart 1976). However, the notion of c-command is not sufficient to account for the range of structural positions that a bound variable's antecedent may occupy in Mandarin. To account for the Mandarin data, Tang (1989) formulated the subcommand condition as in (5.6), which generalizes the c-command constraint to include examples such as (5.5) above. Tang notes that subcommand is furthermore subject to the constraint that embedded NPs cannot antecede *ziji* if they are contained within a potential binder of *ziji*. Thus in (5.5), because *shishi* 'fact' is not a potential binder of *ziji*, *Zhangsan* may bind *ziji* in virtue of it being contained in a commanding subject position.

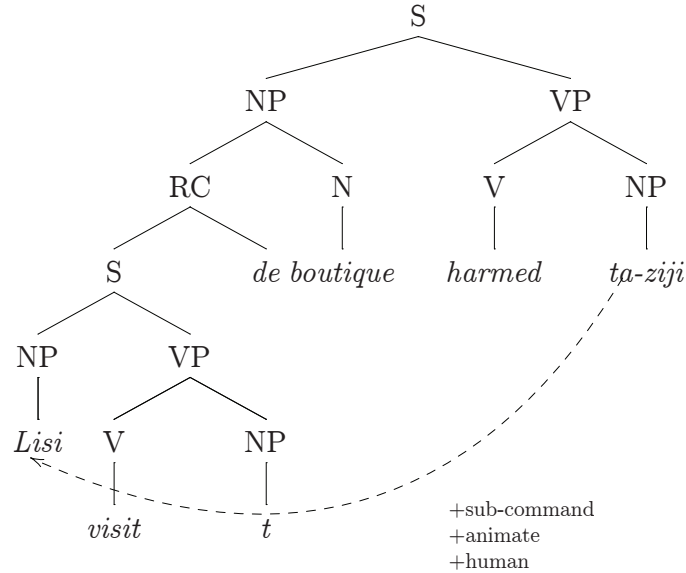
- (5.6)  $\alpha$  sub-commands  $\beta$  iff  
 a.  $\alpha$  c-commands  $\beta$ , or  
 b.  $\alpha$  is an NP contained in an NP that c-commands  $\beta$  or that sub-commands  $\beta$ , and any argument containing  $\alpha$  is in subject position.

Because binding from embedded nouns is only possible when the local subject is not a potential binder for either *ziji* or *ta-ziji*, the binding possibilities for embedded NPs is dependent on the identity of the head noun, as seen in examples (5.7) and (5.8). In these sentences the binding possibilities for *ziji* and *ta-ziji* are the same. When the local subject is animate ('seamstress' in 5.7), binding from the embedded NP is not possible for either anaphor. However, for both anaphors when the local subject is inanimate, the embedded subject *Mrs. Zhang* is a licit binder.

The pair of sentences in (5.7) and (5.8) provide an interesting test case that could be used to test the manner in which the two anaphors *ziji* and *ta-ziji* access their antecedents. Note that when the local noun is the anaphor's antecedent, as in (5.7), structural and semantic prominence are aligned. Because structural cues (i.e. *local subject*) and semantic/discourse prominence cues (i.e. *salient animate entity*) align in this case, both *ta-ziji* and *ziji* should behave identically. Even if *ziji* and *ta-ziji* use qualitatively distinct mechanisms to access their antecedent, for sentences like (5.7) they should display similar behavior: they both immediately access the local subject.



**Figure 5.1:** Structured search for *ziji* forces the parser to consider the local antecedent position before the sub-commanding antecedent.



**Figure 5.2:** Direct access for *ta-ziji* based on semantic or discourse prominence allows the parser to immediately access the sub-commanding antecedent.

Interestingly, the predictions of feature-based direct access and structured access diverge when it comes to processing binding from a subcommand position, as in (5.8). If structured search is deployed, the embedded subject position cannot be directly accessed. Instead, the local subject *boutique* in (5.8) must be checked, and only after its incompatibility is verified can the embedded position be accessed. This is schematized for sentence (5.8) in Figure 5.1. Alternatively, if feature-based direct access is employed, then access should occur in a single processing step, as in Figure 5.2. Thus in order to determine whether structural or feature-based access of antecedents is engaged for either anaphor, sentences like (5.8) provide the key piece of data. The structured access pattern in Figure 5.1 is expected to hold for *ziji* based on the structured search strategy findings in Chapter 4. Since only structural

information is used to initially bind *ziji*, then the local subject position should be first checked and rejected, requiring extra processing to access the embedded sub-commanding position.

- (5.7)  $ziji_{i/*i}/ziji_{i/*i}$ . [Zhang taitai<sub>i</sub> guanggu de] nücaifeng<sub>j</sub> hai-le ta-  
[Mrs. Zhang visit DE] seamstress harm-PERF ta-ziji /  
ziji.  
“The seamstress that Mrs. Zhang visits harmed herself.”
- (5.8)  $ziji_{i/*i}/ziji_{i/*i}$ . [Zhang taitai<sub>i</sub> guanggu de] shizhuangdian<sub>j</sub> hai-le ta-  
[Mrs. Zhang visit DE] boutique harm-PERF ta-ziji /  
ziji.  
“The boutique that Mrs. Zhang visits harmed her.”

The crucial question for the current purposes is whether or not *ta-ziji* accesses the embedded position in the same manner. As described above, there is good reason to suspect that *ta-ziji* is preferentially processed as an intensified or emphatic pronoun. If semantic prominence is the primary constraint on licit *ta-ziji*-antecedent relationships, then it is reasonable to expect that *ta-ziji* accesses its antecedent directly using prominence information. A direct-access mechanism for *ta-ziji* would leverage discourse or semantic prominence to recall its antecedent, and thus would access the embedded position in a single processing step (Figure 5.2).

Determining whether *ta-ziji* engages a structured search as in Figure 5.1, or a direct-access retrieval for accessing its antecedent as in Figure 5.2, provides a first step in determining the range of application of a structured access mechanism in comprehension. If *ta-ziji* patterns like *ziji* in accessing sub-commanding antecedents in a structured manner, then the hypothesis that structured access is employed in

interpreted dependencies is supported. However, if *ta-ziji* accesses its antecedents in a direct-access manner, then it would suggest that structured access is deployed to satisfy a narrower range of grammatical constraints.

Following arguments laid out in Chapter 2 and 3 above, in order to determine if feature-based or structured access to a distant or inaccessible antecedent is deployed, it is crucial to contrast the processing of the anaphor with a distant antecedent (5.9a) to a baseline where there is no antecedent linguistically represented (5.9b).

- (5.9) a. [*Zhang taitai<sub>i</sub> guanggu de*] *shizhuangdian<sub>j</sub> hai-le ta-*  
*ziji<sub>i/\*i</sub>/ziji<sub>i/\*i</sub>*  
 [Mrs. Zhang visit DE] boutique harm-PERF *ta-ziji* /  
*ziji*.  
 “The boutique that Mrs. Zhang visits harmed her.”
- b. [*Meiti baodao de*] *shizhuangdian<sub>j</sub> hai-le ta-*  
*ziji<sub>i/\*i</sub>/ziji<sub>i/\*i</sub>*  
 [Media report DE] boutique harm-PERF *ta-ziji* /  
*ziji*.  
 “The boutique that the media reported on harmed her.”

The logic of this comparison is identical to that in the experiments presented in Chapters 2 and 3. If direct access employed discourse or semantic features is deployed for *ta-ziji*, then (5.9a) should be easier to process than (5.9b) at the point of processing the anaphor’s reference. On the other hand, if structured search is used, they should be at least equally difficult upon reaching the anaphor, reflecting the fact that they both appear unacceptable when only the local subject is considered.

Experiment 8 tests the processing impact of sub-commanding binding of *ziji* using self-paced reading, and Experiment 9 tests identical environments with *ta-ziji*. I delay discussion of either experiment until after both have been presented.

#### *Experiment 8: Ziji and sub-commanding antecedents*

##### Participants

41 students from the University of Maryland community participated in the experiment. All participants were native Mandarin Chinese speakers from mainland China, and all had normal or corrected-to-normal vision. They were paid \$10 per hour for their participation in the experiment.

##### Stimuli

The experimental materials consisted of four conditions designed to investigate the effect of semantically coherent, but sub-commanding antecedents on the processing of *ziji*. There were two potential antecedent positions: the local subject position and the embedded subject position. As in Experiments 6 and 7, binding possibilities were manipulated by manipulating the animacy of the two subject positions. The conditions are summarized in Table 5.2; the *local* subject factor refers to the animacy of the local subject, and the *distant* subject factor refers to the animacy of the embedded subject. I refer to the [+local,+distant] condition as the *multiple*

*match* condition. The [+local,-distant] and [-local,+distant] conditions were the *local match* and *distant match* conditions, respectively. The *no match* condition had no animate NPs ([-local,-distant]). The distant (sub-commanding) antecedent position was the subject of an object relative clause that modified the main clause subject. This antecedent position was always the first word in the sentence, due to the head-final order of Mandarin relative clauses. The local (main clause) subject was always in fourth position.

In order to avoid wrap-up effects at the critical region, the *ba* construction was used. This construction uses a particle *ba* to mark the direct object, which is then moved to preverbal position. So that the two arguments of the main clause were not linearly adjacent, a temporal adverbial was placed between the local subject and the *ba*-marked *ziji*. Between *ziji* and the final verb, a manner adverbial was placed in order to provide an extra spillover region.

Condition	Example
Multiple match [+local,+distant]	[张太太 / 经常逛顾的] / 那个 / 女裁缝 / 上个星期 / 把 / 自己 / 不小心 / 弄伤了. Mrs. Zhang / often visit DE / that-CL / seamstress / last week / BA / ziji / not careful / harm-PERF. “The seamstress that Mrs. Zhang often visits carelessly hurt herself last week.”
Local match [+local,-distant]	[媒体/ 报告的] / 那个 / 女裁缝 / 上个星期 / 把 / 自己 / 不小心 / 弄伤了. Media / report on DE / that-CL / seamstress / last week / BA / ziji / not careful / harm-PERF. “The seamstress that the media reported on carelessly hurt herself last week.”
Distant match [-local,+distant]	[张太太 / 经常逛顾的] / 那个 / 时装店/ 上个星期 / 把 / 自己 / 不小心 / 弄伤了. Mrs. Zhang / often visit DE / that-CL / boutique / last week / BA / ziji / not careful / harm-PERF. “The boutique that Mrs. Zhang often visits carelessly hurt her last week.”
No match [-local,-distant]	[媒体/ 报告的] / 那个 / 时装店/ 上个星期 / 把 / 自己 / 不小心 / 弄伤了. Media / report on DE / that-CL / boutique / last week / BA / ziji / not careful / harm-PERF. “The boutique that the media reported on carelessly hurt <i>self</i> last week.”

**Table 5.2:** Critical conditions from Experiment 8. Region breaks are indicated by slashes.

Of the four experimental conditions, only the no match condition required *ziji* to take an extra sentential antecedent. The multiple-match, local match, and distant match conditions were all grammatical with a bound *ziji* interpretation.

18 sets of these four conditions conditions were produced, and distributed into four lists in a pseudo-Latin square fashion. They were combined with 77 fillers, including materials from an unrelated experiment, for a total of 95 sentences. The ratio of acceptable-to-unacceptable sentences varied slightly from list to list due to the pseudo-Latin square list, but remained between 45 and 55 % acceptable. The fillers included 10 sentences that contained *ba* followed by non-anaphoric NPs in order to prevent anticipation of *ziji*.

### Procedure

Sentences were presented using a moving-window self-paced reading paradigm, using the Linger software. Each sentence was presented in black characters on a white screen, and no sentence was more than one line long. All sentences were presented using simplified Chinese characters. The sentences were segmented into 9 regions according to native speaker intuitions about where best to insert boundaries (as marked in Table 5.2), resulting in regions that ranged from one character (i.e. *ba*) to at most 6 characters (i.e. *yishuticaoguanjun*, “gymnastics champion”). Sentences initially appeared as a series of dashes that obscured the words, and by pressing the space bar participants were able to sequentially reveal each region for a self-determined amount of time. Each region was remasked after the participant finished



reading it. After each sentence, a comprehension question was presented in its entirety on the screen, and participants were instructed to press *f* for yes, and *j* for no. In the critical experimental sentences, the comprehension question queried a part of the sentence that did not have to do with *ziji*'s referential dependency. Feedback was given for incorrect responses.

### Offline judgments

In order to confirm that the embedded subjects are reliably accepted as antecedents for *ziji* (as reported in Tang 1989), the experimental materials were used for an offline judgment study. 22 participants were asked to judge the acceptability of the sentences they read on a 7-point scale, where 7 was completely acceptable and 1 was completely unacceptable. Participants were instructed to judge the sentences with regard to whether or not they were acceptable in colloquial speech. The results are presented in Table 5.2. Data was gathered using the IbexFarm over the internet; students from Beijing Normal University's psychology department were recruited as participants.

<b>Multiple match</b> <b>[+local,+distant]</b>	<b>Local match</b> <b>[+local,-disant]</b>	<b>Distant match</b> <b>[-local,+distant]</b>	<b>No match</b> <b>[-local,-distant]</b>
4.77 ( $\pm 0.29$ )	5.08 ( $\pm 0.29$ )	4.15 ( $\pm 0.41$ )	3.29 ( $\pm 0.43$ )

**Table 5.2:** Mean judgments and standard error by subjects for *ziji* Experiment 8 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable.

A two-way repeated measures ANOVA by subjects revealed a significant main effect of local noun animacy ( $F(1,21) = 22.2, p < 0.001$ ), as well as a significant interaction of local noun animacy with distant noun animacy ( $F(1,21) = 10.2, p < 0.01$ ). In addition, there was a marginal main effect of distant noun animacy ( $F(1,21) = 3.08, p < 0.1$ ). Resolving this interaction further using planned pairwise  $t$ -tests revealed a marginal difference between multiple match and local match conditions ( $t(21) = -1.92, p < 0.8$ ), and a reliable difference between distant match and no match conditions ( $t(21) = 2.87, p < 0.01$ ).

In addition, a repeated measures ANOVA on judgment times revealed that conditions with local antecedents were judged more quickly than conditions where the local antecedent was inanimate ( $F(1,21) = 4.50, p < 0.05$ ).

These results confirm that sentences with embedded antecedents are considered to be more acceptable than sentences with no linguistically represented antecedent for *ziji*. Importantly, embedded antecedents for *ziji* were judged significantly better than no antecedent conditions, indicating that participants were willing to consider the embedded position, confirming the data in Tang (1989).

### Data Analysis

Reading times at the critical anaphor and the spillover region were submitted for statistical analysis. It is common practice in self-paced reading data to reject outlying data points based on a cutoff criterion (e.g., more than 2.5 standard deviations from the mean for a given region / condition pairing). This is necessary to

deal with outliers, which can significantly distort estimates of average reading time (Ratcliff 1993). Instead of adopting a rejection threshold, however, I instead log-transformed all reaction times prior to analysis. In addition to minimizing the impact of reaction time outliers, this method has the additional benefit of making the reaction time data close to normally distributed, satisfying an important assumption of linear model analysis (Gelman & Hill 2005). No data was excluded from analysis.

Statistical analysis was performed using mixed-effect linear regressions to assess the magnitude, direction, and reliability of the experimental factors on reading times. The experimental fixed effects in the models were the factors LOCAL (whether or not the sentence was grammatical), DISTANT (whether or not the embedded NP was plural), and their interaction. The fixed effects were coding using simple difference sum coding (inanimate levels were coded as  $-.5$ , animate levels as  $.5$ ). Thus all reported coefficients reflect the magnitude of the difference between levels of a given factor (in log-transformed RT space). In addition to these fixed effects, I additionally considered random intercepts for subjects and items, as well as random slopes for the experimental fixed effects by subjects and by items. In all cases, the significance of non-experimental fixed effects and random effects was assessed, and I report in all cases the best-fit model (following Baayen et al 2008; Jaeger 2008). I leave the experimental fixed effect structure constant across all models, because these effects were theoretically motivated by the design of the study. For most analyses, in addition to the experimental fixed effects, the best model included only random intercepts for subjects and items, as well as all a fixed effect for trial order. Word by word reading times are presented in Figure 5.3; for presentational purposes, pre-

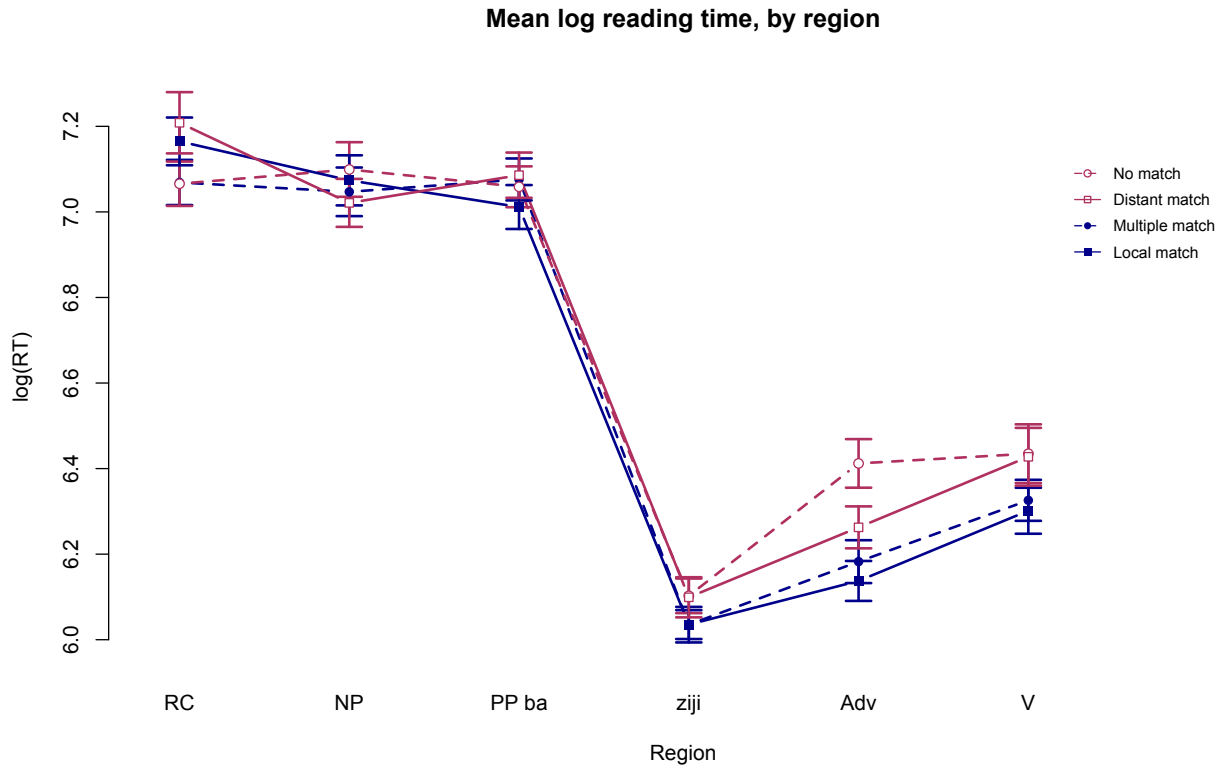
critical regions are collapsed into three regions: relative clause (RC), head noun phrase (NP), and pre-critical prepositional phrase (PP *ba*).

### Results

Mean reading times are presented in Figure 5.3. Statistical analysis of the critical and spillover regions is summarized in Table 5.3. At the critical *ziji* region, modeling revealed that conditions with a local inanimate subject were read significantly slower than those with animate local subjects ( $\beta = -0.063$ , SE: 0.026,  $pMCMC < 0.05$ ). This effect was also observed at the spillover region ( $\beta = -0.178$ , SE: 0.034,  $pMCMC < 0.0001$ ), in addition to a main effect of the embedded noun's animacy ( $\beta = 0.094$ , SE: 0.034,  $pMCMC < 0.01$ ). Planned comparisons revealed that the animate embedded nouns caused a significant slowdown when the head noun was inanimate ( $\beta = 0.149$ , SE: 0.049,  $pMCMC < 0.01$ ), but not when it was animate ( $\beta = 0.040$ , SE: 0.049,  $pMCMC < 0.5$ ).

	$\beta$	SE	$t / z$
<i>ziji</i>			
LOCAL	-0.063	0.026	-2.44*
DISTANT	0.008	0.026	0.31
LOCAL×DISTANT	-0.011	0.052	-0.21
Spillover			
LOCAL	-0.178	0.034	-5.19***
DISTANT	0.094	0.034	2.73**
LOCAL×DISTANT	-0.110	0.069	-1.58

**Table 5.3:** Summary of fixed effects for best-fit models at the critical *ziji* and spillover regions, including  $t$ -values



**Figure 5.3:** Region-by-region mean log reading times for Experiment 8. Error bars represent  $\pm 1$  standard error, by participants, corrected for between-participant variance.

#### *Experiment 9: Ta-ziji and sub-commanding antecedents*

##### Participants

70 students from the University of Maryland community participated in the experiment. All participants were native Mandarin Chinese speakers from mainland China, and all had normal or corrected-to-normal vision. They were paid \$10 per hour for their participation in the experiment.

## Stimuli

The materials were largely identical to those from Experiment 8; for convenience they are repeated in Table 5.4. Two important changes were made to these materials. First, all instances of *ziji* were replaced with *ta-ziji*. *Ta* is a third-person pronoun that does not distinguish male, female, animate, and inanimate referents phonologically. However, gender and animacy distinctions are maintained in the writing system: *ta* (他) is used for male humans, *ta* (她) for female humans, and *ta* (牠) for animals and inanimates. In order to prevent this extra morphological information from biasing selection of *ta-ziji*'s antecedent, the materials were additionally modified so that within an experimental item set, the animate nouns in each position were of the same gender. Thus, as in Experiment 8, the acceptability of the two referents as antecedents for the anaphor turned solely on their structural status. Half of the revised materials had male nouns, and the other half had female nouns.

Condition	Example
Multiple match [+local,+distant]	[张太太 / 经常逛顾的] / 那个 / 女裁缝 / 上个星期 / 把 / 她自己 / 不小心 / 弄伤了。 Mrs. Zhang / often visit DE / that-CL / seamstress / last week / BA / ta-ziji / not careful / harm-PERF. “The seamstress that Mrs. Zhang often visits carelessly hurt herself last week.”
Local match [+local,-distant]	[媒体/ 报告的] / 那个 / 女裁缝 / 上个星期 / 把 / 她自己 / 不小心 / 弄伤了。 Media / report on DE / that-CL / seamstress / last week / BA / ta-ziji / not careful / harm-PERF. “The seamstress that the media reported on carelessly hurt herself last week.”
Distant match [-local,+distant]	[张太太 / 经常逛顾的] / 那个 / 时装店 / 上个星期 / 把 / 她自己 / 不小心 / 弄伤了。 Mrs. Zhang / often visit DE / that-CL / boutique / last week / BA / ta-ziji / not careful / harm-PERF. “The boutique that Mrs. Zhang often visits carelessly hurt herself last week.”

No match	[媒体/ 报告的] / 那个 / 时装店 / 上个星期 / 把 / 她自己 / 不小心 / 弄伤了.
[-local,-distant]	Media / report on DE / that-CL / boutique / last week / BA / ta-ziji / not careful / harm-PERF. “The boutique that the media reported on carelessly hurt herself last week.”

**Table 5.4:** Critical conditions from Experiment 9. Region breaks are indicated by slashes.

As in Experiment 8, 18 sets of these four conditions were produced, and distributed into four lists in a pseudo-Latin square fashion. They were combined with 77 fillers for a total of 95 sentences. The ratio of acceptable-to-unacceptable sentences varied slightly from list to list due to the pseudo-Latin square design, but remained between 45 and 55 % acceptable. The fillers included 10 sentences that contained *ba* followed by non-anaphoric NPs in order to prevent anticipation of *ziji*.

### Procedure

The experimental procedure was identical to Experiment 8.

### Offline judgments

19 additional participants were asked to judge the acceptability of the experimental sentences on a 7-point scale, where 7 was completely acceptable and 1 was completely unacceptable. Participants received the same instructions as in the judgment pre-test in Experiment 8. The results are presented in Table 5.5. Data collection and participant recruitment was identical to Experiment 8.

<b>Multiple match</b> <b>[+local,+distant]</b>	<b>Local match</b> <b>[+local,-disant]</b>	<b>Distant match</b> <b>[-local,+distant]</b>	<b>No match</b> <b>[-local,-distant]</b>
4.12 ( $\pm 0.34$ )	4.96 ( $\pm 0.27$ )	5.06 ( $\pm 0.27$ )	4.03 ( $\pm 0.33$ )

**Table 5.5:** Mean judgments and standard error by subjects for *ziji* Experiment 8 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable.

A two-way repeated measures ANOVA by subjects revealed only a significant interaction of local noun animacy with distant noun animacy ( $F(1,18) = 32.5$ ,  $p < 0.001$ ). Resolving this interaction further using planned pairwise  $t$ -tests revealed significant differences between multiple match and local match conditions ( $t(18) = -3.26$ ,  $p < 0.01$ ), and a reliable difference between distant match and no match conditions ( $t(18) = 4.88$   $p < 0.001$ ).

A repeated-measures ANOVA on judgment times revealed no significant differences between the conditions.

The judgment results show that sentences with embedded antecedents for *ta-ziji* are considered to be as acceptable as sentences with local antecedents; both are more acceptable than sentences with no linguistically represented antecedent. Interestingly, the multiple match condition was rated as significantly worse than the local match condition. One interpretation of this result is that the intensified pronominal is less felicitous in an out-of-the-blue discourse context where there are two equally acceptable antecedents it might refer to. If true, this finding may be taken to support the view that *ta-ziji* is preferentially perceived as an intensified, contrastive pronoun.



## Data Analysis

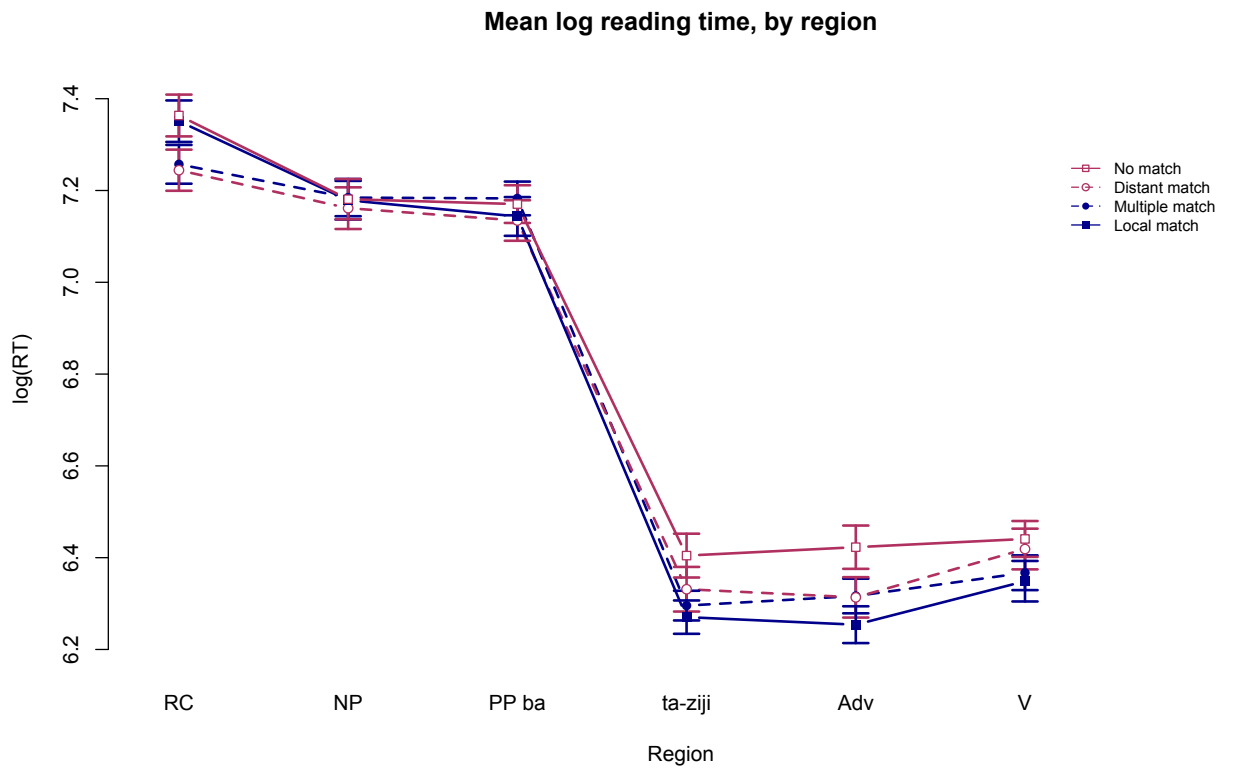
Data analysis was identical to Experiment 8.

## Results

Mean reading times are presented in Figure 5.4; as in Figure 5.3, pre-critical regions are collapsed for purposes of presentation. Statistical analysis of the critical and spillover regions is summarized in Table 5.6. At the critical *ziji* region, there was a significant effect of local noun animacy, such that inanimate local nouns led to a slowdown ( $\beta = -0.089$ , SE: 0.027,  $pMCMC < 0.001$ ). At the spillover region there was a main effect of local noun animacy ( $\beta = -0.081$ , SE: 0.023,  $pMCMC < 0.001$ ), as well as interaction of local and distant noun phrase animacy ( $\beta = 0.165$ , SE: 0.046,  $pMCMC < 0.001$ ). In addition, there was a marginal effect of the embedded noun's animacy ( $\beta = -0.062$ , SE: 0.033,  $pMCMC < 0.07$ ). Planned comparisons revealed that the animate embedded nouns caused a significant speed-up when the head noun was inanimate ( $\beta = -0.103$ , SE: 0.032,  $pMCMC < 0.01$ ), but there was a marginal slowdown for embedded animates when the head noun was animate ( $\beta = 0.062$ , SE: 0.033,  $pMCMC < 0.07$ ).

	$\beta$	SE	$t / z$
<i>ta-ziji</i>			
LOCAL	-0.089	0.027	-3.32***
DISTANT	-0.030	0.023	-1.29
LOCAL×DISTANT	0.076	0.047	1.63
Spillover			
LOCAL	-0.081	0.023	-3.51***
DISTANT	-0.020	0.023	-1.90†
LOCAL×DISTANT	0.165	0.046	3.58***

**Table 5.6:** Summary of fixed effects for best-fit models at the critical *ziji* and spillover regions, including *t*-values

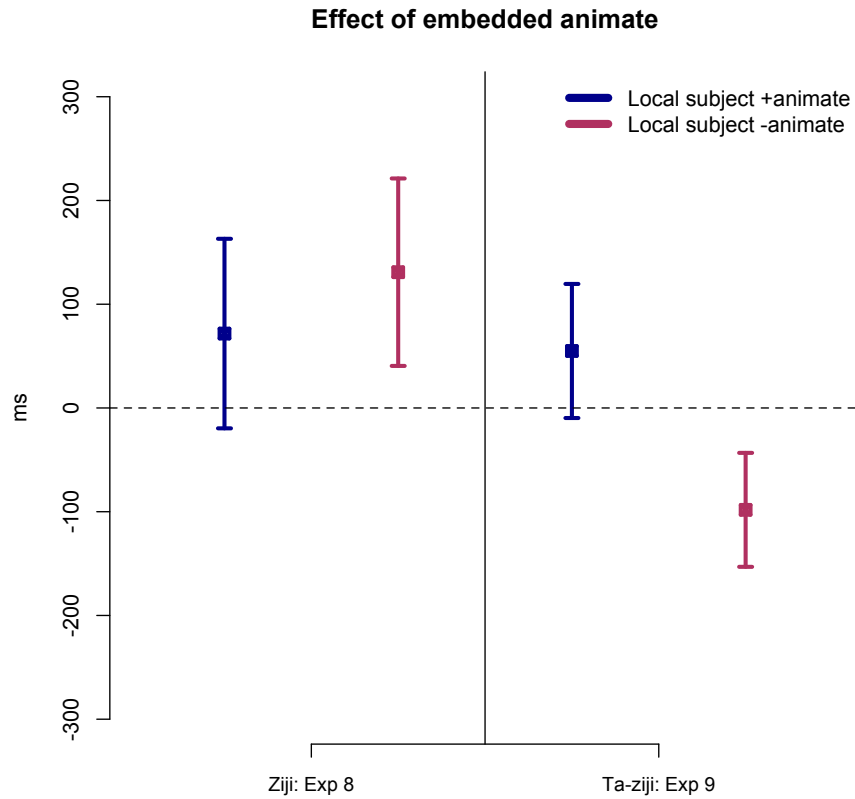


**Figure 5.4:** Region-by-region mean log reading times for Experiment 9. Error bars represent  $\pm 1$  standard error, by participants, corrected for between-participant variance.

## Discussion

In Experiment 8, increased reading times were observed for both distant and no match conditions. At the point of the anaphor, there was a significant main effect of local subject animacy, and in the spillover region, there were main effects of both local and distant subject animacy, with animate distant subjects causing longer reading times in the spillover region. In Experiment 9, there was also an impact of local subject animacy, followed by an interaction of local and distant subject animacy. As in Experiment 8, inanimate local subjects caused longer reading times, but in contrast to the pattern of reading times observed in Experiment 8, there was a significant facilitation for distant animate subjects when the local subject was inanimate.

These results suggest that *ziji* and *ta-ziji* contrast with respect to the difficulty of recovering a sub-commanding antecedent. This pattern of results suggests that the two anaphors recruit qualitatively different strategies for accessing their antecedents online, despite superficially similar locality requirements. Results suggest that *ta-ziji* is able to use semantic or discourse prominence to directly access its antecedent, regardless of structural position. This pattern was not observed for *ziji*, which again appeared to check the semantically inappropriate local subject position before considering a wider range of antecedents.



**Figure 5.5:** Effect of embedded animate (embedded [+animate] subject – embedded [-animate] subject) on spillover reading times in Experiments 8 (*ziji*) and 9 (*ta-ziji*). Error bars represent 95% confidence interval, by participants.

This difference in access strategy is apparent in the summary of the reading time differences between the critical comparisons for *ziji* and *ta-ziji* in Figure 5.5. This summary presents the effect of an embedded animate subject, which is obtained by subtracting the reading time for (5.10b) from (5.10a). It can be seen that the effect of a distant animate subject for *ziji* dependencies is similar regardless of whether or not the head noun is animate. This observation is confirmed by statistical analysis: at the anaphor, there was only a main effect of local subject animacy, followed by main effects for local and distant subject animacy in the spillover region. In the spillover

region, an embedded animate subject caused a slowdown in reading times that was indistinguishable between animate and inanimate head nouns. However, for *ta-ziji*, there was a significant interaction of local subject and distant subject animacy, such that animate embedded subjects with inanimate local subjects caused facilitated processing.

- (5.10) a. [*Zhang taitai<sub>i</sub> guanggu de*] *shizhuangdian<sub>j</sub> hai-le ta-*  
*ziji<sub>i</sub>/\*<sub>i</sub>/ziji<sub>i</sub>/\*<sub>i</sub>*  
 [Mrs. Zhang visit DE] boutique harm-PERF *ta-ziji* /  
*ziji*.  
 “The boutique that Mrs. Zhang visits harmed her.”
- b. [*Meiti baodao de*] *shizhuangdian<sub>j</sub> hai-le ta-*  
*ziji<sub>i</sub>/\*<sub>i</sub>/ziji<sub>i</sub>/\*<sub>i</sub>*  
 [Media report DE] boutique harm-PERF *ta-ziji* /  
*ziji*.  
 “The boutique that the media reported on harmed her.”

The results of Experiment 8 are consistent with the structured search account of *ziji* given in Chapter 4: comprehenders are initially sensitive to the feature mismatch between the local subject and the anaphor, suggesting that the local subject is selectively reaccessed based on its structural position. In the spillover region, the embedded animate exerts an approximately equal slowdown in reading times regardless of the local subject’s animacy. This suggests that whatever processes drive the slowdown observed for embedded animates, they are independent of the local noun’s animacy. Importantly, sub-commanding antecedents did not lead to faster reading times relative to a no antecedent condition. This provides further evidence against a semantic feature-based access account for *ziji*-antecedent dependencies, and stands in contrast to the results for *ta-ziji*. If *ziji* made use of the animacy information

in retrieving its antecedent, then an embedded antecedent should have caused *ziji* to be read more quickly than the no antecedent baseline.

Interestingly, these present results provide another example where online processing difficulty diverges from offline acceptability findings: *ziji* with embedded antecedents was considered more acceptable than the no antecedent condition in the judgment task, but appeared to require more processing in the online reading task. This is surprising in itself, as processing difficulty is generally highly correlated with acceptability.

One possible objection to these findings is that the different processing profiles observed for *ziji* and *ta-ziji* may not be due to differences in access procedures, but rather due to differences in processing difficulty in the no match baseline condition. In Chapter 4 I suggested that comprehenders might be able to use a discourse-based heuristic to avoid the need for structured search when an indexical interpretation of *ziji* is needed. The possibility of this alternative route to resolving *ziji*'s reference may mean that the no match condition is more difficult for *ta-ziji* than it is for *ziji*. No such strategy is possible for *ta-ziji*, as the 3<sup>rd</sup>-person pronoun *ta* blocks the possibility of a first-person construal of the anaphor. The SPR data presented here do not rule out this possibility, but this interpretation seems unlikely in light of the relationship between the local animate subject conditions and the distant subject conditions. Participants clearly spend more time reading the distant match condition than the local match conditions for *ziji*, whereas for *ta-ziji*, the distant match condition patterns with the local match conditions. Since distant match *ta-ziji* is not associated with an appreciable increase in reading times or decrease in

acceptability over grammatical baselines, there does not appear to be any extra processing difficulty associated with recovering a sub-commanding antecedent for *ta-ziji*. This is consistent with a direct access mechanism for recovering the distant antecedent for *ta-ziji* as in Figure 5.2. Conversely, the processing difficulty and degraded acceptability for distant antecedents for *ziji* is consistent instead with a structured access account as in Figure 5.1

Another alternative explanation for the difference between *ziji* and *ta-ziji* is that the information contained in the written form of *ta* provided an extra cue to antecedent identity. As mentioned, *ta* distinguishes male, female and inanimate antecedents when written, although its spoken form does not. This cannot be ruled out on from the present data, but it seems unlikely that written cues to gender would be used while linguistic constraints (on animacy) are not for *ziji*. Even if true, however, the basic argument I present here holds: structured access is not a property of interpreted dependencies. Gender cues for *ta-ziji* serve as pointers to its antecedent, but they do not serve the same role for English *himself*.

It seems plausible that the difference in access strategies for *ziji* and *ta-ziji* stems from deeper underlying differences between the two. This supports the claim that the two anaphors are only superficially similar. In order to draw inferences about grammatical constraints from the online processing profiles, however, clear linking assumptions between grammatical constraints and access strategies need to be made. Thus before taking up further discussion of the two types of anaphor here, it is necessary to first revisit the question of structured access that I started this chapter with.

### *Structured access as syntactic parsing*

Since *ta-ziji* appears to use semantic information as a pointer to the memory representation of its antecedent, while *ziji* does not, we are led to reject the first intuitive hypothesis that structured access is a property of interpreted or referential dependencies. With the new results concerning the access properties of these different anaphors in hand, the empirical scope of structured access can now be slightly expanded, as in Table 5.7 (from Table 5.1 above). In addition to the dependencies that I have considered in this thesis, there are a number of other results that may bear on the development of a theory of structured access. These include English VPE (Martin & McElree 2008, 2009) and English cross-sentential anaphora (Foraker & McElree 2007). Though SAT results have shown that a range of dependencies proceed in a feature-based direct access fashion, these two dependencies are especially interesting in the current context because they are best described as retrospective dependencies; the head of the dependency (the antecedent NP / VP) does not signal that it will be retrieved later on. Thus no prospective processing is likely to be engaged, and the processing of these dependencies may be taken to reflect retrieval processes. Interestingly, both cross-sentential anaphora and VPE in English appear to retrieve their antecedents in a feature-based direct access fashion.

However, in addition to reflexives, there is processing evidence that suggests another dependency that may engage a structured access mechanism: relative clause (RC) attachment in Dutch (van Berkum, Brown & Hagoort 1999b; Brysbaert &



Mitchell 2000). It appears that when retrieving a noun phrase to attach a relative pronoun to, extra-syntactic information is not used to guide this parsing decision. Mitchell et al (1995) argued that this is because abstract information such as major syntactic category is the correct ‘grain size’ over which decisions about RC attachment height should be stated. If the RC attachment decision is modeled a retrieval for an NP attachment site that is engaged by the relative pronoun, then the observation that grammatical gender is not used in this retrieval process is similar to the claim of structured access for reflexives advanced here.

	<b>Facilitatory Interference</b>
English subject-verb agreement	✗
Spanish subject-verb agreement	✗
English reflexives	✓
Mandarin Chinese <i>ziji</i>	✓
German, English NPIs	✗
English VP ellipsis	✗
English cross-sentential anaphora	✗
Mandarin Chinese <i>ta-ziji</i>	✗
Dutch relative clause attachment	✓

**Table 5.7:** Summary of access properties of long-distance dependencies.

Dutch RC attachment processing provides an interesting parallel to the present work. This is because the relative pronoun inflects for gender, imposing a clear formal constraint on its attachment site above and beyond the structural conditions on attachment; the head noun and the relative pronoun must agree in gender features. Nonetheless, it appears that gender information is not used in finding an attachment site for the relative pronoun, just as number was not used in determining a reflexive’s

antecedent in Experiments 1 and 3. For example, Brysbaert and Michell (1996) showed that comprehenders do not apply gender constraints to resolve attachment ambiguities in situations of where the relative pronoun can be attached to one of two NP hosts. As in English, the Dutch sentence in (5.11a) is globally ambiguous: the relative pronoun *die* can be construed as modifying the higher or lower phrase. Brysbaert and Mitchell noted that speakers preferred to construe the RC as modifying the higher NP (*de zoon*).

- (5.11)      a.      *De zoon van de actrice*      [*die*      *op het balkon zat...*]  
                  The son-COM of the actress-COM [that-COM on the balcony sat]  
                  “The son of the actress who was on the balcony...”
- a.      *Het zoontje van de actrice*      [*die*      *op het balkon zat...*]  
                  The son-NEU of the actress-COM [that-COM on the balcony sat]  
                  “The son of the actress who was on the balcony...”

Interestingly, Brysbaert and Mitchell found that the preference for attaching the relative pronoun to the higher noun was the same for (5.11b), even though the head noun *zoontje* mismatches in grammatical gender. Thus (5.11b) was significantly more difficult to process than (5.11a), but not until later disambiguating information was received (Brysbaert & Mitchell 1996, 2000). Comprehenders did not appear to use the gender information early enough in parsing to eliminate the ambiguity of (5.11b).

Van Berkum and colleagues (1999) presented an additional argument for a purely structural parsing strategy for attaching relative pronouns to their NP hosts. Using event-related potentials, they showed that a mismatch in gender features between the relative pronoun and its host NP did not aid comprehenders in

disambiguating a relative / complement clause ambiguity. *Dat* is both a relative pronoun for neuter gender nouns, and a complementizer that introduces an embedded clause. For a string such as *David vertelde de actrice dat* (David told the actress-COM that-NEU / that), the gender mismatch between the common gender *actrice* and the relative pronoun interpretation of *dat* should preclude pursuing a relative clause interpretation. However, using event-related potentials, van Berkum and colleagues argued that participants nonetheless attempted to parse *dat* as a relative pronoun in these cases (given a supportive discourse context), encountering difficulty later in the sentence when further input disambiguated the sentence.

This finding is in line with the structured access account advocated here. The parsing decision and attachment of the relative clause pronoun proceeds blind to gender information, which is only evaluated once the dependency has been constructed. This surprising finding parallels the results in Chapter 2: despite its usefulness, the morphological feature information in reflexive and relative clause pronouns is not deployed for early parsing decisions. Brysbaert and Mitchell (2000) describe this surprising finding in a way that parallels the results in Chapters 2 and 3:

[F]or some as yet unexplained reason, grammatical gender information does not appear to play as rapid and efficient a role in guiding syntactic processing as might have been expected from the formal constraints such cues place on the structures of sentences. (Brysbaert & Mitchell pp 465)

In light of the findings presented here, this conclusion may be sharpened. It is not simply gender features that the parser selectively ignores, but number features in English reflexives as well. More generally, however, a consideration of the dependencies in Table 5.7 suggests that this general statement does not hold across all

of the syntactic dependencies that have been studied. In particular, for dependencies such as subject-verb agreement, these features are not ignored. Instead, they are directly and immediately deployed in the construction of subject-verb agreement dependencies. The narrow focus on syntactic information is thus not a feature of all syntactic dependencies, nor is it a feature of interpreted dependencies.

Given the evidence that has accumulated to date, a strong but simple hypothesis about the relation between syntactic structure and memory access online may be entertained. Setting aside subject-verb agreement for the moment, it may be that structured access is simply a feature of all processing of syntactic dependencies. This claim makes a direct link between syntactic constraints and the information used to access memory online.

A direct link between the information needed to satisfy structural constraints for syntactic dependencies and the information used to access memory and generate structure online is arguably the null hypothesis. For long-distance dependencies that have are primarily structural in nature, the parser appears to use only structural constraints to access and index memory. Reflexive dependencies are unambiguously structure-dependent on a wide range of approaches. Most grammatical theories of anaphoric interpretation treat reflexive anaphor-antecedent dependencies as ‘core’ phenomena whose interpretation is immediately constrained by their syntactic context, despite a wide variety of views about the role that syntax plays in constraining non-reflexive pronominal interpretation (Wasow 1972; Chomsky 1981; Reinhart & Reuland 1993; Fiengo & May 1994; Pollard & Sag 1994; Büring 2005).

Likewise, relative clause attachment (i.e. the relation between the relative pronoun and its NP host) is similarly constrained by structural factors.

However, if a long-distance dependency does not involve building direct syntactic relations between two elements, there is no reason to expect that it should use structural access based on (5.12) above. The dependency between a pronoun and its antecedent, for example, is not generally modeled as a syntactic relation between the two; instead, the two elements are simply taken to denote the same entities in the discourse model (Büring 2005). Although there are apparent structural constraints on where a pronoun's antecedent could be found, it is not clear that these are constraints on where the antecedent cannot be located syntactically, rather than constraints on possible coreference (Lasnik 1976). In particular, a principle B violation cannot be alleviated by placing the desired antecedent in a licit structural position: compare *\*John<sub>i</sub> hit him<sub>i</sub>* to *I saw John<sub>i</sub>*. *\*John<sub>i</sub> hit him<sub>i</sub>*. This suggests that describing syntactically accessible antecedent positions is not the correct way to capture the distribution of coreferential pronouns. There are potentially even fewer constraints on where the antecedent for VP ellipsis may be found (Johnson 2001), with the only plausible constraints being that there be some VP in the discourse that can fill the ellipsis site. As with constraints on coreference, apparent structural constraints on where VPE's antecedent may be found may be reduced to other factors. It seems unlikely that either VPE or coreference relations are dependent on a direct syntactic dependency built between the proform and its antecedent. In light of this, it is perhaps not surprising that both of these comprehenders engage feature-based direct access mechanisms when resolving the reference of these dependencies (Foraker & McElree

2007; Martin & McElree 2008, 2009). Structured access would simply be a poor strategy for these dependencies, inappropriately narrowing search to a subset of positions in a potentially very large discourse space. Likewise, if *ta-ziji* is more appropriately construed as a contrastive or intensified pronoun whose reference is not syntactically represented, then the constraint imposed by structured access does not hold.

To a first approximation, this generalization covers existing empirical findings: syntactic dependencies use only syntactic information to access memory, and no such constraint is applied to non-syntactic anaphoric dependencies. However, one dependency that does not obviously fit this characterization is subject-verb agreement in English. One clear finding is that morphological features are used as retrieval cues in the construction of the dependency. This access profile is not obviously consistent with the statement of structured access given above. From the point of view of structured access, the behavior of subject-verb agreement in comprehension is puzzling.

There are two ways that the exceptional nature of subject-verb agreement might be understood. One possibility is that the characterization of subject-verb agreement as a syntactic dependency is incorrect. If this line of reasoning is correct, then the direct link between structured access and syntactic dependencies is maintained, and understanding the exceptional behavior of agreement requires further elaborating the correct grammatical model of agreement. An alternative possibility is that the direct link between structured access and syntactic dependencies is not the correct generalization. This would suggest that there is an extra condition on the

deployment of structured access in online processes that that subject-verb agreement does not meet.

*Agreement as uninterpreted syntax*

The nature of the syntactic representation of agreement has been the subject of an intense amount of linguistic research in recent years (Chomsky 2000; Wechsler & Zlatic 2000; Bhatt 2005; Corbett 2006; Badecker 2007; Bobaljik 2008; Baker 2008, among many others). Depending on which model of the subject-verb agreement dependency one assumes, agreement may not present an exception to the structured access hypothesis after all. Thus whether or not agreement is a true exception to the generalization in (5.12) turns on assumptions about its syntactic representation. Government & Binding models of agreement modeled the dependency as fundamentally similar to an anaphoric relation (e.g. Chomsky 1981), or a chain formed in the syntax between the agreeing element and the functional head hosting the agreement morphology (Rizzi 1990). On this view, it is difficult to explain the differences between reflexives and agreement, as they are more similar than not in terms of grammatical representation. However, other models of agreement model the dependency in quite a different manner. A range of current theories of agreement, together with the hypothesis of structured access, may be compatible with the apparent exceptional behavior of agreement.

For instance, agreement features may in fact be more central to the construction of syntactic structure than suggested by early GB models. In Minimalist

models, the agreement operation has been elevated to the status of a crucial structure-building operation (Chomsky 2000; Bhatt 2005; Baker 2008). This model suggests that in Table 5.7, agreement is exceptional in that it is the one dependency where phi-features are explicitly linked in the syntax between the head (the subject) and tail (T) of the dependency. If the dependency between the subject and the inflectional morphology is directly licensed by the feature match, as suggested by this model, then this has the effect of pushing the phi-features involved in agreement into an essential, structural role. On this view subject-verb agreement is exceptional in that it the phi-features are syntactically represented in the construction of the dependency. Since they are directly involved in licensing the syntax, the phi-features may be used to index information in subject-verb dependencies in line with (5.12).

This stands in contrast to other syntactic dependencies in Table 5.7, where the role of the phi-feature in structure building is somewhat more indirect. For instance, no model of reflexives directly involves a syntactically represented feature match between the reflexive and its antecedent in the same fashion. The closest model is that of Reuland (2001), who proposes that the anaphor relies on the same AGREE operation to resolve its antecedent. However, even on this AGREE-based model of reflexive dependencies, the agreement relation is not directly between the antecedent and the reflexive; rather, the two elements are indirectly related through verbal morphology. Thus the phi-features in subject-verb agreement may arguably be more narrowly syntactic than in other dependencies, licensing the use of these morphological features in first-pass dependency construction in a structured access model.



Alternatively, one might relegate agreement to a post-syntactic operation, as proposed by Bobaljik (2008). Bobaljik suggests that rather than being a fundamental structure-building operation, verbal agreement is actually an entirely post-syntactic phenomenon. If true, this would exempt agreement processes from the structured access hypothesis, as they would no longer represent strictly syntactic processes. Bobaljik's argument is constructed on cross-linguistic observations about the distribution of agreement processes in language. In particular, when syntactic position and morphological case diverge, agreement tracks morphological case rather than syntactic position, almost without exception. Bobaljik casts this as an 'order of operations' argument: if agreement makes reference to information that is picked up post-syntactically (phonological expression of case), while ignoring information that is presumably present in the syntax (syntactic position), then this may be taken as evidence for the conclusion that agreement morphology is licensed post-syntactically. For example, he notes that agreement in Hindi is blocked whenever an NP is marked with an overt case marker. However, on the assumption that the case marker is added in a post-syntactic morphological component of the grammar, then Hindi agreement must be calculated post-syntactically. Compared to the standard Minimalist model (Chomsky 1998), this claim posits a rather different role for agreement in the organization of the grammar. For the hypothesis of structured access, however, either model has the same impact of the relationship of subject-verb agreement to the statement of structured access in. They both make sense of its exceptional use of phi-features in accessing memory, either by making phi-features directly syntactically

active (Chomsky 2000) or relegating them to an extra-syntactic computation (Bobaljik 2008).

Deciding between narrowly syntactic and post-syntactic models of agreement is of interest for theories of online processing as well as grammatical models. However, it is difficult to understand the exceptional online behavior of subject-verb agreement only by reference to its role in a grammatical model; to do so would require additional linking assumptions about the relationship between the grammar and its online implementation, and this level of separation makes it difficult to derive predictions about online processing directly from the grammatical model.

Alternatively, it may be the case that the direct link between syntactic dependencies and structured access suggested above is not the correct generalization about the range of dependencies that employ structured access. Instead, it may be necessary to critically reconsider this generalization, which could prove insightful in constructing a theory of the use of structure in sentence processing.

#### *Structured access as an optimal access strategy*

Although the predictions of the structured access hypothesis with respect to the processing of subject-verb agreement do depend on the exact model of subject-verb agreement that is assumed, reconsidering the generalization about structured access may generate greater understanding the exceptional status of subject-verb agreement in English. The initial generalization about the deployment structured access appears to reflect a hard constraint that provides a direct link between

grammatical constraints and online syntactic generation. However, there is no direct evidence for this immutability of structured access as a parsing strategy.

One way of understanding the origin of structured access is to note that it is often a good parsing strategy. This follows from the reasonable assumptions that a) the parsing architecture is prone to memory interference, and b) comprehenders are functionally motivated to choose retrieval cues that maximize parsing speed and accuracy. In choosing a set of optimal retrieval cues for any given dependency, there is functional pressure to minimize use of cues that are potentially more misleading than they are helpful. This could be understood as a rational strategy, wherein comprehenders deploy the retrieval cue set that jointly minimizes retrieval error and processing time. If this hypothesis is correct, then the structured access generalization may be as in (5.12):

(5.12) **Structured access:** in constructing a long-distance dependency, the parser employs all and only the information needed to jointly minimize interpretive error and processing time. Structured access arises for interpreted, syntactically constrained dependencies.

On this view, structured access mechanisms for memory retrieval might be understood as reflecting good parser design. The two crucial elements that seem to be implicated in cases of demonstrable structured access are i) interpreted content and ii) structural constraints on the position of the retrieval target. Neither appears to be a sufficient condition on its own. (5.12) suggests this may be ‘rational’ parsing, but this notion needs to be understood with caution. What counts as ‘optimal’ or ‘rational’ parsing necessarily depend on the exact formulation of the objective function to be

optimized, and absent a formalization of such a function for the task of memory retrieval in parsing, it's difficult to mount such an argument. Nevertheless, it seems reasonable to assume that the parser is seeking to maximize both speed and interpretive accuracy, and for interpreted, structurally constrained dependencies, speed and accuracy trade off against each other. Structured access can be slow, as seen in Chapter 4, but it is possible that the average slowdown due to structured access is offset by the gain in interpretive accuracy. This is not true for structurally unconstrained dependencies (where structured access simply slows down processing), or for uninterpreted dependencies (which by hypothesis have no impact on measures of interpretive accuracy). If this is true, then structured access may be understood as a good strategy for the parser to pursue for certain dependencies.

Note that cases where the parser appears to disregard helpful information contained in formal gender or number cues have been used to motivate syntax-first or modular models of comprehension to varying degrees (see, e.g., Frazier & Clifton 1996; De Vincenzi 1999; van Berkum et al 2000; Brysbaert & Mitchell 2000). The informational encapsulation suggested by these cases appears to run counter to the architectural commitments of highly interactive, constraint-based parsing models (MacDonald et al 1994). However, it has been argued previously that these data do not necessarily cut in favor of innately modular architectures. In their consideration of the 'grain size' problem for experience-based models of parsing, Mitchell and colleagues (1995) argued that it is a mistake to assume that the 'right' thing for a parser to do is to use all information it has available to it. The generalization in (5.12) suggests a very similar conclusion: the best strategy is not to use all possible sources

of information for all parsing decisions. The best strategy is to use only those sources of information that are necessary to accomplish the task at hand; in the case of parsing, this means recovering the intended interpretation of the sentence quickly and reliably.

Mitchell, Cuetos, Corley and Brysbaert (1995) argued that structural abstraction in parsing routines might reflect the best strategy given a language user's experience. In particular they reasoned that for certain parsing decisions, it is rational for the parser to track statistics at an abstract structural level, rather than at a fine-grained lexical level. This is because tracking statistics at a lexical level and using that to guide parsing decisions (as suggested in McDonald et al 1994) risks leaving the comprehender with a sparse data problem. For example, there just are not enough instances of all lexical items in an ambiguous NP-P-NP-RC structure to get a solid estimate of the lexical bias for nouns in these structures. This lack of data would suggest that the best strategy for the parser is to back off to more robust estimates about the most probable attachment height over more abstract categories, for which a more robust set of data is available.

Though cast in a different framework, the generalization presented in (5.12) may be understood as supporting a very similar argument. Retrieval-and-interference based models of comprehension offer another means of understanding the biases at work in shaping the parser's decision metrics. For long-distance dependencies that rely on memory retrieval, the richer the information contained in the cue set, the more likely it is that some element in memory will interfere with the retrieval process. It is a bad strategy to use superfluous or highly correlated cues in retrieval, as they

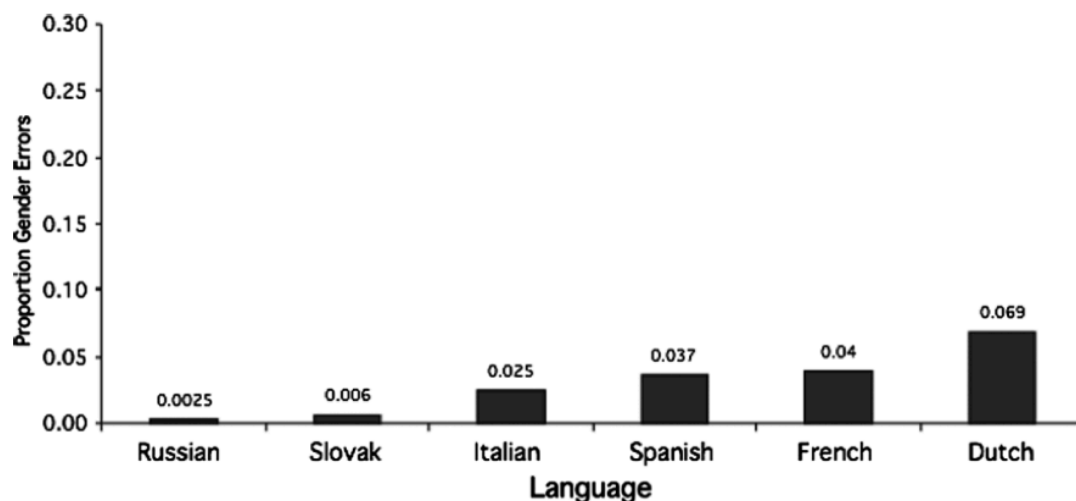
actually have the potential to negatively impact parsing through interference and retrieval of non-target memories. In such a model it is advantageous for the parser to adopt a small number of highly distinctive cues. For a reflexive dependency, lexical cues such as gender, or semantic cues such as animacy, are very highly correlated with structural cues. By deploying these superfluous cues in retrieval, the parser unnecessarily risks memory interference from non-target memory elements that carry those gender or animacy features. These superfluous increase the risk of interference, with no apparent benefit in memory retrieval. If the parser is optimized to use the minimal, distinct set of cues for retrieval, then it will display structured access behavior for reflexive dependencies, and a bias for structural over lexical cues in parsing decisions.

While this approach makes sense of the general use of structured access in parsing, it does not yet address the puzzling behavior of subject-verb agreement in English. As suggested above, the key features for a dependency appear to be interpretive content and structural constraints on retrieval positions. If this generalization is correct, then the question becomes why agreement's lack of interpretive content makes structured access no longer a good strategy. As suggested above, since agreement in English is plausibly inert with respect to any measure of interpretive accuracy, speed and accuracy don't trade off against each other in the same way that they do for reflexive dependencies. If comprehenders are trying to maximize interpretive accuracy, subject-verb agreement in English may just not be processed in a particularly structurally sensitive way.

A secondary question is how comprehenders come to distinguish the access strategies employed for interpreted structural dependencies (such as reflexives) and uninterpreted or nonstructural dependencies (subject-verb agreement or cross-sentential anaphora). It may be that these are universal principles of parsing that could be made to follow from the role of the dependencies in the grammar. Another possibility is that comprehenders develop the distinction through a process of cue optimization. English agreement carries a low functional load, marking only a subset of person / number distinctions in a subset of verbal paradigms. The somewhat restricted nature of English agreement might make agreement interference benign from the point of view of constructing an interpretation from a syntactic parse. The generalization in (5.12) suggests that this is the key reason that agreement in English does not employ structured access. If a cue optimization process is at work, then in the absence of clear interpretive effects, agreement in English might not generate the same sort of error signal necessary to cause the parser to adopt a structured access strategy. In general, structured access is slower than feature-based direct access. There is no pressure to minimize interference as there is no risk of generating an incorrect interpretation, and so the English parser may be optimized to compute agreement as quickly as possible even if the occasional retrieval error arises. This conjecture is consistent with empirical findings about the interpretive reflex of agreement attraction in English. It has been argued that incorrect agreement computation does not adversely impact comprehenders' interpretations (Lau, Wagers, Stroud & Phillips 2008).

If the exceptional behavior of English agreement is due to its low functional or interpretive load, then cross-linguistic investigation should reveal that a greater functional load for agreement causes a narrower, more accurate set of retrieval cues to be used in constructing agreement dependencies. Structured access should be evident in agreement for languages where there is functional pressure to minimize agreement interference. Preliminary investigation suggests that this prediction is borne out crosslinguistically. Lorimor, Bock, Zalkind, Sheyman & Beard (2008) investigated attraction errors in Russian, a free word order, pro-drop language. They argued that the rich morphology in Russian aids comprehenders in filtering out agreement errors, a point that was also argued by Badecker & Kuminiak for Slovak (2007). A cross-linguistic comparison of gender attraction effects reveals that attraction is inversely correlated with functional load, as shown in Figure 5.6. Languages that may rely on verbal agreement as a cue to interpretation, due to scrambling or argument dropping properties, show a diminished rate of attraction; languages with all but moribund gender systems, such as Dutch, are at the complete opposite end of the spectrum, and they show a high rate of gender agreement interference.





**Figure 5.6:** Rate of occurrence of gender attraction errors across languages. Figure from Lorimor, Bock, Zalkind & Sheyman (2008).

If comprehenders optimize their retrieval cue sets to increase accuracy in response to disruptive interference, these patterns of cross-linguistic variation are expected. On this view agreement interference should be all but absent in languages like Hindi, which has an agreement system that alternately agrees with the subject or the object. The variation in the target position, the system's dependence on overt case cues, and the scrambling, pro-drop nature of Hindi makes a narrow, entirely structured access procedure for constructing agreement dependencies a smart choice for accurately constructing the agreement dependency. Native speaker intuition suggests that no agreement attraction obtains for Hindi, across a range of structures and feature configurations (Rajesh Bhatt, Shravan Vasishth, p.c.). In contrast, agreement attraction for English speakers is intuitively evident. If further investigation upholds this intuition about attraction in Hindi, then a optimized-cue account of structured access is supported.

The crucial generalization that appears to emerge from the body of evidence thus far is that structured access appears to be engaged for dependencies that are structurally constrained and interpreted. The accounts for the apparent exceptional nature of agreement processing: although it is structurally constrained, it is not interpreted and so does not engage structured access. I suggested that this generalization might be understood as reflecting an optimal parsing strategy: because interpretive accuracy matters for online parsing, there may be pressure to minimize interference through structured search. However, at this point the source of this strategy remains unclear: it may be a general principle relating different sorts of grammatical rules to online processing routines, or it may reflect an experience-based cue optimization process. Further research on the relationship between interpretive load of agreement and its interference profile is necessary before drawing any firm conclusions.

#### *Structured access and Mandarin anaphors*

The hypothesis of structured access as an adaptive strategy endorsed here weakens the direct link between grammatical processes and online parsing procedures suggested in the original statement of structured access. Morphological or semantic constraints can either be satisfied in first-pass structure generation (e.g. English subject-verb agreement) or after the construction of a dependency (e.g. English reflexives). If features are classified in one way or another based on an extra-grammatical decision principle, then it is not clear that the processing evidence

presented here necessarily entails that *ziji* and *ta-ziji* are qualitatively different anaphors, one (broadly speaking) semantic or pragmatic, and the other syntactic.

The contrast between *ziji* and *ta-ziji* observed in Experiments 8 and 9 instead means that whatever the grammatical distinctions between the two there might be, the optimal operationalized content of *ta-ziji*'s structure-building cues should include the notion of animacy. If, as Bergeton's (2007) account suggests, *ta-ziji* is more likely to be understood as a contrastive, discourse-bound anaphor, then using non-structural, discourse-based cues in the retrieval cue set is a very reasonable strategy. On this view, the relevant constraint is not necessarily stated in terms of syntactic position. If a structured access strategy were employed, very low accuracy would result. Specifically, the under-sampling of the possible antecedents that would come from structurally narrowing the search space would lead to very low recall (though precision could in principle be quite high).

The contrastive pronoun model of *ta-ziji* was supported by the acceptability judgment task, where it was seen that multiple animate noun phrases for *ta-ziji* actually led to a decrement in acceptability relative to conditions with only a single antecedent. This may reflect the fact that contrastive pronouns such as *ta-ziji* are less felicitous in an out-of-the-blue discourse context with two possible antecedent antecedents. A similar contrast obtains in emphatic reflexive usage in English:

- (5.14)           a. John told Mary that she herself had failed the exam.  
                  b. #John told Bill that he himself had failed the exam.

However, even accounts that maintain that *ta-ziji* is entirely bound in the syntax may be compatible with the feature-based direct access view. For example, Pan (1998) explicitly argues that contrastive *ta-ziji* is a distinct element from bound *ta-ziji*. He notes that bound *ta-ziji* can take long-distance antecedents just in case there are no local, animate antecedents, and that the relevant constraints are best understood as tree-geometric in nature. He proposes that the unique feature of *ta-ziji* is that its binding domain is computed relative to the phrase-structurally closest, most prominent (on Pan's account, this means animate and human) antecedent. On this view, it seems also very likely that the optimal access procedure would include a notion of prominence in the memory retrieval cues. If this is the relevant statement of the constraints, then it seems that good parsing accuracy could be achieved by retrieving on a mixture of structural and prominence-based cues. On this view, though *ta-ziji* is a syntactically licensed anaphor, the range of structures in which it is licensed is determined by the semantic feature content of its antecedents. This is unlike reflexives or agreement, where the feature content and the structural restrictions are orthogonal. For this reason, in this model of *ta-ziji*, the structure and feature content are not in any sense redundant, and so leveraging both sources of information to compute its reference is an effective strategy.

#### *The footprint of structured access*

In the absence of a formal model of the cue optimization process, it is difficult to rigorously evaluate the claim of structured access as an adaptive strategy based on

the data given above. For now, the level of explanation for various patterns of access for phenomena such as English subject-verb agreement, Dutch RC attachment, and Mandarin *ta-ziji* is somewhat superficial. I have offered plausible explanations for why they pattern the way they do, but these depend on assumptions about the distribution of these dependencies, and the utility of various access strategies (i.e. the specific objective function to be optimized). A more thorough account for any given dependency involves investigating the level of accuracy (precision and recall) obtained by using any given combination of retrieval cues to complete a long-distance dependency. Presumably, any claims of ‘optimality’ about the best cue set for a given dependency should be derived from statistics of use, accuracy of retrieval, and a cost function for incorrect retrievals. In the interest of providing clear predictions, formally developing this insight is an important goal for future research.

In (5.12), I offered a generalization that appears to describe the dependencies that engage structured access for memory retrieval: structurally constrained, interpreted dependencies. This generalization generates predictions about other dependencies where one might expect to find evidence of a structured access strategy. As suggested above and elsewhere (Wagers 2008; Lau 2009), it is likely that predictive strategies may dominate for dependencies that can be reliably anticipated ahead of the putative memory access site. This feature makes it difficult to draw clear conclusions about the nature of retrieval in these dependencies (Wagers 2008; Martin & McElree 2008, 2009). A productive strategy is to search for evidence of structured access for structurally constrained, retrospective dependencies. Bound variable anaphors of all sorts provide an excellent case in point: they are subject to significant

structural constraints, and are not generally predictable ahead of the integration site (the variable). Structured access and search is a good strategy in this situation. In order to get an accurate parse, and to minimize disruptive interference, it is worth more narrowly searching licit syntactic positions.

There are a number of other dependencies that I have not considered that might provide further evidence for structured access in comprehension. One example is *wh*-in-situ dependencies, which could be understood as the retrospective counterpart of *wh*-movement dependencies in languages like English. *Wh*-in-situ occurs in languages like Mandarin Chinese, who leave *wh*-elements in their base position (i.e. *Mary saw who?*) rather than move them to a scope position. On most models the in-situ *wh*-dependency involves a long-distance relationship between the site of the interrogative operator that determines the scope of the *wh*-element, and the position where the *wh*-element is thematically integrated. One way to model this dependency is with covert LF movement (Huang 1982; Cheng 2009), where the *wh*-element is moved covertly to occupy the scope position. Alternatively, it may be modeled as unselective binding (Aoun & Li 1993; Tsai 1994), where the *wh*-element is converted to a variable in situ and bound by a *wh*-operator with the appropriate scope. On either account, the position of the scope operator must be retrieved at the point of processing the in-situ *wh*-element. In this sense, both accounts model *wh*-in-situ as a retrospective dependency between the *wh*-element and the structural position where *wh*-scope is marked. It has been shown that in Chinese (Xiang, Dillon & Wagers 2010) and Japanese (Sprouse, Fukuda, Ono & Kluender 2011) that processing *wh*-in-situ is costlier for embedded *wh*-elements than it is for local ones.

Sprouse and colleagues argue that this is due to a backward structured search. However, using SAT, Xiang et al (2010) found no evidence of structured search in long-distance *wh*-in-situ dependencies. In their data, the difficulty associated with long-distance *wh*-construals was reflected in decreased accuracy, not decreased speed or dynamics parameters.

However, languages such as Chinese or Japanese, for which *wh*-in-situ is the primary method of constructing *wh*-dependencies, might not provide the best test case for structured access in *wh*-in-situ due to comprehenders simply optimizing their processing to access the matrix [Spec,CP] in all cases where interrogative force is required. In light of this, French is interesting in that it permits a mixture of *wh*-fronting and *wh*-in-situ strategies for question construction (Oiry 2011). Oiry (2011) notes that there are conflicting reports on the acceptability of long-distance (embedded) *wh*-in-situ questions such as (5.14), with some authors claiming they are not grammatical and others claiming they are (see discussion in Oiry 2011). Native speaker informants appear to have conflicting judgments about these sentences in out-of-the-blue contexts, though they appear to be accepted more often than not.

- (5.14) ?? *Tu penses que le policier a fait quoi?*  
You think that the policeman has done what  
“What do you think the policeman did?”

One possibility is that the mixture of strategies in French has led to a situation where comprehenders have not fully optimized the access procedure for *wh*-words in embedded clauses. If they simply searched for [spec,CP] to construct the appropriate scope for the *wh*-word, then they would have perfect performance in parsing matrix

*wh*-in-situ questions, but in examples like (5.14), there would be disruptive interference from the embedded [spec,CP] position. This leads to the possibility that the mixed pattern of results that has been reported on sentences like these reflects interference-related processing difficulty due to multiple potential landing sites for the *wh*-element. This suggestion remains highly speculative, but the contrast between French and Chinese-type *wh*-in-situ could shed light on the structured nature of *wh*-in-situ processing, as well as the role for cue optimization in this process.

Another interesting area of investigation is the attachment of relative clauses to complex NPs. Cuetos and Mitchell (1988) originally suggested that languages vary arbitrarily with respect to their preferences in attaching relative clauses in ambiguous NP-PP-NP environments such as *the servant of the actress who was on the balcony*. They showed that while English speakers reliably interpret the RC as modifying *the actress* (in accordance with *late closure*, Frazier 1978), Spanish speakers preferred to attach the RC to the higher of the two NPs in Spanish. This cross-linguistic different behavior was later modeled as a sort of parsing non-determinism (Frazier & Clifton 1996, 1997; Kamide & Mitchell 1997), whereby comprehenders are able to delay full commitment to a parse of the relative clause's attachment pending extra disambiguating information. Frazier and Clifton (1996) noted that semantic and discourse factors, such as focus accent or referentiality, impacted attachment preferences.

An alternative possibility is suggested by the structured access hypothesis. It may simply be that the retrieval cues in these environments reflect only the fact that in a certain context, the RC must be attached to an NP in a certain structural



environment (e.g., a thematic domain, Frazier & Clifton 1996). This opens up the possibility that non-determinism impacts the attachment process at the level of memory retrieval, which is by hypothesis stochastic and indirectly impacted by a number of processes that modulate the relative activation of the two NPs in memory. On this account, the reason that comprehenders prefer to attach to the focused NP is that the accent conveys an activation advantage, which translates into a greater probability of retrieval at the point of constructing the NP-RC dependency.

Interestingly, there seems to be some evidence that ‘general’ memory dynamics are at work in selecting the site of attachment for RCs. Gibson, Pearlmutter, Canseco-Gonzalez & Hickok (1996) showed that in sentences where an RC can be attached to one of three NP positions in a noun phrase (*the key to the house on the hill*), the most recent and most distant NPs are preferred over the middle NP position. This is compatible with the serial position effects noted in recall, where it is known that the first and last elements are more accurately recalled due to effects of primacy and recency (Murdock 1962). Gibson and colleagues argue that there are problems with applying the simple list recall model to the RC data. This is in part due to different preferences for primacy and recency across constructions within Spanish; for VP-level attachments, such as temporal adverbials like *yesterday*, comprehenders prefer to attach low, whereas for NP-level RC attachments, they prefer to attach high. It is unclear if this is a real challenge, as it is unclear the processes involved in the two cases are the same. They furthermore note that it is hard to reconcile this view with the between-languages differences. However, if this can be tied to independently observable differences in how Spanish and English speakers distribute their attention

across a complex noun phrase structure, then a memory dynamics account of this ambiguity could actually provide a principled way of determining which languages will show high or low attachment preferences. This is a speculative possibility, and further work is necessary to determine whether or not RC attachment preferences track activation in memory.

### *Conclusion*

In this chapter, I have attempted to sharpen the hypothesis of structured access in comprehension. By comparing similar but distinct anaphors in Mandarin Chinese, I showed that structured access is not a property of interpretive dependencies. Instead, the correct generalization appears to be that structured access is engaged for dependencies that are both interpreted and structurally constrained. I suggested instead that this may reflect an adaptive strategy that reflects a pressure to minimize interference error for interpreted dependencies. By adopting the minimal set of distinctive cues necessary to arrive at a grammatical interpretation of a sentence, comprehenders minimize interference and guard against misinterpretation. On this account, agreement does not use structured access mechanisms in retrieval because it is an uninterpreted dependency, and thus does not impact interpretive accuracy (at least in English).

It has long been argued that abstract parsing principles might reflect functional pressures from the memory architecture of the parser (Frazier 1978; Frazier & Fodor 1978; Mitchell et al 1995), and the arguments provided here make a

similar point. If comprehension is prone to interference, a good strategy is to minimize the information used to access memory in the interest of minimizing distraction from irrelevant memory elements. Interference effects have been of intense research interest in recent years (Gordon et al 2001; Lewis & Vasishth 2005; Vasishth et al 2008), and the arguments provided here suggest that they may in fact provide another functional pressure for the parser to construct structure with reference only to abstract structural categories and positions. Thus structured access might reflect a simple pressure for the parser to get it right as often as possible, and in the shortest amount of time possible.

## Chapter 6: Conclusion

The main claim of this thesis was that for certain long-distance dependencies, comprehenders engage a purely structural mechanism for accessing linguistic working memory. This conclusion is compatible with various architectures: in the context of a content-addressable memory architecture, for example, structured access could be implemented by only utilizing structural information as retrieval cues. More generally, structured access is a claim about the type of information that comprehenders use to assemble long-distance dependencies online. In this thesis I provided computational and experimental experiments in support of this claim, demonstrating that comprehenders selectively attend to structural information over morphological or semantic information while accessing linguistic memory in the course of sentence comprehension.

At the outset, one of the major questions of interest was the relationship between the grammar and the parser. There have been a number of important challenges to the idea that fully grammatically elaborated structure is constructed online, and the challenge from constraints in working memory was the focus of this work. I have argued that despite some apparent limitations of the memory

architecture of sentence comprehension, the parser is nonetheless able to effectively target and retrieve particular syntactic positions in order to construct long-distance dependencies online. More generally, however, the arguments I have presented here form a bounding condition on the grammatical fidelity of the structures constructed by the parser: they must be elaborated sufficiently for the parser to access particular elements in a grammatically sophisticated manner, even if the parser does not always make full use of that structure. If true, this fidelity poses a challenge for models that assume a more indirect or heuristic relationship between online parsing processes and grammatical knowledge (Townsend & Bever 2001; Ferreira et al 2002; Ferreira & Patson 2007).

There were two specific arguments that I presented to make the case for a narrowly structural access mechanism in linguistic comprehension. I review these first, then turn to a brief consideration of the scope of structural access, and implications for the memory architecture of language comprehension.

#### *No consideration of illicit antecedents: Experiments 1-5*

The first empirical argument for structured access focused on the processing of reflexive anaphors in English. An experimental investigation of the processing profile of English reflexives suggests that the initial candidate antecedent set considered by the parser is narrower than expected if a feature-based access mechanism is used; only the local subject is retrieved, regardless of the feature content of other NPs in the sentence. This profile was contrasted with English

agreement, and the impact of feature-matched, inaccessible nouns for both dependencies was investigated in three eye-tracking experiments (Experiments 1-3). In immediate online processing, agreement dependencies retrieved inaccessible noun phrases that had the correct morphological features, building illicit agreement dependencies and causing illusion of grammaticality effects in behavioral measures. In contrast, reflexive anaphors showed a qualitatively different access profile: they were only sensitive to a feature match with the local c-commanding noun phrase.

This provides the first empirical argument for structured access: for certain dependencies, there is no consideration of inaccessible noun phrases during memory access. Possible objections to this conclusion were addressed with computational modeling in Experiments 4-5, which supported the claim that reflexives build the dependency with their antecedent using only structural information.

*No immediate access to distant but accessible antecedents: Experiments 6-9*

Experiments 6-9 advanced the second major empirical argument for structured access. In cases where there are multiple nouns that could contain an antecedent for an anaphor, the initial candidate antecedent set is not narrowed by the feature content of those NPs. Thus even in the presence of structurally accessible, feature-matched antecedents for the Chinese reflexive *ziji*, the parser instead initially retrieves local, feature-mismatched antecedents. The *ziji* findings provide a complementary piece of evidence for structured access. To account for the fact that the local subject is considered first, memory access must make reference to only structural information,

rather than incorporating semantic or morphological feature constraints immediately. Thus results from the processing of both Chinese and English reflexives provide an argument in favor of a uniquely structural access mechanism in processing reflexive dependencies.

Experiments 8 and 9 directly contrasted the processing of two distinct Chinese anaphors: *ziji* and *ta-ziji*. There were two main conclusions from this comparison. First, it was shown that it is not interpretive content alone that leads to structured access; the anaphor *ta-ziji* immediately accesses structurally distant antecedents based on their semantic features. In addition, this comparison provided experimental evidence that it is not linear position in the string that causes structured access effects, as *ziji* and *ta-ziji* occupied the same linear position in the string,

#### *Interpreting interference effects*

In Chapter 3 a consideration of the computational basis of the interference model here led to the conclusion that a demonstration of facilitatory interference is necessary to conclude that the parser has constructed illicit structure. Due to the rational basis of the ACT-R model, it was seen that this is true of any model that links probability of a structure or retrieval with behavioral indices such as reaction time. One way of understanding this result is that the consideration of grammatically unlicensed parses leads to a situation of ‘spurious ambiguity’, where, due to interference, the parser is jointly considering grammatical and ungrammatical parses alike. Facilitation occurs in this situation for the same reason that ambiguity eases

processes in truly ambiguous situations: given multiple options and a ‘race’ to adopt the fastest option to be constructed, the observed reaction time distribution represents a mixture of the fastest processing times from either option, showing facilitation effects on average.

Given this finding, I argued that the evidence for online consideration of ungrammatical parses is sparser than is generally assumed. Of the range of interference results that have been reported, there are only two dependencies that reliably show facilitatory interference: subject-verb agreement and negative polarity item dependencies. The interpretation of NPI interference as the result of partial-matching facilitation has been questioned (Xiang et al 2009), leaving subject-verb agreement (primarily in English) as the parade case of ungrammatical structure construction online. Most online effects that have been attributed to interference are inhibitory, which does not provide online evidence for consideration of illicit structures. Instead, much of the experimental evidence that has been taken to indicate incorrect structure generation comes from offline measures of interpretation, which is a more indirect measure.

### *Structured access and the architecture of comprehension*

I argued at several points that the claim of structured access is compatible with a variety of architectures. In Chapter 3 I presented an implemented computational model of structured access in a content-addressable memory architecture, and in Chapter 5 I argued that structured access might be understood as



an optimal adaptation to memory interference in this framework. The need to provide a grammatically faithful parse provides a functional pressure for abstract structural cues, as they are on average the most reliable and effective cues to memory access or structure generation; similar arguments about the functional advantage to abstraction in structure generation were presented by Mitchell and colleagues (1995).

However, an alternative possibility is that structured access reflects the deployment of a qualitatively different memory access mechanism. That is, it may reflect the deployment of a memory indexing system that indexes memories only according to their position, serially traversing all positions to find the target memory for any given retrieval operation. This possibility is similar to the observation that there are two types of retrieval mechanisms that are deployed for general memory tasks, depending on the nature of the information that needs to be recalled (McElree & Doshier 1993).

If one assumes that structural cues can be defined to target arbitrarily precise syntactic positions (an assumption that is far from obvious), the predictions of a content-addressable architecture that uses narrowly structural search cues and a serial architecture that literally traverses a parse tree node-by-node align in almost all cases. There are a number of possibilities for deciding between these possibilities, however.

#### *Structured access as an optimal adaptation*

One argument for narrowly structural search cues depended on the assumption of a content-addressable architecture. I suggested in Chapter 5 that they could

represent an optimal behavioral policy for a parser that a) is prone to interference but b) nonetheless needs to recover only grammatically licensed parses. On the assumption that the cues used to access memory in a content-addressable architecture are subject to tuning or optimization, I suggested that the parser should only adopt abstract structural cues for certain dependencies, such as English reflexives and Mandarin *ziji*.

In order to make a stronger case for this, this informal reasoning should be complemented by formal modeling. This would more firmly demonstrate that the narrow use of syntactic cues is the best course of action to take, given the constraints on the parser. The question of the optimal behavioral policy to maximize reward is a general problem that has been intensely studied in computer science under the name *reinforcement learning* (see, e.g. Sutton & Barto 1998). These techniques can be used as an informal proof by simulation that a given behavioral policy is optimal given a space of possible actions. If the reasoning in Chapter 5 is correct, then such an analysis should show that the optimal policy is to ignore the correlated but potentially disruptive information contained in morphological cues.

A formal demonstration of this argument using existing computational tools would provide a proof of principle that would significantly bolster the informal arguments I have provided here. This would provide support to the content-addressable architecture implementation of structured access by demonstrating that the arguments are in fact sound. More generally, however, it would suggest that adopting a content-addressable architecture provides a mean of explaining the origin of structured access in the parser.

### Blocking effects

Another potentially illuminating source of evidence about the memory architecture of the parser is the existence of *blocking effects*. The term ‘blocking effect’ is often used to refer to a particular class of constraints on Mandarin reflexives, but the crucial ‘blocking’ nature of these constraints is apparent in a range of other dependencies. To demonstrate the usefulness (and generality) of these effects for the present discussion, let us use ‘blocking effect’ to refer to any instance in which a long-distance dependency between two elements is disrupted by an element along the path between the two elements. Examples include the blocking of *wh*-movement by non-bridge verbs (Erteschik-Shir 2006) or definite noun phrases (Chomsky 1973), definite island effects in NPI licensing (Ladusaw 1979), quantifier intervention effects in NPI licensing (Beck 1996), and person blocking effects in Mandarin reflexives (Huang & Li 2009).

These effects have the potential to be informative with respect to the question of memory architecture. In retrieval-driven content-addressable architectures, dependencies are modeled as direct relations between two constituents in a parse. Interference in this relation is only expected to occur from other constituents that share feature content with either the probe or the target memory. For the blocking effects mentioned above, however, the dependency is disrupted by an intervening element that is not feature-matched to either the probe or the target. For example, non-bridge verbs such as *lisp* do not contribute interfering [+wh] features, so it is

unclear on what dimension they would interfere with the construction of the *wh*-dependency.

Given the nature of these constraints, an important empirical question is whether or not comprehenders are immediately sensitive to blocking elements in constructing a long-distance dependency. If so, then this would provide evidence for a serial architecture. This is because in order to be immediately sensitive to these blocking constraints, comprehenders must be consider syntactic positions that are not directly involved in constructing the dependency. In a serial architecture, this can be made to easily follow from the requirement that the parse tree be traversed node-by-node. In a content-addressable architecture, on the other hand, only constituents that have interfering content should be able to impact early dependency building processes. Thus, blocking effects such as the quantifier intervention effect should not be immediately evident to comprehenders. In a content-addressable architecture, it seems that these structures must be generated and then filtered out at a later stage of comprehension.

### Negative constraints

In addition to blocking constraints, an investigation of negative constraints on parsing operations has the potential to be informative about the architectural source of structured access. Negative constraints are restrictions on which syntactic positions cannot be considered for participation in a long-distance dependency. Principle B of

the binding theory (Chomsky 1981) is the most prominent example: the antecedent for a pronoun can be any NP that is not in its local binding domain.

Although positive syntactic constraints can arguably be accurately stated in a content-addressable implementation of structured access, it is not clear that such negative constraints can. In these architectures, the retrieval cues consist of a positive set of features that are matched against the content of items in memory; it is not clear that negative constraints are well-formed in such a system. An intuitive way of seeing this is to consider the effect of the instructing someone not to direct their attention to a prominent event; a more likely outcome is that this will increase the chances that they will in fact attend to the event.

Thus the control structure needed to accurately and quickly exclude positions from consideration is likely to require an alternative architecture. For this reason, the degree of accuracy in the online implementation of Principle B effects has the potential to be highly informative. There are mixed empirical results concerning how quickly comprehenders can exclude the local subject position from consideration, with some studies suggesting that there is no consideration of the local subject for free pronouns (Nicol 1988; Clifton, Kennison & Albrecht 1997; Chow, Lewis, Lee & Phillips 2011), and others suggesting that the local subject is in fact considered (Badecker & Straub 2002; Kennison 2003).

If comprehenders are able quickly exclude the local subject from consideration, then it is possible to mount an argument for the deployment of a more structured architecture to account for the immediate application of this negative constraint. Otherwise, if the parser is ultimately found to be unable to exclude the

local subject position from consideration in Principle B configurations, then support for a serial, complementary memory architecture is weakened; instead, the inability to quickly implement a negative syntactic constraint would point to a more general use of a content-addressable memory architecture.

### *Conclusion*

The link that has been made in recent years between working memory architectures and parsing has generated much insight into the fine computational properties of representation and memory during sentence comprehension. The present thesis contributes to this line of research; in particular, I have presented several experiments that demonstrate that despite the apparent widespread existence of interference effects, the parser can effectively target particular syntactic positions for retrieval. Recent advances in understanding of the memory architecture of the parser do not entail that the parser is architecturally constrained to entertain ungrammatical parses. On the contrary, an investigation into the scope of structured access raises the possibility that narrowly syntactic memory access actually arises as a functional adaptation to exactly this sort of noisy cognitive architecture.

## Appendix A: Retrieval schedules for models in Chapter 3

For each model simulation, a schedule of constituent creation times and a schedule of hypothesized retrievals was constructed. The time  $t$  at which a given constituent was created was estimated from the empirical reading times in Experiment 1; this time reflects the total amount of time that readers spent reading material to the left of the relevant constituent (i.e., average cumulative regression path duration). Retrievals tied to the processing of any given constituent, such as the retrieval of a subject upon processing a verb, occurred 200ms after the creation of the constituent that triggered the retrieval. The sole exception was the reflexive retrieval, which occurred at 300ms post processing to allow for extra time to attach the reflexive as the object of the verb.

The differences between the agreement and reflexive conditions were modeled only as differences in the feature makeup of the parse constituents and the feature specifications of the retrievals.

*Agreement conditions:*

[*The executive*]<sub>DP1</sub> *who* [*oversaw*]<sub>VP1</sub> [*the middle manager(s)*]<sub>DP2</sub> [*definitely*]<sub>ADV</sub> [*was*]<sub>BE</sub> ..

	DP1	VP1	DP2	Adv	BE
$t$	0	590	1116	1906	2330
Gender	<i>masc</i>	-	<i>masc</i>	-	-
Number	<i>sing</i>	<i>sing</i>	<i>sing/pl</i>	-	-
Category	DP	VP	DP	ADV	T
Role	[ <i>spec,T</i> ]	[ <i>head,T</i> ]	[ <i>comp,V</i> ]	[ <i>adj,T</i> ]	[ <i>head,T</i> ]
Local	+	-	-	+	+

**Table A.1:** Constituent creation times and feature makeup for agreement conditions.

	R1:VP1	R2:DP2	R3:BE
$t$	790	1316	2530
Gender	-	-	-
Number	-	-	<i>sing/pl</i>
Category	DP	VP	DP
Role	[ <i>spec,T</i> ]	[ <i>head,T</i> ]	[ <i>head,T</i> ]
Local	-	-	+

**Table A.2:** Schedule of retrievals and cue sets. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = critical retrieval to attach DP1 to BE.

*Reflexive conditions:*

[*The executive*]<sub>DP1</sub> *who* [*oversaw*]<sub>VP1</sub> [*the middle manager(s)*]<sub>DP2</sub> [*definitely*]<sub>ADV</sub> [*hurt*]<sub>VP2</sub> [*himself*]<sub>REFL</sub> ...

	DP1	VP1	DP2	Adv	VP2	REFL
<i>t</i>	0	580	1091	1850	2285	2624
Gender	<i>masc</i>	-	<i>masc</i>	-	-	-
Number	<i>sing</i>	<i>sing</i>	<i>sing/pl</i>	-	<i>sing/pl</i>	-
Category	DP	VP	DP	ADV	VP	REFL
Role	[ <i>spec,T</i> ]	[ <i>head,T</i> ]	[ <i>comp,V</i> ]	[ <i>adj,T</i> ]	[ <i>head,T</i> ]	[ <i>comp,V</i> ]
Local	+	-	-	+	+	+

**Table A.3:** Constituent creation times and feature makeup for agreement conditions.

	R1:VP1	R2:DP2	R3:VP2	R4:REFL	R5:REFL
<i>t</i>	780	1316	2530	2824	2924
Gender	-	-	-	-	-
Number	-	-	-	-	-
Category	DP	VP	DP	VP	DP
Role	[ <i>spec,T</i> ]	[ <i>head,T</i> ]	[ <i>head,T</i> ]	[ <i>head,T</i> ]	[ <i>spec,T</i> ]
Local	-	-	+	+	+

**Table A.4:** Schedule of retrievals and cue sets for the structured access reflexive model. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = attachment of DP1 to VP2; R4 = attachment of REFL to VP2; R5 = critical retrieval of REFL's antecedent.

	R1:VP1	R2:DP2	R3:VP2	R4:REFL	R5:REFL
<i>t</i>	780	1316	2530	2824	2924
Gender	-	-	-	-	<i>masc/-</i>
Number	-	-	-	-	<i>sing/pl</i>
Category	DP	VP	DP	VP	DP
Role	[ <i>spec,T</i> ]	[ <i>head,T</i> ]	[ <i>head,T</i> ]	[ <i>head,T</i> ]	[ <i>spec,T</i> ]
Local	-	-	+	+	+

**Table A.5:** Schedule of retrievals and cue sets for the feature-based access reflexive model. R1 = attachment of DP1 to VP1; R2 = attachment of DP2 to VP1; R3 = attachment of DP1 to VP2; R4 = attachment of REFL to VP2; R5 = critical retrieval of REFL's antecedent.



## References

- Alcocer, P., & Phillips, C. (2009). A cross-language reversal in illusory licensing. Poster presented at the 22<sup>nd</sup> Annual Meeting of the CUNY Conference on Human Sentence Processing, Davis, CA: March 26-28.
- Anand, P. (2006). *De de se*. Doctoral dissertation, Massachusetts Institute of Technology.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 5, 451-474.
- Anderson, J. R. (1989). A rational analysis of human memory. In Roediger III, H., and Craik, F., (eds.) *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving*: 195-210. Erlbaum, Hillsdale, NJ.
- Anderson, J. R. (1990). *The Adaptive Character of Thought*. Erlbaum, Hillsdale, NJ.
- Anderson, J. R. & Milson, R. (1989). Human Memory: An Adaptive Perspective. *Psychological Review*, 96, 703-719.
- Anderson, J. R., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Erlbaum, Mahwah, NJ.
- Anderson, J. R. & Reder, L. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128, 186-197.
- Antón-Mendez, I., Nicol, J., & Garrett, M. (2002). The relation between gender and number agreement processing. *Syntax*, 5, 1-25.
- Aoun, J., & Li, Y-H A. (1993). *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry*, 24, 199-238.
- Baayen, H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press, Cambridge, UK.
- Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Badecker, W. A feature principle for partial agreement. *Lingua*, 117, 1541-1565.
- Badecker, W., & 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.
- Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, 56, 65-85.

- Baker, C-L. (1995). Contrast, discourse prominence, and intensification, with special reference to locally-free reflexives in British English. *Language*, 71, 63-101.
- Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge University Press, Cambridge, UK.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5-17.
- Beck, S. (1996). Quantified structures as barriers for LF-movement. *Natural Language Semantics*, 4, 1-56.
- Bergeton, U. (2007). *The independence of binding and intensification*. Doctoral dissertation, University of Southern California.
- Berwick, R., & Weinberg, A. (1984). *The Grammatical Basis of Linguistic Performance*. MIT Press, Cambridge.
- Bever, T. (1970). The cognitive basis for linguistic structures. In Hayes, R. (ed.), *Cognition and Language*. Wiley & Sons, New York.
- Bhatt, R. (2005). Long-distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory*, 23, 757-807.
- Bobaljik, J. (2008). Where's phi? Agreement as a post-syntactic operation. In Harbour, D., Adger, D., & Béjar, S. (eds), *Phi-Theory: Phi Features Across Interfaces and Modules*. Oxford University Press, Oxford, UK.
- Bock, K. & Miller, C. (1991). Broken Agreement. *Cognitive Psychology*, 23, 45-93.
- Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, 40, 330-346.
- Bock, K., Eberhard, K., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, 51, 251-278.
- Bock, K., Butterfield, S., Cutler, A., Cutting, J. C., Eberhard, K., & Humphreys, K. Number agreement in British and American English: Disagreeing to agree collectively. *Language*, 82, 64-113.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: dissociating phrase structure from relational structure. *Journal of Memory and Language*, 51, 495-522.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113, 787-821.
- Brysbaert, M., & Mitchell, D. (2000). The failure to use gender information in parsing: A comment on van Berkum, Brown & Hagoort (1999). *Journal of Psycholinguistic Research*, 29, 453-455.
- Büring, D. (2005). *Binding Theory*. Cambridge University Press, Cambridge.

- Burkhardt, P. (2005). *The Syntax-Discourse Interface: Representing and Interpreting Dependency*. John Benjamins, Amsterdam.
- Caplan, D., & Waters, G. (1998). Verbal working memory and sentence comprehension. *Brain and Behavioral Sciences*, 22, 77-126.
- Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, 46, 2028–2040.
- Chandler, J. P. (1969). Subroutine STEPIT—finds local minimum of a smooth function of several parameters. *Behavioral Science*, 14, 81–82.
- Cheng, L-S L. (2009). *Wh-in-situ*, from the 1980s to now. *Language and Linguistics Compass*, 3, 767-791.
- Chomsky, N. (1973). *Conditions on Transformations*. In Kiparsky, P., & Peters, S. (eds), *A Festschrift for Morris Halle*. Mouton, The Hague.
- Chomsky, N. (1977). *Essays on Form and Interpretation*. Elsevier, North-Holland.
- Chomsky, N. (1981). *Lectures on government and binding*. Mouton de Gruyter: Berlin.
- Chomsky, N. (1986). *Knowledge of language: its nature, origins and use*. New York.
- Chomsky, N., & Lasnik, H. 1993. The theory of principles and parameters. In Jacobs, J., Von Stechow, A., & Sterneveld, W., (eds.) *Syntax: An International Handbook of Contemporary Research, Vol. 1*. Mouton de Gruyter, Berlin.
- Chomsky, N. (2000). Minimalist Inquiries: The Framework. In Martin, R., Michaels, D., & Uriagereka, J. (eds), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*. MIT Press, Cambridge, MA.
- Chow, W-Y., Lewis, W., Lee, S., & Phillips, C. (2011). Immediate structural constraints on pronoun antecedent retrieval. Poster presented at the 24<sup>nd</sup> Annual Meeting of the CUNY Conference on Human Sentence Processing, Stanford, CA: March 24-26.
- Clifton, C., Kennison, S., & Albrecht, J. (1997). Reading the words *her*, *his*, *him*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36, 276-292.
- Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, 11-39.
- Cole, P., Hermon, G., & Sung, L-M. (1990). Principles and parameters of long-distance reflexives. *Linguistic Inquiry*, 21, 1-22.
- Cole, P., & Sung, L-M. (1994). Head movement and long-distance reflexives. *Linguistic Inquiry*, 25, 355-406.
- Cole, P., Hermon, G., & Lee, C-L. (2001). Grammatical and discourse conditions on long-distance reflexives in two Chinese dialects. In Cole, P., Hermon, G., & Huang, J. (eds), *Long Distance Reflexives*. Academic Press, New York.

- Corbett, G. (2006). *Agreement*. Cambridge University Press, Cambridge, UK.
- Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.
- Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, 73-105.
- De Vincenzi, M. (1999). Differences between the morphology of gender and number: Evidence from establishing coreferences. *Journal of Psycholinguistic Research*, 28, 537-553.
- Den Dikken, M. (2001). "Plurilinguals," pronouns, and quirky agreement. *The Linguistic Review*, 18, 19-41.
- Doshier, B. A. (1984). Discriminating pre-experimental (semantic) information from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519-555.
- Doshier, B. A., & Rosedale, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory and Language*, 30, 125-160.
- Drenhaus, H., Frisch, S., & Saddy, D. (2005). Processing negative polarity items: When negation comes through the backdoor. In: Kepser, S. and Reis, M. (eds.), *Linguistic Evidence-Empirical, Theoretical, and Computational Perspectives*. Mouton de Gruyter, Berlin.
- Eberhard, K., Cutting, J., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, 112, 531-559.
- Erteschik-Shir, N. (2006). Bridge phenomena. In Everaert, M., & van Riemsdijk, H. (eds.), *The Blackwell Companion to Syntax*. Blackwell/Wiley, New York.
- Fedorenko, E., Babyonyshev, M., & Gibson, E. (2004). The nature of case interference in online sentence processing in Russian. *NELS 34 Conference Proceedings*.
- Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.
- Ferreira, F., & Patson, N. (2007). The 'good-enough' approach to language comprehension. *Language and Linguistics Compass*, 1, 71-83.
- Fiengo, R., & May, R. (1994). *Indices and identity*. MIT Press, Cambridge, MA.
- Fodor, J.A., Bever, T., & Garrett, J. (1974). *The psychology of language: an introduction to psycholinguistics and generative grammar*. McGraw-Hill, New York.
- Foot, R., & Bock, K. (2011). The role of morphology in subject-verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*, forthcoming.

- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability versus accessibility. *Journal of Memory and Language*, 56, 357-383.
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, 17, 371-404.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Doctoral dissertation, University of Connecticut.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519-560.
- Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, 27, 123-146.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 291-325.
- Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: decision principles and structure in sentence comprehension. *Cognition*, 13, 187-222.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Frazier, L., Flores-d'Arcais, G. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331-344.
- Frazier, L., & Clifton, C. (1996). *Construal*. MIT Press, Cambridge.
- Frazier, L., & Clifton, C. (1997). Construal: Overview, motivation and some new evidence. *Journal of Psycholinguistic Research*, 26, 277-295.
- Friederici, A. (1995). The time course of syntactic activation during sentence processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, 50, 259-281.
- Friederici, A., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Brain Research: Cognitive Brain Research*, 1, 183-192.
- Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, 31, 45-63.
- Gallistel, R. (2009). The importance of proving the null. *Psychological Review*, 116, 439-453.
- Gallistel, R., & King, A. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Blackwell/Wiley, New York.
- Gao, L., Liu, Z., & Huang, Y.. (2005). Who is *ziji*? An experimental research on Binding Principle. *Linguistic Science*, 2, 39-50.
- Gelman, A., & Hill, J. (2005). *Data analysis using regression and multilevel /*

*hierarchical models*. Cambridge University Press, Cambridge, UK.

- Gernsbacher, M. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156.
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, 59, 23-59.
- Gibson, E., & Warren, T. (2004). Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55-78.
- Gillespie, M., & Pearlmutter, N. (2011). Hierarchy and scope of planning in subject-verb agreement production. *Cognition*, 118, 377-397.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-65.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during sentence processing. *Psychological Science*, 13, 425-430.
- Gordon, P.C., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 27, 1411-1423.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97-114.
- Gordon, P.C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 1304-1321.
- Gronlund, S. D., Edwards, M. B., & Ohrt, D. D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 1261-1274.
- Greene, S.B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 266-283.
- Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage*, 20, 18-29.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.
- Hagoort, P., Wassenaar, M. E. D., & Brown, C. M. (2003). Syntax-related ERP effects in Dutch. *Cognitive Brain Research*, 16, 38-50.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194-205.
- Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7, 391-

- Harris, T., Wexler, K., & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language*, 75, 313–346.
- Hartsuiker, R., Schriefers, H., Bock, K., & Kikstra, G. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, 31, 1316-1326.
- Häussler, J., & Bader, M. (2009). Agreement checking and number attraction in sentence comprehension: Insights from German relative clauses. *Travaux du cercle linguistique de Prague* 7.
- Hopf, J-M., Bayer, J., Bader, M., & Meng, M. (1998). Event-related brain potentials and case information in syntactic ambiguities. *Journal of Cognitive Neuroscience*, 10, 264-280.
- Hopfield, J.J. (1982). Neural networks and physical with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79, 2554-2558.
- Hornstein, N. (2007). Pronouns in a minimalist setting. In Corver, N., & Nunes, J. (eds.), *The Copy Theory of Movement*. John Benjamins, Philadelphia.
- Huang, C-T J. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Doctoral Dissertation, Massachusetts Institute of Technology.
- Huang, C-T J., & Tang, C-C, J. (1991). The local nature of the long-distance reflexive in Chinese. In Koster, J., & Reuland, E. (eds.), *Long Distance Anaphora*. Cambridge University Press, Cambridge, UK.
- Huang, C-T J., & Liu, L. (2001). Logophoricity, attitudes, and *ziji* at the interface. In: Cole, P., Huang, C-T J., and Hermon, G. (eds.), *Long Distance Reflexives, Syntax and Semantics* 33, Academic Press, New York, 141-195.
- Huang, C-T. J., Cole, P., & Hermon, G. (2006). Long-Distance Reflexives: an East Asian Perspective. In: Everaert, M., van Riemsdijk, H., Goedemans, R., and Hollebrandse, B (eds.) *The Blackwell Companion to Syntax Volume III*, pp. 21-84. Location: Blackwell.
- Huang, C-T J., Li, Y-H A., & Li, Y. (2009). *The Syntax of Chinese*. Cambridge University Press, Cambridge, UK.
- Huang, Y. (2000). *Anaphora: A Cross-linguistic Study*. Oxford University Press, Oxford, UK.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA.
- Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.
- Jarvella, R. (1970) Effects of syntax on running memory span for connected discourse. *Psychonomic Science*, 19, 235-236.

- Jarvella, R., & Pisoni, S. (1971). Relation between syntactic and perceptual units in speech processing. *Journal of the Acoustical Society of America*, 48, 84-88.
- Jin, Z.H., 2003. Verb restraint function to *ziji* long-distance binding. *Chinese Language Learning*, 4, 9-12.
- Johnson, K. (2001). What VP ellipsis can do, what it can't, but not why. In Baltin, M., & Collins, C. (eds), *The Handbook of Contemporary Syntactic Theory*. Blackwell/Wiley, New York.
- Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., & Moore, K.S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193-224.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137-194.
- Kaan, E. & Swaab, T. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15, 98-110.
- Kaiser, E. (2006). *The quest for a referent: A cross-linguistic look at reference resolution*. Doctoral dissertation, University of Pennsylvania.
- Kaiser, E., Runner, J., Sussman, R., & Tanenhaus, M. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, 112, 55-80.
- Kamide, Y., & Mitchell, D. (1997). Relative clause attachment: Nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research*, 26, 247-254.
- Kayne, R. (1989). Notes on English agreement. *Central Institute of English and Foreign Languages Bulletin*, 1, 46-67.
- Kennsion, S. (2003). Comprehending the pronouns *her*, *him* and *his*: Implications for theories of referential processing. *Journal of Memory and Language*, 49, 335-352.
- Kennsion, S., & Trofe, J. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355-378.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.
- Kimball, J. & Aissen, J. (1971). I think, you think, he think. *Linguistic Inquiry*, 2, 241-246.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15-47.
- King, J., & Kutas, M. (2005). Who did what and when? Using world- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 17, 376-395.



- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196-214.
- Kohonen, T. (1980). *Content-addressable memories*. Springer Verlag, Berlin.
- Kolk, H., Chwilla, D., van Herten, M., & Oor, P. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1-36.
- König, E., & Siemund, P. (1999). Intensifiers and reflexives: A typological perspective. In Frajzyngier, Z., & Curl, T. (eds), *Reflexives: Forms and functions*. John Benjamin, Amsterdam.
- Kratzer, A. (2009). Making a pronoun: fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry*, 40, 187-237.
- Kuno, S. (1987). *Functional syntax: anaphora, discourse and empathy*. University of Chicago Press, Chicago.
- Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Kutas, M. and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463-470.
- Ladusaw, W. A. (1979). *Negative polarity items as inherent scope relations*. Doctoral dissertation, University of Texas, Austin.
- Lago, S., Alcocer, P., & Phillips, C. (2011). Agreement attraction in Spanish: Immediate vs. delayed sensitivity. Poster presented at the 24<sup>th</sup> Annual Meeting of the CUNY Conference on Human Sentence Processing, Stanford, CA: March 24-26.
- Lasnik, H. (1976). Remarks on coreference. *Linguistic Analysis*, 2, 1-22.
- Lau, E. (2010). *The predictive nature of language comprehension*. Doctoral dissertation, University of Maryland.
- Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain & Language*, 98, 74-88.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9, 920-933.
- Lees, R., & Klima, E. (1963). Rules for English pronominalization. *Language*, 39, 17-28.
- Levelt, W. (1974). *Formal grammars in linguistics and psycholinguistics*. John Benjamins, Philadelphia.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.

- Lewis, R. (1996). Interference in short-term memory: the magical number two (or three) in Sentence Processing. *Journal of Psycholinguistic Research*, 25, 93-115.
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375-419
- Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10, 447-454.
- Li, X. & Zhou, X. (2010). Who is *ziji*? ERP responses to the Chinese reflexive pronoun during sentence comprehension. *Brain Research*, 1331, 96-104.
- Lidz, J., & Idsardi, W. (1999). Chains and phono-logical form. *University of Pennsylvania Working Papers in Linguistics*, 5, 119-125.
- Liu, Z. (2009). The cognitive process of Chinese reflexive processing. *Journal of Chinese Linguistics*, 37, 1-27.
- Li, C., & Smith, P. (2009). Comparing time-accuracy curves: beyond goodness-of-fit measures. *Psychonomic Bulletin & Review*, 16, 190-203.
- Lødrup, H. (2009). Animacy and long distance binding in Norwegian. *Nordic Journal of Linguistics*, 32, 111-136.
- Logacev, P. & Vasishth, S. (to appear). Morphological ambiguity and working memory. In Lamers, M., & de Swart, P., (eds.), *Case, Word Order and Prominence*.
- Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, 23, 769-799.
- MacDonald, M., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29, 469-492.
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). Lexical nature of ambiguity resolution. *Psychological Review*, 101, 676-703.
- Martin, A. E. & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58, 879-906.
- Martin, A. E. & McElree, B. (2009). Memory operations that support sentence comprehension: evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35, 1231-1239.
- Marcus, M. P. (1980). *Theory of Syntactic Recognition for Natural Languages*. MIT Press, Cambridge.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, New York.
- McCloskey, J. (2001). The morphosyntax of *WH*-extraction in Irish. *Journal of Linguistics*, 37, 67-100.

- McElree, B. (1990). *The time course of recency discrimination: A speed accuracy analysis*. Doctoral dissertation, Columbia University.
- McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, 32, 536-571.
- McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory & Language*, 38, 225-252.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29, 111-123.
- McElree, B., & Doshier, B. (1989). Serial position and set size in short-term memory: time course of recognition. *Journal of Experimental Psychology: General*, 118, 346-373.
- McElree, B., & Doshier, B. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, 122, 291-315.
- McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence processing: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 432-460.
- McElree, B., Foraker, S. & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67-91.
- McRoy, S.W., & Hirst, G. (1990). Race-based parsing and syntactic disambiguation. *Cognitive Psychology*, 14, 313-353.
- Miller, G., & Chomsky, N. (1963). Finitary models of language users. In Luce, R.D., Bush, R.R., & Galanter, E. (eds), *Handbook of Mathematical Psychology, Volume II*. John Wiley, New York.
- Mitchell, D., Cuetos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469-488.
- Mudock, B.B. Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482-488.
- Nairne, J. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, 16, 343-352.
- Nairne, J. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251-269.
- Neville, H. J., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151-165.
- Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research*, 1164, 81-94.

- Nicol, J. (1988). *Coreference processing during sentence comprehension*. Doctoral dissertation, MIT.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5-19.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601-626.
- Oiry, M. (2011). A case of true optionality: *Wh*-in-situ patterns like long movement in French. In Roussou, A., & Vlachos, C., (eds) *Linguistic Analysis: Optionality in wh-movement*,
- Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 6, 785-806.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 786-803.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273-285.
- Öztek, I., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 36, 383-397.
- Pan, H. (1998). Closeness, prominence, and binding theory. *Natural Language and Linguistic Theory*, 16, 771-815.
- Pan, H. (2000). Why the blocking effect? In Cole, P., Huang, C-T J., & Hermon, G (eds), *Syntax and Semantics Vol. 33: Long Distance Reflexives*. Academic Press, New York.
- Parker, D., & Phillips, C. (2011). Illusory negative polarity item licensing is selective. Poster presented at the 24<sup>th</sup> Annual Meeting of the CUNY Conference on Human Sentence Processing, Stanford, CA: March 24-26.
- Pearlmutter, N., & MacDonald, M. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521-542.
- Pearlmutter, N., Garnsey, S., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427-456.
- Phillips, C., Wagers, M., & Lau, E. F. (2010). Grammatical illusions and selective fallibility in real-time language comprehension. In: J. Runner (ed.) *Experiments at the Interference*, Syntax & Semantics, vol. 37, Bingley, UK: Emerald Publications.

- Pica, P. (1986). On the nature of the reflexivization cycle. In McDonough, J., & Plunkett, B. (eds), *Proceedings of the 17<sup>th</sup> Annual Meeting of the North East Linguistic Society*.
- Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer Verlag, New York.
- Pollard, C., & Sag, I. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry*, 23, 261-303.
- Pritchett, B. (1993). *Grammatical Competence and Parsing Performance*. University of Chicago Press, Chicago.
- Pulvermüller, F., Shtyrov, Y., Hasting, A., & Carlyon, R. (2008). Syntax as a reflexive: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, 104, 244-253.
- Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 21, 178-184.
- Ratcliff, R. (1983). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.
- Ratcliff, R., & McKoon, G. (2008). Passive parallel automatic minimalist processing. In Engel, C., & Singer, W., (Eds.), *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*. MIT Press, Cambridge.
- Reed, A. V. (1976). The time course of recognition in human memory. *Memory and Cognition*, 4, 16-30.
- Reinhart, T. (1976). *The Syntactic Domain of Anaphora*. Doctoral dissertation, Massachusetts Institute of Technology.
- Reinhard, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657-720.
- Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry*, 32, 439-492.
- Ricker, T., Aubuchon, A., & Cowen, N. (2010). Working memory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 573-585.
- Rizzi, L. (1990). On the anaphor-agreement effect. *Rivista di Linguistica*, 2, 27-42.
- Ross, J. (1967). *Constraints on variables in syntax*. Doctoral dissertation, Massachusetts Institute of Technology.
- Runner, J., Sussman, R., & Tanenhaus, M. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition*, 89, 1-13.
- Sag, I. (1976). *Deletion and logical form*. Doctoral dissertation, Massachusetts Institute of Technology.
- Sag, I., & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy*, 7, 325-345.
- Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry*, 18, 445-479.

- Sprouse, J., Fukuda, S., Ono, H., & Kluender, R. (2011). Reverse island effects and the backward search for a licenser in multiple wh-questions. *Syntax*, 14, 179-203.
- Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, 114, 447-454.
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421-457.
- Sturt, P. (2003a). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Sturt, P. (2003b). A new look at the syntax-discourse interference: The use of binding principles in sentence processing. *Journal of Psycholinguistic Research*, 32, 125-139.
- Sturt, P., Pickering, M.J., & Crocker, M.W. (2000). Search strategies in syntactic reanalysis. *Journal of Psycholinguistic Research*, 29, 183-194.
- Suckow, K., Vasishth, S., & Lewis, R. (2005). Interference and memory overload during parsing. In *Proceedings of AMLaP 2005*, Ghent, Belgium, September 2005. Ghent University.
- Sutton, R., & Barto, A. (1998). *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355-370.
- Tang, C-C J. (1989). Chinese reflexives. *Natural Language and Linguistic Theory*, 7, 93-121.
- Townsend, D. & Bever, T. (2001). *Sentence comprehension: the integration of habits and rules*. MIT Press, Cambridge.
- Traxler, M., Pickering, M., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558-592.
- Tsai, W-T D. (1994). On nominal islands and LF extraction in Chinese. *Natural Language and Linguistic Theory*, 12, 121-175.
- van Berkum, J., Brown, C., & Hagoort, P. (1999a). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28, 555-566.
- van Berkum, J., Brown, C., & Hagoort, P. (1999b). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.
- 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. & McElree, B. (2006). Retrieval interference in sentence processing. *Journal of Memory and Language*, 55, 157-166.
- Van Gompel, R., Traxler, M., & Pickering, M. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage Models. *Journal of Memory and Language*, 45, 225-258.
- Vasishth, S., Brüssow, S., Lewis, R. & Drenhaus, H. (2008). Processing Polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32, 685-712.
- Vigliocco, G., & Nicol J. (1998). Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear? *Cognition*, 68, 13-29.
- Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, 45, 368-390.
- Vigliocco, G., & Hartsuiker, R. (2002). The interplay of meaning, sound, and syntax in language production. *Psychological Bulletin*, 128, 442-472.
- Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic comprehension*. Doctoral dissertation, University of Maryland, College Park.
- Wagers, M., & McElree, B. (2009). Focal attention and the timing of memory retrieval in language comprehension. Talk given at the Architectures and Mechanisms for Language Processing Conference, September, Barcelona.
- Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language*, 61, 206-237.
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In Halle, M., Besnan, J., & Miller, G.A. (eds.), *Linguistic Theory and Psychological Reality*. MIT Press, Cambridge.
- Wasow, T. (1972). *Anaphoric relations in English*. Doctoral dissertation, Massachusetts Institute of Technology.
- Wechsler, S., & Zlatić, L. (2000). A theory of agreement and its application to Serbo-Croatian. *Language*, 76, 799-832.
- Wickens, T. (2001). *Elementary Signal Detection Theory*. Oxford University Press, New York.
- Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67-85.
- Wickelgren, W. A., Corbett, A., & Doshier, B. (1980). Priming and retrieval from short-term memory: A speed-accuracy tradeoff analysis. *Journal of Verbal Learning and Verbal Behavior*, 19, 387-404.
- Xiang, M., Dillon, B., & Phillips, C. (2006). Testing the strength of the spurious licensing effect for negative polarity items. Talk given at the 19<sup>th</sup> Annual CUNY Conference on Human Sentence Processing, New York, NY: March.

- Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing across dependency types: ERP evidence. *Brain and Language*, 108, 40-55.
- Xiang, M., Dillon, B., & Wagers, M. (2010). Processing *wh*-movement dependencies in a language without *wh*-movement. Poster presented at the 23<sup>rd</sup> Annual Meeting of the CUNY Conference on Human Sentence Processing, New York, NY: March.
- Xue, P., Pollard, C., & Sag, I. A new perspective on Chinese *ziji*. *Proceedings of the 13th Annual Meeting of the West Coast Conference on Formal Linguistics*, 432-447.
- Yngve, V. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104, 444-466.
- Zribi-Hertz, A. (1995). Emphatic or reflexive? On the endophoric character of French *lui-meme* and similar complex pronouns. *Journal of Linguistics*, 31, 333-374.