

ABSTRACT

Title of dissertation: THE STRUCTURE OF MEMORY MEETS MEMORY
FOR STRUCTURE IN LINGUISTIC COGNITION

Matthew Webb Wagers
Doctor of Philosophy, 2008

Dissertation directed by: Professor Colin Phillips
Department of Linguistics

This dissertation is concerned with the problem of how structured linguistic representations interact with the architecture of human memory. Much recent work has attempted to unify real-time linguistic memory with a general content-addressable architecture (Lewis & Vasishth, 2005; McElree, 2006). Because grammatical principles and constraints are strongly relational in nature, and linguistic representation hierarchical, this kind of architecture is not well suited to restricting the search of memory to grammatically-licensed constituents alone. This dissertation investigates under what conditions real-time language comprehension is grammatically accurate. Two kinds of grammatical dependencies were examined in reading time and speeded grammaticality experiments: subject-verb agreement licensing in agreement attraction configurations (“The runners who the driver wave to ...”; Kimball & Aissen, 1971, Bock & Miller, 1991), and active completion of *wh*-dependencies. We develop a simple formal model of agreement attraction in an associative memory that makes accurate predictions across different structures. We conclude that dependencies that can only be licensed exclusively retrospectively, by searching the memory to generate candidate analyses, are the most prone to grammatical infidelity. The exception may be retrospective searches with especially strong contextual restrictions, as in reflexive anaphora. However dependencies

that can be licensed principally by a prospective search, like *wh*-dependencies or backwards anaphora, are highly grammatically accurate.

THE STRUCTURE OF MEMORY MEETS
MEMORY FOR STRUCTURE IN LINGUISTIC COGNITION

by

Matthew Webb Wagers

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
2008

Committee

Professor Colin Phillips, Chair

Professor Norbert Hornstein

Associate Professor Jeffrey Lidz

Professor Amy Weinberg

Associate Professor Michael Dougherty, Dean's Representative

© Copyright by
Matthew Webb Wagers
2008

Preface

Chapter 2 reports research that was jointly conducted with fellow graduate student Ellen Lau. Parts of Chapter 2 were submitted together with Ms. Lau as a co-authored journal article.

Acknowledgments

I would like to express my deepest gratitude to Colin Phillips. I always felt that he took a chance on my graduate career, and I thank him for his willingness, attention, and patience to guide me. I am grateful for the insights we have been able to share with one another these past years. (Also: despite Colin's best, if not dogged, efforts over the years, I do not think I have yet been cured of my 'parentheticalitis.' I am nonetheless indebted to him for many improvements in clarity of the presentation of the dissertation.)

I wish also to thank my committee, who were both flexible and generous with their time.

To all my friends, at Maryland and beyond, there is no adequate or comprehensive expression of gratitude for the support and joy you have provided me. In particular, I want to thank Ellen Lau. Some back-of-the-room whispering with Ellen during Mark Baker's 2005 Blackwell lectures led to the agreement attraction studies reported here. Working with Ellen since then has been a constant pleasure, and I have learned much from being her collaborator and her friend. I also thank Clare Stroud for her friendship. Clare was my officemate for five years, and I have no doubt that our many Noodles excursions were a crucial part of both our graduate careers. I also thank Brian Dillon & Ming Xiang, who have been puzzling out many of the same issues as I have. This dissertation is improved in no small part because of our many discussions in the lab. Fortunately we were also able to spend much time together outside the lab.

Many of the ideas in this dissertation sprung from Colin Phillips and David Poeppel's Cognitive Neuroscience of Language seminar; Amy Weinberg and Philip Resnik's Computational Psycholinguistics seminar; and Colin Phillips and Jeff Lidz's Psycholinguistics seminar. And I cannot imagine how those ideas would have germinated, were it not for frequent cookies and discussions with Norbert Hornstein. I am also appreciative of the feedback and encouragement I have received from researchers outside the University of Maryland community, including Bill Badecker, Lyn Frazier, Rick Lewis, Gary Marcus, Brian McElree, Neal Pearlmuter, Julie Van Dyke and the CUNY Psycholinguistics Supper group.

I would also like to acknowledge the Graduate School. During the writing of the dissertation, I was supported by a Wylie Dissertation Fellowship.

Finally, Phil Longo played noun-noun compounding with me (yes, it is a game), taught me to eat carrots, and always made sure I went to the movies. I cannot imagine finishing this research, or doing much else really, without his love and support.

Table of Contents

1	INTRODUCTION	1
1.1	THE CHALLENGE OF NAVIGATING STRUCTURE IN REAL-TIME	1
1.2	GRAMMATICAL FIDELITY, GRAMMATICAL FALLIBILITY	6
1.3	OUTLINE OF THE DISSERTATION	11
1.3.1	<i>Chapter 2: Agreement attraction and selective fallibility.....</i>	<i>11</i>
1.3.2	<i>Chapter 3: The trouble with subjects</i>	<i>12</i>
1.3.3	<i>Chapter 4: Active dependency formation and mechanisms for the accurate recognition of grammatical dependencies.....</i>	<i>14</i>
2	AGREEMENT ATTRACTION AND SELECTIVE FALLIBILITY BINDING AND ACCESSING FEATURES IN COMPLEX SYNTACTIC OBJECTS, PART I.....	16
2.1	INTRODUCTION	16
2.1.1	<i>Agreement attraction: what's at stake</i>	<i>16</i>
2.1.2	<i>Outline of the chapter.....</i>	<i>20</i>
2.2	AN OVERVIEW OF AGREEMENT ATTRACTION	20
2.3	PREVIOUS STUDIES OF AGREEMENT PRODUCTION AND THE HIERARCHICAL NATURE OF ATTRACTION	23
2.3.1	<i>Hierarchical, not linear distance, matters.....</i>	<i>26</i>
2.3.2	<i>The attractor's 'structural domain' matters.....</i>	<i>27</i>
2.3.3	<i>Ordering of verb and attractor does not matter.....</i>	<i>28</i>
2.4	THE FEATURE PERCOLATION ACCOUNT OF AGREEMENT ATTRACTION	29
2.5	IMPLICATIONS OF FEATURE PERCOLATION AND OBJECTIONS	33
2.5.1	<i>Erroneous feature percolation as erroneous rule application</i>	<i>35</i>
2.5.2	<i>Erroneous feature percolation as uncontrolled spreading activation</i>	<i>38</i>
2.5.3	<i>Summary of erroneous feature percolation</i>	<i>41</i>
2.6	THE SIMULTANEITY ACCOUNT OF AGREEMENT ATTRACTION	42
2.6.1	<i>Simultaneity in planning a complex subject interferes with verb formulation.....</i>	<i>42</i>
2.6.2	<i>The relationship between planning order and hierarchy</i>	<i>44</i>
2.6.3	<i>Disentangling representation and process-based accounts</i>	<i>45</i>
2.7	ATTRACTION IN COMPREHENSION	46
2.7.1	<i>The Symmetry Prediction.....</i>	<i>46</i>
2.7.2	<i>Pearlmutter, Garnsey, & Bock (1999).....</i>	<i>52</i>
2.7.3	<i>Summary.....</i>	<i>59</i>
2.8	TESTING PERCOLATION I: RELATIVE CLAUSE ATTRACTION	59
2.8.1	<i>Kimball & Aissen (1971), Relative Clause Attraction & Experiment 1-2 Rationale.....</i>	<i>59</i>
2.8.2	<i>Experiment 1</i>	<i>64</i>
2.8.3	<i>Experiment 2</i>	<i>71</i>
2.8.4	<i>Experiment 3</i>	<i>78</i>
2.9	TESTING PERCOLATION II: THE GRAMMATICAL-UNGRAMMATICAL ASYMMETRY IN COMPREHENSION	84
2.9.1	<i>Wagers, Lau, & Phillips (2008) and On-line Comprehension.....</i>	<i>84</i>
2.9.2	<i>Controlling for RT correlations among adjacent regions: a mixed-effects models analysis .</i>	<i>87</i>
2.9.3	<i>Experiment 4: Speeded grammaticality tests of complex subject attraction</i>	<i>89</i>
2.10	CONCLUSIONS	93
3	THE TROUBLE WITH SUBJECTS BINDING AND ACCESSING FEATURES IN COMPLEX SYNTACTIC OBJECTS, PART II.....	95
3.1	INTRODUCTION	95
3.2	SEARCHING STRUCTURE WITH UNSTRUCTURED SEARCHES	97
3.2.1	<i>Content-addressable search</i>	<i>97</i>
3.2.2	<i>The restricted focus of attention.....</i>	<i>116</i>

3.2.3	<i>Implications</i>	120
3.3	AGREEMENT ATTRACTION IN COMPREHENSION	122
3.3.1	<i>Intuition</i>	122
3.3.2	<i>Formalization</i>	126
3.3.3	<i>Agreement & Case (Experiment 5)</i>	146
3.3.4	<i>Clause-boundedness</i>	158
3.3.5	<i>Next to the cabinets</i>	163
3.3.6	<i>Linking comprehension and production</i>	163
3.4	INTERFERENCE AND SUBJECTS	166
3.4.1	<i>Introduction</i>	166
3.4.2	<i>Van Dyke & Lewis (2003), Van Dyke (2007)</i>	168
3.4.3	<i>Replicating and extending Van Dyke (Experiment 6)</i>	180
3.4.4	<i>NPI Licensing v. Reflexive Anaphora</i>	188
3.5	CONCLUSIONS	198
4	ACTIVE DEPENDENCY FORMATION AND MECHANISMS FOR THE ACCURATE RECOGNITION OF GRAMMATICAL DEPENDENCIES	201
4.1	INTRODUCTION	201
4.2	THE ROLE OF PREDICTABILITY	203
4.2.1	<i>Forwards v. backward anaphora</i>	203
4.2.2	<i>Reconsidering agreement attraction</i>	208
4.3	PROCESSING WH-DEPENDENCY CONSTRUCTIONS	210
4.3.1	<i>Active dependency formation</i>	210
4.3.2	<i>Mechanisms of active dependency formation</i>	213
4.3.3	<i>Similarity-based interference and wh-dependency completion</i>	219
4.3.4	<i>Three studies</i>	228
4.4	THE GRAMMAR'S ROLE IN TRIGGERING WH-DEPENDENCY FORMATION	229
4.4.1	<i>The motivation for active dependency formation and island constraints</i>	229
4.4.2	<i>The Coordinate Structure Constraint and Active Dependency Formation I (Experiment 7)</i>	237
4.4.3	<i>Materials and Methods</i>	237
4.4.4	<i>The Coordinate Structure Constraint and Active Dependency Formation II (Experiment 8)</i>	249
4.4.5	<i>General discussion of Experiments 7 & 8</i>	257
4.5	THE FIDELITY OF RETRIEVAL IN WH-DEPENDENCY FORMATION	258
4.5.1	<i>Introduction</i>	258
4.5.2	<i>Experiment 9</i>	260
4.5.3	<i>Experiment 10</i>	267
4.5.4	<i>Accurately identifying the head of a dependency</i>	274
4.6	CARRYING INFORMATION FORWARD IN TIME	278
4.6.1	<i>Lexically-specific features (Experiments 11a, 11b)</i>	279
4.6.2	<i>Lexical identity (Experiment 12)</i>	292
4.6.3	<i>FG (Pied-piping) (Experiment 13)</i>	300
4.6.4	<i>Conclusions</i>	309
4.7	CONCLUSIONS	311
5	CONCLUSIONS	314
5.1	SPECIFIC CONCLUSIONS	314
5.1.1	<i>Agreement attraction</i>	314
5.1.2	<i>Wh-dependency formation</i>	316
5.2	BROADER CONCLUSIONS	318
6	APPENDICES	322
7	REFERENCES	335

List of Tables

Table 2-1	The Symmetry Prediction for Feature Percolation	48
Table 2-2	Interpretations of reading time patterns in relative clause agreement comprehension	62
Table 2-3	Sample materials for Experiment 1	65
Table 2-4	Sample plural subject materials for Experiment 2.....	72
Table 3-1	Speeded grammaticality judgments of complex subject attraction From Experiment 4	123
Table 3-2	Retrieval structure and outcomes Singular-headed subjects, grammatical continuations	135
Table 3-3	Retrieval structure and outcomes Singular-headed subjects, ungrammatical continuations	136
Table 3-4	Retrieval structure and outcomes Plural-headed subjects, grammatical continuations	137
Table 3-5	Retrieval structure and outcomes Plural-headed subjects, ungrammatical continuations	139
Table 3-6	Retrieval structure and outcomes for RC Attraction Plural RC head, Singular RC subject, Ungrammatical.....	141
Table 3-7	Revised retrieval structure and outcomes, RC attraction Plural RC head, Singular RC subject 143	143
Table 3-8	Speeded grammaticality judgments of RC attraction.....	144
Table 3-9	Revised retrieval structure and outcomes, RC attraction Singular RC head, Plural RC subject 145	145
Table 3-10	Subject and attractor sampling probabilities for Subj-attached RCs.....	148
Table 3-11	Subject and attractor sampling probabilities for Obj-attached RCs.....	148
Table 3-12	Sample materials set for Experiment 5	150
Table 3-13	Comprehension accuracy from Van Dyke (2007) Experiment 3	179
Table 3-14	Erroneously chosen nouns in Experiment 3 cloze comprehension task	180
Table 3-15	Comprehension accuracy for Experiment 6.....	183
Table 4-1	Predictions for active dependency formation in multiple dependency constructions.....	236
Table 4-2	Sample materials set for Experiment 7	238
Table 4-3	Experiment 7 Acceptability Ratings Summary	242
Table 4-4	Sample materials set for Experiment 8	250
Table 4-5	Sample materials set for Experiment 9	261
Table 4-6	Comprehension question accuracy for Experiment 9.....	262
Table 4-7	Sample materials set for Experiment 10	270
Table 4-8	Comprehension question accuracy for Experiment 10	271
Table 4-9	Sample materials set for Experiment 12	294
Table 4-10	Comprehension accuracy for Experiment 12.....	295
Table 4-11	Sample materials set for Experiment 13	302
Table 4-12	Comprehension accuracy for Experiment 13.....	303

List of Figures

Figure 2-1	Percolation of number features in a complex subject	30
Figure 2-2	Valuing subject number for ‘the key to the cabinets’	33
Figure 2-3	The effects of subject-verb disagreement in reading	50
Figure 2-4	Pearlmutter, Garnsey & Bock (1999) Experiment 1 Results	54
Figure 2-5	Pearlmutter, Garnsey & Bock (1999) Experiment 2 Results First-pass reading times.....	57
Figure 2-6	Experiment 1: Relative Clause Attraction Reading Time Results	68
Figure 2-7	Experiment 2: RC Attraction Reading Time Results, Singular Subject	74
Figure 2-8	Experiment 2: RC Attraction Reading Time Results, Plural Subject	75
Figure 2-9	Experiment 3: Relative clause attraction, Singular subjects Speeded grammaticality, proportion ‘yes’ responses	81
Figure 2-10	Experiment 3: Relative clause attraction, Plural subjects Speeded grammaticality, proportion ‘yes’ responses	82
Figure 2-11	Wagers, Lau, & Phillips (2008): Complex Subject Attraction	86
Figure 2-12	Wagers, Lau, & Phillips (2008) Experiment 4, Residual RTs Two previous Region RTs regressed out	88
Figure 2-13	Experiment 4a: Complex subject attraction, singular subjects Speeded grammaticality, proportion ‘yes’ responses	91
Figure 2-14	Experiment 4b: Complex subject attraction, plural subjects Speeded grammaticality, proportion ‘yes’ responses	92
Figure 3-1	Accessible and inaccessible licensors in an abstract tree.....	97
Figure 3-2	Hypothetical SAT functions	102
Figure 3-3	McElree, Foraker, & Dyer (2003), Experiment 2 Average SAT Function	105
Figure 3-4	Implicit encoding of serial order	112
Figure 3-5	Comparison of cue convergence rules	127
Figure 3-6	Phrase structure tree for ‘the man with the hat’	130
Figure 3-7	Experiment 5: Relative clause attraction, Subject-attached RCs Speeded grammaticality, proportion ‘yes’ responses	151
Figure 3-8	Experiment 5: Relative clause attraction, Object-attached RCs Speeded grammaticality, proportion ‘yes’ responses	153
Figure 3-9	Van Dyke & Lewis (2003), Experiment 4.....	171
Figure 3-10	Van Dyke & Lewis (2003), Experiments 2 & 3.....	173
Figure 3-11	Van Dyke (2007) Experiment 3, Critical region reading times	176
Figure 3-12	Van Dyke (2007) Experiment 3, Pre-critical region reading times	178
Figure 3-13	Experiment 6 Self-paced reading results	185
Figure 4-1	Van Dyke & McElree (2006) Critical verb region.....	223
Figure 4-2	Experiment 7 Region-by-Region Reading Times	246
Figure 4-3	Experiment 8 Region-by-Region Reading Times	254
Figure 4-4	Experiment 9 Reading time results: Wh-clause Conditions	264
Figure 4-5	Experiment 9 Reading time results: If-clause Conditions	265
Figure 4-6	Experiment 10 Reading time results: Relative clause conditions	271
Figure 4-7	Experiment 10 Reading time results: Coordinated clause conditions.....	272
Figure 4-8	Experiment 11a Region-by-region reading times	283
Figure 4-9	Experiment 11b Region-by-region reading times (Long:clause).....	287
Figure 4-10	Experiment 12 Reading time results	296
Figure 4-11	Experiment 12 Follow-up results Speeded grammaticality, proportion ‘yes’ responses ...	299
Figure 4-12	Experiment 13 Region-by-region reading times: Short conditions	304
Figure 4-13	Experiment 13 Region-by-region reading times: Long/PP conditions	305
Figure 4-14	Region-by-region reading times: Long/Clause conditions	306
Figure 6-1	Sg [Sg] Grammatical RT Distribution: estimated RT_G	324
Figure 6-2	Sg [Sg] Ungrammatical RT Distribution: estimated RT_U	325
Figure 6-3	Simulation results: $RT_{A/G}/RT_{U/G}$ means shift symmetrically.....	327

1 Introduction

1.1 The challenge of navigating structure in real-time

This dissertation is concerned with the problem of how structured linguistic representations interact with the architecture of human memory. In particular it addresses the problem of the how constituent encodings of a sentence are accessed by processes of grammatical dependency formation. Nearly all contemporary theories of grammar share a commitment to richly structured mental representations as necessary components of mature linguistic competence (Bresnan, 2001; Chomsky, 1981, 1995; Pollard & Sag, 1994, Steedman, 1997). Though these theories may deploy different kinds or numbers of representations, they all posit abstract categories that can be combined in regular ways to form hierarchically-ordered, compositional objects. This hierarchical order underlies many important generalizations about grammatical dependencies. For example, in (1), intrasentential reference for the pronoun ‘her’ is restricted to ‘Laura’:

(1) Laura’s friend embarrassed her at the wedding.

The sentence can only mean something like “Laura’s friend embarrassed Laura,” and not “Laura’s friend embarrassed herself.” In (2), though, ‘her’ can refer either to ‘Laura’ or ‘Laura’s friend’ (but not ‘Peggy’):

(2) Laura’s friend was afraid that Peggy would embarrass her at the wedding.

A standard formal description of the facts in (1) and (2) is that co-reference between a noun phrase and pronoun is blocked if the noun phrase is both in the same clause as the

pronoun and the noun phrase ‘c-commands’¹ the pronoun (Principle B, Chomsky, 1981). This description would not be possible if the representation in (2) did not allow reference to notions like hierarchical order and domains of rule applicability. A simplified phrase structure representation of (2) that includes both of these concepts is given in the bracketed sentence:

(3) [S [NP Laura’s friend][VP was afraid [S that [S [NP Peggy][VP would embarrass her]]]]]

The explanatory benefit of abstraction over structured representations comes with computational challenges, however. On the timescale of comprehension, tens and hundreds of milliseconds to seconds, the comprehender must deploy the abstract facts about grammatical categories and relations to recognize and understand actual expressions. At the sentence level, pairings of words to structure must be recognized and encoded as part of the current, novel utterance. The representation in (3) contains a considerable amount of information that was simply not in the input. Because natural language expressions can be of considerable complexity and temporal extent, these novel structures must be encoded semi-durably so that they are accessible to later operations. For example, in sentence (2), the reference of the pronoun ‘her’ must be resolved with respect to the syntactic context provided by the preceding parts of the sentence. So those parts, in their hierarchical order, must be retained for ‘her’ to be interpreted. We must

¹ A category A c-commands a category B if A does not dominate B, and the first node dominating A also dominates B (Chomsky, 1981).

The whole phrase ‘Laura’s friend’ c-commands the pronoun within the same clause in (1), and thus coreference is blocked, but its subconstituent ‘Laura’ does not. However, in (2), the whole phrase is far enough away from ‘her’ (outside of the same clause) such that the c-command restriction is voided.

I am describing here the basic Binding Theory account of these facts (Principle B; Chomsky, 1981), but the point is more general: any account that describes the patterns of acceptability (e.g. Reinhart & Reuland, 1993) will need a representational vocabulary with comparable terms.

consequently not only worry about how the hierarchical order of a sentence is encoded, but also what operations are available to the comprehension system for targeting parts of these large, complex representations, well after they have left immediate attention.

A natural way of accessing constituents in a grammatical fashion would be to follow the hierarchical relations as the phrase structure gives them, like links in a chain. Navigating the representation in order of the hierarchical relations provides a means for restricting reference to only grammatically-relevant constituents in the syntactic context. Indeed this is exactly how tree-like data structures are searched in computer science (Knuth, 1965/1997). For example, take another kind of grammatical dependency, one which we will consider in great detail in this dissertation: subject-verb agreement. A verb must agree with the subject in the same clause, as in the following sentence, and its associated bracketing:

(4) [S [NP The path [PP to the monuments]][VP was littered with bottles]].

If the agreement between subject and verb must be verified online, then following the phrase structural relations will lead directly from a verb to the entire phrase that is the subject. The path for verifying agreement (in this simple representation) could be given succinctly with the following chain of dominance statements – Start at V: (VP dominates V, S dominates VP, S dominates NP): End at NP. There is also a plural noun in this sentence, ‘monuments,’ which is grammatically inaccessible to the agreement relation. It would remain irrelevant if only the dominance pathways are followed to the subject NP and some other search order were not employed (i.e. linear).

Recently there have been a number of arguments that the memory architecture in which language processing is embedded is similar in many ways to that of general

episodic memory: it is context-dependent and content-addressable (Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005; McElree, 2006; Martin & McElree, 2008). The many pieces that compose a sentence are still thought to be encoded as linked together in a phrase structure tree, by the dominance relations the grammar generates. But it has been argued that the resolution of grammatical dependencies does not follow a search procedure ordered by those relations. Instead it proceeds in a content-addressable fashion, by probing the memory with features that match the desired constituent. For example, if the constituent in subject position is needed to establish a grammatical relation (like agreement), it would not be accessed by successively following the dominance relations up to that position. Instead it would be retrieved by probing for features characteristic of a subject, like +Nominative case. Crucially content-addressable retrieval grants direct access to just those constituents whose information matches features in the probe. On the one hand, this means fast, position-constant access times (McElree, Foraker, & Dyer, 2003). On the other hand, because the search is unordered, it means that multiple candidate constituents could be returned. One of these candidates may be the grammatically-licensed constituent, but others may not. For example, in a biclausal sentence like (5), there are two subjects, and so there are two candidate matches to a simple +Nominative cue.

(5) The park ranger was dismayed that the path to the monuments was littered with bottles.

If the system were to maintain full fidelity to the principles of the grammar, then on-line comprehension processes would have to have a structurally-sensitive decision metric for which of multiple candidates was the right one. This concern over grammatical fidelity has generated two kinds of responses: (1) that online processing does exhibit grammatical

infidelity, and it is particularly exacerbated when there are many similar constituents in a structure (Van Dyke & Lewis, 2003; Van Dyke, 2007); (2) the right combination of cues might be found (at least for a subset of relations) to target the unique, grammatically licensed constituent (Martin & McElree, 2008).

In this dissertation we will address these claims by broadly surveying under what scenarios grammatical accuracy is reliably observed and under what scenarios grammatical accuracy seems hard to achieve.

In sentence comprehension experiments on subject-verb agreement and the formation of *wh*-dependencies we can infer what kind of analyses the comprehender is entertaining by looking at patterns of error detection, reaction time measures of processing difficulty, and reaction time measures of interpretation. Interestingly we find that subject-verb agreement, the arguably simpler relation, is prone to grammatical infidelity of exactly the kind predicted by a content-addressable architecture. The formation of *wh*-dependencies, on the other hand, is highly grammatically accurate, even though it should be liable to interference from grammatically unavailable constituents.

We defend the view that the memory architecture does burden comprehenders with a major limitation on inducing structured analyses over linguistic input. Making decisions about how to structure and interpret new input is highly dependent on having hierarchical order information about what has come before. The content-addressable memory is not generally amenable to accurate structural reference for fundamental reasons. Important structural relations, like c-command, can be stated over any arbitrary pair of constituents, so there's no reasonable way to make the property of c-command part of constituent encodings. That is, the property of c-command cannot be the content

of an encoding. The fact that relational notions cannot restrict the search of linguistic context introduces inaccuracy into non-local decisions. This general outlook, however, predicts more fallibility than is generally observed. Many processes and phenomena, like that of *wh*-dependency formation, simply seem impervious to ungrammatical analyses. Instead of rejecting the architecture outright, however, we propose adaptations that optimize grammatical accuracy. Chief among these are constituent encoding strategies that permit reference to be restricted to major grammatical domains; and the predictive recognition of dependencies that performs as much grammatical licensing as is possible left-to-right. The online structure building system, we conjecture, is reasonably well-adapted to the memory architecture.

1.2 Grammatical fidelity, grammatical fallibility

In the 1960s and 1970s there emerged a basic consensus that the perceptual or mnemonic representation of a sentence reflects the gross properties of the constituent structure (or thematic structure) assigned to it by the grammar (cf. Fodor, Bever & Garrett, 1974; Levelt, 1974). Most of the studies arriving at this conclusion used techniques that would be considered ‘off-line’: for example, asking experimental participants to recall the location of noise burst in a recorded stimulus (Bever, Lackner, & Kirk, 1969); or, to assign pairs of words scores based on how related they were felt to be (Levelt, 1974). Nonetheless they are informative about what might be called the ‘steady-state’ encoding, the representation that persists once major comprehension processes have concluded at the sentence level. With the advent of experimental measures and designs that can probe on-going processing on the time-scale of single word or morpheme processing (self-paced reading, eye-tracking, EEG, MEG, etc.), it has become

possible to test not just whether the steady-state encoding reflects grammatical distinctions, but whether the instantaneous, on-going encoding is also grammatically accurate. In the past 20 years, using the finer measures, and examining a broader collection of relationships, the facts of the matter are, perhaps unsurprisingly, mixed. Some kinds of real-time comprehension processes are tightly regulated by the grammar and never show evidence that anything but a grammatical analysis is entertained. Those processes we'll refer to generally as *grammatically faithful*. Some processes, however, seem to entertain analyses of the expression which the grammar cannot generate or must exclude. Such processes we'll refer to generally as *grammatically fallible*. Let us review two cases here.

The example of subject-verb agreement nicely illustrates the nature of the problem. Subject-verb agreement is a wide-spread phenomenon among the world's languages and refers to the covariation of verbal morphology with syntactic or semantic properties of the subject phrase (Corbett, 2003). For example, in English the verb form must match the subject phrase in number features:

- (6) (a) The path was/*were littered with bottles.
- (b) The paths *was/were littered with bottles.

If we replace the simple subject “the path” with a more complicated one, like “the path to the monument” or “the path that Mary took with her father” the agreement pattern remains the same.

- (7) (a) The path to the monument was/*were littered with bottles.
- (b) The paths to monument *was/were littered with bottles.
- (8) (a) The path that Mary took with her father was/*were littered with bottles.
- (b) The paths that Mary took with her father *was/were littered with bottles.

Agreement is determined hierarchically: the verb form must match with a property of the entire subject constituent. It does not, for example, simply match the number on an adjacent noun, as (7b) illustrates. The number property of the entire subject depends on a distinguished element contained inside it: its head ‘path’. The notion of ‘head’ is central in many theories of phrase structure (Jackendoff, 1977; Pollard & Sag, 1994; Chomsky, 1995) but more generally reflects the idea that the properties of the whole are determined by an ordered composition of the properties of its parts. Given this core facet of natural language grammars, we can ask whether real-time comprehension is sensitive to a notion of headedness in the same way. That is, does the comprehender form a representation from the input that projects complex syntactic objects from lexical items in a way that distinguishes a head?

Agreement is a useful probe for addressing this question, because we know that speakers make well-defined errors in producing agreement. For example, speakers commonly produce sentences like the following:

(9) The path to the monuments were littered with bottles.

In producing such a sentences, a speaker selects a verb form whose number matches not the head of the subject projection, ‘path,’ but a more deeply embedded noun, ‘monuments.’ The occurrence of forms like these is widely documented by grammarians and other observers in both written and spoken English (e.g., Trollope, 1883; Jespersen, 1924; Strang, 1966; Quirk et al., 1972; Kimball & Aissen, 1973; Francis, 1986). The phenomenon, called *agreement attraction*, has drawn the most attention in research on language production (Bock & Miller, 1991, et seq). Perhaps the most prominent contemporary explanation for agreement attraction, encompassing a sizable body of

observations, is that syntactic objects are encoded such that features of individual lexical items can erroneously ‘percolate’ from different parts of the representation in a manner that would not be grammatically sanctioned (e.g., Eberhard, Cutting & Bock, 2005, Franck, Vigliocco & Nicol 2002; Vigliocco & Nicol 1998). A complex subject, like “the path to the monuments,” is misassigned plural number, because it contains a plural noun whose plural feature has (stochastically) migrated up from a more deeply embedded node in the structure.

Agreement attraction represents a case of *grammatical fallibility*. Under the feature percolation account, attraction occurs because the binding of features in a structured representation is endogenously error-prone: lexical items are initially correctly ordered in a syntactic frame, but it is impossible to stably maintain their initial feature composition because of the way they are structurally related to one another. (Note we argue against the feature percolation mechanism in Chapter 2).

Let us now turn to a case of *grammatical fidelity*, a case where we do not observe plausible but ungrammatical analyses. Here consider the *wh*-dependency formed inside a relative clause. A typical adult speaker of English draws a distinction between the acceptability of the relative clauses in (10a) and (10b):

- (10) (a) The singers that Laura hoped her fiancé would agree to hire ___ ...
(b) *The singers that Laura hoped hiring ___ would be agreeable to her fiancé ...

Relativization in English involves the formation of an A'-dependency: the head of the dependency occupies an (overt) syntactic position in the periphery of the clause, while the foot of the dependency is associated with an argument position which can be embedded an unbounded distance inside the clause. The relative clauses in (10) differ

with respect to the syntactic position of the foot: in (a), it is a right-branch, complement position; but in (b), it is a complement position within a left-branch subject projection. As a class, domain restrictions on the foot of A'-dependencies are referred to as island constraints (Ross, 1967).

A'-dependency formation has been studied extensively in comprehension. Through a combination of reaction time and electrophysiological measures, we know that comprehenders attempt to link the head of the dependency with potential foot locations as soon as possible. This processing occurs in advance of direct evidence that there is an open foot position, a phenomenon referred to as active dependency completion (see Aoshima et al., 2004, for a review, or Chapter 4). The question arises whether comprehenders make errors in locating potential foot locations. Faced with the first seven words in (10b), does the comprehender ever posit a foot for the dependency, where it cannot grammatically occur? The majority of the evidence suggests that comprehenders obey island constraints in online comprehension, and do not construct A' dependencies that the grammar does not allow (e.g., Stowe, 1986; Traxler & Pickering, 1996; see Phillips, 2006, for a review). We provide evidence that strengthens this position in Chapter 4.

When we look at a large set of phenomena, classified by how grammatically faithful they are in real-time processing, it may be that there are no broader-scale patterns. Grammatical fidelity could be an idiosyncratic property of a given construction or process. However, we will argue that there are broader patterns. In particular, a key determinant of accuracy seems to be predictability. Grammatical relations that announce their presence early on, like *wh*-dependencies, can be completed in principally top-down

fashion. This makes sense, we argue, in terms of the memory architecture. Licensing dependencies predictively allows the system to either avoid searching through the syntactic context or to do so with highly targeted, restrictive information. In this way, the grammar can recognize and license dependencies in a way that is adapted to the structural imprecision introduced by the memory.

1.3 Outline of the dissertation

1.3.1 Chapter 2: Agreement attraction and selective fallibility

The goal of Chapter 2 is to test the feature percolation account of agreement attraction in comprehension.

The feature percolation account holds that agreement attraction stems from a faulty encoding of the subject's number features. We first discuss the basic facts of agreement attraction drawn from the production literature which have motivated the feature percolation account. We then detail several variants of the feature percolation account along with their implications for the encoding of grammatical features in constituent representations. We contrast the feature percolation with other accounts, which concern the order in which constituents are accessed in agreement licensing.

In comprehension, feature-percolation makes a strong and falsifiable prediction, which we refer to as the Symmetry Prediction. Specifically, it predicts that agreement attraction should lead to illusions of grammaticality for ungrammatical sentences and illusions of ungrammaticality for grammatical sentences, in equal measure. After reviewing previous comprehension research, we present two reading time experiments and two speeded-grammaticality experiments, both of which fail to uphold the Symmetry Prediction. We present novel evidence from relative clause attraction sentences (Kimball

& Aissen, 1971) as well as data from canonical complex subject attraction, of the type discussed above.

We find that agreement attraction only improves the perception of ungrammatical sentences in comprehension and does not impact grammatical sentences. Agreement attraction, we conclude, exhibits selective fallibility, and one which implicates a process-based account instead of a representational one.

The experimental and theoretical work reported in this chapter represents the fruits of a several-year, joint collaboration with Ellen Lau. Experiments 1-4, and some of the analysis, are reported in Wagers, Lau, & Phillips (submitted to *Journal of Memory and Language*).

1.3.2 Chapter 3: The trouble with subjects

The goal of Chapter 3 is two-fold: (1) to provide an account of agreement attraction in comprehension that encompasses the selective fallibility result of Chapter 2; (2) in doing so, to introduce in greater details the properties of a content-addressable memory and the challenges it poses for hierarchically ordered information.

We first discuss the architectural commitments which we accept as the basis for the rest of the dissertation. We introduce the concept of content-addressable memory and the key properties of a direct access search (Clark & Gronlund, 1996; McElree, 2006), including similarity based interference (Anderson & Neely, 1996). We argue that this architecture presents inherent difficulties for recovering relational properties like c-command, but we also discuss some strategies for overcoming its limitations.

Using assumptions from Shiffrin's Search of Associative Memory (Gillund & Shiffrin, 1984) we develop a formal model of agreement attraction in comprehension

which predicts the patterns we observe in our experiments. We argue agreement checking in comprehension stems from a retrieval operation initiated by the verb. This operation is guided by cues that do not converge on a single constituent representation, but partially match multiple constituents in the representation. It is the interference of the inaccessible constituent that leads to spurious agreement licensing. In a speeded-grammaticality experiment, we test and confirm a prediction of this model. We relate our account to other process-based accounts of production (Solomon & Pearlmuter, 2004) and a formally-similar account developed simultaneously by Badecker & Lewis (2007) for production.

The influence of grammatically-inaccessible constituents has been argued to stem from partial match in several other domains (Van Dyke & Lewis, 2003; Van Dyke, 2007; Vasishth, Drenhaus, Saddy & Lewis, 2005). We review this evidence and find it largely equivocal. We examine claims that complex subject attachment is liable to interference from inaccessible constituents (Van Dyke & Lewis, 2003; Van Dyke, 2007). We argue that much of the online evidence for interference is confounded with the number of clauses in the experimental conditions. We provide support for this contention in a reading-time experiment which we modeled off Van Dyke & Lewis (2003). We also discuss Negative Polarity Item licensing (Drenhaus et al., 2005) and reflexive anaphora (Sturt, 2003). A lack of partial match effects has been documented in the resolution of reflexive anaphora (Sturt, 2003; Xiang, Dillon, & Phillips, submitted), which is seemingly unexpected. We argue that resolution of reflexive anaphora is guided by cues which do not strongly activate embedded constituents.

1.3.3 *Chapter 4: Active dependency formation and mechanisms for the accurate recognition of grammatical dependencies*

The goals of Chapter 4 are (1) to review and document evidence for grammatical fidelity observed in active dependency formation; and (2) provide an account of why active dependency formation is faithful. On the basis of three comprehension studies, we argue that the licensing of *wh*-dependencies is guided principally top-down. In doing so it is able to largely avoid retrieving information from the syntactic context and consequently avoids the influence of grammatically-inaccessible constituents. When retrieval is ultimately necessary, it can be guided by highly restrictive information.

Experiments 7-8 provide evidence that the decision to complete a *wh*-dependency and retrieve the information from the head of the dependency is guided principally top-down, and not from information provided by the local environment. We show that the Coordinate Structure Constraint (Ross, 1967) is respected in online processing. We contrast the processing of *wh*-dependencies inside coordinate structures, in which active dependency formation is observed, with processing inside potential parasitic gap environments, in which active dependency formation is not observed. We conclude that the incentive to satisfy global well-formedness constraints drives active dependency formation more so than an incentive to satisfy local licensing requirements.

Experiments 9-10 provide evidence that identifying the head of a *wh*-dependency is grammatically accurate and not liable to interference. We test whether the resolution of a *wh*-dependency is impacted by other irrelevant dependency heads in the same sentence, both in embedded *wh*-questions and relative clauses. Because other dependency heads have similar structural and featural properties, it is predicted they should interfere with dependency resolution. Results from Van Dyke & McElree (2006) seem to support this

prediction, but we argue that their experiment lacked a crucial control condition. We outline two mechanisms to account for the fidelity we observe in our experiments: one, an encoding scheme that marks dependencies complete or incomplete; two, maintenance of a small amount of unique information about the dependency head that could be used to precisely target the correct head in a retrieval operation.

Finally, Experiments 11-13 provide evidence that most lexically-specific information about a dependency head is lost over increasing dependency lengths, whereas abstract categorial information is not. This finding supports the idea that some information remains available to the comprehender to guide dependency formation. Experiment 11 tests whether the plausibility of a candidate dependency can be evaluated over longer dependencies. Experiment 12 tests whether a verb-PP selectional restriction can be evaluated over longer dependencies. Finally Experiment 13 tests whether the identity of the dependency head is retained over longer dependencies.

Experiments 7, 8 and 11a are reported in Wagers & Phillips (submitted to *Journal of Linguistics*).

2 Agreement attraction and selective fallibility Binding and accessing features in complex syntactic objects, Part I

2.1 *Introduction*

2.1.1 *Agreement attraction: what's at stake*

In Chapter 1, we discussed a major challenge for real-time structure building: how to navigate novel encodings of structure, which are potentially large and complex. One strategy for understanding the mechanisms by which this occurs is to ask how grammatically accurate those online processes are. On the timescale of tens to hundreds of milliseconds, how faithfully can the processing system encode a representation with respect to grammatical principles and constraints? And can it direct its attention between different constituent encodings of that representation in a structure-sensitive fashion? In this chapter, we will turn our attention to the phenomenon of agreement attraction, as a case study in how to tease apart facets of the encoding itself from facets of how the encoding is accessed in real-time. Agreement attraction is an error in the formulation of subject-verb agreement, best introduced by way of example:

- (11) The function of the ducts are unknown.
from J.E. Stevens, "The delicate constitution of sharks," *Bioscience*, 44, 661-4

The subject of the clause is the singular DP "the function of the ducts," but the verb is in its plural form. It fails to match the grammatical number of the subject projection, as

determined by its head, and instead matches one of its subconstituents: “the ducts.” In agreement attraction the control of agreement seems to be wrested away by a nearby but grammatically irrelevant constituent. Accounts of agreement attraction have largely appealed to the notion that multiple nouns (or noun phrases) in a complex subject have independent specifications for number, which compete to value the entire noun phrase. The differences among these accounts lie in why those features compete. There are two major proposals:

- In the first kind of account, the number features bound are erroneously transferred to the agreement controller; (Eberhard, 1997, Eberhard, Cutting & Bock, 2005; Franck, Vigliocco & Nicol, 2002; Hartsuiker et al. 2001; Nicol, Forster, & Veres, 1997; Vigliocco & Nicol, 1998). This account is fundamentally representational, and assigns blame for agreement attraction to an encoding of structure that is grammatically inconsistent or imprecise.
- In the second kind of account, there is no problem in encoding the representation, and assigning features to specific categories in the structure, but there is fallibility in how categories are accessed in real-time (Solomon & Pearlmuter, 2004; Badecker & Lewis, 2008). This account is fundamentally process-based, and assigns blame for agreement attraction to a (partially) structurally insensitive means of accessing component encodings of syntactic structure.

In this chapter, we examine this question, whether agreement attraction is due to the erroneous encoding of the features in the subject, or rather to errors in how these features are accessed in real-time. We turn to the comprehension analog of the agreement

attraction production error to make this case, and ask how comprehenders perceive subject-verb agreement in potential attraction configurations. Based on several real-time reading studies and a complement of speeded grammaticality tests, we argue that it cannot be an unfaithful encoding of features on the agreement controller that is responsible for agreement attraction.²

In this chapter we will carefully examine an account of grammatical infidelity in the encoding of the subject. Our empirical argument has two parts. First we examine an agreement attraction configuration that has received relatively little attention in previous experimental literature, in which attraction occurs inside a plurally-headed relative clause (Kimball & Aissen, 1971):

(12) The ducts_{PL} [_{RC} that the scientist_{SG} study_{PL}] have an unknown function.

We provide evidence that these RC configurations yield exactly the same error profile as the more canonical complex subject agreement controllers, the configuration attested in example (11). We show that agreement attraction exemplifies what we call ‘selective fallibility’ in comprehension: when there is a fully grammatical analysis available to the comprehension system, the comprehension system pursues this analysis nearly all of the time. For example, subject-verb agreement checking in comprehension is not disrupted when the verb agrees with the head of the subject, even though an attractor is present, as in (13):

(13) The function of the ducts is unknown.

² The reading time studies are first reported in Wagers, Lau & Phillips (2008a, submitted) and the grammaticality studies in Wagers, Lau & Phillips (2008b).

Only when a fully grammatical analysis is unavailable is the system tempted into an attraction error. This selective fallibility is incompatible with a system in which binding of number features is endogenously leaky or inaccurate.

In Chapter 3, we show how a means of accessing features via cue-based retrieval (as in Lewis & Vasishth, 2005) can encompass our results. The analysis we offer points to the second kind of account for why features compete: structurally-insensitive access procedures. The problem engendered by subject-verb agreement formulation in production, and subject-verb agreement checking in comprehension, is one of regulating access to number features that are properly bound in a syntactic representation. We will argue these results are germane to other cases of observed fallibility in comprehension. Complex subjects have concerned syntacticians for some time (Huang, 1982; Kayne, 1984; Uriagereka, 1999, *inter alia*). More recently psycholinguists have been discussing and documenting cases of fallibility involving processing subjects in real-time (e.g., Badecker & Straub, 2002; Kluender, 2005; Sturt 2003; Van Dyke & Lewis, 2003; Xiang, Dillon, & Phillips, submitted). We will expand the insights from subject-verb agreement to understanding more generally why subjects, and particularly complex subjects, are troublesome for encoding and accessing hierarchical structure in real-time. As we'll see, just as within the phenomenon of agreement attraction, different kinds of subject relationships are fallible to different extents in real-time processing, and this is a function of the both way in which information is preserved over time prospectively and accessed retrospectively.

2.1.2 *Outline of the chapter*

In section 2.2 we introduce the basic properties of agreement attraction. Then, in section 2.3, we review much of the existing production literature which suggests that attraction is sensitive to structural factors, justifying its interest as a diagnostic of the encoding and navigation of hierarchically ordered information. In section 2.4, we discuss the predominant representational account of agreement attraction, erroneous feature percolation, and, in section 2.5, detail what we believe are the severe implications of this account for grammatical fidelity in structural encoding. In section 2.6, we sketch out the major contemporary alternative to erroneous feature percolation. In section 2.7, we then argue that comprehension data, rather than production data, can decide the question of whether erroneous feature percolation is active in the system. In sections 2.8 and 2.9, we present five comprehension experiments and draw on related data from Wagers, Lau, & Phillips (2008) to argue that erroneous feature percolation is inactive in comprehension.

2.2 *An overview of agreement attraction*

By far the most prominent example of agreement attraction involves complex singular subjects that contain a PP complement or adjunct with a plural DP subconstituent. The attested examples in (14) further illustrate this pattern, summarized schematically in (15):

- (14) (a) [The order of the **tasks**] **were** counterbalanced ...
Reffel, J.A., *Current Psychology*, 16, 308-315
(b) [Rise in **email viruses**] **threaten** net.
Headline, *The Guardian*, 4 Aug 2001, 2
(c) [The sheer weight of **all these figures**] **make** them harder to understand.
Ronald Reagan, 13 Oct 1982, quoted in (Francis, 1986)

- (15) [CP ... [DP D_{SG} [NP N_{SG} [PP P DP_{PL}]]]]_{SG} ... V_{PL} ...]

This agreement pattern has long been a subject of concern to grammarians (Francis, 1986; Quirk, Greenbaum, Leech, & Svartvik, 1972; Jespersen, 1924, *inter alia*), where it has often been referred to as ‘proximity concord.’ Only occasionally has it featured in research in the generative tradition (den Dikken, 2001; van Gelderen, 1997; Kimball & Aissen, 1971). However, even as early as 1883, we find the prolific Victorian novelist Anthony Trollope describing the phenomenon, as he reflects in his autobiography upon his experience as a writer:

Rapid writing will no doubt give rise to inaccuracy, -- chiefly because the ear, quick and true as may be its operation, will occasionally break down under pressure, and, before a sentence be closed, will forget the nature of the composition with which it was commenced. **A singular nominative will be disgraced by a plural verb, because other pluralities have intervened and have tempted the ear into plural tendencies ...** Speaking of myself, I am ready to declare that, with much training, I have been unable to avoid them.

(Trollope, 1883)³

Trollope’s explanation that ‘the ear ... will occasionally break down under pressure ... and forget the nature of the composition with which it was commenced,’ is essentially the intuition shared by most observers. For example, Jespersen (1924) referred to a tax on ‘mental energy’ incurred by the production of complex subject (p. 345). Quirk et al. (1972), in their extensive descriptive grammar, attribute this ‘tax’ to the increased distance between the subject head and the verb, reflecting the view that it is the relative proximity of the intervening (non-head) noun to the verb that is the controlling factor in the production of the error.

³ I thank Norbert Hornstein for bringing this passage to my attention. To the best of my knowledge, it is the earliest description of (and explanation for) agreement attraction yet discovered. Anthony Trollope was evidently a man of substantial talent, and is also remembered as the inventor of the pillar box (cf. Trollope, 1883).

What can we conclude about the encoding of syntactic structure from observations of agreement attraction? It is important to note that the error is characteristic not only of production, but also comprehension (Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2008): comprehenders experience an ‘illusion’ of grammaticality for exactly the forms that are erroneously produced, a phenomenon that we’ll expand upon below. Therefore it is likely that agreement attraction reflects deep architectural properties, and not just ones specific to production. If the agreement attraction errors truly reflect ‘proximity concord,’ – that is, the verb sometimes simply agrees with a serially or temporally nearby or adjacent noun phrase, because of its nearness or adjacency – then this phenomenon either suggests a rather weak real-time encoding of structure, or a set of real-time processes, which, unlike those postulated for the grammar, are insensitive to hierarchical structure. Such an account would cohere with some contemporary models of structural encoding in which string-adjacent or nearby elements in an expression can compete to form local structures that conflict with global cues (e.g., Stevenson, 1994; Tabor, Galantucci, & Richardson 2004).

However the apparent generalization of proximity turns out to be false: it is not the nearness of the intervening noun to the verb that leads to agreement attraction errors. Rather, the hierarchical distance between the head of the subject and intervening noun overwhelmingly determines the rate of error production. Beginning with Bock & Miller’s seminal 1991 paper, “Broken Agreement,” the phenomenon of agreement attraction was brought under experimental control in language production studies, and much later, in comprehension studies. We will now step through a number of these studies, which show that agreement attraction is highly sensitive to the structural properties of the subject

projection. Consequently we will argue that agreement attraction is less indicative of a hierarchy that is weakly structured, but rather diagnostic of how individual feature tokens are accessed in an otherwise well-formed structure.

2.3 Previous studies of agreement production and the hierarchical nature of attraction

Rates of error production in subject-verb agreement are typically studied using an elicitation paradigm. In Bock & Miller's 1991 study, participants were presented with a recorded sentence preamble, consisting of a complex subject like: "The key to the cabinets." Participants then had to repeat the preamble along with a full-sentence completion as quickly as possible⁴. If the preamble was correctly reproduced, the verb form was then scored for whether it was inflected in a number, and, if so, whether the correct inflection was used. In Bock & Miller's experiments, and in subsequent agreement attraction experiments, the experimental design typically manipulates the number of the intervening noun (henceforth, the 'attractor') with respect to the number on the head noun. For example, in Bock & Miller (1991)'s Experiment 1, preambles all contained singular head nouns, and the number on the attractor was manipulated to either match or mismatch the subject head noun. In this experiment, the 'Match' factor was crossed with the serial length of the preamble. An example set of materials is reproduced below:

⁴ In Bock & Miller (1991) participants were simply required to supply a completion to the sentence. In other studies, particular sentence types have been encouraged, in order to elicit more agreeing verb forms: for example, passives, which necessarily require an auxiliary. Subsequent studies have also used the visual modality (e.g., Vigliocco & Nicol, 1998). Since these manipulations do not influence patterns of attraction, we omit further discussion of these issues.

- (16) (a) MATCH/SHORT: The key to the cabinet
 (b) MISMATCH/SHORT: The key to the cabinets
 (c) MATCH/LONG: The key to the ornate Victorian cabinet
 (d) MISMATCH/LONG: The key to the ornate Victorian cabinets

The ‘Match’ terminology is standard in the literature on agreement attraction, but, it is one which we and many outside consumers of this literature find somewhat confusing. In this review of the literature, and the experiments presented below, we will therefore refer to conditions directly by the number on head and attractor nouns. So, in Bock & Miller’s Experiment 1, Conditions (16a) & (16c) are both ‘singular attractor’ conditions, whereas (16b) & (16d) are ‘plural attractor’ conditions. As a short hand, we will use bracketed labels like the following: Sg [Pl], where the outermost symbol indicates the number on hierarchically superior noun, i.e., the head noun in a complex subject, and the innermost symbol the number on the embedded noun, i.e. the attractor in a complex subject.

The strongest generalization about patterns of agreement with complex subjects is referred to as the ‘plural markedness’ generalization: the attractor is most likely to exert an influence on agreement when it is plural and the head noun is singular, i.e.: the Sg [Pl] form (Bock & Eberhard, 1993; Eberhard, 1997). When the head noun is plural, error rates in agreement do not depend on the number of the attractor. In their extensive meta-analysis of English agreement production experiments, Eberhard, Cutting & Bock (2005) found that, on average, plural attractors in singularly-headed complex subjects (Sg [Pl]) elicit plural agreement on the verb 13% of the time (N = 16; compared to a baseline of erroneous plural agreement for Sg [Sg] preambles of 1%; N = 14). In contrast, singular attractors in plurally-headed complex subjects (Pl [Sg]) elicit erroneous singular agreement only 3% of the time (N = 11; compared to a baseline for Pl

[Pl] preambles of 2%; N = 11). The plural markedness generalization is cross-linguistically robust and has been attested in many other languages, including ones that are morphologically richer than English: Spanish (Vigliocco, Butterworth & Garrett, 1996), Italian (Vigliocco et al. 1995), German (Hartsuiker et al., 2001), Dutch (Hartsuiker et al., 2001), French (Fayol, Largy & Lemaire, 1994) & Slovene (Harrison, 2004). The plural markedness effect restricts the class of explanations for agreement attraction, and points to a syntactically-sensitive mechanism. Firstly, it undercuts the simplest class of explanations based on adjacency or linear proximity of the attractor to the verb. If only string-locality mattered, then we should expect to see attraction errors for Pl [Sg] subjects as well as Sg [Pl] subjects. What is striking is not that the occurrence of attraction errors is merely reduced for Pl [Sg] subjects, but that it is virtually absent. Previous observers were aware of the plural markedness effect (e.g., Trollope, 1883, Strang, 1966), but it is only with the experimental studies of the past two decades that it has been possible to establish that Sg [Pl] subjects are essentially the *only* context in which the embedded number matters for error rates. Secondly, the plural markedness generalization aligns with observations of morphosyntactic markedness in the nominal system for non-singular forms (Greenberg, 1966). It is consistent with an underspecified or privative feature system for number (e.g., Harley, 1994). In such systems plural features are the ‘active’ grammatical features, in the sense that morphosyntactic rules can refer to them, but not a singular feature or singular value. Singular nouns simply bear no number feature, or no value for number, being singular ‘by default.’ It is therefore not surprising, once feature composition is taken into account, that singular attractors should be ineffective in inducing attraction.

The plural markedness generalization may suggest that agreement attraction is a grammatically-sensitive phenomenon. However, it does not entirely preclude non-structural explanations, however: for example, a more complicated adjacency explanation could hold that verb agreement can sometimes be determined by an adjacent noun if that noun bears a marked feature. More direct tests have established that the attractor influences the verb via a hierarchical syntactic representation, and not a string-linear one. The evidence supporting this claim comes from three kinds of complex subject experiments.

2.3.1 Hierarchical, not linear distance, matters.

The first kind of experiment on the effect of hierarchy shows that the hierarchical distance between the head of the subject and the attractor strongly influences the rate of attraction. Franck, Vigliocco & Nicol (2002) showed that when the subject phrase contained two stacked PP modifiers, like, “The inscription on the door(s) of the toilet(s)”, a plural noun inside the medial prepositional phrase led to more attraction errors than the most deeply embedded one. That is to say, participants were more likely to use a plural verb form when the preamble was “The inscription on the doors to the toilet” than when it was “The inscription on the door to the toilets.” In this case, a plural noun that is hierarchically closer to the subject head is more likely to induce attraction than one which is linearly more adjacent to the verb.⁵ Compared to this hierarchical manipulation,

⁵ There are other properties that align with hierarchical order in this case that could be at play. For example, “the inscription on the doors of the toilet” lends itself more easily to a distributive interpretation (in which there are multiple inscriptions”) than does “the inscription on the door of the toilets.” Intuitively there is a more direct correspondence between inscriptions and doors than between inscriptions and toilets. There is evidence in production that the notional number matters in selecting a verb form for collective nouns like ‘fleet’ or ‘gang’, though the evidence is less clear cut for typically non-collective

increasing the linear distance from the head to the attractor has relatively little effect. Bock & Miller's Experiment 1 (1991), using the materials in (16), failed to show a reliable effect of the linear distance between the head and the attractor on error rates, when structure was held constant. These sets of results, taken together, highlight the importance of the hierarchical prominence of the attractor within a complex subject, and minimize the contribution of the relative proximity of the attractor and head to the verb⁶.

2.3.2 *The attractor's 'structural domain' matters.*

The second kind of experiment demonstrates that the nature of the syntactic domain that contains the attractor impacts rates of attraction. Bock & Cutting (1992) showed that when a potential attractor was embedded in a relative clause, as in the preambles in (17), it had a smaller effect than when embedded in a PP (and when serial distance between head and attractor was controlled).

- (17) (a) The boy [_{RC} that liked the snakes]
(b) The editor [_{RC} who rejected the books]

In Bock & Cutting (1992), Sg [Pl] complex PP subjects induced 6% more attraction errors than did Sg [Pl] relative clauses: leading to plural agreement roughly 17% of the time, compared to 11% of the time for the relative clauses⁷. In Solomon & Pearlmutter

nouns like 'inscription' (cf. Vigliocco, Butterworth & Garrett, 1996; Vigliocco, Hartsuiker, Jarema & Kolk, 1996; Eberhard, 1999; Bock et al., 1999; Eberhard, Cutting, & Bock, 2005). Apart from how number is construed notionally, the agreement controller in the stacked PP cases, 'inscription', is restricted by its relation to the 'doors,' and not directly by its relationship to 'toilets'. In Solomon & Pearlmutter (2004)'s account it is the tightness or directness of the semantic relation that controls attraction (discussed in section 2.6.1).

⁶ But see Haskell & MacDonald, 2005, who argue that while linear effects are small, they are not absent.

⁷ These are the rates for responses that could be scored as agreeing or not. Sg [Sg] baselines for PP and RC subjects respectively are 4% and 1% respectively. The error

(2004), the disparity was even greater: complex subjects led to plural agreement 20% of the time, compared to only 10% of the time for relative clauses. Both Bock & Cutting (1992) and Solomon & Pearlmutter (2004) tested sentential complements, like ‘The report that Megan described the traffic accidents.’ In Solomon & Pearlmutter’s study, these showed essentially no effects of attraction (1% plural forms for Sg [Pl], compared to 3% for Sg [Sg]). Bock & Cutting (1992) showed a slightly more modest effect: 2% more errors in Sg [Pl] condition, which is still smaller than an approximate 6% attraction rate for complex PP subjects observed in that experiment (where, overall, error rates were lower).

The fact that attraction errors are relatively more infrequent in relative clauses, and even more so for sentential complements, presents a serious challenge to an account of agreement attraction which does not take the syntactic representation, or at least some structure beyond linear order, into account.

2.3.3 Ordering of verb and attractor does not matter.

The final kind of experiment directly demonstrates that the relative order of (subject head – attractor – verb) is not necessary for attraction. Vigliocco & Nicol (1998) compared rates of attraction in simple declaratives with complex subjects, as in (18a), to the polar question version in (18b), in which subject and auxiliary are inverted (comparison was between experiments). The auxiliary-inverted configuration disrupts the adjacency between the attractor and the verb, but keeps the (relative) hierarchical distance between the auxiliary and the elements inside the subject constant. Polar questions with complex subjects were elicited by instructing subjects to form questions

analysis in Bock & Cutting (1992) is not over Sg [Pl] – Sg [Sg] comparisons, which we have been reporting, but rather over Sg [Pl] – Pl [Pl] comparisons.

given a visually presented adjective (like ‘safe’) followed by a separately presented complex subject (like ‘the helicopter for the flights’).

- (18) (a) The helicopter for the flights are safe.
(b) Are the helicopter for the flights safe?

Rates of attraction were found to be nearly identical for both kinds of configuration: for declaratives, 13% (Sg [Sg] baseline: 0.3%); for questions, 14% (Sg [Sg] baseline: 2%).

The preponderance of evidence suggests that what matters most in inducing attraction is the relationship between the head of a subject projection and the attracting nouns. The likelihood of attraction occurring is (experimentally, at least) dominated by that relation; and it seems that the terms that matter are not string-linear but hierarchical. Put succinctly, the older label for agreement attraction, ‘proximity concord’ is a misnomer.

2.4 The feature percolation account of agreement attraction

Based on the existence of hierarchical distance effects, it has been proposed that agreement attraction results from feature movement or ‘percolation’ within a syntactic representation (Nicol, Forster, Veres, 1997; Franck, Vigliocco & Nicol, 2002; Eberhard, Cutting & Bock, 2005). In percolation, features on a daughter syntactic constituent can be inherited by its mother node. In the typical agreement attraction case of a subject with a PP modifier (“the key to the cabinets”), the number features bound to nouns within the PP percolate upward, valuing higher phrasal projections for number. In some proportion of cases, it can percolate up to the highest projection, that of the subject noun phrase. The verb or verb phrase, by hypothesis, is reliably valued by the number on the subject phrase, and so will be inappropriately valued in just that proportion of cases when the PP-object’s number percolates far enough to value the subject phrase. Figure 2-1, taken from

Vigliocco & Nicol (1998), shows how features on a deeply embedded nominal within the subject can be transferred to the entire subject, via the immediate dominance relations that connect it to the entire subject projection.

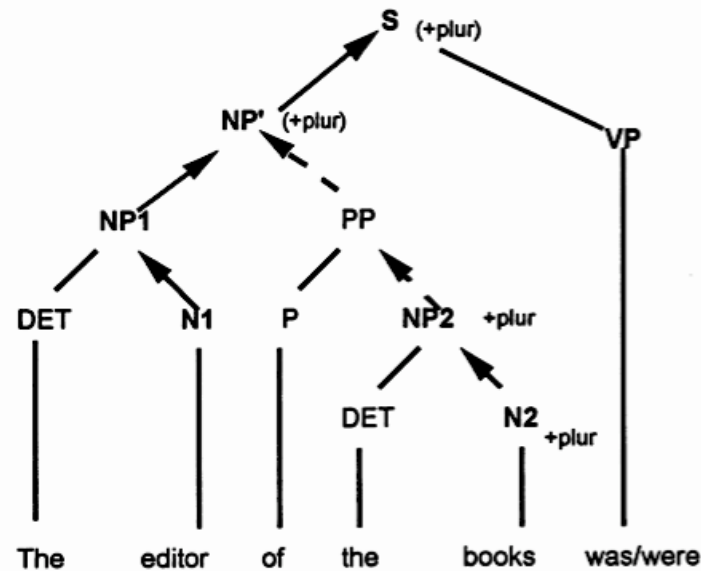


Figure 2-1 Percolation of number features in a complex subject

Upward arrows indicate the path of percolation. The [plural] feature on the most deeply embedded noun, N2, is inherited by categories along the dominance path that links it to the subject projection, NP. Taken from Vigliocco & Nicol (1998).

Percolation of an embedded nominal's features is assumed to be erroneous and probabilistic. There is some likelihood, p , that the features on a non-head node will percolate from a non-head to the projection immediately dominating it. The likelihood of traveling a given hierarchical distance is a product of the likelihood of percolating along each immediate dominance link. Consequently the overall likelihood of traversing longer syntactic distances decreases geometrically. Because the mechanism of feature percolation has this property, it can account for structural depth effects like those demonstrated by Bock and Cutting (1992) and Franck et al. (2002): the most deeply

embedded noun in a stack of PPs (like toilets in “the inscription on the door on the toilets”) or in a relative clause (like books in “the editor who rejected the books”) is uncontroversially more distant from the subject head noun than the noun in a single PP modifier (like cabinets in “the key to the cabinets”)⁸. Furthermore, feature percolation naturally encompasses the plural markedness effect if the feature system is privative, (i.e., singular number is represented by default in the absence of any number feature, whereas plural number is represented by a plural feature). Then in the case of embedded singulars there will be no feature to percolate upwards, and therefore there should be no possibility for attraction to occur.

Most recently a variant of the feature percolation account has been quantitatively formalized by Eberhard, Cutting & Bock (2005), in their ‘Marking and Morphing’ model. Their model provides a framework for marking number on the subject, by pooling the contributions of different noun sources into a continuously-valued parameter which they dub “SAP” (for ‘Singular-and-Plural’). A plural noun has lexical number specification of 1, whereas a singular noun, with no collective interpretation, has lexical number specification of 0. SAP is determined by weighting the lexical specification with the

⁸ Notice, however, that feature percolation is less successful at explaining a contrast between relative clauses and sentential complements. In both these cases, the attractor is in direct object position and equidistant from the subject’s maximal projection (modulo the presence of a *D'* in the case of sentential complements). One way to capture the fact that relative clauses induce more attraction, consistent with a feature percolation account, is to suppose that the matrix clause subject head is misassigned number in its gap position within the relative clause, which is nearer the direct object position. This account would predict disagreement within the relative clause (which has not been tested); it would also seem to depend on how the relationship between the gap position in the relative clause and the head of the DP that contains it is formally mediated.

contrastive frequency of the plural.⁹ Each head contributes its own ‘SAP’ to the overall SAP value of the subject projection, weighted by its structural distance to the subject root; see equation (19a). The continuous SAP values are mapped onto the probability of producing plural agreement on the verb via the logistic transformation (19b). Figure 2-2 summarizes this process for the phrase ‘the key to the cabinets’ (showing the values for the free parameters, which were fit from the 17 studies Eberhard, Cutting & Bock surveyed).

(19) **Marking and Morphing: Verb Model** (Eberhard, Cutting & Bock, 2005)

- (a) Root SAP $S(r)$ is the sum of the SAP, $S(m)$, of each head, j , in the subject, weighted by its structural distance, w , to the root. $S(n)$ is the default marking on the subject, which we can assume here to be 0 (i.e., singular).

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

- (b) $S(r)$ is mapped onto the probability of producing plural agreement by the logistic transform. The term b is a constant bias, capturing the baseline production of plural agreement, when only singulars are present.

$$\frac{1}{1 + e^{-[S(r)+b]}}$$

⁹ The contrastive frequency weighting captures the fact that the more frequent the plural form of a given lexical item, the less effective it is for inducing attraction (Bock, Eberhard, & Cutting, 2004). In the limit are words like ‘suds,’ which (virtually) have no corresponding singular form.

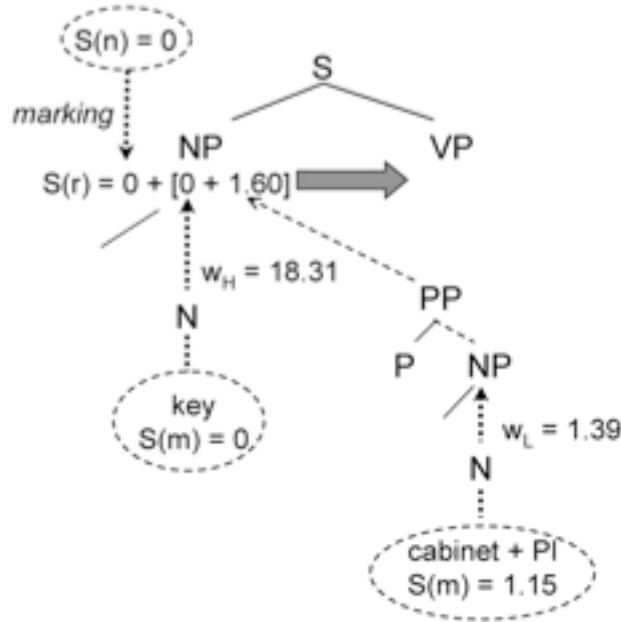


Figure 2-2 Valuing subject number for ‘the key to the cabinets’

The head ‘key’ contributes nothing to SAP since it is singular. The attractor ‘cabinets’ contributes its lexical SAP value 1.15, multiplied by the structural distance weight for nouns in that position ($w_L = 1.15$). The subject SAP value of 1.60 translates into a probability of 13.9% of producing a plural verb. Taken from Eberhard, Cutting & Bock (2005), Figure 6.

2.5 Implications of feature percolation and objections

The erroneous feature percolation model of agreement attraction has been one of the most influential models of how a non-head noun within the subject can influence number on the verb. The idea of features being transferred via immediate dominance relations seems to elegantly capture the two major generalizations discussed above: plural markedness and hierarchical distance. Eberhard, Cutting & Bock’s specific implementation of feature percolation fits with four free parameters data from seventeen experiments in which different sets of nouns, with different sets of properties, were used

(and, in addition, is extended to the analogous phenomenon of pronoun attraction).

However the representational commitments of an erroneous feature percolation model have perhaps not been adequately appreciated in the literature. There are severe implications both for the grammatical fidelity of the representations as well as resource consumption concerns for the encoding machinery. We will outline both of these concerns and consider some apparent empirical shortcomings of the percolation model. This discussion will lead to two direct tests of the model, accomplished by studying agreement processing in language comprehension.

First it bears emphasizing that what follows is not a critique of the concept of percolation, as a grammatical device. Most major theories need something like percolation to capture the fact that phrases inherit their properties from a distinguished member of the phrase, i.e., the head. There are well-understood types of rules in syntactic theory that permit feature percolation. In some theories, like *X'*-theory (Jackendoff, 1977), the phrase structure rewrite rules are (schematically) constrained so that the features of the left-hand term match the features of one right-hand term. Bare Phrase Structure (Chomsky, 1994) guarantees a similar outcome via the labeling convention for the Merge operation. Other grammars use mechanisms like implication (Lambek-style categorial grammars; Bayer & Marcus, 1996) or unification (Shieber, 1986). For example, Head-driven Phrase Structure Grammar (HPSG; Pollard & Sag, 1994) is a prominent unification grammar and it is constrained by the Head Feature Principle, which ensures the agreement features of a phrase are inherited from its head(s).

2.5.1 *Erroneous feature percolation as erroneous rule application*

What is at stake is the claim that erroneous feature percolation can occur in complex subjects, roughly in the way outlined in Figure 2-1. There are two ways of construing this claim. Under the first construal, the rule that governs feature inheritance by a mother node from its daughters is misapplied somehow, such that during the structure building process the plural feature of the embedded noun is successively inherited by its ancestors, until it reaches the subject node itself. A derivation for an erroneously plural-marked subject phrase is outlined in (20), with erroneous steps starred.

- (20) An erroneous derivation for ‘the key to the cabinets’
- (a) Build embedded DP: [DP the cabinets]_{PL}
 - (b) Build embedded PP: [PP to DP_{PL}]
 - (c) * Erroneously assign the number feature of DP_{PL} to PP: PP_{PL}
 - (d) Build highest NP: [NP key PP_{PL}]
 - (e) * Erroneously assign the number feature of PP_{PL} to NP: NP_{PL}
 - (f) Build subject DP: [DP the NP_{PL}]
 - (g) Assign plural to the highest DP: [DP the key to the cabinets]_{PL}

This rule-based construal of the erroneous feature percolation account therefore imputes grammatical ill-formedness to the encoding of the subject in two senses. First, features that are supposed to be inherited by a projection from its head, are instead inherited from its complement. Secondly, these features must either be able to pass through the PP or they must be allowed to be assigned to it. In the former case, percolation is allowed when there is not immediate dominance, as the subject projection inherits the number features

of a DP that is not its daughter node. In the latter case, the PP ends up bearing number features, which is otherwise unattested in the grammar of English.¹⁰

Feature percolation may very well not be step-by-step, along the dominance path. Den Dikken (2001) has proposed a movement analog of percolation that satisfies this condition. Den Dikken assumes that the features of an NP are promoted, via a Generalized Quantifier Raising-like mechanism, to adjoin with the determiner the NP merges with. He further assumes that successive cyclic movement is possible. Consequently, the following derivation is assumed to be possible, in which the formal features of the attractor NP, φ_2 , can QR to adjoin with the highest determiner¹¹. These

¹⁰ If percolation must occur via immediate dominance, then there are at least *two* errors of rule application necessary to get the plural feature out of the most deeply embedded DP. If that were true, then the encoding of the subject is even sloppier than it appears on the basis of error rates. The logic of this conclusion is as follows. Assume that the likelihood of producing plural agreement is a monotone decreasing function of the number of erroneous rule applications. This assumption captures the fact that increasing hierarchical distance between subject head and attractor leads to fewer errors, since the extra distance requires more erroneous rule applications. Assume that plural agreement is produced only when the final erroneous rule application occurs, which assigns the number feature of the least embedded PP to the least embedded NP. If the 10-15% of attraction errors observed for complex subjects reflects likelihood of *both* erroneous steps occurring in the same derivation, then it follows that there is a higher likelihood that only the first step occurred. More than 15% of the time, then, the feature composition of the subject is as follows, which would have no visible reflex:

(i) [DP the [NP key [PP to the cabinets]*PL]]

If erroneous rule applications are independent and equally likely events, occurring with probability p , then the ultimate error rate observed in complex subjects is p^2 . If p^2 is the rate of attraction, 15%, then the likelihood of only one erroneous rule application is approximately 39%. In such a system, over half of all complex subject of the form [DP D [NP N_{SG} [PP P DP_{PL}]]] contains ungrammatical number specifications (39% + 15% = 54%).

¹¹ While den Dikken (2001)'s justification for why the formal features of the embedded DP can licitly move upwards is somewhat obscure, the motivations for assuming it are clear. I repeat one of the arguments here. Following a claim attributed to Richard Kayne in class lectures (New York University, 1998), the author offers the following judgment: when the verb agrees with a quantified attractor inside a Sg [Pl] subject, only the wide-scope interpretation of the attractor is possible:

features c-command the formal features of the higher NP, φ_1 , and thus (by hypothesis) value the entire DP.

(21) **Formal feature movement: scoping out of embedded DP**

- (a) [DP the [NP2 cabinets]] \rightarrow_{FF} movement
- (b) [DP [D φ_2 [the]] [NP2 cabinets]]
- (c) [NP1 key [PP to [DP [D φ_2 [the]] [NP2 cabinets]]]] \rightarrow_{FF} mvmt
- (d) [DP [D φ_1 [the]][NP key [PP to [DP [D φ_2 [the]] [NP2 cabinets]]]]] \rightarrow_{FF} mvmt
- (e) [DP [D φ_2 [D φ_1 [the]]][NP key [PP to [DP [D φ_2 [the]] [NP2 cabinets]]]]]

Den Dikken assumes that there really is no error here, just an option that the grammar's generative machinery provides. However the likelihood of formal feature QR (from landing site to landing site; or the likelihood of checking agreement at LF (see fn. 11) must somehow be conditionalized on distance, to account for effects of hierarchy.

(i) The key to all the doors are missing: $\forall > \exists$: many keys; $\# \exists > \forall$: one key
 Plural number on the verb and the wide-scope interpretation seem to go hand-in-hand, which might follow if the quantifier *and* agreement features move to a higher position in the DP – and if agreement checking can take place at LF. However in both experimental studies and corpora of attested attraction, most examples do not involve an overt quantifier. Furthermore, the judgments here seem delicate to me. To my knowledge, they have not been studied experimentally, though there is a clear experimental prediction. The author points out (and attributes to Anastasia Giannakidou) that quantifiers restricted to narrow scope should not trigger attraction, as in (ii):

- (ii) *The key to few doors are missing.

I have no difficulty with this sentence on a distributed reading. But the claim could be tested, by contrasting rates of attraction between sentences like (i) and (ii). Den Dikken (2001) also wishes to account for a lack of attraction when the attractor is pronominal, as in (iii):

- (iii) *The identity of them are to remain secret.

In the same paper, it is argued that (weak) pronouns are invisible at LF, and consequently cannot trigger agreement attraction. It has recently been argued that attraction is most pronounced when both higher and lower nominals are case ambiguous (see Badecker & Kuminiak, 2007 for review). A case ambiguity requirement can explain the lack of attraction for English pronouns.

2.5.2 *Erroneous feature percolation as uncontrolled spreading activation*

There is another construal of erroneous feature percolation, in which valuation of the subject by the attractor noun's number specification does not result from the misapplication of a grammatical rule at the symbolic level. Rather, it stems from lower-level properties of how structured representations are encoded in a particular cognitive architecture. Eberhard, Cutting & Bock (2005) are most clearly proponents of this view. Features erroneously 'percolate' in structures because structures are encoded in spreading activation networks (Dell, 1986), and consequently, number features can passively migrate from node to node along active connections in any direction. In discussing their own model they write:

When a source of number information is bound to a temporary structural network for an utterance, it transmits its information to the structure. Within the structure, the information moves or spreads according to principles of structural organization, assembly, and dissolution (543).

The authors further clarify:

[T]he transmission of SAP [*the model's continuously-valued plural feature –MWW*] was treated as an activation-like process, with the weights of the connections or branches in the structural network modulating the amount of SAP that is transmitted across them (Stevenson, 1994) ... Because SAP may flow unobstructed throughout a structural network, number information bound anywhere within a structure has the potential to influence agreement processes. For this reason, even number information outside a subject or antecedent noun phrase (as in Hartsuiker et al., 2001)¹² can affect agreement, to a degree that is negatively correlated with its structural distance from the locus of agreement control (544).

¹² Hartsuiker, Antón-Méndez, & van Zee (2001) found that in production of Dutch V2 order, Subj_{SG} – Obj_{PL} – V sequences can induce agreement attraction, though at lower rates than complex subjects. These results fit uncomfortably with the feature percolation model. One option that does seem consistent, however, is to hold fast to the assumption that the number features on the object must interfere with agreement by overwriting the number features on the subject or at least by percolating up to the S level node where NP and VP unify. The difference in error rates would therefore reflect the fact that the hierarchical distance between the subject projection and the object is greater than the distance between the subject projection and a DP it contains.

The assumption that number features can be ‘transmitted’ to the structural network is not innocent. In the SAP-model, it is not explicitly stated that the ‘structural network’ is isomorphic to the phrase structure representation, but it is strongly implied in the claim that the spread of activation follows “principles of structural organization, assembly, and dissolution” (543). Spreading activation models feature in several connectionist accounts of linguistic structure. They are based on a correspondence between the shape of a tree representation and the connectivity of the network: units in the network represent nodes in a tree, and connections between nodes represent edges between tree nodes (Selman & Hirst, 1985; van der Velde and de Kamps, 2006). Such an architecture is directly relevant to the binding problem laid out in Chapter 1: how can all the familiar primitive pieces of a syntactic representation be combined together in a novel fashion, on the fly? Spreading activation models that can instantiate any arbitrary pairing of words to structure require considerable resources. The reason is simple: since connectivity between nodes is assumed to be a fixed component of the architecture, all connections must be present in the network between any two nodes that may need to be related (van der Velde & de Kamps, 2006; cf. Plate, 1994). Moreover, to encode a significant extent of structure, it is necessary to have multiple nodes for each category type, since multiple tokenings may be necessary. For example, multiple DP and NP nodes are necessary to encode a complex subject, since a complex subject contains two DP category tokens and two NP category tokens. Any lexical item that projects to a particular category must therefore be connected to multiple tokens of that category, since it cannot be known in advance which tokens of a category will be available for binding. Crucially, the same can be said of the detailed feature structure of a category: any given NP node must be able to bind any given

nominal feature token, like a number feature¹³.

The construal of erroneous feature percolation as evidence for the passive spread of information in a structural network only requires more resource demands, in the form of enriched and likely useless connectivity. Because representations are themselves structured, a model in which an activated plural feature is able to share its activation with a nearby constituent in the structure requires the existence of pathways between those constituents specifically for the transmission of number information. If feature percolation occurs strictly via dominance paths, then in order for a plural feature to spread from an embedded DP upwards, two conditions must be met. Firstly, each category along the way must be able to bind a number feature. Consequently, each category node must be connected to number nodes. As before, the number feature may be otherwise grammatically inaccurate or uninterpretable, as in the case of binding a plural feature to PP. Secondly, for number activation to spread and thus cause the percolation of the plural feature, the number node bound to one category must be connected to the number node bound to the next category in tree. Consequently, there must be rich connectivity between all the number nodes in the architecture.

Though it is tempting to think of erroneous feature percolation as the passive spread of activation between connected nodes, it seems to require positing many more nodes and connections than would be necessary otherwise. Even if we put resource consumption worries aside, we see that the network must be structured in a very particular way to

¹³ Since any given structure instantiates only a small subset of possible relations, spreading activation models must have an effective gating system, so that activation only spreads between some nodes. Van der Velde & de Kamps (2006) solve this problem by connecting category or lexical nodes not directly, but through special gating circuits. These circuits must be activated by the parsing module for activity to flow between two nodes.

permit the occurrence of feature percolation: it is by no means an automatic property of embedding a syntactic structure in a spreading activation network, despite an apparent similarity between the mechanism and the metaphor.

2.5.3 Summary of erroneous feature percolation

In the previous sections we outlined some possible mechanistic pathways for erroneous feature percolation, and argued that these mechanisms lead us to make commitments about the encoding of structure that are suspect. However even if these arguments leave us unperturbed, the fact remains that the feature percolation model of agreement attraction leads us to posit a sloppy and grammatically unfaithful encoding of the subject constituent. In the production work discussed thus far, evidence for a deficient encoding comes from the sheer fact that errors in agreement production occur. That is, we observe ungrammatical forms at particular rates controlled by structural properties of the subject. It is, however, not a necessary conclusion that the encoding of the subject is to blame for the production of these errors. Production requires the interaction of a complicated set of processes, and is not simply the direct translation of a syntactic tree into a string of words. While the fact that structural properties of the syntactic representation strongly influences agreement attraction points to an encoding problem, patterns of access involved in sequencing the planning of constituents may also play a role. Indeed, the second, process-based class of explanations for agreement attraction has more explicitly tied attraction errors to the order in which the constituents are planned in production or are made available to determine the verb form (Bock & Cutting, 1992; Solomon & Pearlmutter, 2004; Badecker & Lewis, 2008). In the next section, we'll consider how a properly encoded representation of the subject could nonetheless give rise

to agreement attraction.

2.6 The simultaneity account of agreement attraction

2.6.1 Simultaneity in planning a complex subject interferes with verb formulation

The basic idea behind process-based accounts of agreement attraction relies on some notion of simultaneous access to distinct constituents. Recall that Bock & Cutting (1992), and later Solomon & Pearlmuter (2004), observed a disparity between rates of agreement attraction induced by complex subjects containing PPs and those containing RCs. Under a percolation account, the higher rates observed in PP subjects compared to RC subjects are due to the decreased structural distance between the head and attractor in PP subjects. The embedded plural feature has to migrate a shorter distance in the PP case than in the RC case. However, Bock & Cutting originally attributed the difference to a clause-boundedness constraint on planning: features inside a relative clause are less available to interfere with the selection of the verb form, because the target verb is in a different clause. Recall that attraction errors are not totally eliminated in RC cases, so such a clause-boundedness constraint must be weakened somehow.

Solomon & Pearlmuter (2004) explain agreement attraction in the same spirit as Bock & Cutting's clause boundedness model. They assume that constituents within closely-related semantic or syntactic domains may be simultaneously activated during the mapping of a conceptual structure onto a syntactic frame¹⁴. If two similar constituents are

¹⁴ One can get the basic intuition that there is simultaneity in planning by considering another type of speech error: exchange errors (e.g., Garrett, 1980). A classic exchange error is illustrated below:

(i) There are [DP memories of theory] that do not assume storage costs.

comment by William Badecker,

10 May, 2008, Mayfest, University of Maryland

co-active, then it is possible that features in one constituent will interfere with features of the same type in the other constituent, influencing agreement processes before the subject representation is completed. What is key here is simultaneity of information in the processing system. At the same time that the complex subject is being structured, the verb form is being prepared: interference is possible to the extent that multiple sources of similar information are available as the verb morphology is being computed. Solomon & Pearlmuter (2004) do not view simultaneity as being determined by a discrete property, like whether or not two constituents belong to the same clause, but instead simultaneity is determined (at least in part) by a concept they call ‘semantic relatedness.’ To get a sense of what they mean by semantic relatedness, consider two kinds of DP (their own examples): a phrase like “the ring of silver,” consisting of a noun head and its PP modification, and a phrase like “the ketchup and the mustard,” which is a conjunction of two NPs. In the first DP, ‘ring’ and ‘silver’ are considered to be more semantically related, because this DP refers to a ring which has an inherent property characterized by silver (namely, its material composition). Consequently understanding the exact sense of ‘ring’ expressed by this phrase is contingent upon the modification with the noun ‘silver.’ On other hand, in the second DP, while ‘ketchup’ and ‘mustard’ are related by the conjunction and have a real-world relationship in the sense that ketchup and mustard often co-occur in the world, understanding what ketchup means in this phrase is not contingent upon understanding that it co-occurs linguistically with ‘mustard.’ ‘Ketchup’

Clearly the phrase “theories of memory” was intended to be produced in this case, but the speaker exchanged the hierarchically lower (and linearly later) head noun ‘theory’ with the NP head ‘memory,’ implying that production planning does not map clearly onto serial or hierarchical order. Note that the plural morphology was not exchanged (and the verb form reflects this as well).

and ‘mustard’ are less semantically related. In the planning of these DPs, Solomon & Pearlmuter propose that more semantically related nouns are more likely to be simultaneously active: therefore, ‘ring’ is more likely to be co-active with ‘silver,’ than ‘ketchup’ and ‘mustard.’

Solomon and Pearlmuter provide evidence for their view by showing a dependence between agreement attraction and semantic relatedness when structural distance is held constant. The more closely related two nouns inside a complex subject are (rated to be), the more likely that subject is to induce attraction. Nouns inside PPs that express an accompaniment or locative relations (e.g. “the chauffeur with the actors”, “the pizza with the tasty beverages”) induce fewer agreement errors than those in a functional or attributive role (e.g. “the chauffeur for the actors”, “the pizza with the yummy toppings”). Likewise, direct object nouns in relative clauses induce more agreement attraction than direct object nouns in sentential complements. The explanation of that contrast was somewhat mysterious for the percolation account because the relative distances between the subject head and the attractor are the same (see fn. 8); and it is perhaps no less mysterious in the absence of a theory of semantic relatedness (which the authors do not offer). However, ratings of ‘felt relatedness’ between two nouns in an expression do show that an RC head noun and the direct object of the RC are felt to be more closely related than the direct object of a sentential complement. Perhaps the co-argument relationship between an RC head and the RC direct object is partially to blame.

2.6.2 The relationship between planning order and hierarchy

The major challenge that a simultaneity account faces is how to capture the structural effects outlined in section 2.3. The strategy for addressing this challenge is to

posit that the order of planning, and consequently the likelihood of simultaneity between two head nouns, tracks hierarchical prominence, at least approximately. Consider Franck et al.'s (2002) contrast between rates of attraction induced by medial versus most deeply embedded NPs.

- (22) (i) the inscription [on the doors [of the toilet]] > more attraction
(ii) the inscription [on the door [of the toilets]]

It seems reasonable in this case to assume a higher likelihood of co-activation between 'inscription' and 'doors' than 'inscription' and 'toilets', since there is a more direct relationship between 'inscription' and 'doors' than 'inscription' and 'toilets' (indeed it is 'doors' that mediates the relationship between the two). However, then consider Vigliocco & Nicol's (1998) finding in question formation, that there is roughly equal occurrence of forms like "The helicopter for the flights are ready" and "Are the helicopter for the flights ready?" We can maintain a correspondence between planning order and hierarchical prominence of the head and attractor nouns, but how the order of planning those nouns is ordered with respect to the Aux will matter. Solomon & Pearlmuter (2004) are not explicit about how the Aux is selected. But, intuitively, if there is a constant relationship between the planning of the Aux with respect to the subject head noun, regardless of the Aux's output order, then the same simultaneity mechanism should apply.

2.6.3 Disentangling representation and process-based accounts

The erroneous feature percolation model of agreement attraction attributes agreement attraction errors to a faulty representation of the subject. The simultaneity account attributes agreement attraction errors to a planning process that has difficulty distinguishing between similar constituents that are simultaneously available to the

system. The consequence of both of these accounts is measured in terms of the rates of error production: that is, what is observed is that the system produces an error. However, we propose to disentangle the two accounts by looking at comprehension data. In trying to understand the relation between agreement attraction and syntactic representation, evidence from agreement processing in comprehension provides a useful complement to evidence from production. Comprehension experiments provide greater control over the representations we can force participants to entertain, and crucially allow us to observe reactions to both grammatical and ungrammatical agreement. The feature percolation account makes a clear prediction about how grammatical and ungrammatical strings containing attractors should be processed: crucially, grammatical processing should be made more difficult to the extent that ungrammatical processing is eased. By comparing word-by-word reaction times in sentence reading, we can test this prediction. First, in the next section, we spell out this prediction in greater detail.

2.7 Attraction in Comprehension

2.7.1 The Symmetry Prediction

Although the comprehension literature is smaller, its findings are generally convergent with those from production; such studies find that in the same scenarios in which individuals are likely to produce an agreement error, they experience less difficulty in processing an agreement error (Nicol, Forster & Veres, 1997; Pearlmutter, Garnsey & Bock, 1999). The types of experiments that comprise this literature, i.e., speeded judgment, reading time and electrophysiological studies, provide interestingly different measures than are found in production studies. Whereas the key measure in production studies of agreement attraction is the proportion of agreement errors that an individual

produces, the key measure in comprehension studies is an index of processing difficulty, as reflected in a localized reading time difference (e.g. Pearlmutter, Garnsey & Bock, 1999), or a signature pattern of evoked potentials (e.g. Kaan, 2002). A number of studies have also used grammaticality judgments under time pressure following rapid serial visual presentation (Haussler & Bader, submitted) or word-by-word grammaticality judgments (Clifton, Frazier & Deevy, 1999). A crucial difference from production studies is that it is possible in a comprehension experiment to compare the language processing response to both grammatical and ungrammatical agreement in the context of agreement attraction environments.

The feature percolation account of agreement attraction makes a very clear prediction for comprehension of grammatical agreement when the subject would typically induce agreement attraction in production. The feature percolation account states that a complex subject like:

(23) [DP D [NP N_{SG} [PP P DP_{PL}]]] e.g., *The key to the cabinets*

is encoded with the plural feature in some predictable percentage of cases; say, 15%.

When the input string includes this DP and an ungrammatical verb, i.e. a plural form, in 15% of cases the verb and subject should appear to agree (and appear to disagree in 85% of cases). But, likewise, when the input includes this DP and a grammatical verb, i.e., a singular form, in 15% of the cases the verb and subject should appear to disagree (and appear to agree in 85% of the cases).

This prediction we dub the symmetry prediction: The faulty encoding should lead grammatical input to be perceived as ungrammatical – to induce *illusions of ungrammatical* – just as it should lead ungrammatical input to be perceived as

grammatical – to induce *illusions of grammaticality*. The rates at which these ‘illusions’ occur should be significant and comparable. Table 2-1 works out this prediction for a specific example.

		<u>Subject</u>				<u>Verb</u>		
		The key to the cabinets				was Were		
NUMBER					PERCEIVED GRAMMATICAL?			
	Grammatical	Encoded	Percentage of cases				Yes	No
	Sg	Sg	85%		No	Yes		
		Pl	15%					

Table 2-1 The Symmetry Prediction for Feature Percolation

“The key to the cabinets” is a grammatically singular phrase. If it is erroneously encoded as plural phrase in some percentage of cases, then it should be perceived to agree with “were” in that percentage of cases, and to disagree with “was” in a similar percentage of cases.

We can thus test the erroneous feature percolation account in comprehension if we can measure perception of grammaticality across attractor and non-attractor containing sentences in which the verb matches the head of the complex subject. As mentioned above there are several methodologies that could potentially serve this purpose. We could ask participants to classify sentences as acceptable or not, and directly measure the distribution of responses, in a speeded grammaticality test. Such a test is relatively off-line however, so we also turn to self-paced word-by-word for a finer time course measure. Reaction time studies in eye-tracking and self-paced reading have shown that reaction times increase sharply at the verb or on the word immediately following it. (Pearlmutter, Garnsey & Bock, 1999). And in electrophysiology Coulson et al. (1998), Hagoort et al. (1993), and Osterhout & Mobley (1995) have all shown that subject-verb agreement failures lead to a P600, the evoked response potential that is often taken to

index syntactic processing difficulties, on the verb signalling disagreement¹⁵. Wagers, Lau & Phillips (2008) confirm that the online behavioral response to agreement violations is substantial. In Experiment 1, participants read sentences containing simple subjects, either singular or plural, and the copular or auxiliary BE, which either matched or mismatched the subject in number:

- (24) (a) GRAMMATICAL: SG/PL SUBJECT
The old key/s unsurprisingly was/were rusty from many years of disuse.
(b) UNGRAMMATICAL: SG/PL SUBJECT
The old key/s unsurprisingly were/was rusty from many years of disuse.

Figure 2-3 shows word-by-word reading times of such sentences. When participants read the verb, and immediately thereafter (Regions 5-6), there is a large main effect of grammaticality. Verbs that fail to agree with their subjects appear to substantially disrupt processing.

¹⁵ The pattern of ERPs can shift if a more meta-linguistic task is added. Osterhout & Mobley (1995) found different components when an acceptability judgment task was added: an increased P2, and a Left Anterior Negativity.

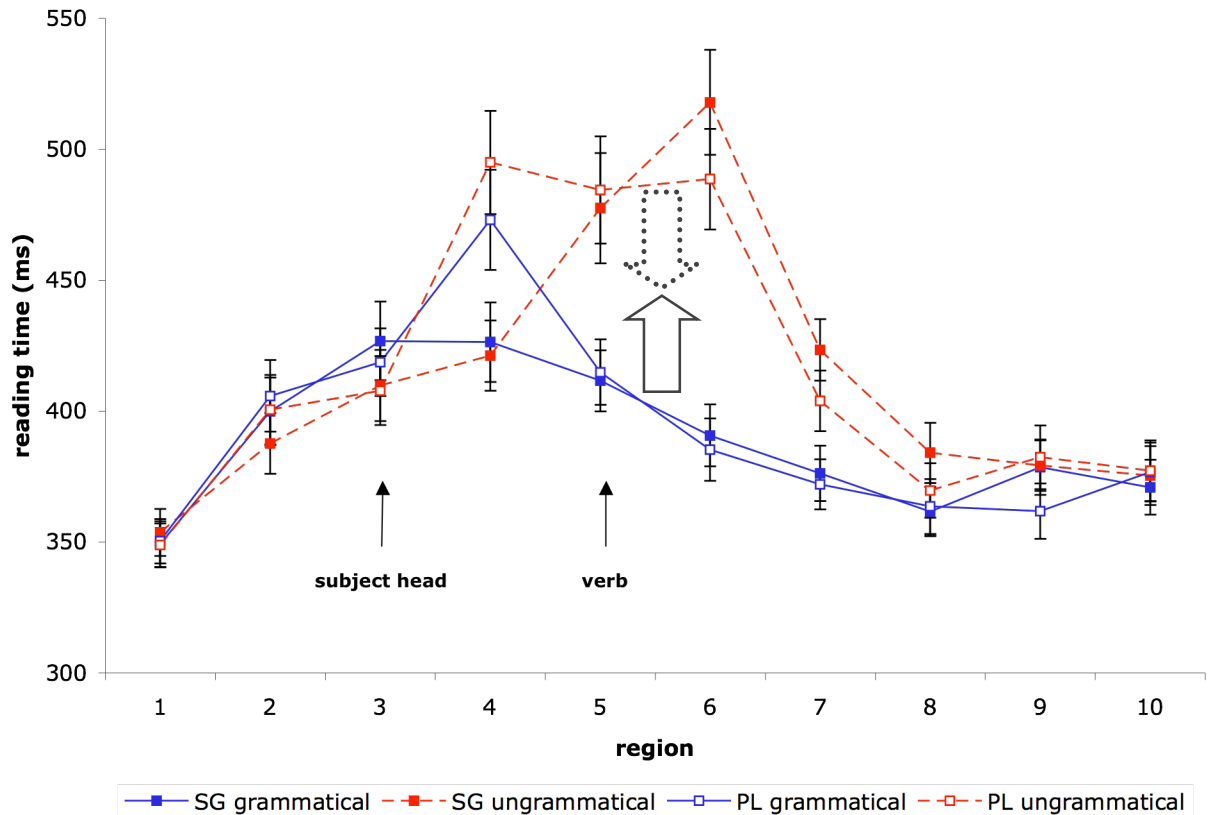


Figure 2-3 The effects of subject-verb disagreement in reading

Reading times from Wagers, Lau, & Phillips (2008) Experiment 1. Sample sentence with subscripts to indicate region coding:

The₁ old₂ key(s)₃ unsurprisingly₄ was/were₅ rusty₆ from₇ many₈ years₉ of₁₀ ...

The large block arrows indicate the expected effects based on the Symmetry Prediction for feature percolation. Error bars indicate standard error of the mean (in each cell).

Superimposed on Figure 2-3 are the expected effects based on the Symmetry Prediction for feature percolation leads us to expect the following. The Sg [Pl] sentences are predicted to be liable to both illusions of grammaticality and ungrammaticality. Therefore, reading times in the verb region should speed up for the ungrammatical sentences (dashed arrow), since more sentences should seem grammatical. But likewise,

reading times should increase for the grammatical sentences (solid arrow), since more sentences should seem ungrammatical¹⁶.

The most complete (and until the present work, nearly unique) online data set for processing agreement attraction sentences in English comes from Pearlmutter, Garnsey & Bock (1999). Do their data support the Symmetry Prediction? We will examine results from their experiments, which superficially seem to support Symmetry (as the authors concluded). However, after pointing out some problems which we believe block drawing this conclusion from their data, we report eight of our experiments (four of which are reported in Wagers, Lau, & Phillips 2008) that systematically fail to confirm the Symmetry Prediction. These results require an explanation that cannot be stated in terms of the faulty encoding of the subject; but like Solomon & Pearlmutter's (2004) simultaneity account of production, require that the verb select the wrong feature from an otherwise properly encoded representation. Structure insensitivity thus arises from how online processes navigate the structured representation.

¹⁶ One might worry that the symmetry prediction does not hold, due to the characteristically skewed distribution of RTs in SPR studies. In Appendix A, we show that the symmetry prediction does hold in mean reaction times. To spell out this logic briefly, suppose that the reading time on an individual trial for a verb which matches the number encoded on the subject projection is drawn from RT distribution M; and that the reading time for a verb which fails to match the encoded number is drawn from distribution M'. The Symmetry Prediction for reading times can be modeled explicitly if we assume that the distribution of reading times observed experimentally is a linear mixture, sampling from M and M'. The mixing proportions depend on the likelihood of a faulty encoding of the subject DP. In Appendix A, we report the simulations showing that this predicts linear effects on the means of reaction time distributions, even for the highly skewed Ex-Gaussian RT distributions that are typical of reading. The difference in means between a 85/15% M/M' distribution and a 100% M distribution is identical to the difference in means between a 15/85% M/M' distribution and a 100% M' distribution. Therefore the feature percolation account does predict a mirror effect in RTs for attraction conditions.

2.7.2 *Pearlmutter, Garnsey, & Bock (1999)*

Pearlmutter, Garnsey, & Bock (1999; henceforth, also PGB) have conducted the most extensive experiment on agreement attraction in comprehension. They considered both word-by-word self-paced reading and eyetracking measures of comprehension for grammatical and ungrammatical sentences with complex subjects containing attractors. In their experiments, they considered only singularly headed complex subjects, and crossed the number of the attractor (Sg or Pl) with grammaticality. An example materials set is given below:

- (25) (a) SG [SG] / GRAMMATICAL: The key to the cabinet was rusty.
(b) SG [SG] / UNGRAMMATICAL: The key to the cabinet were rusty.
(c) SG [PL] / GRAMMATICAL: The key to the cabinets were rusty.
(d) SG [PL] / UNGRAMMATICAL: The key to the cabinets was rusty.

Reaction time results from their Experiment 1, a word-by-word self-paced reading experiment, are plotted in Figure 2-4 (their Figure 1). Results are given as residual reading times with 95% confidence intervals (for graphical inference)¹⁷. Consider Region

¹⁷ Residual reading times linearly regress out the relationship between reading time and word length. In doing so it also takes out overall differences in means between participants. This measure is used widely in the sentence processing literature, since there is a monotone increasing relationship between word length in characters (or some property that correlates with length) and reading times. See Ferreira & Henderson (1991). However residual RTs are often used somewhat indiscriminately: a better solution is to counterbalance materials for length, so that the distribution of character lengths for a region of interest is identical across conditions. Almost all researchers, including Pearlmutter, Garnsey, & Bock (1999), do this as matter of course. In an analysis of a large corpora of our own RT experiments (including experiments not reported here), we found that there was indeed a reliable, linear RT trend in word length, when collapsed across participants and experiments: each character contributed 3-5 ms of extra reading time. However, there is considerable variability among participants (and experiments), and only a subset exemplify the overall pattern. Some participants exhibit non-linear relationships, like a stepwise increasing function; others an opposite pattern, like a decreasing linear function, and many no apparent relationship at all. More worryingly, there are widespread spillover effects in self-paced reading. The benefit of the regression may be greatly diminished, if effects of word-length are also strong in the spill-over region. If the experimenter must take to word length into account, it is best accomplished

7, the region immediately following the verb (which is outlined). There is a large grammaticality effect – compare SG [SG] UNGRAMMATICAL sentences (Symbol: ●), compared to their grammatical counterpart (○). However, the attraction sentences, SG [PL] / UNGRAMMATICAL (■), were read significantly faster than the SG [SG] UNGRAMMATICAL sentences (□). SG [PL] / UNGRAMMATICAL was still read significantly more slowly than either grammatical sentences (SG [SG] or SG [PL]), but it is a smaller effect exactly for those sentences that are produced in the elicited production tasks. By the Symmetry Prediction, we should find a comparable effect in the opposite direction, when we compare SG [PL] / GRAMMATICAL to SG [SG] / GRAMMATICAL. Indeed we find the prediction upheld in Region 7: SG [PL] / GRAMMATICAL (□) is read more slowly than SG [SG] / GRAMMATICAL (○). Consistent with the Symmetry Prediction, RTs in both attraction conditions have moved symmetrically to an intermediate position between the non-attraction conditions: there is a significant interaction between grammaticality and attractor number.

in the context of a mixed-effects model to account for individual subject variability and prior region dependencies (see Vasishth, 2006 and our section 2.9.2).

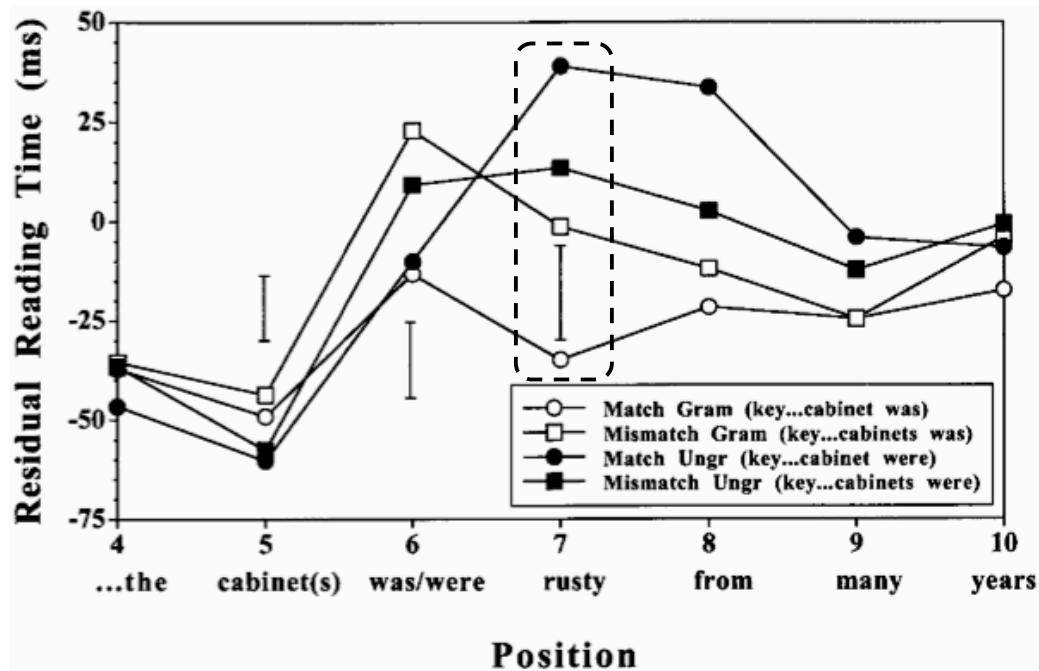


Figure 2-4 Pearlmutter, Garnsey & Bock (1999) Experiment 1 Results

Grand mean residual reading time by word position. Error bars are 95% confidence intervals for differences between cell means (over participants analysis). The Match factor refers to attractor number: 'Mismatch' corresponds to plural attractors. Region 7, showing the attractor effect, is outlined. Figure and text adapted from PGB Figure 1.

The authors conclude:

This pattern across the four conditions is consistent with the idea that the effect of a head/local NP-mismatch is to increase the probability of an error in computing the number of the subject NP, resulting in more mismatch-induced seeming errors in the grammatical conditions, but fewer mismatch-induced seeming errors in the ungrammatical conditions (436).

They also interpret the data to support Symmetry. What is worrisome about this dataset, however, is that the interaction between grammaticality and attractor number is observed following a large main effect of attractor number on the verb. The fact that there is no effect of grammaticality on the verb itself is not troublesome, as effects of processing difficulty are often delayed by a region or two in self-paced reading. However the fact

that an effect of attractor number is observed before the interaction between number and grammaticality creates a serious ambiguity for interpreting the data: does the slowdown observed in grammatical attractor conditions really reflect an illusion of ungrammaticality? or does it reflect a (temporary) shift in the baseline RT for Sg [Pl] subjects? Noun number is well-known to affect lexical decision times, although the size of this effect is modulated by root and surface frequency (New, Brysbaert, Segui, Ferrand & Rastle, 2004; Lehtonen, Niska, Wande, Niemi & Laine, 2006). There is now evidence that many of the effects from lexical decision experiments can impact techniques employed in sentence comprehension experiments, like self-paced reading and eye-tracking (Niswander, Pollatsek, & Rayner, 2000; Bertram, Hyönä, & Laine, 2000; Lau, Rozanova, & Phillips, 2007). Indeed the simple subject-verb agreement experiment summarized in Figure 2-3 shows an apparent morphological complexity effect in Region 4, one word beyond the plural DP. The main effect of attractor number observed on the verb may thus be a consequence of increased morphological complexity in that condition. If such effects can spill over to the following region, then the data do not unambiguously support the Symmetry Prediction. Notice that the same concerns do not apply the Ungrammatical attractor condition, because a baseline shift due to morphological complexity would only help to obscure any attractor effect. Since the attractor effect, in the form of a downward shift in reading times, is present, then we can still conclude that attractor conditions are easier to process when they are ungrammatical.

Pearlmutter, Garnsey, & Bock (1999) replicated the same symmetrical pattern in eye-tracking (their Experiment 2), where RT effects are typically reported to be more tightly aligned to manipulated regions. In addition to finer temporal resolution, eye-tracking

experiments also provide more measurements of processing difficulty, since it possible to track (separately) initial fixations on a word of interest, re-fixations, and total fixations¹⁸. The total fixation time measure, which adds up time spent in a region across an entire trial, shows the same interaction between attractor number and grammaticality as in the self-paced reading experiment, on the verb, the post-verbal region, as well as on the attractor itself. Because this measure incorporates re-reading (as the presence of the interaction on the attractor indicates), it is difficult to address the concerns we raised above. First-pass reading times, in which only the initial time spent in a region (that is, before moving out of that region, in any direction) is measured, are more germane. Interestingly, in first-pass reading times, there were no reliable effects of experimental condition, except a marginal by-items interaction in the region following the verb, where the SG [SG] / GRAMMATICAL condition was read reliably faster than *all* other conditions. SG [SG] / GRAMMATICAL (“The key to the cabinet is ...”) thus appears to be a distinguished experimental condition in very early measures (see Figure 2-5). Whether this has anything to do with agreement is unclear. This condition is also the only one that has no plural morphology whatsoever, nominal or verbal. As the authors stress, though, readers may adopt one of (at least) two strategies to violations in reading: spend more time on the violation region, or regress and re-read. So first-pass measures in a reading-

¹⁸ Eye-tracking experiments also putatively involve a different kind of language processing, since experimental participants have access to an external memory aid. In this respect, reading in eye-tracking experiments is more similar to naturalistic reading; one might then debate whether self-paced reading is likewise more like natural language processing, at least mnemonically.

It is possible, for example, that the nature of the encoding and navigation problem may change with an external memory. Indeed, an interesting aspect of Pearlmutter, Garnsey, & Bock (1999), which we do not discuss, is where in the text readers are likely to saccade to following a violation, and whether such saccades are under linguistic control (the short answer appears to be no).

time experiment that contains violations may not be as diagnostic of early processes as they might initially seem (see also fn. 18).

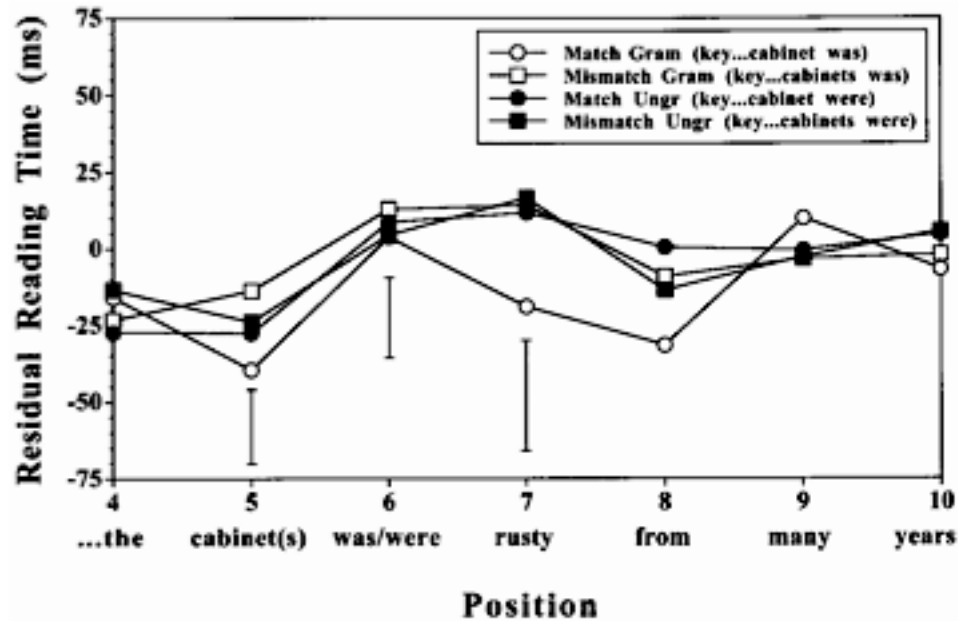


Figure 2-5 Pearlmutter, Garnsey & Bock (1999) Experiment 2 Results First-pass reading times

Grand mean residual reading time by word position. Error bars are 95% confidence intervals for differences between cell means (participants analysis). The Match factor refers to attractor number: ‘Mismatch’ corresponds to plural attractors. Figure and text adapted from PGB Figure 2.

In Experiment 3, another self-paced reading experiment, the authors replicated the grammatical slowdown observed in Experiment 1. In that experiment there were no ungrammatical conditions, only a comparison with plurally-headed complex subjects (in order to test whether the plural markedness generalization holds in comprehension). This slowdown could be liable to our morphological complexity concern, except that no slowdown is observed for PI [PI] conditions in that region (relative to PI [Sg]. But there

does appear to be at least a trend for slower RTs following plural attractors; and the PI [PI] condition is read reliably slower than PI [Sg] in the following region.

In sum, Pearlmutter, Garnsey, & Bock (1999) provide evidence that is consistent with the Symmetry Prediction of the feature percolation account. If the interactions of grammaticality and attractor number that they observe in verbal and post-verbal reading times reflect only processes of agreement, then we conclude that agreement attraction configurations can lead both to illusions of ungrammaticality and to illusions of grammaticality in comprehension. However there are concerns with the timing of the attraction effect *with respect to* the timing of putative morphological complexity effects. It is difficult for us to make this case conclusively, because there is not complete alignment across their three data sets. The first-pass measures in PGB's Experiment 2 are too non-selective for the manipulations of interest, because they do not segregate the ungrammatical conditions. If morphological complexity can explain the slowdowns observed in PGB's Experiment 3, then it must be part of an account that is sufficiently nuanced to explain why plurals in singularly-headed subjects offset reading times with a delay of one region, whereas plurals in plurally-headed subjects do so only after two regions. It is worth considering though that feature percolation is not unambiguously supported, either, even if concerns of morphological complexity could be overcome in PGB. For example, if the subject DP's number is simply faultily encoded, then one still wants an explanation for why the attractor effect appears earlier in grammatical sentences than in ungrammatical ones.

2.7.3 Summary

There are good reasons to suppose that a faulty encoding of the subject lies behind the basic agreement attraction phenomena. However as we have tried to stress, there are also good reasons for skepticism, both theoretical and empirical. We will not dwell any longer on interpreting the record, and instead report the results from several experiments which show that fallibility to agreement attraction configurations in comprehension is selective. Overwhelmingly agreement attraction leads to illusions of grammaticality for actually ungrammatical strings, but does not engender (at a rate that we can reliably detect) illusions of ungrammaticality for actually grammatical strings. This kind of pattern cannot be explained on the assumption that the subject encoding is faulty.

2.8 Testing Percolation I: Relative Clause Attraction

2.8.1 Kimball & Aissen (1971), Relative Clause Attraction & Experiment 1-2 Rationale

In the first set of experiments we present testing percolation, we will examine a non-canonical agreement attraction configuration. Thus far the discussion has focused on complex subjects like “the key to the cabinets.” Testing the erroneous feature percolation model in comprehension with complex subjects, we have argued, is difficult. Feature percolation predicts that the processing load in Sg [Pl] grammatical strings should increase, but so does the presence of a local plural also predicts a slow-down for exactly the same strings. Fortunately there is another configuration of agreement controller and attractor, one that behaves similarly to complex subjects in inducing attraction, but does not put a plural in the local environment of the verb: agreement inside relative clauses.

Kimball & Aissen (1971) first brought to our attention the contrast between the clearly ill-formed (26a) and the apparently more acceptable (26b):

- (26) (a) * The politician who the farmer refuse to vote for ...
 (b) ? The politicians who the farmer refuse to vote for ...

Below, we will show that this pattern is quite similar to the canonical agreement attraction phenomenon that obtains with complex subjects. But there are two properties of this construction that we want to first highlight which make it an interesting experimental test of an account based upon erroneous feature percolation. First, the local environment of the verb is identical in grammatical sentences where the number of the relative clause head is varied:

- (27) (a) RC HEAD SG
 The politician who the farmer refuses to vote for ...
 (b) RC HEAD PL
 The politicians who the farmer refuses to vote for ...

This construction therefore offers a potentially ‘purer’ test of whether or not grammatical subject-verb agreement is processed more slowly when an attractor is in the structure. Secondly, this configuration inverts the hierarchical relationship between subject head and attractor. We assume that the relative clause examples have an adjunction structure (see Bhatt, 1999, for a discussion of alternatives), where the NP attractor ‘politicians’ c-commands the RC subject:

- (28) **attractor** **subject**
 [DP the [NP [NP politicians] [CP who [TP [DP the farmer] refuse to vote for]]]]

In contrast, the subject head c-commands the attractor in the canonical complex subject cases of attraction:

- (29) **subject** **attractor**
 [DP the [NP key [PP to [DP the cabinets]]]] were rusty ...

Naturally this raises the question of whether there is a real parallel between the two types of attraction. In the case of complex subjects, the dominance path proceeds from attractor to subject so that the attractor's features are always only inherited by a dominating category (*attractor* DP < PP < NP *subject*). For relative clauses, the dominance path from attractor to subject goes in the opposite direction: (*attractor* NP > CP > TP > DP *subject*). This reversal of dominance relationships is only troublesome (perhaps) for the erroneous rule application construal of percolation. Recall that Eberhard, Cutting, & Bock (2005)'s more comprehensive theory holds that, "number information bound anywhere within a structure has the potential to influence agreement processes ... to a degree that is negatively correlated with its structural distance from the locus of agreement control" (544). There are arguably more major category boundaries that separate subject and attractor in the relative clause configurations (and, in particular, a clausal boundary), yet the dominance path length distance is comparable. Consequently, Eberhard, Cutting, & Bock (2005) should offer parallel explanations to complex subject and relative clause attraction.

Because the local environment around the verb is identical across attracting and non-attracting configurations, and because the hierarchical relationship between attractor and subject is inverted relative to complex subjects, the comprehension pattern for relative clause attraction is informative across a variety of basic results. Table 2-2 details the interpretation of potential reaction time patterns, if the RC sentences were compared in an experiment similar to Pearlmutter, Garnsey, & Bock (1999).

Attractor effect on RT		Erroneous feature percolation	
Ungrammatical	Grammatical	Via rule application	Via spreading activation
↓	↑	✓	Strongly supports feature percolation, both construals. No 'local plural' confound.
↓	~	✗	Contradicts feature percolation, both construals. Verifies 'attracting' property of RC heads.
~	~	✓	Consistent, upwards-only percolation
~	↑	✗	Contradicts, as currently formulated
↑	↓	✗	Consistent with no feature percolation theory. RC attraction may be fundamentally different.
↑	~		

Table 2-2 Interpretations of reading time patterns in relative clause agreement comprehension

Relative clause agreement therefore provides a compatible, and even stronger, test of feature percolation models than complex subjects. Before embarking on an experiment, it is important to consider whether there really are appropriate analogies between the two configurations. Informal intuition suggests the answer is yes. The plural markedness distinction holds: there is no similar amelioration of a plural subject – singular verb mismatch, by a singular relative clause head:

- (30) (a) * The politicians who the farmers refuses to vote for ...
(b) * The politician who the farmers refuses to vote for ...

There is suggestion of a hierarchical distance effect reported in Kimball & Aissen. They report that the amelioration appears go away if the misagreeing subject-verb pair are more deeply embedded¹⁹:

¹⁹ An interesting further claim by Kimball & Aissen, to which we cannot devote much attention to here, gives a more nuanced view of how distance from verb to relative clause head could impact agreement. One embedding seems to neutralize the amelioration of the

- (31) (a) *The politician who the farmer believes his neighbor refuse to vote for ...
 (b) ?*The politicians who the farmer believes his neighbor refuse to vote for...

Kimball and Aissen's study was observational, and they attributed this pattern to a Northeast US/Boston dialect, for whom (26a) is unacceptable, but (26b) is fine. To the (non-Northeastern) ear of myself and others, the contrast holds, however. Moreover, examples of the relative clause-type can be found in well-edited texts, such as the *New York Times*, and a University of Arizona honors thesis²⁰:

- (32) (a) We can live with the [NP errors_{SPL}][RC that classification software_{SG} make_{PL} ...]
 (Nunberg 2003, p. 5)
 (b) These include ... the [NP problem_{SPL}][RC that incrementality_{SG} pose_{PL} ...]
 (Byram 2007, p. 58)

They can also be observed in casual speech:

- (33) In what ways do the [NP hypotheses_{SPL}][RC one_{SG} entertain_{PL}] influence visual information search?

Mike Dougherty, 1 Feb, 2008, University of Maryland Psychology Department talk

subject-verb mismatch, and render the sentence bad again, as in (31b), repeated as (i)-(ii). But if plural agreement is realized on each verb in the relative clause, the contrast is claimed to re-appear, (iii)-(iv).

- (i) *The politician who the farmer believes his neighbor refuse to vote for ...
 (ii) ?*The politicians who the farmer believes his neighbor refuse to vote for...
 (iii) *The politician who the farmer believe his neighbor refuse to vote for ...
 (iv) The politicians who the farmer believe his neighbor refuse to vote for ...

Kimball & Aissen consider this to be a kind of cyclicity effect in agreement. They also note that the same attraction effect seems to hold for other A' dependencies, as in (v), Topicalization, and (vi), Wh-questions

- (v) (a) *This politician, the farmer refuse to vote for
 (b) These politicians, the farmer refuse to vote for
 (vi) (a) *Which politician do the farmer refuse to vote for?
 (b) Which politicians do the farmer refuse to vote for?

To our ear, however, the contrasts also seem to hold in these examples, though the effect inside relative clauses remains the clearest.

²⁰ RC attraction also proved difficult to avoid in the composition of this dissertation. Here is an error made during the writing of Chapter 4 (subsequently corrected):

... testing [NP sentences_{SPL}][RC in which the verbs_{SG} do_{PL} not participate in a filler-gap dependency]

Some recent experimental work using cumulative (word-by-word) grammaticality judgments confirms an amelioration for ungrammatical Sg-Pl agreement, when embedded inside a plurally-headed relative clause (Clifton, Deevy & Frazier 2001). Indeed, in the first significant production study on agreement attraction, Bock and Miller (1991) demonstrated strong attraction effects in production for relative clause constructions in their Experiment 3, and recently Franck and colleagues (2006) observed such effects in a French production experiment (though they did not find a plural markedness asymmetry). Interestingly neither Bock and Miller, nor Franck et al., make much of the structural distinction between complex subjects and the relative clause configuration.

There are clear similarities between RC attraction and complex subject attraction. More importantly, as detailed in Table 2-2, the RC attraction configurations seem capable of resolving the question of whether feature percolation is active, and, potentially, which construal of feature percolation is relevant. Therefore, Experiments 1 & 2 below report the online effects of agreement attraction in object relative clause constructions.

2.8.2 *Experiment 1*

In Experiment 1 participants read singular-subject relative clauses in a moving window, word-by-word, self-paced reading task. This experiment crossed the factor, GRAMMATICALITY, whether or not the verb inside the relative clause matched the subject in number, with ATTRACTOR NUMBER, whether the relative clause head was singular or plural. Therefore, this experiment is the relative clause analog of Pearlmutter, Garnsey, & Bock (1999)'s Experiment 1.

2.8.2.1 Materials and Methods

Note: The experiments reported in this chapter are self-paced reading and speeded grammaticality judgment studies. Because these kinds of studies re-occur in subsequent chapters, we will describe the procedure and analysis techniques in close detail below. In subsequent discussions of similar experiments we will more briefly summarize the methodological aspects of the experiment, unless there are crucial differences.

Standard Errors: Unless otherwise noted, error bars in data figures indicate the standard error of the cell means. Confidence intervals for experimental comparisons are reported in the text.

Participants

Participants were 30 native speakers of English from the University of Maryland community with no history of language disorders.

Sample set of experimental items for Experiment 1		
<i>Grammaticality</i>	<i>Attractor number</i>	
<i>Grammatical</i>	<i>Sg</i>	The musician who the reviewer praises so highly will probably win a Grammy.
	<i>Pl</i>	The musicians who the reviewer praises so highly will...
<i>Ungrammatical</i>	<i>Sg</i>	The musician who the reviewer praise so highly will...
	<i>Pl</i>	The musicians who the reviewer praise so highly will...

Table 2-3 Sample materials for Experiment 1

Materials

Experimental materials consisted of 48 sentence sets arranged in a 2×2 design with relative clause head number (singular/plural) and grammaticality (grammatical/ungrammatical) as factors. An example set is presented in Table 2-3. The first six words always contained a noun phrase modified by a relative clause, following the form *det-noun-‘who’-det-noun-verb*. The subject-verb agreement manipulated here was the agreement between the noun and verb contained in the relative clause, and thus the head noun modified by the relative clause was considered to be the ‘attractor’. The

subject of the relative clause was always singular. Because in this design the noun immediately adjacent to the verb was always singular, effects of morphological complexity were not a concern. The word following the critical verb was usually a short function word and never carried agreement. The 48 sentence sets were distributed across 4 lists in a Latin Square design, and were combined with 24 items of a prepositional-phrase agreement attraction design (all grammatical; these data are reported in section 2.9.1) and 192 filler sentences of similar length. Experiment-wide the percentage of ungrammatical sentences was 13.6%.

Procedure

Sentences were presented on a desktop PC using the Linger software (Doug Rohde, MIT) in a self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Each trial began with a screen presenting a sentence in which the words were masked by dashes, while spaces and punctuation remained intact. Each time the participant pressed the space bar, a word was revealed and the previous word was re-masked. A yes/no comprehension question appeared all at once on the screen after each sentence. The 'f' key was used for 'yes' and the 'j' key was used for 'no'. Onscreen feedback was provided for incorrect answers. Participants were instructed to read at a natural pace and answer the questions as accurately as possible. Order of item presentation was randomized for each participant. 7 practice items were presented before the beginning of the experiment.

Analysis

Only items for which the comprehension question was answered correctly were included in the analysis. Reading times that exceeded a threshold of 2.5 standard

deviations, by region and condition, were excluded (Ratcliff, 1993). Regions 2-10 were examined; the critical verb appeared in region 6. Data for each of these 9 regions were entered into a 2×2 repeated-measures ANOVA with subject number and grammaticality as factors. Using R (R Development Core Team, Vienna), ANOVAs were computed on participant mean reading times across items (F1) and on item means across participants (F2). Min F' statistics (Clark, 1973; Raaijmakers, Schrijnemakers, & Gremmen, 1999) were also computed, although because our items were counterbalanced across lists, this test is probably too conservative (see Raaijmakers et al., 1999 for discussion). These statistics are presented in full in Appendix B. Since it has been argued to be problematic to determine confidence intervals from repeated measures ANOVAs in a way that treats participants as random effects (Blouin & Riopelle, 2005), we performed a complementary analysis by fitting linear mixed-effect models to our data. Models were fit using restricted log-likelihood maximization, simultaneously controlling for subject and item as random factors; 95% confidence intervals (CIs) were then derived from these models by Markov Chain Monte Carlo simulation (Baayen, Davidson, & Bates, submitted). These are presented in the text.

2.8.2.2 Results

Comprehension Accuracy

Mean comprehension question accuracy for experimental items across participants and items was 92.3%, and did not differ across conditions (logistic mixed-effects model, ps n.s.). Two participants showed a comprehension question accuracy rate of less 80% across all items and were thus excluded from further analysis.

Reading Times

Figure 2-6 summarizes the reading time results from Experiment 1. The critical verb region (R6) did not show a main effect of grammaticality ($F_s < 1$). However, the spillover region (R7) showed main effects of attractor number and grammaticality, and crucially, their interaction. Pairwise comparisons showed that there was a significant grammaticality effect only when the relative clause head was singular (grammatical mean = 337 ms; ungrammatical mean = 399 ms; 95% CI = 27.2 ms, $p < .05$), but not when it was plural (grammatical mean = 331; ungrammatical mean = 341; 95% CI = 18.7 ms, $p > .1$). In region 8, both main effects and the interaction persisted. No significant effects were found in regions 9 and 10 ($F_s < 1$), except for a marginal effect of attractor number.

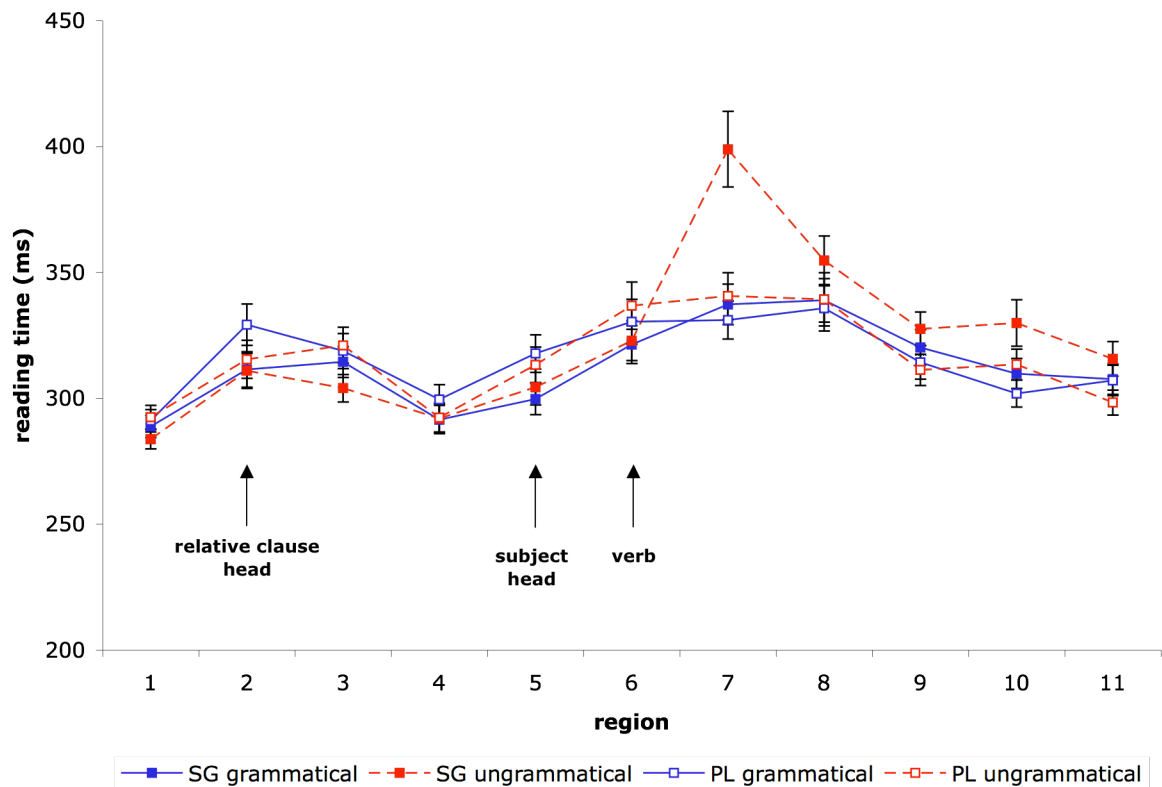


Figure 2-6 Experiment 1: Relative Clause Attraction Reading Time Results

We interpret the specificity of the grammaticality response to singular headed relative clauses, and in particular, the absence of such an effect when the relative clause head is plural, as an online attraction effect. The grammaticality effect, however, reflects the difference between ungrammatical and grammatical conditions, when a misagreeing verb is read. The crucial question is whether the grammatical conditions themselves show any sensitivity to attractor number: that is, was there an illusion of ungrammaticality? To investigate this possibility we conducted a pairwise comparison over the two grammatical conditions in the regions following the verb. In R7, the region immediately following the verb, where the grammaticality effect is strongest, no significant differences were found (singular mean = 337 ms; plural mean = 331 ms; 95% CI = 16.3 ms; $p > .1$). No differences were discovered in subsequent regions.

Also of interest is whether or not there were any effects of morphological complexity in this experiment. Region 2 (the attractor) showed a main effect of attractor number, such that the plural conditions had longer reading times (plural mean = 322; singular mean = 311; 95% CI = 7.68 ms). Since the verb was not encountered until region 6, this effect could not have been driven by agreement, and was therefore plausibly related to the additional length and morphological complexity of the plural nouns relative to the singular nouns. Longer reading times for the plural-attractor conditions persisted to the critical verb (R6); the main effect attractor number was reliable at the relative pronoun (R3) and the relative clause subject noun (R5); however, these later effects were relatively small.

2.8.2.3 Discussion

The results of Experiment 1 suggest that the head of the relative clause can act as a strong attractor for agreement. When a singular subject is followed by a plural verb, and the RC head is singular, there is a large reading time disruption, as compared to a singular subject-singular verb sequence. However, when the RC head was plural, the ungrammatical singular subject-plural verb sequences did not differ from the grammatical singular subject-singular verb condition at the spillover region. Thus, as in previous studies of agreement attraction in comprehension using PP-modified subjects, we found that an RC head attractor matching the RC verb's number significantly reduced the reading time disruption normally seen to subject-verb number mismatch. The main effect of attractor number in region 2 also corroborates the finding in Experiment 1 that plurals result in longer reading times than do singulars in self-paced reading.

These results contrast with the predictions of a feature percolation account. Such an account crucially predicts that the reduction in sensitivity to grammatical violation should be driven symmetrically: ungrammatical conditions should improve, and grammatical conditions should worsen. However, we also found that there was no attractor effect in the grammatical cases. The grammatical plural attractor condition did not show significantly longer reading times than the grammatical singular attractor condition. This result is inconsistent with the hypothesis that attraction effects in comprehension are due to a faulty representation of the subject. This finding contrasts with some previous work which did show grammatical attractor effects (e.g. Pearlmutter et al. 1999; see also Nicol et al. 1997). Consistent with our discussion above, it seems plausible that differences in the local environment were partially responsible for the differences between grammatical conditions seen in those studies. In this experiment, because of the non-intervening nature

of the relative clause configuration, the local environment was matched (singular subject). However, another possibility is that the differences we found in the ungrammatical conditions were due to some phenomenon other than what has typically been described as agreement attraction. The next experiment provides further evidence to argue that these differences do represent agreement attraction.

2.8.3 Experiment 2

We undertook Experiment 2 to further test whether what we have called relative clause attraction behaves similarly to complex subject attraction. To do so, we wanted to test whether the plural markedness generalization observed in complex subject attractions is also present in RC attraction. Kimball & Aissen's observations suggest this to be the case. To test this possibility in an online context, we conducted a more complicated version of Experiment 1 in which we added a further cross to the existing 2×2 design: all sentences in Experiment 1 had singular subjects, but in Experiment 2 we manipulated this as a third factor. If only singular subjects are liable to permit attraction, as is true in complex subjects, then we should see only a main effect of grammaticality when the subject is plural.

2.8.3.1 Materials and Methods

Participants

Participants were 60 native speakers of English from the University of Maryland community with no history of language disorders.

Materials

Experimental materials consisted of the same 48 sentence sets as in Experiment 2, but

this time arranged in a $2 \times 2 \times 2$ design with attractor number (singular/plural), subject number (singular/plural), and grammaticality (grammatical/ungrammatical) as factors. Table 2-4 gives a sample set of materials. The composition of fillers, and the experiment-wide proportion of ungrammatical sentences was the same as in Experiment 1.

Sample set of experimental items for Experiment 2			
RC Subject number	Grammaticality	Attractor number	
<i>SG</i>	<i>Grammatical</i>	<i>Sg</i>	The musician who the reviewer praises so highly will probably win a Grammy.
		<i>Pl</i>	The musicians who the reviewer praises so highly will...
	<i>Ungrammatical</i>	<i>Sg</i>	The musician who the reviewer praise so highly will...
		<i>Pl</i>	The musicians who the reviewer praise so highly will...
<i>PL</i>	<i>Grammatical</i>	<i>Pl</i>	The musicians who the reviewers praise so highly will probably win a Grammy.
		<i>Sg</i>	The musician who the reviewers praise so highly will...
	<i>Ungrammatical</i>	<i>Pl</i>	The musicians who the reviewers praises so highly will...
		<i>Sg</i>	The musician who the reviewers praises so highly will...

Table 2-4 Sample plural subject materials for Experiment 2

Procedure and Analysis

The procedure was the same self-paced reading procedure described in Experiment 1, and the analysis followed similar steps. Only items for which the comprehension question was answered correctly were included in the analysis. Reading times that exceeded a threshold of 2.5 S.D. by region and condition were excluded. Due to experimenter error the distribution of participants across the 8 lists was unbalanced across the first 56 participants. Four additional participants were recruited to balance the design at $n = 56$,

and these are the results that we discuss. However, the pattern of results did not differ from the analysis in which all 60 participants were included.

The same analysis procedures were followed as in Experiment 1. In this experiment, the comparisons of interest were all within a given level of the subject number factor; we were interested in whether the 4 plural subject conditions would show the same pattern relative to each other as the 4 singular subject conditions. In order to examine this question we split the design into two 2×2 repeated-measures ANOVAs, one for each level of subject number (singular/plural), with attractor number and grammaticality as factors. For completeness, we also computed a $2 \times 2 \times 2$ repeated measures ANOVA with attractor number, subject number, and grammaticality as factors. The results are also presented in Appendix B.

2.8.3.2 Results

Comprehension Question Accuracy

Mean comprehension question accuracy for experimental items across participants and items was 93.7%, and did not differ across experimental conditions. However, visual inspection of the means suggested that the plural-attractor/plural-subject conditions were answered less accurately than the other conditions. A post-hoc comparison revealed a reliable effect of the presence of two plural nouns (plural attractor and subject), compared to one or none (mean of zero or one plurals = 94.5%; mean of two plurals = 91.2%; $p < 0.01$).

Self-paced reading

The results of Experiment 2 are plotted in Figure 2-7 (singular subject conditions) and

Figure 2-8 (plural subject conditions). At the verb in region 6 there was a clear and consistent effect of RC subject number (plural mean = 365 ms, singular mean = 348 ms, 95% CI = 7.5 ms, $p < 0.005$), similar to the effect of subject number observed in Wagers, Lau, & Phillips (2008)'s simple subject-verb agreement study (reported in Figure 2-3).

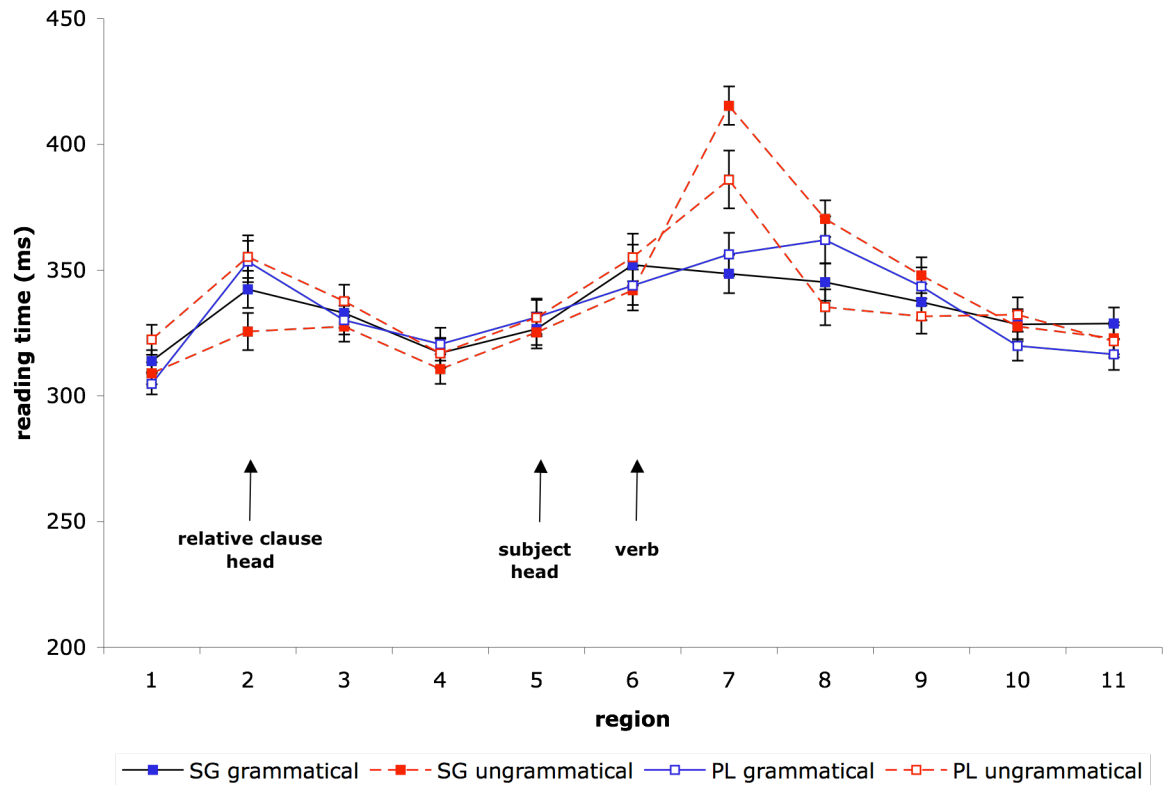


Figure 2-7 Experiment 2: RC Attraction Reading Time Results, Singular Subject

Region 7, the region following the critical verb, showed a main effect of grammaticality (mean ungrammatical = 403 ms; mean grammatical = 355 ms; 95% CI = 13.1 ms, $p < 0.005$), and a marginal interaction of grammaticality and attractor number. Splitting the design by relative clause subject number revealed a pattern of attraction similar to Experiment 1, but only for singular subjects. For singular subjects, as in Experiment 1, the plural attractor conditions showed a smaller grammaticality effect

(ungrammatical mean = 386 ms, grammatical mean = 356 ms, 95% CI = 18.6 ms, $p < 0.05$) than the singular attractor conditions (ungrammatical mean = 415 ms, grammatical mean 348 ms, 95% CI = 21.3 ms, $p < 0.005$). This interaction was marginally significant in region 7 ($p = .09$), and significant in region 8 ($p < .05$).

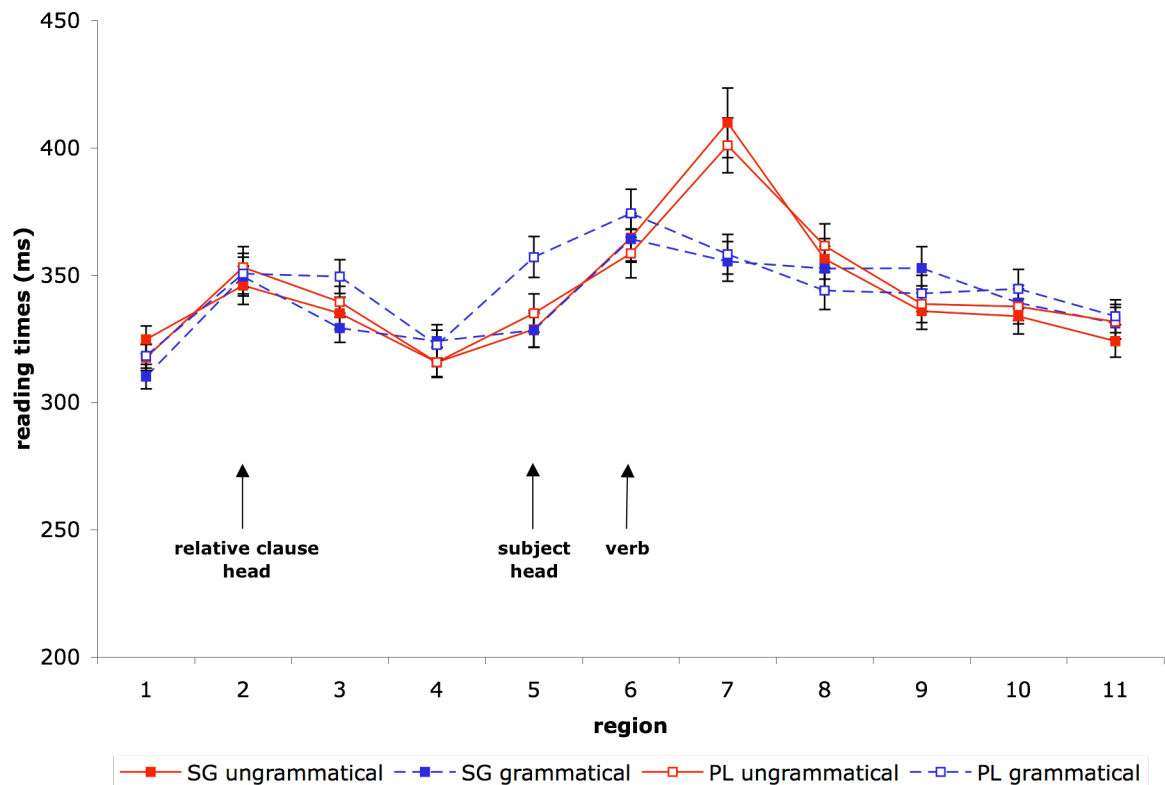


Figure 2-8 Experiment 2: RC Attraction Reading Time Results, Plural Subject

By comparison, attractor number had no impact upon the grammaticality effect for plural RC subjects (Plural attractors: ungrammatical mean = 401 ms, grammatical mean = 358 ms, 95% CI = 17.8 ms, $p < 0.001$); Singular attractors: ungrammatical mean = 410 ms, grammatical mean = 355 ms, 95% CI = 23 ms, $p < 0.0005$).

As in Experiment 1, we further tested for the existence of attraction effects in the grammatical conditions by conducting pairwise comparisons between singular and plural attractor conditions in grammatical sentences, for both singular and plural subject sets, in

the critical verb spillover region (R7). We found no significant effects of attractor in either the grammatical singular-subject conditions (R7: plural mean = 356 ms, singular mean = 348 ms, 95% CI = 15.4 ms, $p > .1$; R8: plural mean = 361 ms, singular mean = 345 ms, 95% CI = 17.1 ms, $p > .1$) or the grammatical plural-subject conditions (R7: plural mean = 358 ms, singular mean = 355 ms, 95% CI = 14.6 ms, $p > .1$; R8: plural mean = 344 ms, singular mean = 353 ms, 95% CI = 15.2 ms, $p > .1$). This pattern of results shows that attraction is asymmetric: it leads to illusions of grammatically, corresponding to the reductions in difficulty observed for ungrammatical conditions, but no illusions of ungrammatically, which would correspond to an increase in difficulty for grammatical conditions.

We examined these results for effects of morphological complexity as well. In regions 2 and 3, the relative clause head and relative pronoun, the omnibus ANOVA showed a main effect of attractor number as in Experiment 1, due to slower reading times for the plural head conditions (R2: plural mean = 353 ms, singular mean = 341 ms, 95% CI = 6.5 ms, $p < 0.01$; R3: plural mean = 339 ms, singular mean = 331 ms, 95% CI = 4.8 ms, $p < 0.05$). However, the 2 x 2 ANOVAs, split by subject number, reveal that singular subject conditions show this effect more strongly in region 2, and plural subject conditions in region 3. Because the differences were small and subject number was not manipulated until later in the sentence, this variation is presumably random. However, it is consistent with the timing of the effects reported in Pearlmutter, Garnsey, & Bock (1999)'s Experiment 1, in which Sg [Pl] subjects showed a slowdown in the region immediately following the plural, whereas Pl [Pl] subjects showed this slowdown two regions downstream. Region 5, the relative clause subject region, showed both main

effects of attractor number and subject number (RC Subj: plural mean = 337 ms, singular mean = 329 ms, 95% CI = 6.5 ms, $p < 0.05$; attractor number: plural mean = 339 ms, singular mean = 327 ms, 95% CI = 6.4 ms, $p < 0.05$). However the effect of subject number in this region appears to be carried by an exceptional value for grammatical plural attractor/plural subject conditions. Since this difference preceded the grammaticality manipulation, it is likely spurious.

2.8.3.3 Discussion

Experiment 2 replicated the basic attraction effect discovered in Experiment 1: the presence of a plural attractor in the RC-head position reduced the disruption due to ungrammatical subject-verb mismatch when the subject was singular. Furthermore, Experiment 2 confirmed the plural markedness generalization that has repeatedly been observed in studies on attraction in complex subject DPs: the attractor manipulation had no effect when the subject was plural. This finding supports the idea that the attraction effect shown here has a similar basis, despite the different ordering of attractor and subject.

Taken together, the results of Experiments 1 and 2 combined argue against a percolation account of agreement attraction in comprehension. In both experiments, we found attractor effects in singular subject, ungrammatical conditions, but we found no ‘mirror’ effect in grammatical conditions, for either singular or plural subject conditions. In other words, the presence of an attractor noun mismatching the verb in number had no effect on reading times as long as the subject and verb did match in number. The presence of a plural attractor was able to create illusions of grammaticality, but not illusions of ungrammaticality. This finding is inconsistent with all models of erroneous

feature percolation.

2.8.4 Experiment 3

Experiments 1 and 2 provide evidence that relative clause agreement is subject to attractor effects, and that the explanation cannot be a faulty encoding of number on the subject. The basis for these conclusions is patterns of reading-time difficulty experienced in the verb and post-verbal regions. An interesting question is how readers would report perceiving these violations. Convergence between online and offline measures would provide stronger evidence for the selective fallibility of agreement attraction. A simple, and usually reliable, means of obtaining acceptability judgments from a large population of informants is a pencil-and-paper rating questionnaire, in which participants rate the acceptability of a sentence on a 5- or 7-point scale. In pilot work we found that this method was insensitive to agreement attraction effects, and generated considerable variability across participants. Informal analysis and debriefings suggested that some participants in this task could be remarkably insensitive to any kind of agreement violation. Others seemed hyper-sensitive (and reported engaging in the task as if it were a proof-reading exercise). We therefore turned to speeded grammaticality judgments. In this task, participants read a sentence in the Rapid Serial Visual Presentation modality (RSVP; Potter, 1988) and must make a choice, ‘Acceptable’ or ‘Not acceptable’, under a deadline.

2.8.4.1 Materials and Methods

Participants

Participants were 16 native speakers of English from the University of Maryland community with no history of language disorders. They received credit in an introductory linguistics course for their participation.

Materials

40 of the 48 sentence sets were taken from the Experiment 2, and so crossed the factors grammaticality, attractor number (RC head: plural/singular) and subject number (plural/singular). They were distributed across 8 lists by a Latin Square. Each participant therefore saw five items per condition.

This experiment was run concurrently with a related experiment on complex subjects (Experiment 4, reported below in 2.9.3.1), which incorporated 24 sentence sets (half of the conditions in which were ungrammatical). 56 further filler sentences were included, 28 of which contained diverse types of violations, including: sequence-of-tense mismatches (“If the careful scientist had tested his data one more time, he finds that his results were wrong all along”), auxiliary selection violation (“Every new intern that the political campaign group hired will doing lots of work”), subcategorization violations in filler-gap dependencies (“The orphan to whom the millionaire inherited his fortune ...”), gender mismatching reflexives (“The businessman who made a record number of sales this year treated herself to a drink.”), and event structure violations (“The goofy clown amused the children in 30 minutes.”) Overall, each participant saw 60 nominally well-formed sentences and 60 nominally ill-formed sentences.

Procedure

Sentences were presented on a desktop PC using the Linger software (Doug Rohde, MIT) in an RSVP paradigm (Potter, 1988). Each trial began automatically, and

sentences were presented at a rate of 300 ms/word. Immediately after the last word, which was marked with a period, participants were prompted to respond whether or not the sentence was an acceptable sentence of English. The 'f' key was used for 'yes' and the 'j' key for 'no.' Participants had 2 seconds to respond. If 2 seconds elapsed with no response, participants were informed they had waited too long. The next trial began 1 second after the participant's response or the time-out. Participants were instructed to pay close attention and respond as quickly as possible. Order of item presentation was randomized for each participant. 6 practice items were presented before the beginning of the experiment.

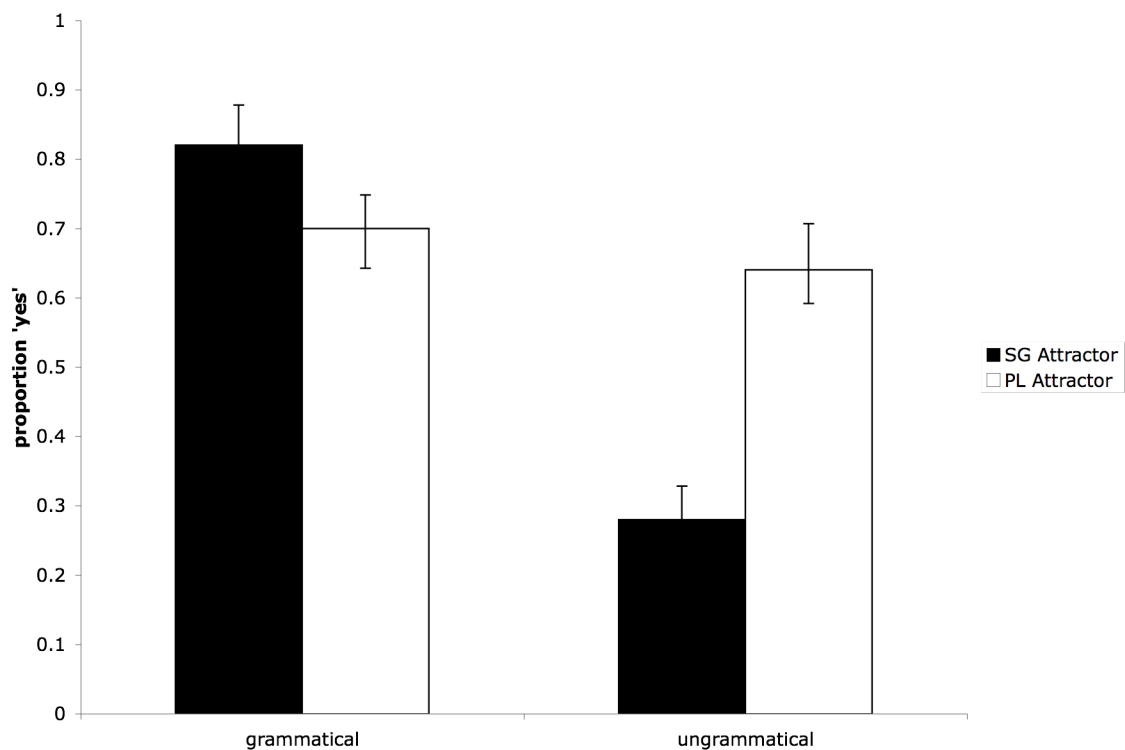
Analysis

Data were analyzed by fitting a logistic mixed-effects model (Agresti, 2002) to the fixed experimental factors of grammaticality, attractor number and relative clause subject number, and the random factors of subject and item. Logistic model coefficients reflect the contribution of the predictor variables to the probability of responding 'yes' in the judgment task. This probability is expressed as a logit, or log-odds, where $\text{logit}(p) = \log [p / (1 - p)]$. Models were fit with R (R Development Core Team, Vienna) and the lme4 package (ver. 0.99875-9; D. Bates, 2007), using restricted log-likelihood maximization. As in Experiment 2, three model fits were performed: a 'full model' that included all main effects and interactions of the complete $2 \times 2 \times 2$ design, and two models split by subject number, one 2×2 each for singular and plural subjects. Fixed effects were initially nested within the random factors, but were found to do no better at explaining the variance than the non-nested models; therefore only non-nested models are reported. 95% confidence intervals were calculated over the coefficients, as described in Experiment 1.

Out of the 640 experimental responses collected overall, there were only 11 timeouts.

2.8.4.2 Results

Figure 2-9 reports the average proportion of ‘yes’ responses for singular subject conditions, and Figure 2-10 the same values for the plural subject conditions. Before reporting the analysis, inspection of the figures reveals two patterns: first, the attractor only has any effect in singular subject conditions, consistent with plural markedness; second, the attractor has a large asymmetrical effect in singular subject conditions, inconsistent with feature percolation.



**Figure 2-9 Experiment 3: Relative clause attraction, Singular subjects
Speeded grammaticality, proportion ‘yes’ responses**

Error bars represent the standard error of the mean proportion across participants.

Full model. We observed a main effect of grammaticality: participants were less likely to respond ‘yes’ when subject and verb failed to agree (fixed effect logit coefficient β : -2.65 ± 0.78 ; $p < 0.001$). There was a marginal effect of subject number, such that participants were slightly less likely to respond ‘yes’ when the subject was plural (β : -0.68 ± 0.77 ; $p < 0.10$). Two interactions were significant: the two-way grammaticality \times attractor number interaction, such that participants were more likely to respond ‘yes’ to ungrammatical sentences when the relative clause head was plural (β : 2.33 ± 1.0 ; $p < 0.001$); and the three-way grammaticality \times attractor \times subject interaction, such that participants were less likely to say ‘yes’ to ungrammatical sentences when both the relative clause head and subject were plural (β : -2.18 ± 1.5 ; $p < 0.005$).

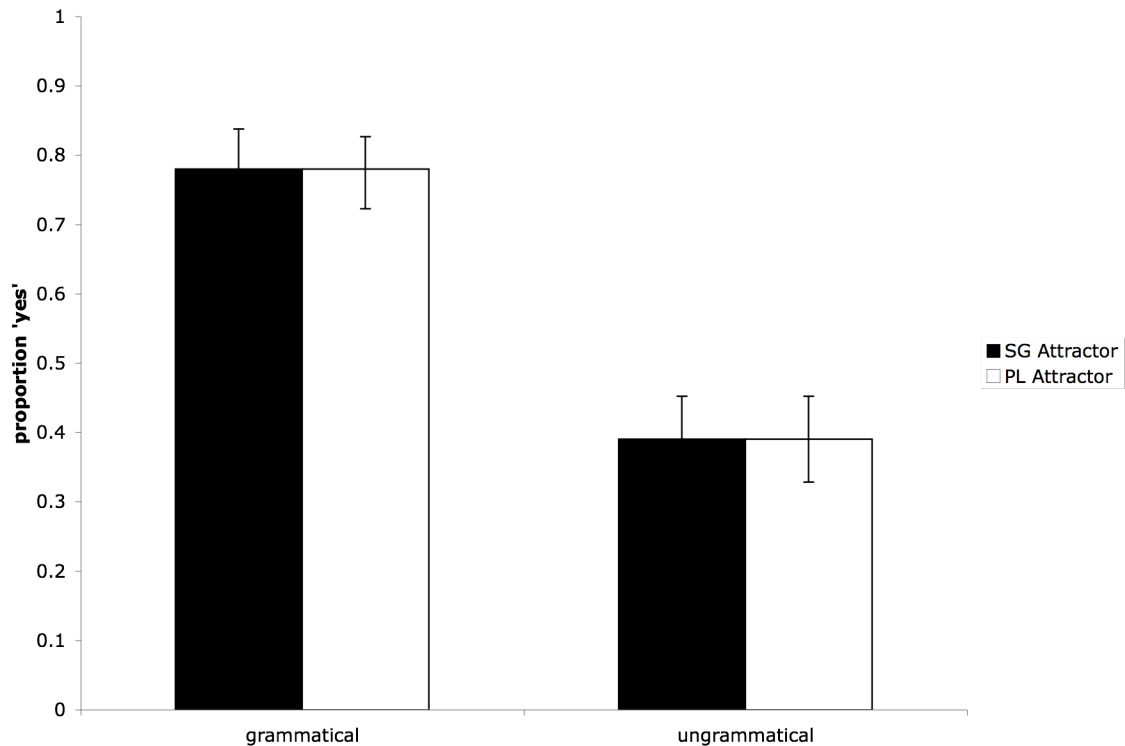


Figure 2-10 Experiment 3: Relative clause attraction, Plural subjects
Speeded grammaticality, proportion ‘yes’ responses

Error bars represent the standard error of the mean proportion across

participants.

Singular subject model. In the 2×2 model that holds subject number constant as singular, there was a significant main effect of grammaticality ($\beta: -2.50 \pm 0.76$; $p < 0.001$) and a significant interaction of grammaticality with attractor number ($\beta: 2.21 \pm 1.0$; $p < 0.001$). Participants are less likely to say ‘yes’ for ungrammatical sentences, but that likelihood increases when the relative clause head is plural. There is a marginal effect of attractor number ($\beta: -0.67 \pm 0.75$; $p < 0.10$), such that participants are slightly less likely to say ‘yes’ for grammatical sentences when the relative clause head is plural.

Plural subject model. In the 2×2 model that holds subject number constant as plural, there is only a significant main effect of grammaticality ($\beta: -2.14 \pm 0.78$; $p < 0.001$), and no reliable effects of attractor number or their interaction.

2.8.4.3 Experiment 3: Discussion

The results of the offline judgment task converge with those found in the online task. When the subject of the relative clause is singular, participants are good at detecting subject-verb disagreement; except when a plural attractor is present. The presence of the attractor in such sentences leads to a strong illusion of acceptability: participants go from rejecting sentences at a overall rate of 72%, when the relative clause head is singular, to rejecting only 36% of them, when there is a plural attractor. When the subject of the relative clause is plural, there is only an effect of grammaticality. The number of the relative clause head does not affect judgment behavior when the subject is plural, consistent with the plural markedness generalization.

The attractor does not have a symmetrical effect on judgments of grammatical sentences. There is some deviation from the online results, reflected in a marginally

increased tendency to reject grammatical sentences, when a plural attractor is present: from a baseline rejection rate of 18% for singular attractors to 30% for plural attractors. In the online task, participants did not slow down when they read grammatical sentences containing a plural attractor. In the 2×2 model, however, this effect was only marginal, so we do not make strong conclusions in the absence of a replication, or an experiment with greater power (but see Chapter 3). On the face of it, this may seem to be support for the feature percolation model, except that the effect observed is far from equivalent in size to the ungrammatical effect. The erroneous feature percolation hypothesis predicts the effects should be symmetrical. Whatever effect the attractor has, it is simply not increasing the proportion of subjects that are erroneously encoded plural.

2.9 Testing Percolation II: The Grammatical-Ungrammatical Asymmetry in Comprehension

2.9.1 Wagers, Lau, & Phillips (2008) and On-line Comprehension

Despite the apparent similarities in the attraction phenomenon between complex subjects and relative clauses, a reasonable response to the conclusions of Experiments 1 – 4 would be that the mechanism of attraction in RC agreement is clearly not feature percolation, but that this does not impact conclusions about complex subjects. This objection is most potent for a construal of feature percolation in which features may only percolate upwards. There is no upwards-only path from relative clause head to relative clause subject (unless the features may percolate from the head's copy or coindexed category in the relative clause object position). However, for Eberhard, Cutting & Bock (2005), any source of nominal number should be able to transmit its 'SAP' to the structural network, irrespective of dominance relations, and impact agreement processes.

Therefore, the lack of a grammatical contrast in relative clause attraction challenges Eberhard, Cutting & Bock (2005)'s model.²¹

As for complex subject attraction in comprehension, Wagers, Lau, & Phillips (2008; WLP) revisited Pearlmutter, Garnsey & Bock (1999)'s experiment. WLP reasoned that if what looked like a grammaticality effect in the previous comprehension experiments was really a confound of morphological complexity, then separating the local plural from the verb should decrease this 'impostor' effect. In a self-paced reading study, they had participants read sentences adapted from PGB, with a sentence-level adverb interposed between the local plural and the verb, as below:

(34) The path to the monument(s)unsurprisingly was/were littered with bottles.

The idea is that the adverb could buffer any of the processing difficulty that might spill-over (or be delayed) from the attractor. Their results are shown in Figure 2-11.

²¹ One might counter that clause boundary between the relative clause head and the relative clause subject somehow obstructs the flow of 'SAP.' However, recall that both Bock & Cutting (1992) and Solomon & Pearlmutter (2004) found that an attractor inside a relative clause could lead to errors in the main clause verb (albeit at lower rates than a PP-contained attractor). Therefore clause boundaries cannot completely block attraction.

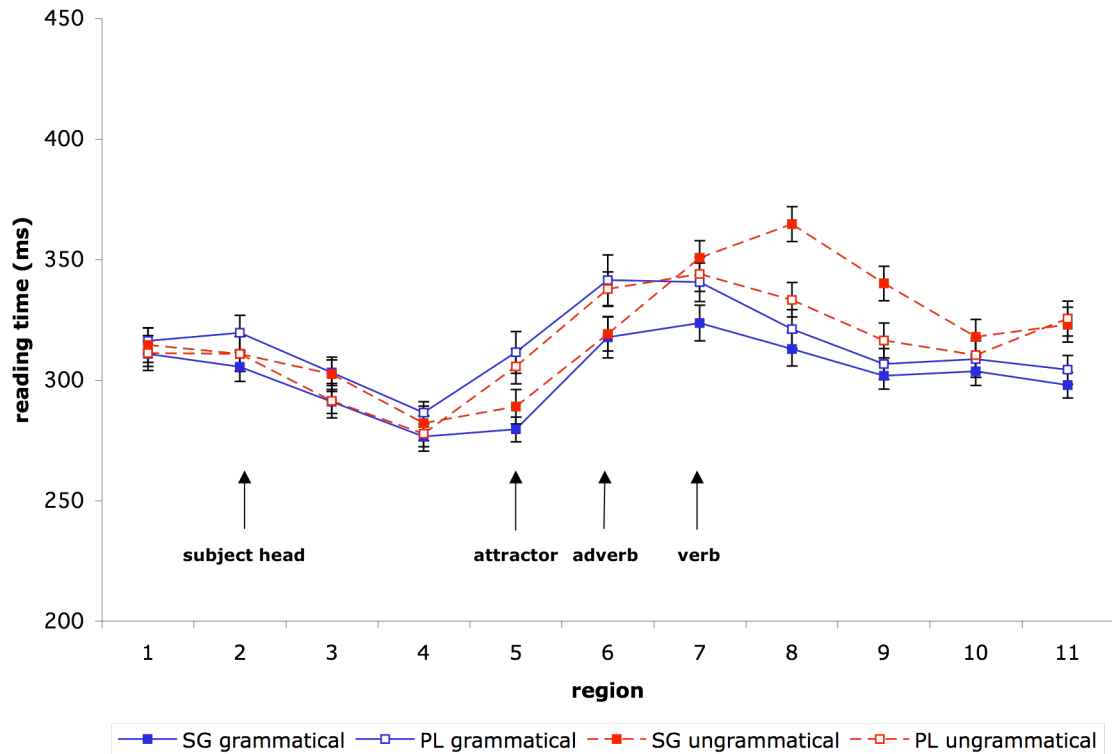


Figure 2-11 Wagers, Lau, & Phillips (2008): Complex Subject Attraction

Reading times from Wagers, Lau, & Phillips (2008) Experiment 4. Sample sentence with subscripts to indicate region:

The₁ path₂ to₃ the₄ monument(s)₅ unsurprisingly₆ was/were₇ littered₆ with₇ bottles₈ ...

Adapted from their Figure 5.

The results show that there is a cost to reading the plural attractor, reflected as main effect of number both in the attractor region itself, and on the adverb (Regions 5-6). In the immediate post-verbal region (Region 8), there is an attraction effect for ungrammatical conditions, but not one for grammatical ones. However, in Region 7, there is still a slow-down observed for grammatical conditions. It is important to note that in that region there is no difference between ungrammatical conditions. Nonetheless, because Region 7 corresponds to the the verb, it raises the possibility attractor could incur difficulty in grammatical sentences, consistent with the feature percolation account. The

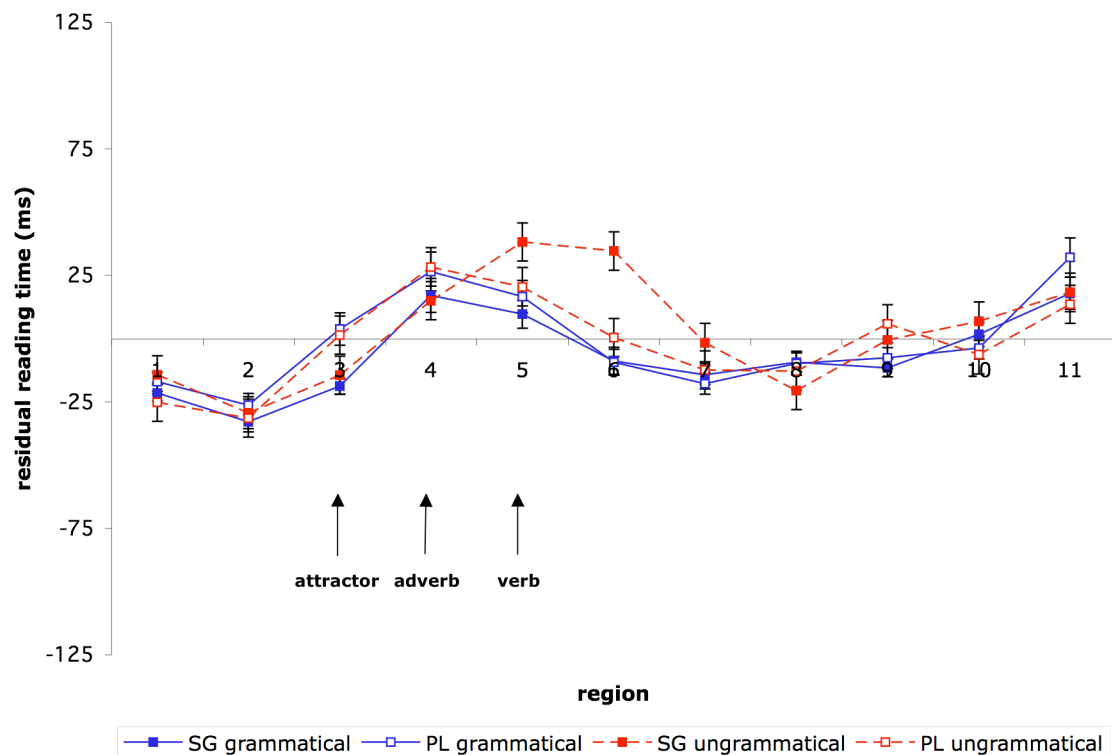
temporal disjunction between when a slow-down is observed for grammatical conditions and when a speed-up is observed for ungrammatical conditions is much stronger in Wagers, Lau & Phillips' data than in Pearlmutter, Garnsey, & Bock (1999). The post-verbal effect of the attractor on ungrammatical conditions is long-lasting, whereas the only slowdown observed in grammatical conditions does not continue past the verb.

Wagers, Lau, & Phillips (2008)'s adverb manipulation did not, on its own, fully resolve the issue of whether agreement processing contributes to the slowdown observed in grammatical conditions. The reading time effects engendered by plurals may be more long-lasting than one region, however. RC attraction Experiment 1 also showed plural effects that extended beyond one region (as did Pearlmutter, Garnsey, & Bock's Experiment 3). Based on the raw data from Wagers, Lau, & Phillips, we conducted a further analysis that more precisely localizes the source of the slowdown in grammatical conditions.

2.9.2 Controlling for RT correlations among adjacent regions: a mixed-effects models analysis

The distribution of reading times in a given region is not independent of its neighbors. There is a strong correlation between reading times within a window of 1-2 regions (unpublished observation from a large reading time corpus). We therefore analyzed the RT data from Wagers, Lau, & Phillips (2008)'s Experiment 4, by regressing out from the verb region data the contribution of the RT in the two regions that precede verb (see Vasishth, 2006). Data were analyzed using a linear mixed-effects model, as described above. Before the previous region RT regression was carried out, an unnested model, incorporating fixed and random factors, estimated the slowdown in grammatical conditions to be $18 \text{ ms} \pm 17 \text{ ms}$ ($p < 0.05$). Recall that a grammatical slowdown due to

the attractor is the crucial effect predicted by feature percolation. In reading time data sets, it is also what we suspect is contaminated by the morphological complexity of the plural attractor. After the previous region RT covariates were incorporated into the model, the slowdown shrunk to $4 \text{ ms} \pm 16 \text{ ms}$ (n.s.). Both previous region covariates were significant ($R(n-1) \beta: 0.27 \pm 0.07$; $R(n-2) \beta: 0.34 \pm 0.10$; $ps < 0.005$). These effects can be visualized by regressing out the two previous region RTs for every region and then plotting the residuals. See Figure 2-12.



**Figure 2-12 Wagers, Lau, & Phillips (2008) Experiment 4, Residual RTs
Two previous Region RTs regressed out**

On the attractor itself there is a main effect of attractor number ($p < 0.01$) but no effect on the adverb or verb. On the verb there is only a main effect of grammaticality ($p < 0.01$). The main effect of grammaticality persists into the following region ($p < 0.005$), and is marginal two regions downstream ($p < 0.10$). On the immediate post-verbal region,

there is a strong interaction between grammaticality and attractor number such that plural attractors considerably speed-up processing in ungrammatical sentences ($p < 0.05$). Most importantly, there is no effect of attractor number on the grammatical conditions in any of the verbal or post-verbal regions.

In this analysis we attempted to assess the contribution that each new word made, independently of the preceding RT baseline. One region beyond the verb we found strong evidence that ungrammatical strings were read considerably faster in sentences containing an attractor. However, we found no evidence at or beyond the verb that grammatical strings were read more slowly in attractor sentences. If agreement checking can only commence once the information on the verb is available, the only effect an attractor has on agreement is in the ungrammatical conditions. We also found an independent contribution of morphological complexity introduced by the attractor. These results strongly contradict the symmetry predictions of the feature percolation model of agreement attraction. They are consistent with our interpretation of previous data: the apparent slow-downs in grammatical attractor conditions observed in Pearlmutter, Garnsey, & Bock (1999) and Wagers, Lau, & Phillips (2008) are not due to the impact of the attractor on agreement checking in those conditions; they reflect the contribution of reading a plural in the attractor position, independently of its relationship to the verb.

2.9.3 Experiment 4: Speeded grammaticality tests of complex subject attraction

As with RC attraction, we can ask what comprehenders report perceiving when they process grammatical and ungrammatical complex subject sentences which contain agreement attractors.

2.9.3.1 Experiment 4a: Singular complex subjects

When we conducted Experiment 3, we also incorporated Wagers, Lau, & Phillips (2008)'s complex subject stimuli in a speeded grammaticality task. These stimuli are the canonical complex subject attraction sentences, without adverbs. The design was a 2×2 cross of grammaticality and attractor number. An example set of materials is given below:

- (35) (a) GRAMMATICAL/SG ATTRACTOR
The path to the monument is littered with bottles.
(b) GRAMMATICAL/PL ATTRACTOR
The path to the monuments is littered with bottles.
(c) UNGRAMMATICAL/SG ATTRACTOR
The path to the monument are littered with bottles.
(d) UNGRAMMATICAL/PL ATTRACTOR
The path to the monuments are littered with bottles.

Participants, Procedure and Analysis details are identical to Experiments 3 above.

Figure 2-13 reports the proportion of 'yes' judgments for the four experimental conditions. The results of the 2×2 logistic mixed-effects model confirm what an inspection of the figure suggests: there is main effect of grammaticality (β : -4.0 ± 1.0 , $p < 0.001$) and an interaction of grammaticality with attractor number (β : 1.7 ± 1.3 , $p < 0.01$). Participants are more likely to accept ungrammatical sentences, when an attractor is present. Acceptance rates more than double (Sg [Sg] ungrammatical: 25%, Sg [Pl] ungrammatical: 55%), which is comparable to the same effect observed in RC attraction (Experiment 3). However, participants are not more likely to reject grammatical sentences, when there is an attractor (β : -0.26 ± 1.1 , n.s.).

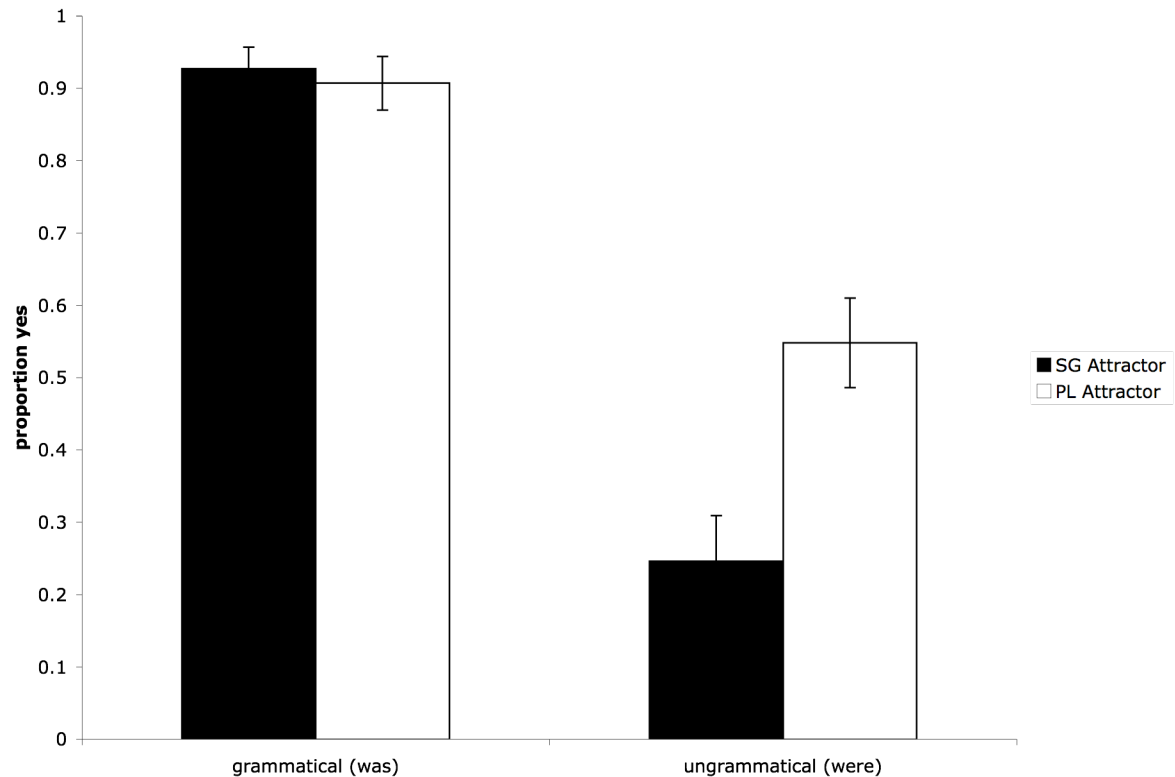


Figure 2-13 Experiment 4a: Complex subject attraction, singular subjects
Speeded grammaticality, proportion ‘yes’ responses

Error bars are standard error of the mean proportion across participants.

2.9.3.2 Experiment 4b: Plural complex subjects

We conducted a version of Experiment 4a using plural complex subjects, to test the plural markedness effect for complex subjects in this task. The Procedure and Analysis were identical to Experiment 4a. A sample set of materials set is given below:

- (36) (a) GRAMMATICAL/PL ATTRACTOR
The paths to the monuments are littered with bottles.
- (b) GRAMMATICAL/SG ATTRACTOR
The paths to the monument are littered with bottles.
- (c) UNGRAMMATICAL/PL ATTRACTOR
The paths to the monuments is littered with bottles.
- (d) UNGRAMMATICAL/SG ATTRACTOR
The paths to the monument is littered with bottles.

In this experiment, there were 24 participants, all members of the University of Maryland community. They were awarded credit in an introductory linguistics course for their participation.

Figure 2-14 reports the proportion ‘yes’ judgments for the four experimental conditions. The results of the 2×2 logistic mixed-effects model confirm that there is only a main effect of grammaticality ($\beta: -3.0 \pm 0.9, p < 0.001$). Consistent with the plural markedness generalization, the attractor has no impact on judging behavior for plurally headed complex subjects.

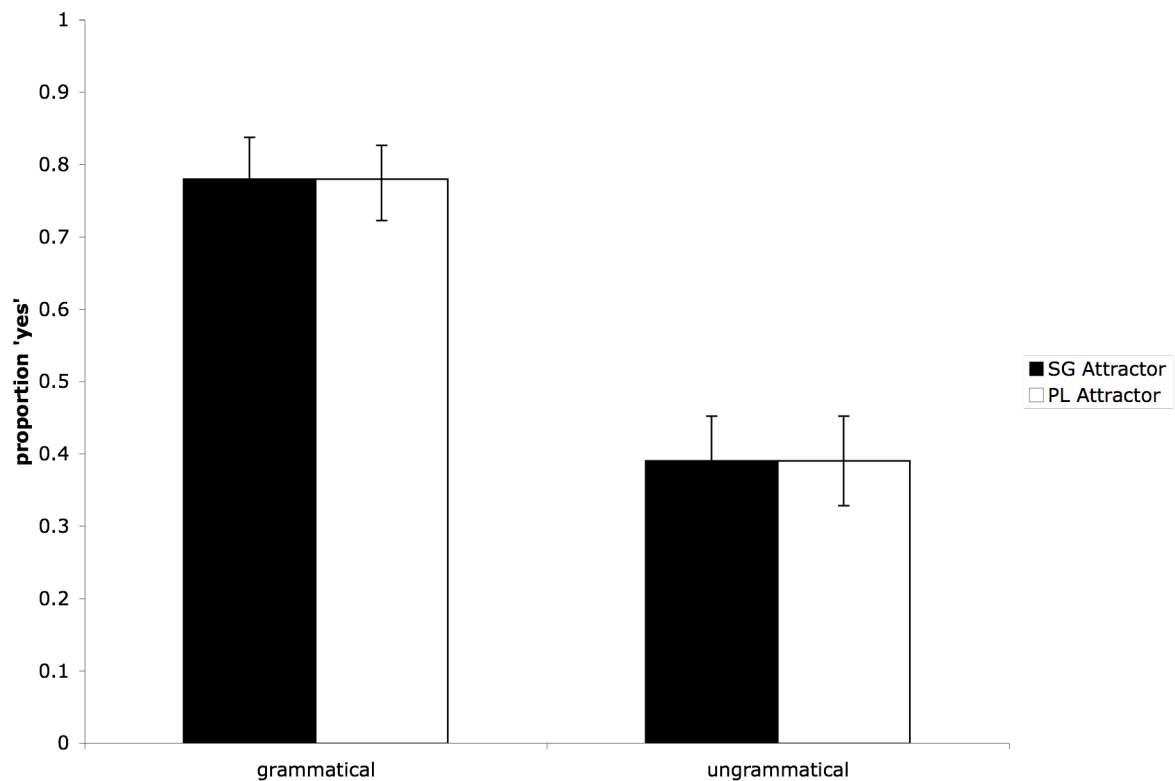


Figure 2-14 Experiment 4b: Complex subject attraction, plural subjects
Speeded grammaticality, proportion ‘yes’ responses

Error bars represent the standard error of the mean proportion across participants.

2.10 Conclusions

Agreement attraction has plausibly been argued to result from the faulty encoding of a complex syntactic object: the subject. What is faulty about this encoding is that it incorrectly binds the number feature of a nearby noun phrase as its own. The feature percolation model explains this ‘mis-binding’ as a property of the encoding architecture: features can be mistakenly passed from the attractor noun to the subject projection by means of the dominance paths between categories in the structure. We have argued against erroneous feature percolation based on skepticism of its mechanistic foundations (I) and on a series of comprehension experiments that show agreement attraction is selectively fallible (II):

- (I) The candidate mechanisms for passing features by means of structural links imply either an encoding system that is frequently insensitive to the core grammatical notion of headedness (erroneous rule application construal), or a mechanism of encoding that must consume resources to make errors (passive spread of activation construal in a structural network).
- (II) The comprehension facts are inconsistent with feature percolation. Because feature percolation determines the grammatical number of the subject projection, it predicts that comprehenders should perceive some ungrammatical strings as grammatical and symmetrically perceive a similar proportion of grammatical strings as ungrammatical. In reading times and in speeded grammaticality judgments, we find clear evidence

that comprehenders perceive some ungrammatical strings as well-formed.

In our own data, there is scant evidence that grammatical strings are perceived as anything but grammatical.

The comprehension results not only rule out erroneous feature percolation models, but, more generally, any model in which the subject encoding and only the subject encoding is to blame for the errors. What is required to explain the results is sensitivity to the match between the information carried by the verb, and information contained by the subject projection. When the verb and the head of the subject match, the attractor does not intrude into agreement checking. When the verb and the head of the subject *fail to match*, the attractor can exert an influence. In the terms in which we have framed the problem, it is a ‘navigation’ question: how the comprehender uses the input to control reference to the preceding syntactic context.

In Chapter 3 we will defend an account of agreement attraction in comprehension based on cue-based retrieval in a content-addressable memory. We will argue that, though the logical space for comparison has been undersampled, comprehension and production attraction phenomena stem from the same kind of memory and control architecture. Finally in an attempt to unify a number of different empirical domains, we compare agreement attraction errors to the other instances of grammatical infidelity that have been documented, as well as instances that are resistant to grammatical infidelity.

3 The trouble with subjects Binding and accessing features in complex syntactic objects, Part II

3.1 Introduction

In Chapter 2 we introduced the phenomenon of agreement attraction, in which the number feature of the verb matches a non-subject projection. We argued that agreement attraction could not result from a faulty encoding of the subject, in which the subject projection binds the wrong number feature. Our strongest evidence against faulty encoding came from a set of comprehension experiments, reported in Chapter 2 and in Wagers, Lau, & Phillips (2008), in which non-subject projections only impact processing when the verb fails to match the subject head in number features. This asymmetry is unexpected if the subject projection characteristically mis-binds the number feature of the attractor as a part of the encoding process.

In this chapter, we will offer an explanation for how this asymmetry could arise. We posit that subject-verb agreement is licensed in part by a retrospective search operation in a content-addressable memory (McElree, 2006). We substantiate this claim by presenting a simple, formal model of agreement licensing that makes the correct predictions with respect to both complex subjects and relative clause attraction. We show that the model extends to a novel variant of relative clause attraction, confirmed by a speeded grammaticality experiment. Agreement attraction is thus taken to exemplify a processing architecture in which partial matches can arise due to retrieval interference.

The concept of interference, and particularly interference stemming from cue-based memory retrieval, has recently achieved prominence in the sentence processing literature (Gordon, Hendrick, & Johnson, 2001, 2004; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). There is naturally interest in identifying the commonalities of sentence processing with other cognitive tasks. The existence of a broadly successful syntactic structure building model (Lewis & Vasishth, 2005), based on John Anderson's general ACT-R model of cognition (Anderson et al., 2004), has strengthened this pursuit. However many of the conclusions in this literature have seemed to be at odds with assumptions, implicit perhaps, in grammatically-based theories of syntactic recognition and processing. There are two closely-related, but distinct, architectural claims that bear evaluating. Before returning to agreement attraction, let us lay out the relevant architectural terrain, though the reader may proceed directly to section 3.3 for the agreement attraction model.

The first claim we will consider is that the search mechanism occurs via cue-directed retrieval in a content-addressable mechanism. This mechanism is fast and affords direct access to target information. Yet it poses a challenge to enforcing structure-sensitive restrictions on grammatical dependencies because it is controlled by item and not relational properties. The second claim is that the effective processing workspace is highly restricted and incapable of maintaining much information. As a consequence most discontinuous information has to be retrieved. This property also presents challenges to structure-sensitivity, as it reduces the effective syntactic context that can be consulted to make parsing decisions.

3.2 Searching structure with unstructured searches

3.2.1 Content-addressable search

3.2.1.1 Fundamental properties

The innovative claim in recent retrieval-based accounts of syntactic processing (McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005) is that access to the syntactic context is not granted by means of a structurally-ordered search process, but rather by means of a direct access, content-addressable search, which renders all relevant constituents simultaneously available. The diagram in Figure 3-1 illustrates an abstract phrase structure, in which a head, X, must be licensed by a c-commanding category in the structure with the feature $+\alpha$. This scenario is a pervasive one in natural language: agreement between subject and verb fit this schema, as does *wh*-movement (or movement dependencies generally).

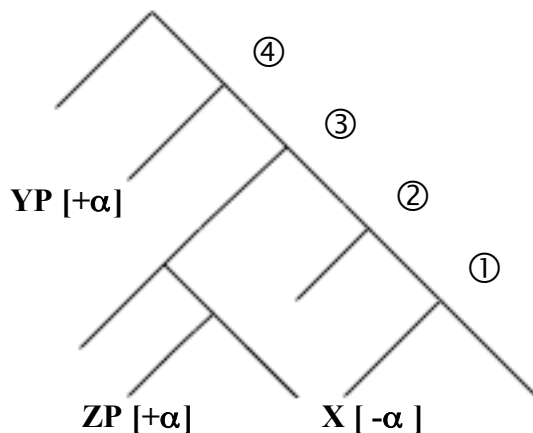


Figure 3-1 Accessible and inaccessible licensers in an abstract tree

Category X must be licensed by the feature α . YP is an accessible licenser, as it is c-commands X. ZP is inaccessible.

Imagine that we are building structure left-to-right and have reached X, which must now be licensed by a $+\alpha$ category that c-commands it. In order to find a c-commanding $+\alpha$ category it is possible to specify a search process that traverses the

dominance paths, starting from X. Here is how a very simple algorithm would work in binary branching structure.

- (37) (a) Begin at the node immediately dominating X: XP.
(b) Move to the node immediately dominating XP.
(c) Inspect its daughters. Do any bear feature $+\alpha$?
(d) Yes: X is licensed by that daughter. Terminate search.
No: Return to step (b).

The order in which this process visits nodes is illustrated in Figure 3-1 with circled numerals. The search terminates on node ④, where it finds YP $+\alpha$. This process can straightforwardly be modified to incorporate appropriate grammatical restrictions. Additional termination conditions can be added: for example, the search could terminate at cyclic nodes, in the search for $[\text{+wh}]$, or terminate at clause nodes to find the antecedent of a reflexive anaphor. Conditions on the inspection process could restrict the search to particular kinds of daughters, e.g., specifiers, not heads. The key point is that the order in which the process generates and then selects candidate licensing categories is governed by phrase structural relations. Our simple procedure belongs to a larger class of well-specified algorithms known in computer science as node visitation algorithms (cf. Knuth, 1965/1997). Node visitation algorithms provide a means for accurately and exhaustively searching a tree structure, by traversing its paths and visiting each node once. Crucially for syntactic processing, a node visitation algorithm permits relational restrictions like c-command to be incorporated into the search. This outcome seems desirable, as grammatical generalizations are often almost always stated over the relational properties that hold between two categories, in addition to their inherent properties.

How would the same licensing procedure illustrated in Figure 3-1 occur in a content-addressable memory? The key idea in such a memory is that search begins with a (sub)set of information that the desired representation should contain. We will refer to this as the retrieval structure or the set of retrieval cues. This information is then compared simultaneously with all representations in memory to generate a set of candidate matches, based on the similarity of the encoding in memory with the retrieval structure. In theories of recognition and recall, this property is exhibited by ‘global matching models,’ and it can be implemented in diverse architectures (Clark & Gronlund, 1996). Models can differ on how the strength of the match maps onto selection, depending on task. For purposes of illustration, we will assume that the (normalized) match strength maps onto the probability that a particular representation will be selected (see section 3.3.2.1 below). In our hypothetical example, it is straightforward to see how to generate candidates based on the inherent features necessary for licensing node X: the retrieval structure would contain the [+ α] feature. But this cue alone would return both the c-commanding and non-c-commanding categories, YP and ZP. It is unavoidable that grammatically-illicit constituents should be returned in some cases, if the search initially identifies constituents’ encodings on the basis of their contents. Natural language representations repeat many of the same motifs and are potentially recursive; consequently they are likely to be self-similar. This outcome could have one of two effects on language comprehension. The presence of multiple candidates could render licensing or dependency formation more difficult if a subsequent process had to then somehow select the grammatically-licit antecedent. Alternatively it could make the process more error-prone, if there were no subsequent process to select the

grammatically-licit antecedent. That is, the system may unknowingly construct illicit representations. Van Dyke (2007) has argued this is exactly what happens when verbs attach to complex subjects (ones that contain other subject projections), a claim that we discuss in great detail in section 3.4.2.

3.2.1.2 Evidence

McElree (2000) and McElree, Foraker & Dyer (2003) have argued that the generation of candidate constituents in memory search reflects direct access, consistent with a content-addressable search architecture. There are three crucial components they use to make their case: (I) a contrastive prediction between ordered and direct access searches, (II) a measure for evaluating this prediction, and (III) a linguistic construction to test the prediction.

The logic of these studies is as follows: if the search process follows a node-by-node algorithm, generating candidates in a structurally-guided fashion, then the time it takes to license a dependency should be an increasing function of the hierarchical distance between the two elements of a dependency. If, on the other hand, the search process involves direct access, generating candidates by a feature match, then the time it takes to begin to license a dependency should be constant

The measure of processing time is crucial. McElree and colleagues argue that simply measuring reaction times in a reading task or grammaticality task is misleading, as participants may make speed-accuracy tradeoffs. For language processing, these tradeoffs can differ not only across sentences, but within sentences. For example, participants may engage in detailed, high accuracy processing to establish the thematic roles of an event expressed in the main clause, but revert to relatively shallow, low-accuracy processing in

a non-restrictive relative clause that contributes little to fixing reference. Similarly depth of processing may vary with clausal embedding or sentence length. McElree advocates the use of a response-signal paradigm, the speed accuracy tradeoff procedure (SAT), which measures task accuracy as a function of processing time (Wickelgren, 1977). SAT provides separate measures of two key properties: the strength of a representation and its accessibility to cognitive processes. Accessibility refers to the speed with which on-going processes can access a representation, i.e. retrieval speed.

During an SAT experiment, participants are trained to discriminate two classes of stimuli; in the case of sentence processing experiments, the discrimination is typically between acceptable versus unacceptable sentences. Participants read sentences word-by-word in RSVP presentation. A response cue follows the sentence, to which participants are trained to respond with yes/no acceptability judgment within 100-300 ms. The onset of the response cue is varied across the experiment, so that data is collected at time points marking predetermined intervals, following the conclusion of the sentence to several seconds thereafter²². For each participant it is possible to construct a response function that shows how sensitivity grows over time. This function yields three measures, illustrated by the schematic SAT functions in Figure 3-2.

²² A minimum of six to eight time samples are required, so experiments are resource-intensive, requiring up to 180-320 trials per condition, and consequently several experimental sessions (e.g. McElree & Griffith, 1995). There is an alternative procedure, the multi-response SAT (MR-SAT) procedure, which provides a more efficient means of deriving SAT functions. During MR-SAT, participants are trained to respond repeatedly within a trial to a series of response cues, dynamically modulating their responses as their judgment of the sentence changes. MR-SAT reduces the number of trials necessary to 30-40 trials/condition (Brian McElree, p.c.)

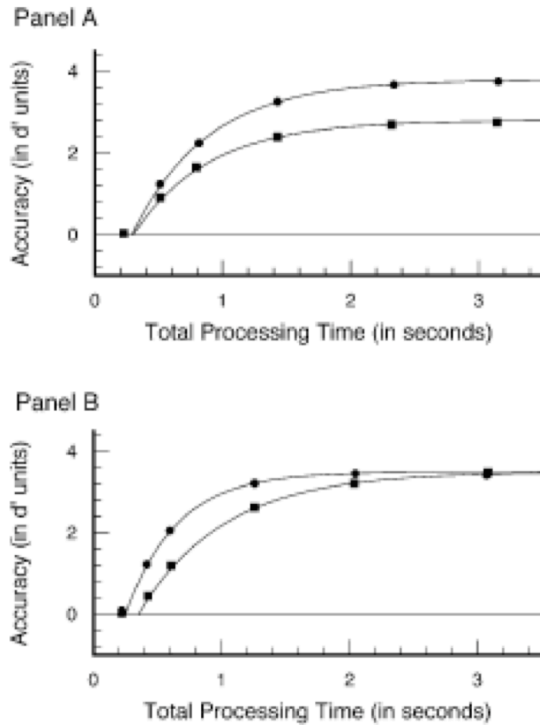


Figure 3-2 Hypothetical SAT functions

The SAT time-course functions show that accuracy initially is at chance, steadily increases for period, and then reaches an asymptote. Time-course functions are usually fit by the exponential approach to a limit, fit with three parameters: the intercept, rate of rise, and asymptote (Wickelgren, 1977 inter alia.; cf. Ratcliff, 1978).

Panel A shows conditions differing by asymptote, while Panel B shows conditions differing by intercept and rate.

Figure provided by Brian McElree.

The first measure is overall accuracy, reflected in the asymptote of the function, and taken as a measure of representational strength; two conditions that differ in asymptotic accuracy are illustrated in Figure 3-2, Panel A. The intercept and rate parameters jointly describe the dynamics of processing. SAT dynamics reflect either the underlying accrual of information if processing is continuous or the underlying distribution of finishing times if processing is discrete or quantal (Doshier, 1976; McElree & Doshier, 1989; Ratcliff, 1988). Panel B illustrates two conditions with disproportional

dynamics, which reach the same proportion of the asymptote at different times. The intercept and rate parameters are the parameters of interest, since they reflect speed of access to a representation, independent of its strength or ultimate availability.

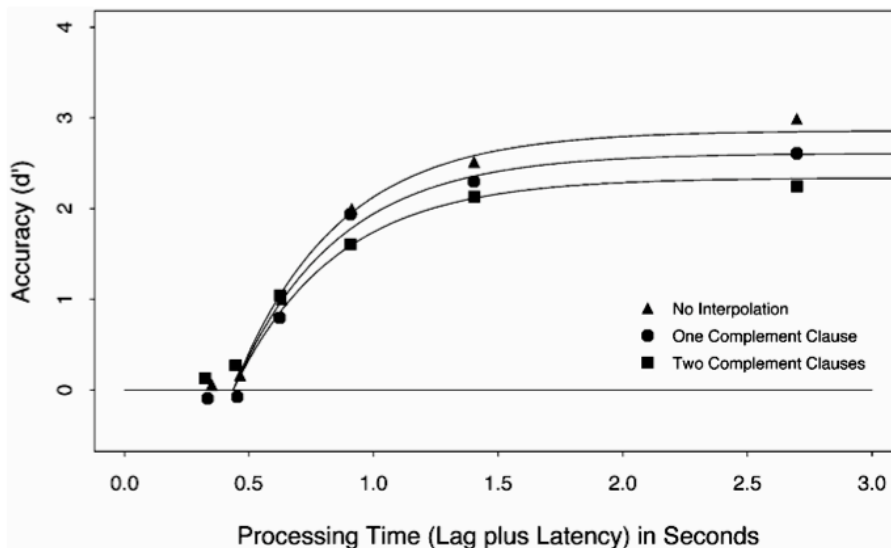
McElree, Foraker & Dyer (2003; henceforth MFD) applied the SAT procedure to sentences containing clefts. Clefting is a species of A' movement, and as such creates an unbounded dependency between a clause-initial phrase and gap site. MFD hypothesized that to comprehend these sentences it would be necessary to retrieve the clefted constituent at the gap site. To test whether increasing distance increased the retrieval time for the clefted constituent, as predicted by a system with ordered search, MFD systematically varied the distance between the clefted phrase and its gap. The gap site was in object position, either in the same immediate clause as the displaced constituent, or was one or two clauses distant. As a signal for whether participants retrieved the constituent and interpreted at the gap site, MFD manipulated the selectional requirements of verb that hosts the gap. In unacceptable sentences, binding the clefted phrase to the gap site would lead to an anomalous interpretation, so participants should reject the sentence. A full example set of materials is given in (38). The clefted constituent is underlined and the gap site marked by an underscore. The acceptable and unacceptable verbs for each sentence prefix are given separated by a slash: in this set, the displaced constituent is “the scandal,” for which the acceptable verb is ‘relish’ and the unacceptable verb is ‘panic.’

- (38) (a) SAME CLAUSE
It was the scandal that the celebrity relished / panicked ____.
- (b) ONE CLAUSE INTERPOLATED
It was the scandal that the model believed the celebrity relished / panicked ____.
- (c) TWO CLAUSES INTERPOLATED
It was the scandal that the model believed that the journalist reported that the celebrity relished / panicked ____.

The critical region is at the end of the sentence. By hypothesis, forming the A' dependency is the last sentence processing event to affect discriminability of acceptable versus unacceptable sentences, and as such it determines the shape of the SAT function. If information about the clefted constituent is gained via a direct access method, then there should be no difference in the intercept/rate parameters of the SAT function. However, if it is obtained by an ordered search, governed by the dominance relations in the superficial structure of the sentence, then the intercept/rate parameters should vary with hierarchical distance.

Figure 3-3 shows the average SAT function for the eight participants in this experiment. SAT data are analyzed for each participant separately, by fitting hierarchically nested models to accuracy as expressed by d' . A null model fits the data with one intercept, one rate, and one asymptote parameter for all conditions; a full model with separate parameters for each condition (thus, nine parameters over all). What turned out to be the best fitting model had three asymptote parameters, one for each condition, but a single intercept and single rate parameter for all conditions. The dynamics of accruing information about the clefted constituents was identical across conditions, and thus did not vary with hierarchical distance. Ultimate accuracy did vary across condition, as a decreasing function of hierarchical distance. There are two possible kinds of explanation for this effect: (1) as sentence length increases, the probability of misanalysis

increases; (2) while the accessibility of the representation in terms of access time may not change, its overall availability, or the availability of the information contained inside it, may. Availability is a joint function of the quality of the encoding and the contents of the retrieval structure, and decrements are to be expected as more constituents compete for cues or offer spurious matches. This observation is consistent with the results from the literature on list memory, in which both recognition and recall decline as more items are added to the list (cf. Dennis & Humphreys, 2001). It is also consistent with other psycholinguistic research, which indicates difficulty in successfully processing A' dependencies as the number of intervening clauses increases (e.g., Phillips, Kazanina, & Abada, 2005²³).



**Figure 3-3 McElree, Foraker, & Dyer (2003), Experiment 2
Average SAT Function**

²³ Phillips, Kazanina, & Abada (2005) found that longer *wh*-dependencies led to later P600s on the verb that subcategorized the *wh*-phrase. The P600 is an evoked response potential that is sensitive to dependency completion (Kaan et al., 2000). Note that such a timing delay is itself neutral about whether the extra length leads to longer search times or simply decreased accuracy.

The best fit model for these data posits one intercept and one rate parameter for all conditions, but three separate asymptote parameters. Figure taken from McElree, Foraker, & Dyer (2003).

The key finding for the present discussion is that access speed for a displaced constituent does not vary with the number of clauses that intervene between the constituent and the site of interpretation. This finding is consistent with a content-addressable memory architecture that affords direct access to representations in memory. What is problematic, however, is that some widely held assumptions about the syntactic representation guarantee the same result. MFD assume that an ordered search must sift through a number of representations that is proportional to the hierarchical distance between where the constituent is pronounced and where it is interpreted. If there were no information about the constituent in the interposed clauses, this assumption would be valid. The bracketing in (39) illustrates a phrase structure consistent with this assumption. Syntactic elements that ‘share’ information about the clefted constituent are in bold font.

(39) [_{DP} **the scandal**_i [_{CP} **Op**_i that ([_{TP} [_{CP} ... [_{TP} ... [_{CP} ...) [_{TP} the celebrity relished **t**_i]]]]]]]

However there is considerable evidence from syntactic research (Chomsky, 1973; Torrego, 1983, 1984; Georgopoulos, 1985; Chung, 1998; McCloskey, 2000, 2001; Bruening, 2004) that dependencies into embedded clauses are not mediated directly between the displaced constituent’s surface position at the highest clause edge and the gap site in the embedded clause. Rather, an intermediate dependency element is created at the edge of each embedded clause. This property of displacement is referred to as successive cyclicity and conforms to the restriction that rule application applies to bounded syntactic domains. A cyclic representation of MFD’s stimuli, given in (40),

which shows that distance between the gap and the first syntactic element that contains information about the displaced constituent is constant across the clausal interpolation manipulation.

(40) [_{DP} **the scandal**_i [_{CP} **Op**_i that ([_{TP} [_{CP} **t**_i [_{TP} ... [_{CP} **t**_i) [_{TP} the celebrity relished **t**_i]]]]]]]

It is an open question whether cyclic representations are constructed in real-time and, if so, whether the information encoded at the intermediate positions is sufficient to judge the selectional fit of the *wh*-phrase with the verb. Therefore, although the clausal interpolation increases the temporal offset from when the displaced constituent was first encoded, it is possible it does not increase the structural distance in the sentence representation.

McElree, Foraker, & Dyer (2003) also considered the dynamics of subject-verb processing. They studied sentences in which they varied the distance between the subject head and the verb by PP and RC modification. There they also found constant dynamics (except when subject and verb were adjacent, and except when two object relative clauses intervened). This manipulation is irrelevant to discriminating between a direct access and an ordered search mechanism, because their manipulations did not increase hierarchical distance between subject projection and the verb. Only serial distance and the complexity of the subject changed across conditions. These results at least do suggest that the order that governs search cannot be linear. It is difficult to construct the right stimuli that truly modify hierarchical distance between dependent elements without the possibility of intermediate representations. Movement dependencies in general present this problem, since the tendency in syntactic explanation has been to suppose that apparently distant

dependencies are really a succession of local ones (cf. Kayne, 1984, Kroch & Joshi, 1985, Pollard & Sag, 1994). We offer some possibilities for future work, however. One is to consider reflexive binding in syntactic alternations:

- (41) (a) Scott gave a book about himself/herself to the library.
 (b) Scott gave the library a book about himself/herself.

The dependency of interest is between the reflexive anaphor and the subject of the sentence. In (41a), this anaphor is contained within the closest argument to the verb, while in (41b) it is in the farthest argument. If we accept the analysis of ditransitive verbs as heading a binary branching projection (e.g., Barss & Lasnik, 1986, Larson, 1988, Pesetsky 1995, Harley, 2002), then the anaphor in (41a) is hierarchically closer to the subject than in (41b)²⁴, as the structures in (42) illustrates (adapted from Harley, 2002):

- (42) (a) [_{VP} Scott [_{v'} give [_{PP} [_{DP} a book about himself] [_P P_{LOC} [_{PP} to the library]]]]]]
 (b) [_{VP} Scott [_{v'} give [_{PP} [_{DP} the library]] [_P P_{HAVE} [_{DP} a book about himself]]]]]

Unfortunately these stimuli are not well-suited to the SAT technique, since the region that putatively triggers retrieval, *himself*, is not at sentence-final positions in both pairs.

Furthermore, it seems unnatural to give the response cue before the sentence is finished, unless participants were trained to judge acceptability-so-far. One possibility would be to shift “a book about himself” in (41a) to the sentence final position, without changing the phrase structure; this may be possible if the DP is sufficiently heavy:

- (43) ? Scott gave to the library a self-aggrandizing book about himself.

Another possibility is to consider *wh*-island violations, as in (44). Such sentences are interesting because the occurrence of an intervening *wh*-phrase prevents the occurrence of an intermediate representation of the target *wh*-phrase (Chomsky, 1977).

²⁴ The distance contrast does not change, regardless of whether there is a transformational relationship or not between the ditransitive’s alternate projections.

The [Spec,CP] position in the embedded clause is filled with another *wh*, (45), potentially forcing truly long-distance movement.

(44) * It was the scandal that the journalist reported how the celebrity relished ____

(45) [DP the scandal_i [CP Op_i that [TP the model wondered [CP how C [the celebrity
relished]]]]]

Such sentences are also ill-formed but not awful. It is therefore not unreasonable to ask participants to discriminate the sentences on the basis of the verb's selection properties, as in MFD:

(46) It was the scandal that the model wondered how the celebrity relished/panicked ____

These stimuli do not lend themselves to more than one clausal interposition. Though one *wh*-island violation can lead to only mild unacceptability, filling multiple [Spec,CP]s seems much worse:

(47) * It was the scandal that the model wondered why the journalist reported how the celebrity relished/panicked.

Most recently, Martin & McElree (2008) have examined VP Ellipsis, as in (48).

(48) The editor admired the author's writing, but the critics did not ____.

In order to interpret this sentence, it is necessary to identify the antecedent VP for the ellipsis site (antecedent underlined, ellipsis site marked with an underscore). Martin & McElree (2008) tested whether or not increasing hierarchical distance between a candidate VP and the ellipsis site led to different access dynamics. In their experiment, antecedent VPs could be either one or two clauses away as in (49a) and (49b), respectively. An acceptability contrast was created by modifying the properties of the subject of the elided VP.

(49) (a) Near antecedent
The editor admired the author's writing, but the critic/*binding did not.

(b) Distant antecedent

The editor admired the author's writing, but everyone at the publishing house was shocked to hear that the critic/*binding did not.

Using the SAT technique Martin & McElree (2008) found that ultimate accuracy was lower for distant antecedent conditions, but that neither rate nor intercept parameters of the SAT function varied. They concluded that accessibility of the antecedent VP representation did not vary as a function of distance. Unlike MFD, this phenomenon is less clearly liable concerns about intermediate representations (but cf. Johnson, 2001), and so constitutes stronger evidence for a direct access mechanisms. However, there are relatively few syntactic constraints that govern where the antecedent VP can be found (Johnson, 2001). Antecedent VPs can be found in non-commanding positions (as in Martin & McElree's stimuli), in c-commanding positions (as in Antecedent Contained Deletion), and extra-sententially. It therefore seems plausible that locating a VP antecedent need not be ordered by dominance pathways in the sentence. On the one hand, a content-addressable mechanism may be well-suited for this kind of search. On the other hand, it does not constitute a strong test of whether any ordered searches are used in language comprehension since VP Ellipsis may not be the right kind of phenomenon to invoke such a search.

In summary, McElree and colleagues have offered the only direct evidence that the search for constituents occurs in parallel, based on the observation of constant response dynamics in discriminating acceptable and unacceptable sentences. The argument in McElree, Foraker, & Dyer (2003) is unfortunately undercut by fairly standard assumptions about the syntactic representation of *wh*-dependencies. It nonetheless deserves to be taken seriously, since it is a unique and theoretically precise argument. The challenge that remains is practical (albeit not trivial), which is to sample a

spectrum of linguistic stimuli that would give the ordered search hypothesis a fair chance of showing its influence in retrieval dynamics. Martin & McElree (2008)'s ellipsis test is one such case. We offered some tentative further suggestions.

3.2.1.3 Bringing structure back

In a content-addressable system there are ways to compensate for the structurally-insensitive search process. Firstly, the matching process could be used merely as a fast and efficient first-pass process, which is then followed by a more controlled, ordered search process. The logic of such an architecture might be as follows: instead of ascending the phrase structure tree node-by-node for a constituent (that may or may not exist), perform a fast parallel search to identify whether or not there is match to begin with. Then, descend along the dominance paths to return to the foot of the tree (on the assumption that constituent encodings point to their mothers and daughters). It should not be overlooked that in many cases of simply structured sentences, the content features are likely sufficient to pinpoint the right candidate, especially if it is unique. For example, consider a subject-seeking head processed inside a simple matrix clause, with no embedded clauses. Probing memory for a constituent bearing nominative case may be more efficient than doing any tree climbing at all.

Secondly, some hierarchical order may be implicitly encoded in analogue properties of the constituent encodings, like in an activation value. This strategy has been pursued in models of simple serial order (e.g. Grossberg, 1978; Page & Norris, 1998). Figure 3-4 illustrates the basic idea. Successive items in a list are encoded with decreasing activation levels. The relative activation levels of any two items retrieved map onto their (relative) order in the list.

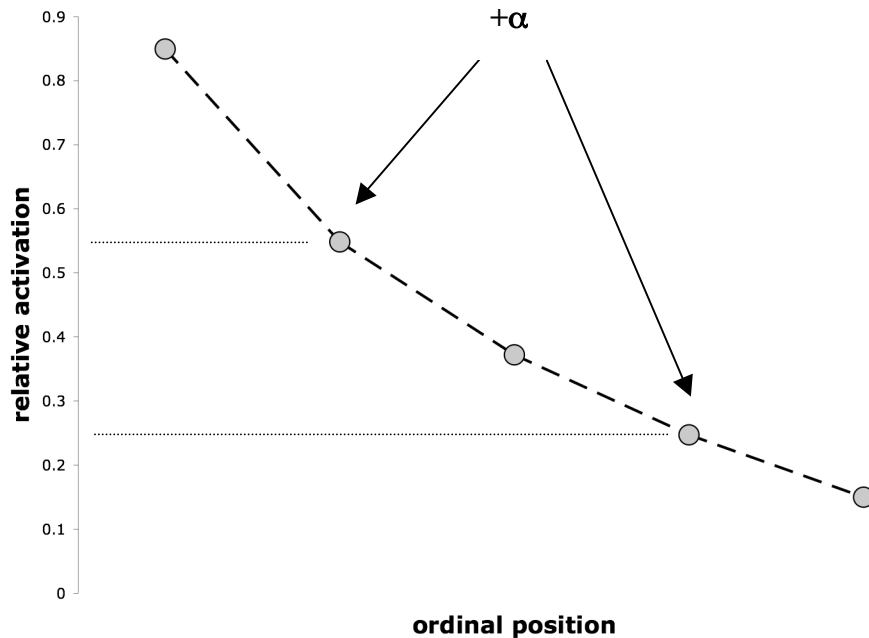


Figure 3-4 Implicit encoding of serial order

The cue $+\alpha$ retrieves two item representations (solid lines with arrowheads). The relative order of the two items can be inferred by their relative activation levels (dotted lines).

Indeed a similar mechanism can be seen at play in the ACT-R model of sentence processing (Lewis & Vasishth, 2005). There each constituent representation has a baseline activation value. This activation value is meant to reflect the likelihood a constituent representation will be used in future processing, consistent with ACT-R's emphasis on being an 'adaptive' architecture (see also Shiffrin & Steyvers, 1997). The more frequently a representation is used by a process, the more easily retrievable it will be. Now consider a complex subject, like "the old man from New England that my father introduced me to." In this phrase, the head noun 'man' is modified by three separate constituents, an AP, a PP, and a CP. Consequently it undergoes relatively more processing and re-encoding than the other heads in the string. ACT-R will therefore

incrementally assign it a higher activation value. Suppose that the ‘man’ and ‘father’ constituents are later identified as candidates in a search (for example, triggered by the reflexive anaphor ‘himself’). If both candidates are equally good matches for the retrieval cues, and both candidates have the same baseline activation (prior to modification), ‘man’ will be more easily retrievable. A heuristic decision metric would choose candidates with higher activation values on the assumption that they are more likely to be hierarchically prominent.

This metric must truly be heuristic though. Any independent modulation of the activation values will disrupt the hierarchical order. It is not difficult to imagine a potential counterexample to the modification example above; for example, think of a complex subject with a modestly modified head, but a heavily modified embedded subject: “the man that my old cousin from New England who knows many famous people introduced me to.” Finally there is a general problem with an analogue encoding of hierarchical order. We can think of the syntactic structure of a sentence as essentially a list of lists. The categories along the main ‘trunk’ of a binary branching phrase structure tree comprise the master list. The relative order of any two categories within the list is sufficient to determine which category dominates the other (or which c-commands which, if it is a list of heads). But each category points to another list, that of the sub-tree it dominates, which may itself contain further lists. The difficulty, therefore, with implicitly encoding relative order in an analogue fashion is that to interpret the outcome, it is necessary to know whether the candidate representations in a comparison set come from the same ‘list’ or not. In section 3.3.4, we argue for a plausible encoding scheme under which all the constituents contained in a given domain are marked as belonging to that

domain, which gives slightly more traction to a more general analogue solution.

However, we have not yet been able to find a satisfactory general scheme that would map an activation-like quantity onto global hierarchical order, but some devices may be available in restricted circumstances.

Finally, a third way to compensate for a structure-insensitive search procedure is to enforce retrieval of the correct constituent by predicting the necessity of retrieving it. For some dependencies, the first encountered member of the dependency is distinctive enough to signal the presence of the dependency. *Wh*-dependencies in English have this character: encountering a *wh*-phrase in the clause periphery signals that a *wh*-dependency exists, and that the phrase must be paired with a gap in the subsequent structure. The parser could preserve the *wh*-phrase by entering it onto a stack, which was one of the earliest suggestions for completing *wh*-dependencies (Wanner & Maratsos, 1978). It may be too much to concede such a mechanism, however. As we discuss in section 3.2.2, the cue-directed retrieval of immediately retained information is closely related to a second architectural constraint, which is an extremely restricted focus of attention (Cowan 1995, 2001; Nairne, 2002; McElree, 2006; Jonides et al., 2008). A stack, conceived as a distinguished memory space for the maintenance of an encoding, is at odds with this viewpoint. If space limitations truly are at issue, then a reasonable compromise would be to maintain a highly stripped-down encoding of the *wh*-phrase, containing hardly any of its content but perhaps some signature property, like a unique I.D., throughout the full-course of processing. When the full information in the phrase needs to be retrieved, as it will be at a gap site, then the “I.D.” of the *wh*-phrase as a retrieval cue will be a highly effective cue for retrieving the structurally-licit constituent because it is highly

distinctive. Suffice it to say, elements that can participate in a dependency do not necessarily signal that they will need to be retrieved in the future. Pronominal anaphora is a good example: the occurrence of a name in a structure does not guarantee that a coreferential pronoun will occur later in the structure²⁵. We expand upon this general idea in much greater detail in Chapter 4, and defend the generalization that constituents that predict their own retrieval are more accurately retrieved.

In summary, a content-addressable memory allows representations to be accessed based on the inherent properties of the representation. On the one hand, this permits the rapid retrieval of potentially relevant representations, without the need to consider wholly irrelevant ones (McElree, 2006). On the other hand, relational properties, like c-command, are difficult to recover in the same fashion. Indeed it is a property of the architecture that relational properties are backgrounded to inherent ones, since access is determined by the match between retrieval cues and encodings. But relational properties like c-command are not inherent features of a node, since it must be determinable whether they hold for any arbitrary pair of constituents in a structure. It is conceivable to imagine an encoding system in which every constituent contains a list of the constituents in c-commands; however, this would require updating a large proportion of the constituents anytime nodes are added to the structure. We sketched out some general alternatives for recovering hierarchical information, and we return to this problem in greater detail below and in Chapter 4. However, we want to consider one further joint architectural claim about syntactic memory, before moving on to the agreement model. The claim is that there is a very restricted amount of information that can be

²⁵ But see Omaki et al. (2007) for arguments that even names have weak predictability for pronominal coreference.

simultaneously maintained in the focus of attention, and, consequently, that most information that is discontinuous, from a time or process perspective, must be retrieved (McElree, 2006). It has been argued that much of syntactic processing is skilled memory retrieval (Lewis & Vasishth, 2005). These claims have a broader corollary in an emerging new perspective in memory theory, that there really is no distinguished working memory for recent events, just long-term episodic memory (Nairne, 2002, Jonides et al., 2008).

3.2.2 The restricted focus of attention

Interpreting an expression depends on coordinating information that enters the processing system widely separated in time. This problem is exemplified not only by the unbounded dependencies found in wh-questions, clefts, comparatives, and relative clauses (e.g., “The tune that John was casually whistling ...”) but also more locally, as in subject-verb agreement (“The songs from the popular movie were playing”) or verb-argument selection (“The orphan inherited a sizable portfolio of securities from the millionaire”). However, it has long been argued that the ability to actively attend to and concurrently process information is very limited (e.g., Broadbent, 1958; Cowan, 1995, 2001, 2005; McElree, 2001, 2006). As a key architectural feature in many contemporary models of memory (Jonides et al., 2008), a focus of attention instantiates this limitation by partitioning representational space into a sharply bounded nucleus of information immediately accessible to cognitive processes and the representations which must be accessed via a retrieval process. For language processing it is relevant to know how much linguistic information can be processed concurrently, since this determines what kind of information, and how often, will have to be retrieved.

Several independent lines of evidence from a variety of cognitive and perceptual tasks support sharp capacity limitations on the focal state (Cowan, 2000, 2005). McElree (1998, 2001, 2006) has argued that measures of the speed of accessing information provide the most direct and unequivocal evidence for a unique representational state associated with focal attention. These measures show a sharply dichotomous pattern: processing dynamics are exceptionally fast for responses based on information actively maintained in awareness, approximately 30-50% faster than responses based on information displaced from focal attention (Doshier, 1981; McElree, 1996, 1998, 2001, 2006; McElree & Doshier, 1989, 1993; McElree et al, 2003; Oberauer, 2002, 2006; Öztekin & McElree, 2007; Verhaeghen et al., 2004; Wickelgren et al., 1980). Responses are argued to be fast because no retrieval operation is needed to access the contents of these representations; hence, that information is immediately available for ongoing operations (McElree, 2006).

The observed discontinuity in processing speed provides a way of empirically measuring the span of focal attention. The available evidence on the processing of sequentially presented information using this estimate suggests a very limited span: in most circumstances, only the representation associated with the last event remains in focal attention. However, there are two crucial qualifications: more than one nominal item can be in focus if the task encourages the encoding of multiple items into a chunk (McElree, 1998) and less recent items may be present if subjects are induced to actively (re-)process those events (McElree, 2001, 2006). These findings set the stage for meaningfully relating focal attention and language comprehension. An adequate system for linguistic understanding would seem to require the ability to entertain at least two

place relations, and thus for focal attention to host at least that many ‘items’ bound as a single representation – what Jonides et al. (2008) refer to as a ‘functional complex’.

We have no current estimate of what counts as a chunk for focal attention, and particularly what counts as a linguistic chunk. McElree, Foraker & Dyer (2003) provided some evidence that a complex subject can quickly occupy all of focal attention. In a SAT experiment identical to the cleft-processing experiment discussed in section 3.2.1.2, MFD successively interposed more material between subject head and verb:

- (50) (a) The book ripped/laughed
(b) The book that the editor admired ripped/laughed
(c) The book from the prestigious press that the editor admired ripped/laughed
(d) The book that the editor who quit the journal admired ripped/laughed

SAT dynamics, measured by the intercept and rate parameters of the accuracy function, showed a discontinuous split, between very fast dynamics in condition (a), where subject and verb heads are adjacent, and conditions (b)-(d), where the subject is complex. These results, they argued, implicate a retrieval operation in conditions (b)-(d) that is not present in condition (a) (or is required less often). It is inferred that in conditions (b)-(d), it is no longer possible to maintain the encoding of the subject projection, or of the subject head itself, concurrently with the incoming information.

These results do not place an especially strong bound on what counts as a chunk in focal attention, since the slow conditions all included an extra clause. There are already many reasons to suspect that clause-boundedness plays a strong role in segmenting linguistic encodings (see section 3.3.4 below for further discussion). It would be useful to know whether task dynamics are fast or slow in simpler cases, like when the subject is PP-modified, or there is an adverb:

- (51) (a) The book from Susan ripped/laughed
(b) The book easily ripped/laughed

In the case of PP modification above, we introduced one more closed-class head and one more lexical head, which bears a close (restrictive) relation to the subject head. Two lexical heads related by a functional projection seems like a reasonable candidate for a minimal ‘functional complex,’ in Jonides et al.’s terms. Despite the present results being only suggestive, it is once again useful to know that effects from the memory literature, with its focus on a laboratory task, such as verbal list learning, show up in more natural, linguistic context. And it gives us confidence that drawing architectural parallels is justified.

There is good evidence to think that there is a small focal state and that representations in this state lead to the fastest processing dynamics. It is actually hard to avoid this conclusion, if the focal state is thought of as the information that is undergoing immediate processing. However, what lies outside that state? A traditional division of the memory space, outside of immediate processing, is into long-term memory and working memory. Working memory is conceptualized as consisting of those memory traces which occupy a distinguished memory state – a ‘workspace’ – either because they are in a special store (e.g., Baddeley, 1986; Shallice & Vallar, 1990, *inter alia*) or because they have intrinsic, persistent activation (Anderson, 1983; Cowan 1995, 2001, *i.a.*). Working memory has a certain capacity, depending on the size of the buffer or how much activation can be shared (e.g., Usher & Cohen, 1999). Representations in working memory are thought to be more accessible than memory traces stored in the long-term store, though accessibility can fluctuate. Accessibility in the short-term depends on either on the trace’s location in the buffer and a process’s ability to cycle through different

locations; or on the inherent activation of the memory trace, which fluctuates over time, generally decays, but can be refreshed through rehearsal.

There is an emerging perspective, however, that eschews the distinction between working memory and long-term memory, in favor of unifying the memory architecture (see Jonides et al., 2008, for a review). It has long been known that at least long-term forgetting cannot be attributed to a decay-like process (McGeoch, 1932; see Anderson & Neely, 1996) and that remembering depends both on the properties of the stored representation and the present information used to recall it (see, e.g., Tulving, 1983). As Nairne (2002) puts it, representations do not have, “‘strength’, or special mnemonic properties, outside of particular retrieval environments.” The characteristics of working and long-term memory would seem to be theoretically very distinct, therefore. However, mounting evidence indicates that the recognition and recollection of immediately retained information seems to depend on the retrieval environment, just as much as long-term information does. Performance depends on the distinctiveness of the information with respect to the retrieval cues (see Nairne, 2002, for review). Moreover, speed of access seems to be constant across studied items, even if they exceed working memory capacity, except for those in immediate awareness (see McElree, 2006, for review).

3.2.3 Implications

There is not a necessary connection between a limited focus of attention and a content-addressable information retrieval mechanism. However, if the capacity of focal attention for linguistic material is very small and representations outside of the focus of attention are contacted by means which are in principle hierarchically insensitive, then we are forced to rethink many issues in the real-time comprehension of language. There

is one way of thinking of syntactic processing as occurring in a ‘workspace’ that makes available a significant amount of structured information over which globally-sensitive parsing decisions can be made. However if all that remains is the focal/non-focal distinction, then the amount of structural information immediately available on which to base parsing decisions is considerably restricted. It seems, furthermore, that recovering syntactic context occurs in an architecture that is inherently not well-suited to structure-sensitive processing. The content-addressable nature of retrieval means that shunting information into the focal state is potentially only weakly constrained by hierarchical order. In a real sense, the recent interest in retrieval-based, interference-prone, content-addressable memory requires a radical re-evaluation of the interface between syntax and memory.

There is an extreme view that all specialized structures for language processing, such as the use of stacks or queues, should be eliminated if possible (e.g., McElree, Foraker, & Dyer, 2003). This view seems premature to us and we will defend the use of some specialized information maintenance as necessary in Chapter 4, which we believe adapts language processing to the memory architecture. It is important to remember that language comprehension is by most measures effortless and accurate. Nonetheless we accept that the information immediately available to the parser’s decision processes is restricted. This is not necessarily a bad thing: making decisions on the basis of a limited amount of information, whose format is known in advance, can make parsing more efficient. Berwick & Weinberg’s account of subjacency (1984) made exactly this point: movement is subjacent because it allows the parser to decide whether or not to posit traces by looking at fixed-size context representation. The alternative is to conduct an

unbounded search over the tree, which is an operation of increased computational complexity.

In the following section, we work out how agreement processing could occur in a content-addressable architecture in which even nearby information has to be retrieved. We discover that a linguistically well-motivated retrieval structure provides considerable flexibility in accomplishing structure-sensitive processing. Nonetheless, agreement attraction, as an error, is a natural consequence of the architecture. More generally, the architecture seems to pose a challenge for processing complex subjects in a structure-sensitive fashion, as we'll see in section 3.4²⁶.

3.3 Agreement attraction in comprehension

3.3.1 Intuition

First we will spell out the intuition behind the retrieval interference model of agreement attraction before formalizing it. The key pattern to capture is the asymmetry between grammatical and ungrammatical sentences. An agreement attractor leads to an illusion of grammaticality when the sentence is ungrammatical. It does not lead to an illusion of ungrammaticality when the sentence is grammatical. Data from the Experiment 4 speeded grammaticality task are repeated in Table 3-1 to emphasize this point. It should be kept in mind that what distinguishes the grammatical and ungrammatical sentences is whether the number on the verb matches the number on the subject head noun.

²⁶ This fact, or at least its analogue in derivational syntax, has arguably already been appreciated in the syntactic literature (Uriagereka, 1998).

		<i>Response</i> (% 'Yes': acceptable)	
		'cabinet'	'cabinets'
<i>Sentence Type</i>			
Grammatical	“The key to the ___ was”	93%	91%
Ungrammatical	“The key to the ___ were”	25%	55%

Table 3-1 Speeded grammaticality judgments of complex subject attraction From Experiment 4

Let us suppose that an (automatic) event in processing the verb is checking for agreement with the subject. This process can be conceptualized as a search of the preceding syntactic context to verify that the right kind of agreement controller exists, in the right configuration, to license the verb’s morphology. Therefore the search process will be guided by two kinds of information: the properties of the agreement controller that can license the agreement relationship as well as the compatible feature value for the number on the verb. The former, which I shall refer to as ‘licensing features’, could include properties like grammatical function (i.e., SUBJ), structural position (i.e., [Spec,TP]²⁷), or Case (i.e., NOM). If the search succeeds in identifying a constituent that bears the licensing features and a compatible number feature, then the agreement is considered ‘checked’ and the process succeeds²⁸.

Here is the key intuition about agreement attraction in comprehension: it is the outcome of a checking process in which no constituent matches both features, but a constituent that matches the agreement feature is identified by the search process. There

²⁷ Encoding the structural ‘coordinates’ of an item would be somewhat more non-standard than grammatical role or case. Though it is not unimaginable on a local level to mark which categories are the head, which the complements, and which the specifier (as HPSG does using attribute-value matrices; cf. Pollard & Sag, 1994).

²⁸ We are, just for the moment, intentionally vague about what counts as ‘identifying’ a constituent: it could be the equivalent of returning a recognition signal, i.e. “I remember encountering this constituent”, or it could be the actual recall of the constituent, bringing it back into the active processing workspace. Clearly for dependencies that require interpretation, like *wh*-dependencies, it must be recall. For more formal dependencies, like agreement, it is less obvious whether recall is necessary.

is a similar account of negative polarity item licensing that works on this partial match principle as well (Vasishth, Drenhaus, Saddy & Lewis, 2005), which we will discuss in section 3.4.4. As well, a concurrently developed model of agreement attraction production relies on the same principle (Badecker & Lewis, 2007), which we discuss in section 3.3.6. This simple formulation goes far in capturing the grammatical/ungrammatical asymmetry. For ungrammatical sentences with a plural attractor, the subject phrase bears the licensing feature, but not the agreement feature; whereas the attractor at least bears the matching agreement feature. There is a partial match with both the grammatical controller and the attractor. For grammatical sentences with a plural attractor, the subject phrase itself bears both the matching agreement feature and the licensing feature; and the attractor bears neither agreement feature or licensing feature. There is a full match with the grammatical controller, and none with the attractor.

The challenge is to come up with a search process that has these properties, which can be consistently specified across grammatical and ungrammatical sentences. We propose a model implemented in a content-addressable memory that seems to deliver the right outcome. Recapitulating section 3.2: a content-addressable memory is one in which information is retrieved by means of the content of the encoding. The key property of content-addressable memory is that it allows direct access to representations in memory. There is no need to search through irrelevant representations, only those that match, in some respect, the retrieval probe that initiates the search process (Clark & Gronlund, 1996; McElree, 2006). Because the search is driven by the similarity between the retrieval probe and what has previously been encoded in the memory, it allows partially similar representations to impact the process. This property of the search, called

similarity-based interference (Anderson & Neely, 1996), is what permits the attractor to intrude in the agreement-checking process in ungrammatical sentences.

To formalize our intuitions, we will first work out our assumptions in the framework of Shiffrin's Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981, Gillund & Shiffrin 1984). Nothing in particular depends on this choice, except that it allows straightforward modeling²⁹. It is only one of several frameworks that captures the major empirical generalization of similarity-based interference: memory retrieval depends on the match between the retrieval probe and representations in memory. For other models that formalize this notion, see, among others, Eich (1982), Hintzman (1984, 1988), Murdock (1982, 1993), Nairne (1990), or Shiffrin & Steyvers (1997).

In the rest of section 3.3 we introduce the retrieval model of agreement attraction, and work through some of its consequence in a speeded-grammaticality experiment. Our goal is to better understand the concept of content-addressable memory in language processing, and some of the issues that it raises, by working through a specific example. However, in section 3.4 we turn to a phenomenon that is closely related to our agreement attraction model, which is the attachment of complex subjects. Recent work by Julie Van Dyke (Van Dyke & Lewis, 2003; Van Dyke, 2007) has argued that complex subject attachment is difficult because it is prone to similarity-based interference. We examine her arguments and try to refine them in a self-paced reading experiment (Experiment 6) in section 3.4.3.

²⁹ It is important to note that we are not engaging in simulation modeling, in the sense of generating a distribution of outcomes that indicate something about the robustness or dynamics of the retrieval process. Rather we are interested in how the formal properties of a content-addressable memory interact with the formal properties of the syntax at a more abstract level.

3.3.2 *Formalization*

The memory model is specified as consisting of three components: (1) the content of stored representations, (2) the retrieval structure, and (3) the task or memory goal. We give first the general outline of these components, and then lay out the assumptions specific to syntactic structure.

3.3.2.1 General Properties

In SAM parlance, the encodings in memory are referred to as *images*. Images consist in feature sets packaged as a single unit. Images can include information about the item representation itself, the context in which it was encountered, and its associations to other images.

The retrieval structure consists of a set of cues, corresponding to item, context, and category properties. Each cue has a particular strength of association to the images in memory. The set of cues differentially activates the images in memory, according to the weighted product of the cue strengths to each image. Let us assume that the memory consists of a set of n images (I_1, \dots, I_n), the cue set's m cues (Q_1, \dots, Q_m), and the strength from the j th cue to the i th image given by $S(Q_j, I_i)$. The equation in (52) specifies that the activation of image i , A_i , is the product of the strengths from each cue in the cue set to that image. The strengths inside the product are raised to the weighting value associated with each cue, w_j . The weight allows the model to assign different saliencies to the cues in the retrieval structure. That is, some cues can count more than others. If the cues are constrained to sum to one, then retrieval effectively becomes a limited-capacity process: adding more cues lowers the expected activation of any given image based on a

single matching cue³⁰.

(52) **Activation equation**

$$A_i = \prod_{j=1}^m S(Q_j, I_i)^{w_j}$$

The non-linear combination of retrieval strengths endows the model with an important property: sensitivity to conjunctions. Figure 3-5 illustrates this property for a hypothetical retrieval scenario with five cues, in which the degree of match, normalized to unity, is shown as a function of the number of convergent cues. Three combination rules are shown: a sum of cue strengths rule, a cube-of-sums rule (cf. Hintzman, 1988) and a product of cue strengths rule (as in SAM). For the non-linear rules, representations that match all the cues are much more highly favored than partial matches.

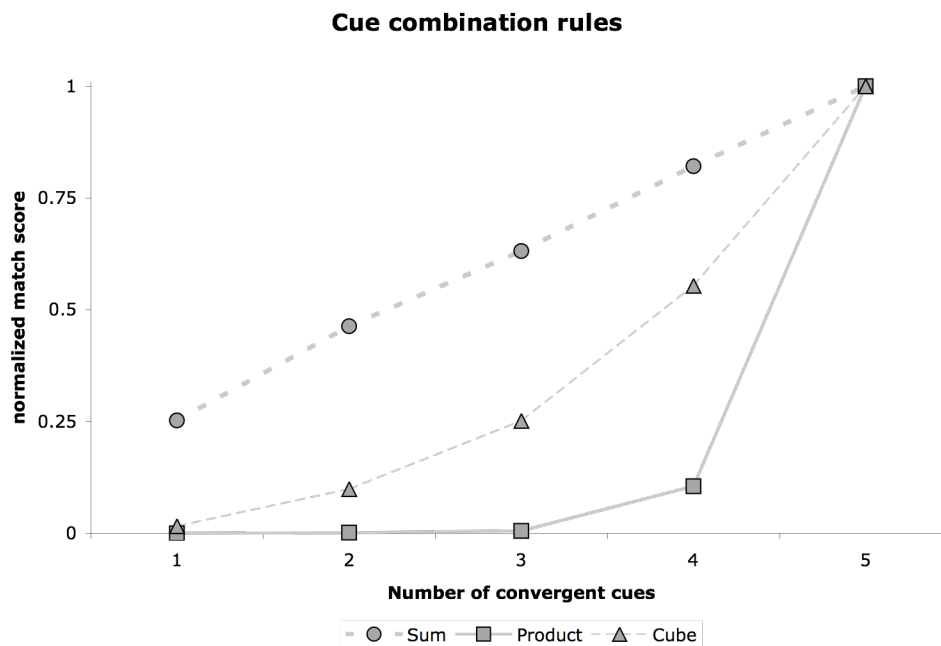


Figure 3-5 Comparison of cue convergence rules

³⁰A useful intuition about the capacity limitation is that, in general, a few highly distinctive cues is better than many mediocre ones at retrieving a target encoding.

Normalized match score is shown as a function of how many of five cues converge on a given representation. Each matching cue has high cue-to-image strength (0.95) while non-matching cues have low strength (0.1). Match scores are normalized to a full match, in which five cues converge.

How the activations map onto the retrieval depend on the goal of probing the memory. SAM countenances two kinds of tasks: recognition and recall. In recognition, the goal of probing the memory is to generate a signal indicating whether or not a match exists. In recall, the goal is to bring the image itself back into the active processing workspace. For recognition, the activations are simply summed to generate a familiarity score, given by the equation in (53):

(53) **Recognition familiarity score**

$$F(Q_1, \dots, Q_m) = \sum_{i=1}^N A_i$$

This familiarity score then feeds a decision process. One simple decision process is to set a threshold. For example, if $F(\bullet)$ exceeds this threshold, then the system deems the cues to correspond to an existing image. For recall, the activations correspond to the probability with which the image will be sampled and recovered. The probability of sampling image i , out of the N images in memory, is given by the equation in (54). This equation says that this probability is determined by normalizing the activation of a given image, by the sum total of activations.

(54) **Recall sampling probability**

$$P(I_i | Q_1, \dots, Q_m) = \frac{A_i}{\sum_{k=1}^N A_k}$$

If an image is sampled, the probability of recovering the information contained therein is given by the equation in (55):

(55) **Recall recovery probability**

$$P_{RECOVER}(I_i | Q_1, \dots, Q_m) = 1 - \exp\left(-\sum_{j=1}^M w_j S(Q_j, I_i)\right)$$

3.3.2.2 Language-specific assumptions

We posit that *images* correspond to syntactic constituents, and particularly maximal projections. This is an assumption that we share with Lewis & Vasishth (2005). It would not be impossible to package an arbitrary extent of a structure as a unit. But there are good reasons for assuming a unit of storage that is something like an XP. The most obvious is that we are considering within-sentence processing, so there needs to be a way of addressing smaller-than-sentence portions of the structure. With respect to agreement attraction, it needs to be the case that the number features of different DPs can be independently accessed; packaging them in separate images seems a natural way to do this. This assumption is not at odds with how a tree representation would be encoded in a standard random-access memory. In such a memory, structured representations like trees are encoded by linking a set of discrete memory locations with pointers. For trees, each node in the tree corresponds to a discrete memory location with a certain number of fields. What makes it a tree representation is that certain distinguished fields point to the next node down (or the next node up) (Knuth, 1965/1997).

How a syntactic representation maps onto individual encodings is one (modest) way in which a candidate processing architecture could guarantee structure sensitivity. By forcing the memory architecture to package its encodings in linguistically-relevant pieces, like maximal projections, retrieval-mediated reference to syntactic encodings is constrained to only return linguistically-relevant pieces. This conclusion is familiar: the

earliest work in psycholinguistics quickly came to the conclusion that the clauses were a salient perceptual and mnemonic unit (see Fodor, Bever, & Garrett, 1974; cf. Shiffrin, Murnane, Gronlund, & Roth, 1989, for more recent evidence from memory paradigms).

In order to encode an arbitrarily complex syntactic representation as a set of features in a unitary encoding, we need a way of specifying recursive feature values. A complex subject is an excellent example to show that point: it is a DP that contains a DP (Figure 3-6). For sake of illustration, suppose that an image is a list of attribute-value pairs, where the attributes are adapted from some standard relations in phrase structure. We can imagine two ways of encoding a complex subject: a ‘single image’ encoding, (56), in which feature values can be recursive feature structures; and a ‘multiple image’ encoding (57), which contains only non-recursive feature values. Instead of recursive feature structures, these images point to other images in the memory (indicated by the → symbol).

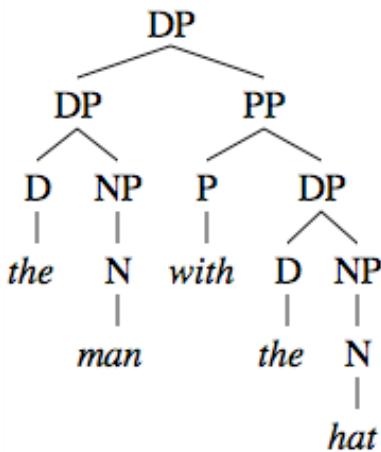


Figure 3-6 Phrase structure tree for ‘the man with the hat’

- (56) **SINGLE IMAGE ENCODING**
 HEAD: <the>
 COMP: <HEAD: man>
DP.1 ADJUNCT: <HEAD:<with>, COMP: <HEAD: <the>, COMP: <HEAD: <hat>>>>

(57) **MULTIPLE IMAGE ENCODING**

		HEAD: <the>
		COMP: → NP.1
	DP.1	ADJUNCT: → PP.1
	NP.2	HEAD: <hat>
		HEAD: <with>
	PP.1	COMP: → DP.2
		HEAD: <the>
	DP.2	COMP: → NP.2
	NP.1	HEAD: <man>

Linguistic theory does not arbitrate a decision between a single or multiple image encoding. We can certainly specify a feature language that is recursive (consider HPSG, for example; Pollard & Sag, 1994). Recursivity is a fact about language structure, and that is not at issue. The trouble with recursively specified features in a memory that is content addressable is that it seems to render opaque the content contained within deeper embeddings. For example, in (57), it is apparent that the head ‘hat’ exists in the structure by inspecting just the HEAD values of the images. The cue { HEAD: <hat> } would activate NP.2 from the image set in (57) by means of its feature composition. In (56), however, ‘hat’ is not visible as a head until the value of the ADJUNCT feature is unpacked. The cue { HEAD: <hat> } would not be an efficacious cue strictly by means of (56)’s feature structure; the association between { HEAD: <hat> } and DP.1 would have to be additionally encoded. If we want to guarantee direct, content-addressable access to the information contained in a structure, therefore, then the multiple encodings seem preferable.

There are clear tradeoffs. Breaking up the encoding into many images renders the pieces more visible to the search process, but it makes recovering relationships more difficult. For example, it raises the expected number of retrieval operations that are

necessary to assemble information about a relationship. If we consider the encoding in (56), knowing that ‘hat’ is contained within that encoding translates into knowing that ‘hat’ is dominated by DP.1. In the encodings in (57), this fact must be deduced by sequentially retrieving the images (cf. McElree, Foraker, & Dyer, 2003; and McElree, 2006; which make a similar point regarding the retrieval of order information). It is an empirical question exactly how much information a given image contains. For example, we might re-encode (57) in a way that does not split DPs and NPs, such that each image encodes something like an extended projection (Grimshaw, 1991). The crucial point is that we assume that multiple images exist in memory for a complex subject. As a consequence, for a complex subject there are two DPs that can be contacted by a retrieval operation.

The retrieval structure is assumed to consist of attribute-value pair cues. If an image contains an attribute-value pair, then that pair will be an effective cue. Although it is conceivable to vary the strength of the cue, based on how confidently that feature has been encoded, we will first assume that cue strength is essentially all-or-none. For the memory goal we will consider two alternative conceptions of agreement checking: agreement checking as recognition and agreement checking as recall. In the case of checking-as-recognition, whether or not agreement is licensed is determined based on a familiarity score. In the case of checking-as-recall, a candidate match must be recovered into the processing workspace, and the feature values inspected to see whether they match. For simplicity, we assume that probability of recovering the information in a sampled image is uniformly high; this does not impact the pattern of results.

3.3.2.3 Complex subjects

First we consider complex subject agreement. We assume that the fragment “the N to the N” corresponds to a memory with three images, the two DPs and the PP. DP images contain information about a single functional head and a single lexical head (F-HEAD and L-HEAD). The contents of the memory are given for both Sg [Sg] and Sg [Pl] complex subjects. We adhere to a privative feature system, in which there simply is no NUM feature for singular nouns. DP.1 and DP.2 are distinguished on the basis of case values: Nominative for DP.1 and Oblique for DP.2.

(58) Sg [Sg]: Images for “the path to the monument”

F-Head: <the> L-Head: <path> L-Comp: → PP.1 Case: Nom DP.1	L-Head: <to> L-Comp: → DP.2 PP.1	F-Head: <the> L-Head: <monument> Case: Obl DP.2
--	--	---

(59) Sg [Pl]: Images for “the path to the monuments”

F-Head: <the> L-Head: <path> L-Comp: → PP.1 Case: Nom DP.1	L-Head: <to> L-Comp: → DP.2 PP.1	F-Head: <the> L-Head: <monuments> Num: Pl Case: Obl DP.2
--	--	---

We evaluate four scenarios, corresponding to an experiment that crosses attractor number with grammaticality. The grammatical continuation to the fragment “the path to the monument(s)” is a singular verb. We assume that a singular verb prompts retrieval with the following cue set: { CASE: Nom }; that is, it prompts with only a licensing feature. The ungrammatical continuation to the fragment is a plural verb. A plural verb prompts retrieval with the cue set: { CASE: Nom, NUM: Pl }; that is, both a licensing

feature and the agreement feature. Consistent with our assumptions, the strength of an attribute-value pair cue to images containing that attribute-value pair is set to near 1: 0.99, and near 0 otherwise: 0.01. We set these strengths to just near the extrema as a means of incorporating noise into the system and to avoid perfect performance. The weights, w , assigned to each cue are assigned to sum to 1, and are split uniformly among cues in the cue set.

The following table, Table 3-2, demonstrates how these cue sets map onto activations for DP.1 and DP.2, when the head of the complex subject is singular, and the sentence is grammatical. The only cue in the cue set is CASE:Nom, which points unambiguously to the subject projection, DP.1. Consequently, in both singular and plural attractor conditions, the activation of this image always dwarfs that of DP.2, which only receives noise activation ($A(\text{DP.1}) = 0.99 \gg A(\text{DP.2}) = 0.01$). Familiarity scores are identical across conditions, so both conditions should behave identically in the decision process linked to a recognition task. The probability of sampling these images in a recall process mirrors the activation values. Consequently, the correct projection would always be recalled.

[DP.1 The path to ...]		[DP.2 the monument]		[DP.2 the monuments]		
		is		is		
		DP.1	DP.2			
<i>Q (Cue)</i>	<i>W(eight)</i>	<i>S(trength)</i>		<i>S(trength)</i>		
CASE:Nom	1	0.99	0.01	0.99	0.01	

				<i>F(amiliarity)</i>			<i>F</i>
<i>A(ctivation)</i>		0.99	0.01	1	0.99	0.01	1
<i>P_{SAMPLE}</i>		0.99	0.01		0.99	0.01	

**Table 3-2 Retrieval structure and outcomes
Singular-headed subjects, grammatical continuations**

The upper cells specify the strength of association between the cues in the cue-set and the stored images. The lower cells show the outcome of probing the memory with that cue set: in terms of the activations of the images, their sum (the familiarity score in a recognition task), and the normalized activation (the sampling probability in a recall task).

The probability of sampling the correct projection, i.e. the subject DP.1, is highlighted in the double-bordered cell.

To see how the model performs for ungrammatical conditions, when the head is singular, we turn to Table 3-3. Here the verb supplies two cues: CASE:Nom and Num:PL. When there is a singular attractor, only CASE:Nom is an effective cue; and consequently the subject projections receives all of the activation. As before, the probability of sampling the correct image considerably exceeds the probability of the attractor image (91% > 9%). When there is a plural attractor, one cue in the cue set points to the subject image, DP.1, and one points to the attractor image, DP.2, but neither points to both. Consequently, each image receives equal activation, and each has an equal likelihood of being sampled. Notice that in both cases, overall activation is lowered (with respect to the previous scenario).

[DP.1 The path to ...]		[DP.2 the monument] *are		[DP.2 the monuments] *are		
		DP.1	DP.2	DP.1	DP.2	
Q (Cue)	W(eight)	S(trength)		S(trength)		
CASE:Nom	0.5	0.99	0.01	0.99	0.01	
NUM:Pl	0.5	0.01	0.01	0.01	0.99	

				<i>F</i>			<i>F</i>
<i>Activation</i>		0.10	0.01	0.11	0.10	0.10	0.20
<i>P_{SAMPLE}</i>		0.91	0.09		0.50	0.50	

**Table 3-3 Retrieval structure and outcomes
Singular-headed subjects, ungrammatical continuations**

The upper cells specify the strength of association between the cues in the cue-set and the stored images. The lower cells show the outcome of probing the memory with that cue set: in terms of the activations of the images, their sum (the familiarity score in a recognition task), and the normalized activation (the sampling probability in a recall task).

The probability of sampling the correct projection, i.e. the subject DP.1, is highlighted in the double-bordered cell.

The limited capacity property of retrieval comes into play here: the presence of an additional cue, even if it is ineffective, lowers the weight of other cues, even if they are effective. As a consequence, for the singular attractor condition, despite the fact that the only effective cue points unambiguously to the correct image, the sampling probability of the correct projection is only 91% (compared to 99% in the previous scenario). However, since neither projection matches the verb in number, this shift in probabilities does not imply a shift in judgment behavior. Finally, the familiarity scores differentiate the conditions in this scenario as well, reflecting the fact that one cue is efficacious in the singular attractor condition, but two are efficacious in the plural attractor condition.

In Table 3-4 and Table 3-5, the model outcomes are specified for grammatical and ungrammatical sentences, when the head of the subject is plural. In the grammatical condition (Table 3-4), the verb is plural, and the cues are assumed to be { CASE:Nom and

NUM:PI }. Regardless of the number on the attractor, both cues will activate the subject. Because the attractor never satisfies both cues, the correct projection receives most of the activation, and has the highest sampling probability. When the attractor does satisfy the plural cue, the balance of sampling between DP.1 and DP.2 shifts. It is unclear whether this would impact judgment behavior, because the attractor does match the verb in number. In the cases where the attractor is sampled (10% of the time), its number feature would still agree. If only that fact mattered, then both conditions would lead to a successful outcome in agreement checking in virtually all cases. The familiarity score is high in both cases, because both scenarios involve a cue set that is maximally effective for one image.

[DP.1 The paths to ...]		[DP.2 the monument]		[DP.2 the monuments]		
		are		are		
		DP.1	DP.2	DP.1	DP.2	
<i>Q (Cue)</i>	<i>W(eight)</i>	<i>S(trength)</i>		<i>S(trength)</i>		
CASE:Nom	0.5	0.99	0.01	0.99	0.01	
NUM:PI	0.5	0.99	0.01	0.99	0.99	

				<i>F</i>			<i>F</i>
<i>Activation</i>		0.99	0.01	1.0	0.99	0.10	1.1
<i>P_{SAMPLE}</i>		0.99	0.01		0.91	0.10	

**Table 3-4 Retrieval structure and outcomes
Plural-headed subjects, grammatical continuations**

The upper cells specify the strength of association between the cues in the cue-set and the stored images. The lower cells show the outcome of probing the memory with that cue set: in terms of the activations of the images, their sum (the familiarity score in a recognition task), and the normalized activation (the sampling probability in a recall task).

The probability of sampling the correct projection, i.e. the subject DP.1, is highlighted in the double-bordered cell.

In the ungrammatical condition (Table 3-5), there is no NUM:PI cue. The only effective cue is CASE:Nom, and it correctly points to the subject projection in both cases.

Notice that the familiarity scores are high and equal in both cases: on the one hand, this indexes the efficacy of the single cue in the cue set for identifying a constituent. On the other hand, it illustrates the point that raw familiarity alone is likely not sufficient to determine a grammaticality response. We raised the question earlier of whether familiarity alone might be sufficient to account for agreement attraction. The intuition was that agreement attraction represents a situation in which the comprehender recognizes the existence of the number features in the syntactic context, but does not pause to examine whether or not they were present in the image corresponding to the structurally-licensed constituent. However, that strategy only seems to gain traction when there are two features at play. In the case of a plurally-headed subject, and an ungrammatical sentence, the only cue is for nominative case; it alone drives the distribution of activation among images. Because there is no singular cue, there is therefore no way in which familiarity is informative about agreement checking. It must be the case that information contained in the returned image is examined, in order to determine that these sentences are ungrammatical.

[DP.1 The paths to ...		[DP.2 the monument] *is			[DP.2 the monuments] *is		
		DP.1	DP.2		DP.1	DP.2	
Q (Cue)	W(eight)	S(trength)			S(trength)		
CASE:Nom	1	0.99	0.01		0.99	0.01	
				<i>F</i>			
<i>Activation</i>		0.99	0.01	1.0	0.99	0.01	1.0
P _{SAMPLE}		0.99	0.01		0.99	0.01	

**Table 3-5 Retrieval structure and outcomes
Plural-headed subjects, ungrammatical continuations**

The upper cells specify the strength of association between the cues in the cue-set and the stored images. The lower cells show the outcome of probing the memory with that cue set: in terms of the activations of the images, their sum (the familiarity score in a recognition task), and the normalized activation (the sampling probability in a recall task).

The probability of sampling the correct projection, i.e. the subject DP.1, is highlighted in the double-bordered cell.

In summary, a very simple cue-driven model seems to re-capitulate the basic patterns observed in agreement attraction comprehension. Crucially we were able to capture the asymmetry between grammatical and ungrammatical conditions, observed for plural attractors. The plural attractor only intruded in the ungrammatical condition, because the efficacy of the cues was divided between the subject projection and the attractor projection. Using the SAM model allowed us to make our initial intuitions precise.

Before moving on to relative clause attraction, we would like to consider one further issue. Suppose that our choice of cues in the scenarios above was misleadingly judicious. Suppose that it was the case that, even in Sg [Pl] grammatical sentences, the verb supplies a cue that contacts both subject and attractor projections, that is, both images DP.1 and DP.2. It is not implausible to imagine such a case: for example, a syntactic category licensing feature, like CAT:DP. This scenario is similar to how the

model treats sentences like “The paths to the monuments are,” where two cues point to the subject (CASE:Nom, NUM:Pl), but one cue points to two images (NUM:Pl). Notice that in that case the disparity in activation of the DP.1 and DP.2 images was still large: 0.91 to 0.10. The reason is that cue combination is not linear, but a weighted product.

Consequently convergent cues are much more effective than single cues at retrieval.

Therefore, there would remain a qualitative difference between the Sg [Pl] ungrammatical sentences and the Sg [Pl] grammatical sentences. In the latter case, the verb triggers retrieval with information that is *always* associated with the subject image (CASE:Nom, CAT:DP, etc. etc.), whereas in the former case, the information is split between the two images (CASE:Nom v. NUM:Pl). Put more succinctly, all of the cues in the grammatical sentences are cooperative, whereas they are in competition in ungrammatical sentences.

3.3.2.4 Relative clause attraction

Next we turn our attention to relative clause attraction. Because the experimental results are qualitatively similar for the two constructions, the aim is to see how far the same explanation will extend. An example of ungrammatical attraction is given in the relative clause in (60), with a partial phrase structure bracketing:

(60) [DP [DP The runners] [CP who [DP the driver] wave to each morning]] ...

Focusing on just the DPs in the structure, we assign the images below:

(61) [Pl [Sg]]: Images for “the runners who the driver”

<p>F-Head: <the> L-Head: <runners> Adjunct: →CP.2 Case: Nom Num: Pl</p> <p>DP.1</p>	<p>F-Head: <the> L-Head: <driver> Case: Nom</p> <p>DP.2</p>
--	--

Notice that there is a problem with the images as given. If the (ungrammatical) verb form supplies the same two cues as in the complex subject examples above, { CASE:Nom, NUM:PI }, then the cues no longer compete, but converge on the attractor. Table 3-6 illustrates the numerical predictions.

[DP.1 The runners ...		[DP.2 the driver]		
		*wave		
		DP.1	DP.2	
<i>Q (Cue)</i>	<i>W(eight)</i>	<i>S(trength)</i>		
CASE:Nom	0.5	0.99	0.99	
NUM:PI	0.5	0.99	0.01	
				<i>F</i>
<i>Activation</i>		0.99	0.10	1.1
<i>P_{SAMPLE}</i>		0.91	0.09	

**Table 3-6 Retrieval structure and outcomes for RC Attraction
Plural RC head, Singular RC subject, Ungrammatical**

The probability of sampling the correct projection, i.e. the subject DP.2, is highlighted in the double-bordered cell.

The reason this scenario differs from the complex subject scenario is that the licensing feature we selected, CASE:Nom, differentiates the two DPs in complex subjects: the embedded DP does not bear the same case feature as the subject DP. However, in the relative clause case, both DPs are nominative DPs. This outcome may not be problematic for the attraction case: for example, we see no evidence of difficulty due to ungrammaticality in the first RC attraction reading time experiment. (Experiment 1). However in both Experiment 2 and the judgment study (Experiment 4) there are grammaticality effects. More troublesome is when we consider an example like the following:

(62) The runner who the drivers waves to ...

Both the online and offline results indicate that comprehenders detect the ungrammaticality in this string, and that the singular relative clause head never intrudes to ameliorate it. However, if the cue set of ‘waves’ is, by hypothesis, just { CASE:Nom }, then we expect retrieval to be split between the matrix subject and the RC subject, which would incorrectly predict an attraction effect. Thus our retrieval structure cannot be consistently implemented in the two scenarios and achieve comparable results.

In the specification above, at retrieval the system uses the licensing properties of the subject to identify its constituent image in memory. We chose Case to serve as the relevant licensing feature. However, the problem the RC construction poses is not which inherent property we choose to identify the subject with. Instead of Case, we might have selected grammatical function (i.e., Subject), or structural position (i.e., [Spec,TP]). But in the RC examples, both DPs are grammatical subjects and both occupy the same position in their respective clause. The problem is that the system needs to be able to identify the subject relationally: it must identify the subject *of the same clause* as the verb. Can we re-configure the retrieval structure to account for this aspect of the licensing?

Suppose that we add another atomic feature to the system, identifying which clause a constituent belongs to. A revised set of images for ‘the runners who the driver’ is given below:

(63) [Pl [Sg]]: Images for “the runners who the driver”

<p>F-Head: <the> L-Head: <runners> Adjunct: →CP.2 Case: Nom Num: Pl Clause: 1</p> <p>DP.1</p>	<p>F-Head: <the> L-Head: <driver> Case: Nom Clause: 2</p> <p>DP.2</p>
--	---

The clauses are differentiated by a scalar value: call the matrix clause 1, and the RC clause 2. These numbers have no special meaning (like depth of embedding) but merely serve as arbitrary indices (see section 3.3.4 below). Suppose now the cue set for the RC verb in an ungrammatical sentence is: { CASE:Nom, NUM:Pl, Clause:2}. We keep our default assumptions that cues are associated with images with near-0 or 1 strength for features present in an image, and that there is an equal division of attention. The outcomes are encouraging. First, we consider the full results for the plural attractor RC fragments.

Table 3-7 reports the results for both grammatical and ungrammatical sentences.

[DP.1 The runners who...]		[DP.2 the driver] waves			[DP.2 the driver] *wave		
		DP.1	DP.2		DP.1	DP.2	
<i>Q (Cue)</i>	<i>W(eight)</i>	<i>S(trength)</i>		<i>W(eight)</i>	<i>S(trength)</i>		
CASE:Nom	0.5	0.99	0.99	0.33	0.99	0.99	
CLAUSE:2	0.5	0.01	0.99	0.33	0.01	0.99	
NUM:PL	--	--	--	0.33	0.99	0.01	

				<i>F</i>			<i>F</i>
<i>Activation</i>		0.10	0.99	1.1	0.22	0.22	0.43
<i>P_{SAMPLE}</i>		0.09	0.91		0.50	0.50	

**Table 3-7 Revised retrieval structure and outcomes, RC attraction
Plural RC head, Singular RC subject**

The probability of sampling the correct projection, i.e. the subject DP.2, is highlighted in the double-bordered cell.

Focusing on the ungrammatical case first (second column set), we see that the cues compete equally for the DP.1 and the DP.2 images, consistent with the pattern achieved for complex subjects. In the grammatical case, the convergence of the Case and Clause cues means that the correct image is sampled overwhelmingly (91% v. 9%). The qualitative pattern is not categorical however and opens the possibility that the attractor may intrude in agreement checking for the grammatical sentences, in a very small

proportion of cases. A similar prediction was obtained for sentences like “The keys to the cabinets are ...” but this prediction is difficult to verify, given that any noun-verb combination correctly agrees. However, in the RC case, the attractor and the verb misagree, so there could be a consequence: 10% of the time, an illusion of ungrammaticality should obtain. On the one hand, no reliable effect of attractor was found in the real-time studies, when grammatical conditions were compared. On the other hand, in the offline studies, there was a marginal effect of attractor number in the grammatical conditions. The data from Experiment 3 are repeated in the table below:

		<i>Response</i> (% ‘Yes’: <i>acceptable</i>)	
		‘runner’	‘runners’
<i>Sentence Type</i>			
Grammatical	“The ___ who the driver sees”	82%	70%
Ungrammatical	“The ___ who the driver see”	28%	64%

Table 3-8 Speeded grammaticality judgments of RC attraction
From Experiment 3

It may be that the increased tendency to report grammatical RC sentences as ungrammatical when there is a plural attractor stems from the kind of retrieval structure we have posited. Case and Clause cues converge to select the correct subject, but the Case cue also partially activates the RC head. The RC head mismatches the verb in number feature, resulting in the impression of ungrammaticality. We explore the consequences of the partial case activation in greater detail in 3.3.3, and accrue some supporting evidence (Experiment 5).

However, there is one more crucial scenario under which we must consider our revised retrieval structure for RC attraction: plural RC subjects. Empirically, the plural subject seems to ‘insulate’ the agreement processes from intrusion by a singular attractor. Table 3-9 reports the model results from such a configuration. The correct image is

overwhelmingly sampled in both cases. The decreased number of cues, however, means that the contribution of partial match of CASE:Nom is greater for ungrammatical conditions. Just as in the grammatical condition above, a small intrusion of the attractor is predicted in ungrammatical sentences here. Experimentally, however, no amelioration of the ungrammatical cases is observed, either in judgments or in reading times.

[DP.1 The runner who...]		[DP.2 the drivers]		[DP.2 the drivers] *waves			
		DP.1	DP.2				
Q (Cue)	W(eight)	S(trength)		W(eight)	S(trength)		
CASE:Nom	0.33	0.99	0.99	0.50	0.99	0.99	
CLAUSE:2	0.33	0.01	0.99	0.50	0.01	0.99	
NUM:PL	0.33	0.01	0.99	--	--	--	

				F			F
Activation		0.05	0.99	1.0	0.09	0.99	1.1
P _{SAMPLE}		0.05	0.95		0.09	0.90	

**Table 3-9 Revised retrieval structure and outcomes, RC attraction
Singular RC head, Plural RC subject**

The probability of sampling the correct projection, i.e. the subject DP.2, is highlighted in the double-bordered cell.

How should we interpret the results in Table 3-7 and Table 3-9? On the one hand, the predictions seem qualitatively correct, and in line with the asymmetrical results obtained in Experiments 1 through 4. The attractor intrudes strongly in ungrammatical sentences. On the other hand, the model does not categorically predict a complete lack of intrusion of the attractor in grammatical sentences (Table 3-7), nor does it categorically predict that a singular attractor should have no impact for ungrammatical singular agreement (Table 3-9). In both of these cases, the CASE:Nom cue points to the attractor, shifting by a few points the baseline activation contributed by the near-0 strength cues. As we mentioned above, there is some experimental indication that the plural attractor can intrude slightly when the sentence is grammatical, which the model captures as the

partial contribution of CASE:Nom as in Table 3-8). However, there is no corresponding experimental effect when the attractor is singular, and the sentence is ungrammatical.

Because the judgment effect for grammatical sentences containing plural attractors was small (and marginal), we would like to replicate it. If this effect is real, then either the model is correct, and we must explain why no corresponding effect is observed for ungrammatical sentences containing singular attractors; or we must revise the model.

3.3.3 *Agreement & Case (Experiment 5)*

In this Experiment we attempted to replicate part of Experiment 3, in which we observed that a plural RC head could intrude upon the checking of grammatical agreement, leading to the slightly decreased acceptability of (64b) with respect to (64a).

- (64) (a) The runner who the driver see ... ~>
(b) The runners who the driver see ...

There was a marked asymmetry between how much better ungrammatical sentences got, when an attractor was present – a lot – and how much worse grammatical sentences got, when an attractor was present – just a little. These results were clearly inconsistent with an account based upon erroneous feature percolation. In our formal, cue-based retrieval model, attraction arises from partial activation by the number cue on the attractor encoding. In that model, it was necessary for some cue to contact the representation of the subject, which we implemented with the Case property. This decision seems motivated, as case and agreement have been linked in syntactic analysis, associating nominative Case with subject-verb agreement (Chomsky, 1995). However, there is a consequence to using an inherent property, like Case, to access subject representations, which is that irrelevant subjects should be subject to retrieval. In grammatical RCs, like

(65), the Case and clausemate requirement of ‘sees’ converge on the correct subject phrase ‘the driver’; the ‘Case’ cue partially contacts ‘The runners’ as well.

(65) The runners who the driver sees each morning on the commute ...

Consequently it is expected that in a small proportion of cases, ‘the runners’ will be retrieved for agreement checking. How small a number of cases depends on the precise numerical properties of the model. The small (marginal) decrement in acceptability observed in Experiment 3 is consistent with this system. However, there is a qualitative distinction with configurations in which the RC head matches none of the verb’s cues. Such a configuration obtains when the relative clause is not attached to a subject-like head: for example, in an object-attached relative clause.

(66) Gerard recognized the runners who the driver sees each morning on the commute.

Here the relative clause head is an object in the main clause. It occupies a different structural position, and bears accusative Case. It therefore looks nothing like an agreement licenser, and should not intrude as a partial match. Consequently, unlike the subject-attached RC, where a small illusion of ungrammaticality is possible in grammatical sentences, the object-attached RC should exhibit no such illusion. In Experiment 5 we performed a direct comparison of subject- and object-attached RCs in a speeded grammaticality task, to determine whether or not the matrix clause subject-hood of the attractor impacts acceptability.

3.3.3.1 Model predictions

First we consider the SAM predictions for object-attached v. subject-attached RC attraction. The retrieval structure is identical as outlined in Table 3-7 and Table 3-9 above: there is a Case cue, a clause-mate cue, and a number cue. The predictions for a subject-attached RC, singular RC subject are re-capitulated below, in Table 3-10, simply as the sampling probability of the two DP images in memory.

Condition			Sampling probability	
<i>Agreement</i>	<i>Number</i>	<i>Example</i>	<i>Subject</i>	<i>Attractor</i>
Grammatical	SG	The runner who the driver sees ...	0.91	0.09
	PL	The runners who the driver sees ...	0.91	0.09
Ungrammatical	SG	The runner who the driver see ...	0.82	0.18
	PL	The runners who the driver see ...	0.50	0.50

Table 3-10 Subject and attractor sampling probabilities for Subj-attached RCs

Three cues are used by the model: a clause-mate cue, a case cue, and a number cue. Cue strength is all-or-none and attentional weighting is distributed uniformly.

The predictions for an object-attached RC are given in Table 3-11.

Condition			Sampling probability	
<i>Agreement</i>	<i>Number</i>	<i>Example:</i> <i>‘Gerard recognized ... ‘ :</i>	<i>Subject</i>	<i>Attractor</i>
Grammatical	SG	... the runner who the driver sees ...	0.99	0.01
	PL	... the runners who the driver sees ...	0.99	0.01
Ungrammatical	SG	... the runner who the driver see ...	0.94	0.06
	PL	... the runners who the driver see ...	0.80	0.20

Table 3-11 Subject and attractor sampling probabilities for Obj-attached RCs

Three cues are used by the model: a clause-mate cue, a case cue, and a number cue. Cue strength is all-or-none and attentional weighting is distributed uniformly.

There are two important comparisons to make between the two tables: first, as we've been discussing, for grammatical RCs with a plural head there is no likelihood of sampling the attractor when the RC is object attached, whereas there is an increased likelihood of doing so when the RC is subject attached (compare the double-bordered cells across tables). Secondly, however, the attraction effect is predicted to grow smaller for ungrammatical object-attached RCs (compare the wavy-bordered cells tables). The CASE:Nom cue is no longer effective in contacting the attractor, so the increased activation it provided is missing: NUM:Pl is the only cue activating the attractor, competing against a near-full match with the subject, CASE:Nom and CLAUSE:2. Note, however, that the attraction effect is not absent: the model simply predicts a more lopsided split in the sampling of subject and attractor: 80/20 instead of 50/50.

In an experimental design that crosses grammaticality, attractor number and attachment site, the model makes the usual, familiar predictions: a decrease in 'yes' responses when the subject and verb fail to agree, but an increase in 'yes' responses when a plural attractor is present. However, there are two signature predictions: (1) a decrease in 'yes' responses to grammatical sentences containing a plural attractor, but only for subject-attached RCs; and (2) a smaller ungrammatical attractor effect for object-attached RCs.

3.3.3.2 Materials and Methods

Participants

Participants were 24 native speakers of English from the University of Maryland community with no history of language disorders. They received credit in an introductory linguistics course for their participation.

Materials

40 sentence sets were adapted from Experiment 3. Each sentence crossed grammaticality, attractor number (RC head: plural/singular), and the site of attachment for the relative clause (Subj/Obj). The attachment site manipulation affects the Case property of the attractor: when subject-attached, the Case of the attractor is nominative; accusative when it is object-attached. A sample set of materials is given below:

Sample set of experimental items for Experiment 5			
<i>Attachment site</i>	<i>Grammaticality</i>	<i>Attractor number</i>	
<i>Subj</i>	<i>Grammatical</i>	<i>Sg</i>	The musician who the reviewer praises so highly will probably win a Grammy.
		<i>Pl</i>	The musicians who the reviewer praises so highly will...
	<i>Ungrammatical</i>	<i>Sg</i>	The musician who the reviewer praise so highly will...
		<i>Pl</i>	The musicians who the reviewer praise so highly will...
<i>Obj</i>	<i>Grammatical</i>	<i>Sg</i>	Phil met the musician who the reviewer praises so highly ...
		<i>Pl</i>	Phil met the musicians who the reviewer praises so highly ...
	<i>Ungrammatical</i>	<i>Sg</i>	Phil met the musician who the reviewer praise so highly ...
		<i>Pl</i>	Phil met the musicians who the reviewer praise so highly ...

Table 3-12 Sample materials set for Experiment 5

The materials were distributed across 8 lists by a Latin Square. Each participant therefore saw five items per condition.

This experiment was run concurrently with a related experiment on complex subjects, which incorporated 24 sentence sets (half of the conditions in which were ungrammatical). 56 further filler sentences were included, which were the same fillers run in Experiment 3. Overall, each participant saw 60 well-formed sentences and 60 ill-formed sentences.

Procedure and Analysis

Presentation and analysis details were as described for Experiment 4.

3.3.3.3 Results

Figure 3-7 reports the proportion of ‘yes’ response in the *Subj*-attached conditions, and Figure 3-8, the same values in the *Obj*-attached conditions.

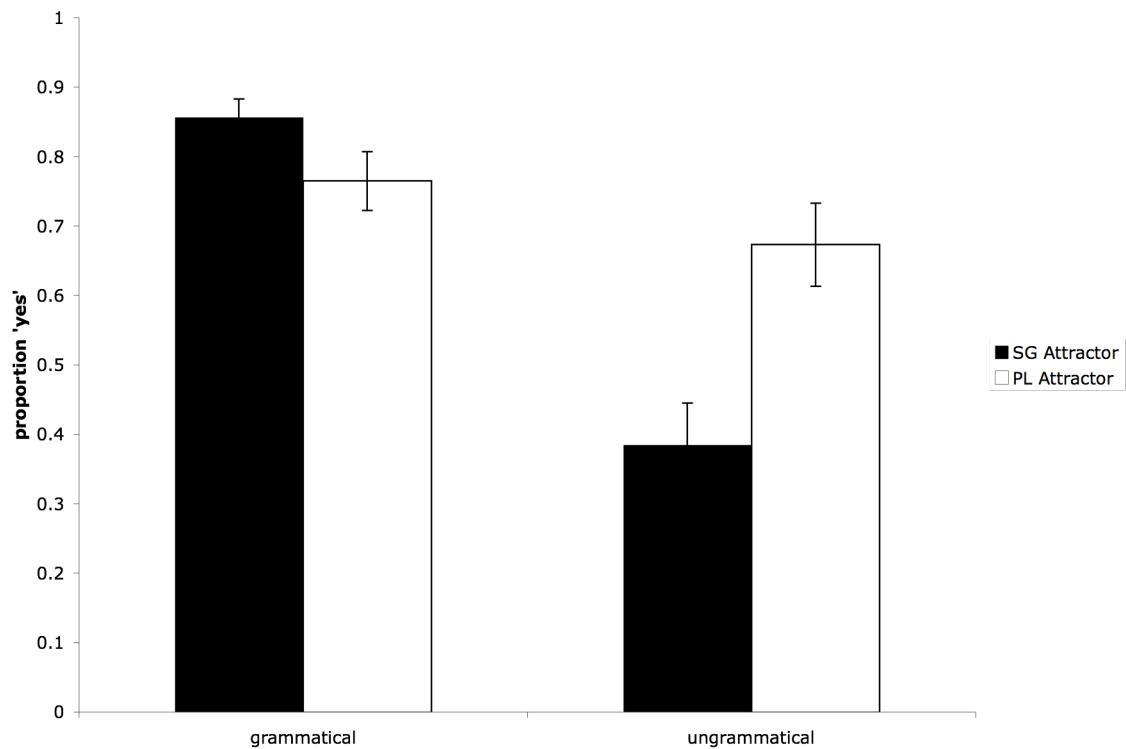


Figure 3-7 Experiment 5: Relative clause attraction, Subject-attached RCs
Speeded grammaticality, proportion ‘yes’ responses

Full model: GRAMMATICALITY × ATTRACTOR × ATTACHMENT. We observed a main effect of grammaticality: participants were less likely to respond ‘yes’ when subject and verb failed to agree (fixed effect logit coefficient β : -2.11 ± 0.65 ; $p < 0.001$). We also observed a main effect of attachment: participants were more likely to respond ‘yes’ when the relative clause was attached to the subject (β : 1.00 ± 0.66 ; $p < 0.005$). Two interactions were significant: the two-way grammaticality × attractor number interaction,

such that participants were more likely to respond ‘yes’ to ungrammatical sentences when the relative clause head was plural ($\beta: 1.12 \pm 0.80; p < 0.001$); and a three-way grammaticality \times attractor \times attachment interaction, such that participants were more likely to respond ‘yes’ to ungrammatical sentences, when both the relative clause head was plural and the attachment-site was the subject ($\beta: 1.27 \pm 0.60; p < 0.05$).

Partial model/Subject Attached: GRAMMATICALITY X ATTRACTOR. Considering only the half of the conditions in which the RC was subject attached, there was a main effect of grammaticality ($\beta: -2.73 \pm 0.67; p < 0.001$), indicating a decreased likelihood of responding ‘yes’ to ungrammatical sentences. There was also a grammaticality \times attractor number interaction ($\beta: 2.37 \pm 0.88; p < 0.001$), such that participants were more likely to respond ‘yes’ to ungrammatical sentences when the relative clause head was plural. Crucially, participants were also slightly less likely to say ‘yes’ to sentences containing a plural, even if they were grammatical ($\beta: -0.91 \pm 0.67; p < 0.01$), such that participants were less likely to say ‘yes’ to sentences containing a plural attractor. This effect corresponds to the marginal effect observed in Experiment 3; it confirms that grammatical agreement within RC sentences is perceived as less acceptable, when the RC head is plural, the RC subject is singular, and the RC is attached to a phrase in subject position.

Partial model/Object Attached: GRAMMATICALITY X ATTRACTOR. The pattern of results changes when only the object-attached conditions are considered. The only reliable effects were grammaticality ($\beta: -2.02 \pm 0.58; p < 0.001$) and the grammaticality \times attractor number interaction ($\beta: 1.06 \pm 0.40; p < 0.01$). Participants detected ungrammaticality, but were much more likely to say ‘yes’ to ungrammatical sentences if

the sentence contained a plural RC head. Unlike subject-attached relative clauses, participants did not show a decreased likelihood to say ‘yes’ to grammatical sentences, however. In other words, grammatical agreement is impervious to the attractor in object-attached RCs, but not in subject-attached RCs.

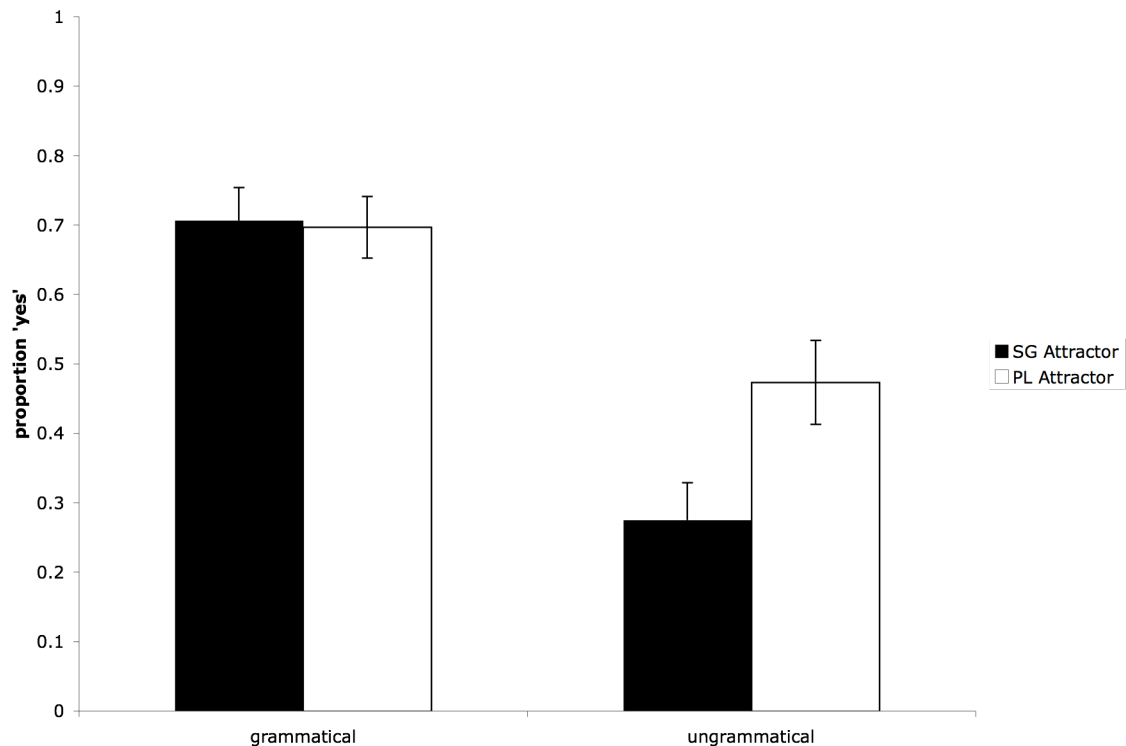


Figure 3-8 Experiment 5: Relative clause attraction, Object-attached RCs
Speeded grammaticality, proportion ‘yes’ responses

3.3.3.4 Discussion

This experiment once again replicated the basic attraction effect: comprehenders detect subject-verb disagreement robustly, except when there is a plural attractor.

However, it also revealed a tendency to judge grammatical sentences containing a plural attractor as ungrammatical. This small tendency was seen in Experiment 3, but there it was only marginally significant; in this experiment, which included more participants, it

was reliable. What this experiment further showed is that this tendency was only present in subject-attached relative clauses. The cue-based model is consistent with the existence of a small effect, but crucially predicts that it should be absent when the attractor is not subject-like. The experiment bore out this prediction.

The experiment also bore out the model's second prediction, which is that the illusion of grammaticality that an attractor induces should shrink when the attractor is in matrix clause object position. In both subject and object-attached cases, a baseline attraction effect is guaranteed by the NUM:Pl cue; but in the subject-case, it is reinforced by the CASE:Nom cue. The discrepancy between the two effects is not great, though it is reliable. In subject cases, the attractor increased acceptability judgments from an ungrammatical baseline of 38% to 67%, a raw difference of 29%. In object cases, the attractor increased judgments from a baseline of 28% to 47% acceptable, a raw difference of 19%. In the full logit model, the difference in odds ratios between subject-attached and object-attached attraction was reliable ($\beta: 1.27 \pm 0.60; p < 0.05$).

The difference in baselines raises concerns for why the sentences containing object-attached relative clauses were judged less acceptable overall. Even for singular-attractor, grammatical sentences, the acceptability rate was 71% for object-attached RCs, compared to 86% for subject-attached RCs. The raw rates are misleading, however: indeed, participants were overall more sensitive to agreement mismatches in object-attached RCs. The odds ratio of saying 'no' in ungrammatical v. grammatical sentences was 6.3 in object-attached RCs, which is slightly greater than the odds ratio of 6.0 for subject-attached RCs. The differences observed in attractor effects between subject- and object-attached RCs cannot therefore be attributed to scale differences.

3.3.3.5 Timing differences, task demands and an alternative explanation

The difference between subject- and object-attached RCs appears consistent with the partial match property of the content-addressable retrieval model. However there is another candidate explanation to consider, which concerns the relationship between the act of judgment and the occurrence of the violation (or of processing difficulty). Suppose that there are two ways to perform the grammaticality judgment task: (1) upon noting violations, set and maintain an internal ‘no’ response until the judgment is signaled; (2) re-inspect the representation when the signal to judge is given. An individual may choose and mix strategies depending upon properties of input. Suppose then that the ‘flag and set response’ strategy is more effective when the delay between flagging and responding is relatively modest, and becomes less effective the longer the ‘flag’ has to be maintained. The intuition is that ‘older’ flags are less reliable, and are treated with less confidence as time goes by. The earlier that the flag is set in the sentence, then the less information it is based on; and the more likely it indexes a violation that could have been resolved later.

In the subject-attached RC case, the subject-verb mismatch occurs early in the comprehension of the sentence, and considerable time elapses before the judgment is made. There are on average 10 words that occur between the critical verb and the end of the sentence. Given a presentation rate of 300 ms / word, roughly 3 seconds would have elapsed between the violation and the opportunity to respond. In the object-attached RC case, the judgment occurs relatively soon after the critical verb. There are on average only 5.5 words that are presented in that interval, corresponding to 1650 ms of elapsed time. As a consequence, in object RC cases the judging behavior is controlled largely by the ‘flag’ strategy, since object RC errors are mostly associated with recently set flags. In the subject RC cases, the participant is more prone to re-inspect or re-generate the

syntactic representations. In particular, suppose that the comprehender covertly produces the sentence to judge it. In such a scenario, the agreement attraction error could be recapitulated, just as if the individual were producing the sentence for the first time. Even if the comprehended sentence was grammatical, the presence of the plural attractor could sometimes lead to the (covert) production of an error, which would lower the perception of grammaticality for grammatical sentences. Notice that it must be a property of the stimulus that determines which strategy the comprehender relies on. There would be no ‘no’ flag for grammatical object-attached sentences, yet no decrement in acceptable judgments is observed in the attractor condition. Therefore, it couldn’t be that the absence of a ‘no’ flag triggers regeneration. Rather we would have to posit that the participant notices that subject RCs generate many low-confidence responses, and therefore have a greater tendency to re-inspect all such sentences.

The regeneration scenario is plausible. Syntactic representations have long been seen as labile entities, with short half-lives. This property of the representation has been inferred from the fact that an individual’s ability to discriminate between a recently presented sentence and a semantically-equivalent but syntactically distinct version of the sentence declines rapidly after presentation (e.g., Sachs, 1967). It is a prominent property of verbal memory, however, that sentences can be repeated verbatim with high accuracy immediately after presentation (e.g., Jarvella, 1971). Potter & Lombardi (1992) argued that verbatim memory does not reflect a ‘read-out’ operation over a well-preserved syntactic representation. Rather accurate verbatim performance is the outcome of a new production act, fed by the recently activated lexical items and the recent interpretation. They later extended this explanation to syntactic priming (Potter & Lombardi, 1998). We

have raised the specter of precisely this mechanism in suggesting that re-inspection of a syntactic representation invokes production processes. Based on the present data, however, it is difficult to significantly strengthen this perspective or conclusively argue against it. On the one hand, the notion that more processing is involved in judging the subject-attached RC sentences is potentially consistent with the overall higher accuracy in judgments. On the other hand, it naïvely predicts that judgment times for object-attached RC sentences should be faster than for subject-attached RCs. Our data show this is not the case: the average time to make a judgment of object-attached RC sentences is 727 ms, compared to 613 ms for subject-attached RC sentences ($\Delta\mu$: 114 ms; 95% C.I.: [45 ms, 116 ms]; $p < 0.01$). That prediction probably is too simplistic, however: there is potentially more local processing difficulty when judgments are made immediately following an object RC and that difficulty could spill-over into judgment times.

A more direct test will be necessary to better evaluate this alternative explanation. For example, subject-attached RCs presented in a sentence initial position could be contrasted with those that occur in sentence-final position:

- (67) (a) The runners who the driver sees each morning were next to the busy intersection.
(b) Next to the busy intersection were the runners who the driver sees each morning.

The case properties of the attractor head are putatively identical in both positions, but the difference in timing between the critical verb and the judgment signal is greater in (a) than in (b). Consequently, a decrement in acceptance of grammatical attractor sentences should be more pronounced in (a) than in (b), if the controlling factor is timing and not the inherent properties of the attractor.

3.3.4 *Clause-boundedness*

The retrieval structure proposed to handle RC agreement attraction incorporates a clause-mate licensing feature: the CLAUSE:2 cue. This cue ensures that in an ungrammatical sentence like “The runner who the drivers sees” the Case cue will target the correct image, and not sampling will not be split between the attractor and the subject. There is no evidence for singular DP attraction in these configurations, which places an important constraint on the model. However, is the idea of a “Clause” cue reasonable? Notice that a clause cue is fundamentally different from a case or number cue. The latter are well-motivated item properties of DPs. Case and number features play a role in structure building and licensing. For example, Case assignment (or checking) is tightly connected with syntactic explanations for which structural positions nominals may occur (e.g., Rouveret & Vergnaud, 1980). It is therefore plausible to assume that encodings of DPs include information about their Case properties. Likewise, both the existence of formal covariation between verbal and nominal heads and the semantic import of number implies that each head carries number information. It is a different claim, however, to suppose that the DPs contained within a clause encode information about the clause in which they belong. We will now attempt to substantiate this view.

In our view the Case and number properties of a DP map onto what the memory literature refers to as item features, whereas clause identity aligns with an encoding of context. Item features belong to encodings by virtue of the representational system in which they take part. DPs have Case features because the grammar assigns DPs a Case value. Context features belong to encodings by virtue of the circumstances under which the encodings were instantiated. It has been extensively documented that memory search is affected not only by inherent features of item encodings, but also the circumstances

under which an instance of the item was presented for study. A key paradigm is studying how memory search is studied is the free recall paradigm: individuals study a list of words, and are then prompted to recall as many as they can. An interesting aspect of an individual's response is the order in which the words are recalled, which is taken to reflect the organization of the search. One determinant of order is pre-existing associations between words, or how the intrinsic properties of words relate to one another (Bousfield, 1953; Cofer, Bruce, & Reicher, 1966; and many others). However response clustering can also be observed for extrinsic factors related to the conditions of study, like modality, gender of a speaker's voice, or typeface (Murdock & Walker, 1969; Hintzman, Block & Inskip, 1972; Nilsson, 1974). In addition, internal factors can induce clustering, such as the processing task engaged in during encoding (Craik & Lockhart, 1972) or the mood of the participant (Eich, 1980). Finally, temporal contiguity matters: once pre-existing associations are controlled for, it is observed that words in successive list positions tend to be recalled successively (Kahana, 1996). Generally it can be said that pseudo-arbitrary conditions of encoding impact the means of searching items in memory, and these are broadly termed context effects. Context could affect organization by different mechanisms: memory could be organized into separate kinds of stores that reflect input modality (e.g., an auditory store, a visual store, etc.; e.g., Nilsson, 1974), or attributes of the study conditions could be encoded directly in a memory trace as a set of features (Hintzman et al. 1972, Howard & Kahana (2002). Howard and Kahana (2002) conceive of a context representation as a global state code that slowly evolves in time and is impacted by both the input and internal state variables. It can be thought of as a kind of an elaborate time stamp (containing, however, more than just temporal information).

Items that bear temporally closer time stamps will be more similar than those with temporally more separated states, and consequently they will be more likely to be sampled together. We propose that information about clause membership (or membership in a relevant syntactic domain) is included in all constituent representations as a context encoding, a sort of linguistic time stamp.

The global organization of syntactic representations, we conjecture, is reflected not always in explicit encodings, but in the organization of the parsing mechanism. Segmentation of the sentence at the clausal level has been proposed as a perceptual and planning strategy (cf. Fodor, Bever, & Garrett, 1974, Levelt, 1973). This organization seems logical for a parser which must keep similar kinds of relationships distinct across clauses, as each clause, anchored by a verb or finite tense, introduces its own packet of thematic, event, and information structure. If the processing of separate clauses is reflected in separate epochs of processing, then it seems likely that an internal context representation would be impacted in large degree by which clause is currently being processed. Let us flesh out a few details, in what is essentially a simplification of Howard & Kahana (2002)'s temporal context model. Suppose, therefore, that there is an internal context representation, which can be conceived of as simply a large vector. Constituent encodings include not only their inherent features, like Case, number, tense, X' relations, etc., but also the context vector. The current value of c is determined by inputs from currently processed items and any number of internal state variables. If we imagine that context is updated in a step-wise fashion, then we can assume that context at step i is simply a linear combination of current context, c_{i-1} and the new information that is placed in context, c^{NEW} . The parameter β determines how much strength is assigned to the new

information in context. ρ_i weakens the current representation (as a function of how similar new and old information are, in order to keep the magnitude of \mathbf{c} constant; see Howard & Kahana, 2002, for details)

$$(68) \quad \mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}^{\text{NEW}}$$

Suppose further that when processing shifts from one clausal (or cyclic) domain to another, the Context encoding is shifted by a randomly generated \mathbf{c}^{NEW} . This assumption is a crucial one. It has the consequence that the value of context for images encoded in one clausal processing epoch is more similar than the value of context for images across clausal processing epochs. This similarity translates into different cue strengths in the SAM model. During the agreement checking process, the verb inside the RC will use its own context information as a cue, and this cue will highly activate images encoded in similar contexts. In the formulation of the RC model above, the effect corresponds to the index feature we used, assigning the whole RC DP the feature `CLAUSE:1` and the RC subject the feature `CLAUSE:2`, and putting `CLAUSE:2` in the retrieval structure of the verb.

We have sketched one means by which clause information could be encoded in every representation of a constituent contained within the clause. In effect we have stipulated that each constituent image contains a tag with a code for that clause. What we have attempted to argue, however, is that such a stipulation is not unnatural: a clause tag can be conceived of as one (important) component of an evolving context representation. The update of this representation can be keyed to major processing events in a way that leads to gradations and shifts in similarity of encoded units in memory as a function of which processing epoch a constituent was encoded in.

If we reconsider complex subject attraction, however, the incorporation of context in the retrieval structure is not unproblematic. Recall the case of a grammatical sentence including a plural attractor:

(69) The path to the monuments is littered with bottles.

The retrieval model was able to reliably contact the true subject in virtually all cases, because the only active feature in the retrieval structure is the Case cue. A Case cue points unambiguously to the correct image. However, if now there is both a Case cue and a Clause/context cue, the incorrect image should be subject to sampling in a small number of cases because the Clause cue points to it as well as the correct image. One possible way to avoid this situation is to adapt the attention allocated to the context cue on the basis of its power of discriminability. We have operated under the assumption that attention is allocated uniformly to all cues, but that was a simplifying assumption made to illustrate the general properties of the system. However suppose that the weighting assigned to the context cue changes depending on how quickly or how often context has changed in some recent time window (and suppose that the system is capable of determining this fact). Context cues will therefore be more potent in embedded RCs than in simple matrix clauses. If this system has this property, then it predicts that it should be possible to modulate the effect of the attractor in grammatical sentences by embedding the complex subject in more complex (multi-clause) environments. In multi-clause environments the Clause cue is more salient and would consequently receive greater weight.

- (70) (a) The path to the monuments is littered with bottles.
→ NO ATTRACTION
(b) The lawn is tidy but the path to the monuments is littered with bottles.
→ SOME ATTRACTION

This experiment has not been run. Intuition suggests that there is no difference, but any intrusion would be subtle, so intuition is likely not a reliable guide in this situation.

3.3.5 *Next to the cabinets ...*

A further prediction of the current model is that any plural can in principle affect subject-verb agreement. Imagine a sentence like the following:

- (71) Next to the cabinets the cat were sleeping.

The cues provided by the verb ‘were’ are NUM:Pl and CASE:Nom. No DP satisfies both of these cues, but each should partially activate at least one: NUM:Pl being associated with ‘the cabinets’ and CASE:Nom with ‘the cat.’ The data for such constructions do not exist, both because the comprehension literature is relatively small, but also because the production literature, though large, has sampled only a small portion of the logical space of head-attractor configurations. Other kinds of complex subject attraction are sometimes reported, often simply observationally, as in:

- (72) The participants’ identity are to remain a secret.
(R. Kayne, reported in den Dikken, 2001)

Greater work is needed on non-complex subject attraction to better formulate a model.

Given the unattractiveness of feature percolation, these configurations are of increased interest to determine what counts in principle as an accessible DP for attraction.

3.3.6 *Linking comprehension and production*

The content-addressable model of agreement licensing captures the major characteristics of agreement attraction in comprehension. With a set of common

assumptions across constructions, it was possible to recapitulate the major patterns and to some extent the size of effects observed in our reading time and speeded grammaticality studies. The partial match property of cue-based retrieval is responsible for allowing the attractor to intervene in ungrammatical sentences. However, because of a non-linear cue combination rule, partial matches can be relatively marginalized when cues fully converge on a representation. The results of Experiment 5 supported this conclusion.

However, what is to be made of the link between comprehension and production? One attractive aspect of the feature percolation model is that it can be easily stated independently of the task because the error derives from the syntactic encoding itself. Consequently the explanation for agreement attraction in comprehension and production is fundamentally the same. It is less clear that the content-addressable model has the same extensibility, but we would like to argue that it does. First, consider Solomon & Pearlmuter (2004)'s model of agreement attraction in production, discussed in Chapter 2. What drives the attraction effect in that model is the simultaneity of the head nouns during the production planning process. The attractor can wrest away control of agreement by being co-active with the grammatical controller when conceptual structure is mapped onto syntactic structure. Because both heads are simultaneously accessible, the process of selecting verb form sometimes spuriously pays attention to the attractor's features. The more accessible the attractor, the more likely it is to be spuriously selected. Accessibility is determined by the tightness of the semantic relation between heads: roughly, the degree to which the attracting head characterizes the subject head. In the comprehension model proposed here, simultaneity again plays a role. Because the search occurs in parallel, there is a potential that multiple constituents will be rendered

accessible to comprehension processes. The choice of which set of constituents is made accessible depends on the match between the retrieval cues and constituent features. Therefore, there are formal similarities between the two models, although the mechanisms differ.

More recent work by Bill Badecker and Rick Lewis in the ACT-R framework (2007), conducted simultaneously with our own, has sought to explain production errors of agreement in nearly exactly the same fashion that we are explaining the comprehension errors: attraction arises when cue competition leads to a partial match. The production process is posited to take place in exactly the same kind of architecture as comprehension, one in which most operations must retrieve information from recent memory. Verb marking occurs after the subject has been constructed, and it must retrieve the subject to inspect its number properties. Badecker & Lewis assume a richer set of features than we have assumed. Both the local and subject DPs share a category feature, as well as a nominative case feature in the planning process. The whole subject is distinguished because its dominating category, IP, is encoded. Therefore a set of cues exists that will converge upon the whole subject, but only partially on the local noun. Consequently the production system retrieves the whole subject in the majority of cases. Embedded plurals lead to attraction because the cue structure includes a variable value cue: NUM:*var*. The system is thus biased to return explicitly number marked constituents.

Finally, Badecker & Lewis explain the hierarchical effects of agreement attraction (e.g., Franck, et al., 2002) in terms of distinctions in inherent activation. Hierarchically more dominant categories have higher base rates of activation, because they have undergone more processing, a phenomenon that we discussed in 3.2.1.3. We have not

attempted to give an account of hierarchical effects because the evidence for such effects is simply lacking in comprehension. Pearlmutter (2002) attempted to test for such effects, but the results were equivocal. Indeed the similarity we have discovered between complex subject and relative clause attraction suggests that hierarchical effects may be limited in comprehension. Subject and verb are adjacent in relative clause comprehension, yet the relative clause head manages to intervene. This fact adds to the motivation in 3.3.5 for greater work in comprehension, in order to better sample the logical space of subject head and attractor configurations.

Despite the specificity of our account for comprehension phenomena, the potential for unification with production models seems high: the same abstract intuition underlies Solomon & Pearlmutter's model, and the connection with Badecker & Lewis's model is quite explicit.

3.4 Interference and subjects

3.4.1 Introduction

In Section 3.3 we explained agreement attraction by appealing to the properties of retrieval in a content-addressable memory. Using the Search of Associative Memory model, we worked out the assumptions necessary to make such a model feasible, constrained by experimental data and general properties of linguistic representations. In a larger context our discussion of agreement attraction has revolved around the question of where and how to impute fallibility to the real-time linguistic systems. For the phenomenon of agreement attraction the answer seems to be that the difficulty is not in a faulty encoding apparatus, but in the means by which those encodings are accessed in later processing. We would like to broaden of the scope of the discussion to other

phenomenon, and argue that the means by which previously encoded information is accessed is a major determinant of fallibility in real-time systems.

In the remainder of this chapter, we will consider the cases of complex subject attachment discussed by Van Dyke & Lewis (2003) and Van Dyke (2007). Van Dyke claims that in cases like the following, comprehenders sometimes erroneously select a grammatically inaccessible constituent as the subject of predicate. The erroneous subject and predicate pairing is underlined.

(73) The critic who said that the author was busy with his new novel was writing a comprehensive survey of contemporary literature.

The explanation offered should be by now familiar: the grammatically-inaccessible constituent is subject-like enough to be returned in a retrieval operation. Van Dyke draws upon a number of online and offline data to support this point. We take issue with the materials used to obtain this data. In a more closely controlled self-paced reading experiment, we obtain divergent results in the online data, though not the offline data.

In Chapter 4 we will consider other cases in which computing an analysis with a complex subject is fallible and the system appears to entertain illicit relationships. Secondly, however, we examine cases in which illicit analyses never seem to be formed. We argue that the processing system is adapted to the memory architecture in that it aggressively exploits predictive information to restrict the formation of dependency to grammatically-licensed constituents. This point we will illustrate with reference to the sizable body of literature on anaphora, but will examine it in greater detail for the processing of *wh*-dependencies. We will present further experimental evidence that supports our generalization.

3.4.2 *Van Dyke & Lewis (2003), Van Dyke (2007)*

Julie Van Dyke and colleagues have studied complex subject attachment in considerable detail, and have concluded that the parser sometimes identifies grammatically-inaccessible constituents as the subject. Complex subject attachment refers to the pairing of an incoming verb with the immediately preceding subject, when the subject is complex. Consider the following sentence:

(74)

[_{DP.1} The student who knew [_{DP.2} the exam] was important] was waiting in the hallway.

The subject of the sentence is the entire phrase bracketed by DP.1. However DP.1 itself includes a sentence, which has its own subject, DP.2. Parsing DP.1 requires identifying the entire segment ‘the student who knew the exam was important’ as a DP, as well as assembling an analysis for the RC inside DP.1. When the matrix clause auxiliary ‘was’ is encountered, it must be understood as signalling a predication over DP.1. Van Dyke has argued that if identifying the subject involves cue-based retrieval operations at the verb, identifying DP.1 as the subject will be error-prone if the syntactic left context contains other subject-like constituents. The logic transfers transparently from our previous discussion: imagine *was* providing a CASE:Nom cue. In (74) both DP.1 and DP.2 look inherently subject-like, both having a CASE:Nom feature.

Van Dyke & Lewis (2003) tested this prediction in a self-paced reading experiment. They considered simple subjects, complex subjects containing a lexical subject (High Interference conditions), and complex subjects with no lexically-filled subject position (Low Interference conditions). They embedded sentences with each of these kinds of subjects under verbs like ‘forget’ that can take either simple DP objects or an entire clause. They also crossed the subject type with the initial ambiguity of the

embedding, by varying the presence of the complementizer, to obtain six conditions. An example set of materials is given in (75). The ambiguity cross was performed because in that paper Van Dyke & Lewis were also interested in whether or not cue-based parsing was a general mechanism, or whether it might be restricted to the repairing of mis-analyzed sentences (as would be necessary in the reanalysis of the initially ambiguous sentences).

(75) AMBIGUOUS

The executive assistant forgot _____ was standing in the hallway.

- (a) CONTROL: Simple subject
... the student ...
- (b) HIGH INTERFERENCE: Complex subject, contains a lexical subject
... the student who knew that the exam was important ...
- (c) LOW INTERFERENCE: Complex subject, no lexically-filled subject position
... the student who was waiting for the exam ...

UNAMBIGUOUS

The executive assistant forgot that _____ was standing in the hallway.

- (d) CONTROL: Simple subject
... the student ...
- (e) HIGH INTERFERENCE: Complex subject, contains a lexical subject
... the student who knew that the exam was important ...
- (f) LOW INTERFERENCE: Complex subject, no lexically-filled subject position
... the student who was waiting for the exam ...

Notice that it is not possible to have a relative clause that has no grammatical subject, but if the subject position is the relativized position, as in (75c/f), then the subject is not lexically expressed. The same phrase that occupies a (further embedded) lexical subject position in the HIGH INTERFERENCE conditions, ‘the exam’ occupies an oblique position in the LOW INTERFERENCE conditions, as the object of a preposition. If the parser’s identification of the subject, at the matrix verb, is subject to similarity-based interference, then conditions (b/e), which contain additional overt subject-like phrases, should present greater difficulty than conditions (c/f).

The results from Van Dyke & Lewis' Experiment 4 are presented below in Figure 3-9. We include only data from the critical 'Aux+V' region, i.e. 'was standing.' No differences were observed in preceding regions, except slow-downs that obtained when one region contained a single additional word. Residual reading times are presented, regressed over both word length and ordinal position; raw untrimmed reading times exhibited (approximately) the same patterns. Reading times in the 'Aux+V' region show a main effect of subject type, ambiguity and the interaction of the two. Pairwise comparisons reveal what drives these effects: high interference conditions are always read more slowly than low interference conditions ($\Delta\mu$: 63 ms). There is a strong slow-down due to ambiguity, although only in the interference conditions ($\Delta\mu$: 93 ms). Only in the ambiguous sentences are high interference conditions read reliably more slowly than short sentences; low interference conditions are not read reliably differently than short sentences (though the difference is marginal in unambiguous sentences, and approaching marginal in ambiguous ones).

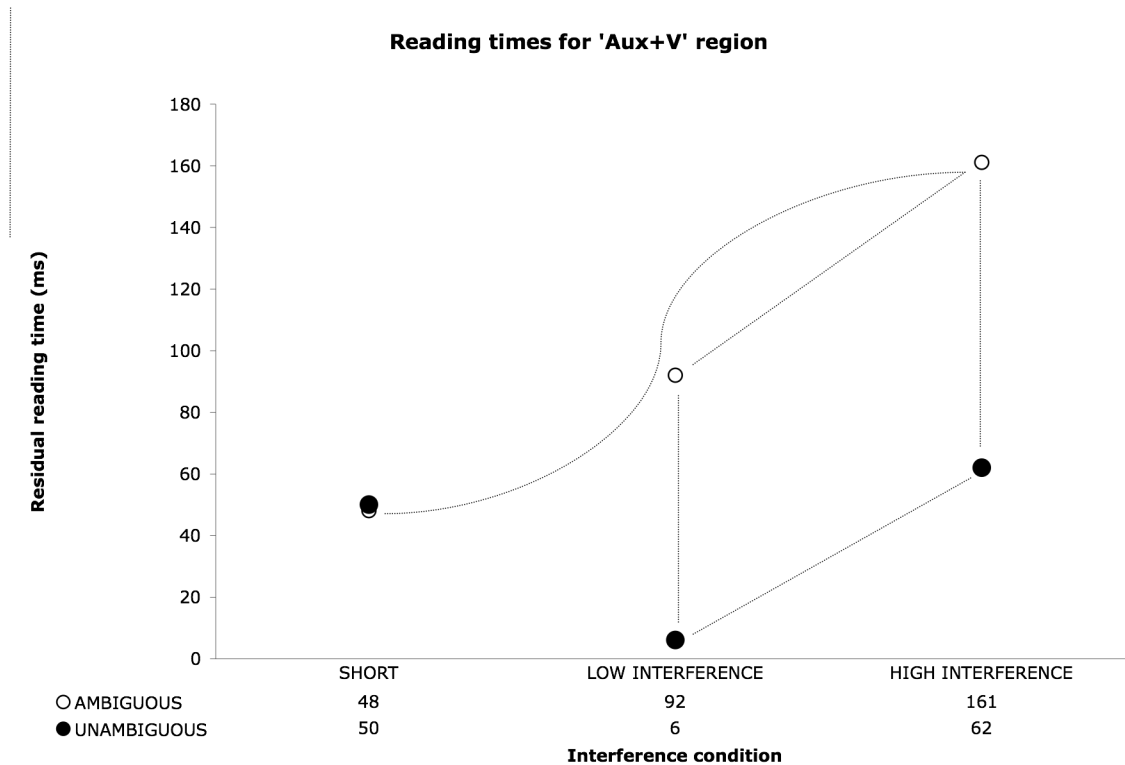


Figure 3-9 Van Dyke & Lewis (2003), Experiment 4

Residual readings times for Experiment 4, Region 3, the two-word ‘Aux + Verb’ region. These data show a main effect of subject type, a main effect of ambiguity, and an interaction. There were insufficient descriptive or inferential statistics reported to construct error bars or confidence intervals. However, results of significant ($p < 0.05$) pairwise comparisons are indicated by lines between the cells compared. $N_{SUBJ} = 36$. $N_{ITEM} = 36$.

The contrast between high and low interference conditions, which is irrespective of ambiguity, is consistent with the idea that the subject is identified at the verb by a general, cue-based process. When a grammatically-inaccessible constituent that is nonetheless subject-like is present in the syntactic context, reading times at the auxiliary and verb increase. It is somewhat troubling, however, that there is no difference between the high interference and short conditions in the unambiguous conditions. On the one hand, it seems plausible that the adjacency of the subject and verb obviates the need for a retrieval operation. On the other hand, we observed attraction effects in RC

configurations where subject and verb were adjacent; so there may not be a necessary connection between string adjacency and the need for retrieval.

However, reassuringly, the online results are mirrored by off-line measures, as well. In Experiments 2 and 3, participants read the same sentences in a self-paced reading task, after which they had to give a grammaticality judgment (on the sentence they had just read).³¹ The results of those experiments, presented in Figure 3-10, show that participants were *always* less accurate on high interference conditions, regardless of ambiguity. Indeed this pattern of results is more consistent with the notion that the inaccessible subject is interfering than the online results: even in the unambiguous conditions, the high interference condition is always more difficult than either the short or low interference conditions.

³¹ The instructions administered to participants are somewhat curious however. From Van Dyke & Lewis (2003, p. 296):

The experimenter explained that for a sentence to be ungrammatical, it should either be missing words or have too many words. This was illustrated with a sentence like “The police gave the citizen who he caught driving too fast on the parkway” or “The student was practicing reviewed his homework.” It seems unusual to explain grammaticality in this fashion, though the examples given are indeed ungrammatical. However by emphasizing a connection between grammaticality and a well-formed phrase structure, it may sharpen the expected contrasts. In the present experiment, sentences are putatively difficult because up to three subject-verb pairs must be matched, and that aspect of the materials may become more salient given the instructions.

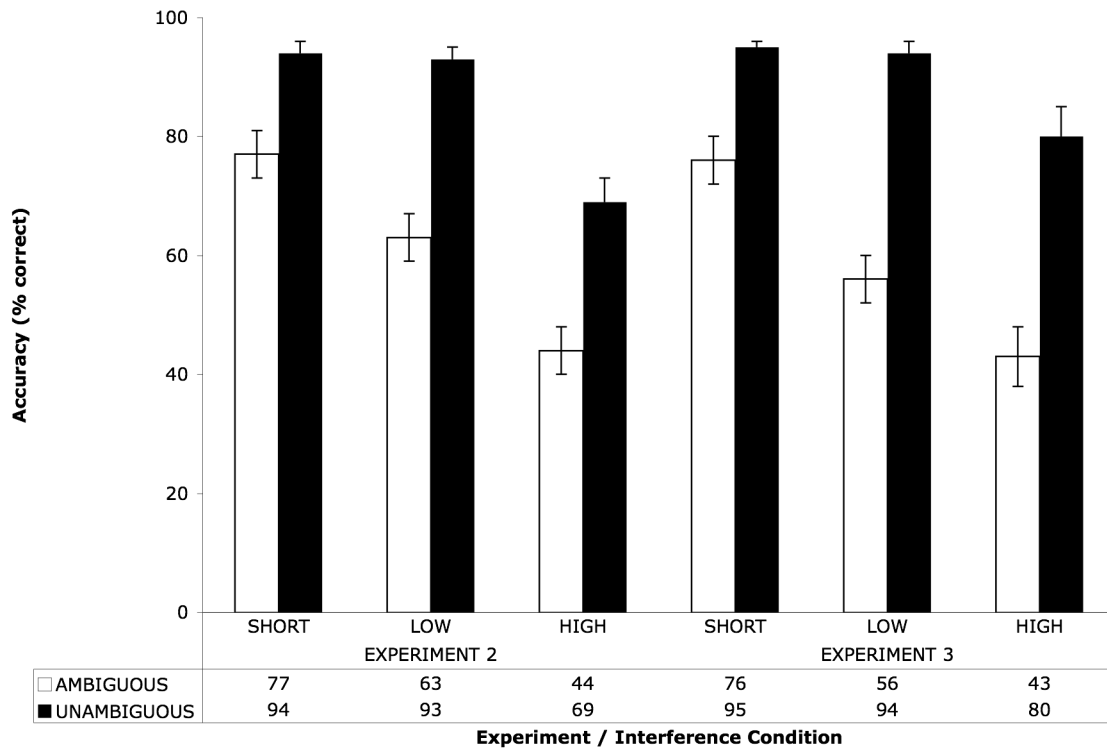


Figure 3-10 Van Dyke & Lewis (2003), Experiments 2 & 3

Accuracy is expressed as percentage of experimental sentences judged grammatical. $N_{SUBJ} = 36$. $N_{ITEM} = 36$.

These materials however contain a serious confound, one which we believe potentially short-circuits the conclusion that it is the presence of a subject-like constituent in the high interference condition that leads to difficulty: the high interference conditions contain one more clause than the low interference conditions.³² Those two conditions are repeated below for comparison, with the intervening clauses bracketed.

³² There are other differences in the example materials given, though it is unknown how representative those materials are of the entire set. The information in the High interference RC is more weakly related to the clause it is contained in. For example, in the Low interference condition, the student was standing in the hallway putatively because she was waiting for the exam. However in the High interference conditions the fact that the student knew the exam was important seems to characterize her independently of the other events she participates in. The Low interference condition is also more imageable. Thanks to Colin Phillips for these observations.

- (76) The executive assistant forgot that
(a) HIGH INTERFERENCE
the student [_{CP} who knew [_{CP} that the exam was important]] was standing ...
(b) LOW INTERFERENCE: Complex subject, no lexically-filled subject position
the student [_{CP} who was waiting for the exam] was standing ...

The region that is supposed to be sensitive to interference, ‘was standing’, occurs at the boundary of two clauses in the high interference conditions, but only one clause in the low interference condition. There are several ways in which this might impact the results. A simple possibility is that grammaticality ratings are a monotone decreasing function of the number of clauses contained in a sentence, or particularly the number of clauses that are not right-embedded. One can imagine a non-retrieval based alternative, in which the subject must be actively maintained until a grammatically-licensed verb is encountered; as the number of subject-verb relationships, or the number of clauses, that intervene increases, the maintenance of this subject may be more error-prone (in the spirit of Wanner & Maratsos, 1978). Moreover, for the reading time results, there may be increased processing difficulty incurred by leaving a doubly embedded clause and returning to the main clause, as opposed to leaving a singly embedded clause. This difficulty would be reflected in the Aux + V region, which signals the clause boundaries, rendering comprehension more error prone in that part of the sentence. Other Spill-over effects are imaginable, such as the difference in the complementation structure of the verbs in the high vs. low interference conditions. Those concerns all center on the high-interference region inducing more difficulty in the course of processing. However even if we put aside those concerns, we can view the results of Van Dyke & Lewis (2003) as reflecting a type of cue-independent interference effect stemming from the presence of an extra clause. Locating any constituent may be more difficult when there are more domains to locate that constituent in.

Van Dyke (2007) goes some way towards addressing the clause number confound. In one of the experiments, participants read sentences similar to Van Dyke & Lewis (2003)'s materials in an eye-tracking experiment, except that in the new experiment an adverbial spill-over region was inserted at the clause boundary. Although this manipulation does not address all of the concerns raised above, it has the potential to relieve some difficulty in the critical region. The double clause boundary problem does still remain, since the adverbial seems to most naturally attach inside the RC. Nonetheless this experiment is worth considering, both to compare how it replicates Van Dyke & Lewis (2003), but also because a second kind of interference manipulation was included: a semantic appropriateness manipulation. In the low semantic interference conditions, the grammatically-inaccessible DP inside an intervening relative clause was manipulated for its plausibility as subject of the critical verb. This semantic interference condition was crossed with the same kind of syntactic interference condition in Van Dyke & Lewis (2003). A sample set of materials is given below:

- (77) The pilot remembered that the lady ____ yesterday afternoon moaned about a refund for the ticket
- (a) LOSYN/LOSEM: Low Syntactic Interference / Low Semantic Interference
who was sitting in the smelly seat
 - (b) LOSYN/HISEM: Low Syntactic Interference / High Semantic Interference
who was sitting near the smelly man
 - (c) HISYN/LOSEM: High Syntactic Interference / Low Semantic Interference
who said that the seat was smelly
 - (d) HISYN/HISEM: High Syntactic Interference / High Semantic Interference
who said that the man was smelly

HISYN conditions included a gapped subject position and a lexically filled subject position embedded inside the intervening relative clause. LOSYN condition RCs were monoclauses which had a single gapped subject position, locating the only overt DP in an

oblique position. In HiSEM conditions, the DP was a good subject for the critical verb, whereas in the LOSEM condition, it was not. With respect to the example given, it is not difficult to conceive of men moaning, but it is highly atypical for seats to moan. On the assumption that the verb supplies a cue relevant to the property of the DP that underlies the sensibility or typicality of the relation, e.g., ANIMACY:ANIM, then the HiSEM verbs pointed ambiguously to the grammatically-licensed subject and the inaccessible one. Both HiSYN and HiSEM conditions provided partial support for the grammatically inaccessible constituent; and in the HiSYN/HiSEM condition these partial cues converged. Consequently, there should be a cline of erroneous processing. Figure 3-11 reports the results of three eye-tracking measures in the critical Aux+V region: first pass, regression path and total reading time.

Reading times in the 'Aux + V' region

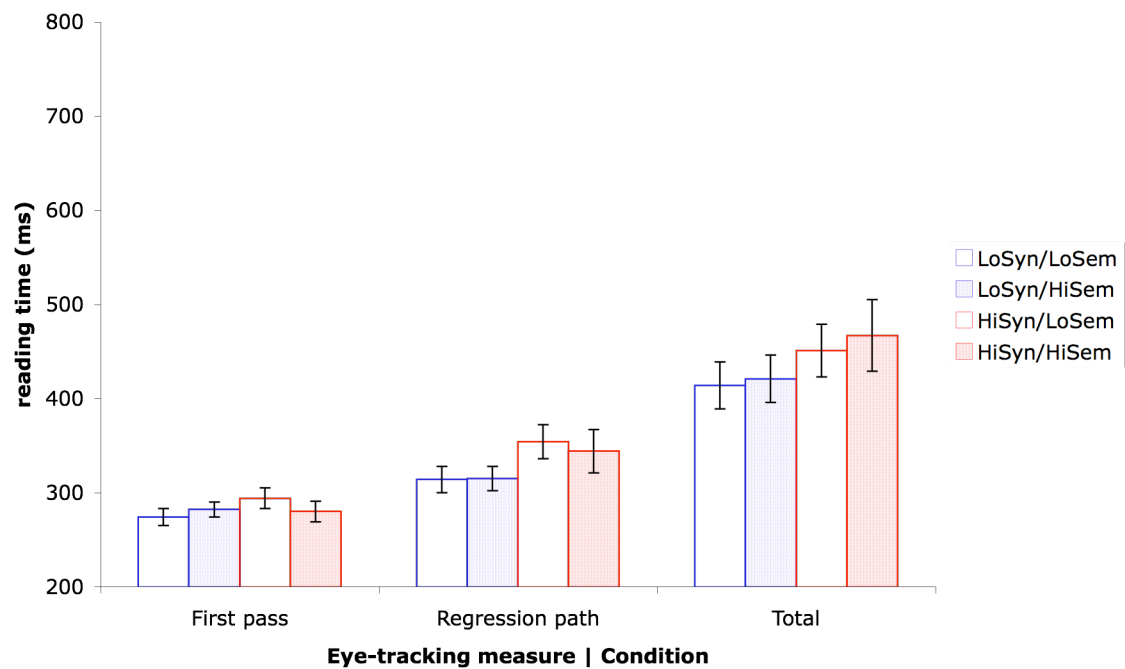


Figure 3-11 Van Dyke (2007) Experiment 3, Critical region reading times

In the first-pass and regression path reading measures, there was an effect of syntactic interference, but only in LOSEM conditions. Since first pass measures reflect the total duration of fixations before any left-ward or right-ward movement from the region, those results are encouraging for the retrieval account. When the dependent measure does not incorporate re-reading of the syntactic context, then we can most confidently assert that the RT measure reflects properties of how the comprehender is consulting constituent representations in memory. In total time, syntactic interference showed a marginal effect overall. No effect of semantic interference was observed until several words downstream, in the sentence final region, corresponding to “for a ticket” in the example set above. There the effect was strongest in the HiSYN condition. These results are consistent with the findings of Van Dyke & Lewis (2003) in the sense that a slowdown that may reflect syntactic interference was observed in the conditions that most closely matched the ones in their experiment (where the semantic fit was low).

What is troublesome is that the adverbial spill-over region showed strong effects of syntactic interference as well. Those data are given in Figure 3-12 below. No significant effects of syntactic interference were observed in first pass measures, but in the regression path measures there was a slowdown in HiSYN/HiSEM conditions. This effect suggests that any difficulty caused by the HiSYN conditions need not necessarily be tied to retrieval processes at the verb or errors of subject-verb binding. This effect may be consistent with the non-specific interference conjecture we offered above for why intervening biclausal RCs cause difficulty. Alternatively, there was also an attachment

ambiguity in just the HiSyn conditions which may explain this effect, though it is unclear why it only shows up in HiSem conditions³³.

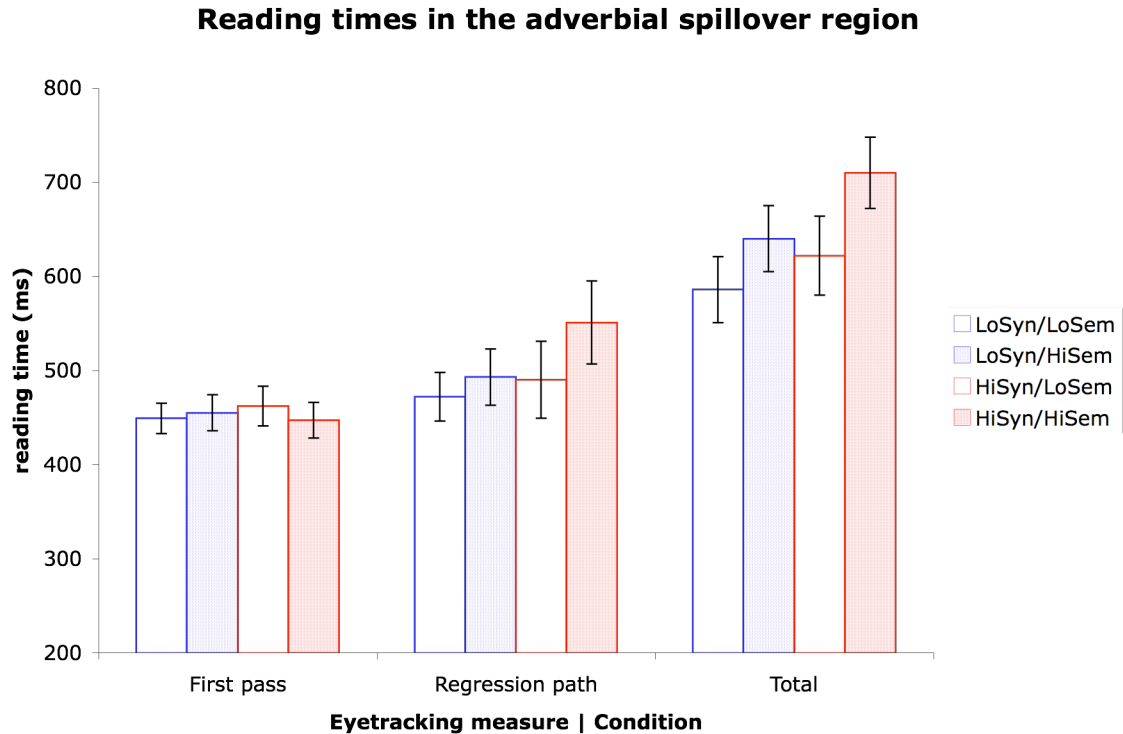


Figure 3-12 Van Dyke (2007) Experiment 3, Pre-critical region reading times

Notably in the online measures there was not a cline of difficulty localized to the verb, with HiSyn/HiSem conditions being the most difficult to process. However, in the offline measures, this pattern was apparent. Van Dyke (2007) reports cloze comprehension measures for the reading time experiments. Following each sentence participants were presented with an open frame like “the _____ moaned about a refund for the ticket,” and then given a three-alternative forced choice task to fill the blank:

³³ The difficulty of the attachment could be determined by properties of the predication. Compare “the lady said the man was smelly yesterday afternoon,” and “the lady said the seat was smelly yesterday afternoon.” In both cases, “yesterday afternoon” could refer to the event time of the bounded ‘saying’ event. Only in the case were ‘the man’ is subject, however, does the downstairs attachment seem felicitous. It seems more typical, that a man might be smelly in a time period bounded by ‘yesterday afternoon,’ than a seat.

{‘pilot’, ‘lady’, ‘man’}. The set of nouns included the grammatically correct subject, the matrix clause subject, and the relative clause DP from the HiSem conditions. The same three nouns were used for each condition since in the LoSem conditions, the relative clause DP could be rejected based on plausibility at test. This task potentially taps more directly into the interpretive outcome of reading. Results of this task are presented in Table 3-13 below.

	<i>Syntactic Interference</i>	
<i>Semantic Interference</i>	Low	High
Low	85%	77%
High	77%	66%

Table 3-13 Comprehension accuracy from Van Dyke (2007) Experiment 3

The HiSYN/HiSEM condition was the most error prone, the LOSYN/LOSEM conditions were the least error prone, and the two Hi*/LO* conditions were in the middle. Both main effects of syntactic interference and semantic interference were reliable. Interestingly, Van Dyke (2007) provided a break-down of error responses. Since the comprehension task was forced choice, participants were either erroneously choosing the matrix subject or the HiSEM DP. The break-down is given as an experiment-wide tally in Table 3-14. When the distractor noun occurred in the subject position (HiSYN), participants chose it in 57% of error responses, compared to 45% in LOSYN position. This shift in proportion was not huge, though it was reliable. It is interesting to note that in many of the HiSEM error trials (on average, 48%), the participants nonetheless choose the matrix subject.

Interference condition	Matrix subject	HiSEM Distractor
LoSYN/LoSEM	41	10
LoSYN/HiSEM	44	36
HiSYN/LoSEM	66	15
HiSYN/HiSEM	52	68

Table 3-14 Erroneously chosen nouns in Experiment 3 cloze comprehension task

Overall the error data suggest that as interference from syntactic and semantic factors increased, the comprehender was more likely to choose a grammatically illicit constituent in the comprehension task. The breakdown of errors shows that the comprehender was more likely to choose a constituent inside the relative clause when it has subject-like features.

3.4.3 *Replicating and extending Van Dyke (Experiment 6)*

We attempted to obtain a subject interference effect in a set of materials similar to Van Dyke & Lewis (2003) and Van Dyke (2007), but one which did not include the clause confound. The data from both of those previous experiments clearly show that the sentences classified as inducing high syntactic interference were harder to process and derive a correct interpretation from. However the cause of the interference remains doubtful because the high interference conditions both contained lexical subjects, but also contained two clauses. In a self-paced reading experiment we tested sentences that manipulated the structural subject-hood of a distracting DP, without introducing extra clauses.

3.4.3.1 Materials and methods

Materials. We constructed sentences that contained a sentential complement taking noun like ‘report.’ The complement of ‘report’ was a copular sentence, either with a lexical projection in [Spec,TP], or the expletive-associate version of that sentence. By hypothesis, [Spec,TP] conditions contained a more subject-like DP constituent than Expletive-Associate conditions. The DP is in the structural position associated with subject in the [Spec,TP] position, in contrast with the Expletive-Associate condition³⁴. These were compared to two types of controls: one in which there was no sentential complement immediately after the head noun, and one in which the sentential complement was replaced by a PP that expressed the same thematic relations. Consequently interveners were either +/- Clausal Complementation, and +/- Filled Subject Position. A sample set is given below:

- (78) The politician was displeased that the report ...
- (a) LEXICAL SUBJECT IN [SPEC, TP] +CLAUSE, +SUBJPOSITION
that support was widespread for her opponent was covered on the evening news.
 - (b) EXPLETIVE-ASSOCIATE +CLAUSE, -SUBJPOSITION
that there was widespread support for her opponent was covered on the evening news.
 - (c) PP COMPLEMENT -CLAUSE, -SUBJPOSITION
of widespread support for her opponent was covered on the evening news.
 - (d) CONTROL/NO INTERVENERS
was covered on the evening news that support was widespread for her opponent.

³⁴ Admittedly the case properties of the DP in Expletive-Associate construction are somewhat murkier. The DP might bear oblique case (cf. Burzio, 1988; Chomsky 1995), though it might also bear nominative case which is checked covertly. Associates do have many subject-position properties; e.g., they control agreement on the verb:

- (i) There was/*were widespread support ...
- (ii) There *was/were widespread rumors ...

The critical regions in each case were the auxiliary and verb: “was covered.” In the three intervener conditions, (a)-(c), the critical region was preceded by the same three-word PP. In the control non-intervener condition, the sentential complement was extraposed beyond the critical VP, so that all sentences had (approximately) the same interpretive content. 24 item sets were created, each with these four conditions.

Each item set was associated with a yes/no comprehension question. The content of the comprehension questions was designed to target information derived from different portion of the sentence, to test for selective deficits in comprehension of these sentence types. 8 item sets had comprehension questions that targeted information obtained from the matrix portion of the sentence. For example, with respect to the sample materials set, “Was it a politician who was displeased?” 8 item sets had comprehension questions that targeted information within the sentential complement. For example, “Was there much support for the politician’s opponent?” 8 item sets had comprehension questions that targeted information requiring the correct pairing of subject and embedded verb. For example, “Was it a report that was on the evening news?”

Participants, procedure and analysis. Protocols and analysis methods are identical to Experiments 1 and 2 reported in Chapter 2. Analysis of variance was not performed, and was replaced exclusively by estimation and simulation of linear mixed-effects models. The regions of interest were Regions 8-12, which correspond to the variable intervener phrases; Regions 13-15, which correspond to the PP common to all interveners; Region 16, the critical auxiliary; Region 17, the critical verb, and Regions 18-21, the VP spillover regions.

Participants were 36 members of the University of Maryland community, who received partial credit in an introductory linguistics course.

3.4.3.2 Comprehension accuracy results

The comprehension results are given in Table 3-15 as percentage correct. Overall accuracy was high, 86%. Control sentences, which had no interveners, had 90% accuracy. Both [Spec,TP] and PP Complement conditions led to significantly lower accuracy (by logistic regression, $p < 0.001$ and 0.05 , respectively), though not Expletive-Associate conditions (n.s.).

Condition	Question-type			
	MatrixSubj	EmbedSubj	EmbedPred	
[Spec,TP]	80%	88%	83%	68%
Exp-Assoc.	88%	99%	83%	81%
PP Comp.	85%	96%	85%	74%
Control	90%	94%	90%	86%
Overall	86%	Overall 94%	85%	77%

Table 3-15 Comprehension accuracy for Experiment 6

Between items there was significant effect of question type. Participants were more accurate on items followed by a question about the matrix subject than either a question about the embedded subject ($p < 0.05$) or about the embedded predicate ($p < 0.001$). We can consider performance on the experiment conditions classified by question type. For questions about the matrix subject, participants did best in the Expletive-Associate and PP Complement conditions, and worst on the control or [Spec,TP] condition. For questions about the embedded subject, there were no reliable differences across conditions. Performance was most degraded on the embedded predication questions, which were written to require that the embedded subject and verb be correctly paired. Performance was best on the control and Expletive-Associate conditions, but worst on the [Spec,TP] and PP Complement conditions. Keep in mind that these

comparisons are not ideal, because they are between items, and so do not counterbalance possible lexical effects. However they do show consistently degraded performance for [Spec,TP] conditions, and particularly in the embedded predicate questions.

3.4.3.3 Self-paced reading results

The results of the self-paced reading task are presented in Figure 3-13. Results are only graphed starting at Region 6 (the determiner of the embedded clause subject). No reliable differences were observed in prior regions (the matrix clause prefix).

Regions 8 – 12: Variable intervener regions. For analysis, word-by-word reading times from the intervener regions were collapsed into one region of interest, excluding the common PP. On average the [Spec,TP] condition was read most slowly. In pairwise comparisons it was reliably slower than the expletive-associate condition ($\Delta\mu$: 17ms, 95% C.I.: [0, 31 ms], $p < 0.05$), though not reliably slower than the PP complement condition ($\Delta\mu$: 12 ms, 95% C.I.: [-3 ms, 29 ms], n.s.).

Regions 13 – 15: Common PP regions. No reliable differences were observed between Conditions in Regions 13, 14, or 15. There was no spill-over of differential difficulty from the previous regions.

Region 16: Critical auxiliary. The control condition was slower than all intervener conditions combined ($\Delta\mu$: 17 ms, 95% C.I.: [3 ms, 30 ms], $p < 0.05$). Crucially no differences were observed in the different intervener conditions: a model with only an intercept (d.f.: 3) captured the variation as well as one with condition coefficients (d.f.: 5; $\chi^2 = 0.7$, n.s.).

Region 17: Critical verb. The control condition was numerically slower than all intervener conditions, but this effect was not reliable statistically. There was no

difference observed in the different intervener conditions, or across all conditions (by model comparison; $\chi^2 = 0.1$, n.s.)

Regions 18-21. No differences between conditions were observed in Regions 18 – 19. In both Regions 20 and 21, the intervener conditions were read reliably more slowly than the control condition (Region 20: $\Delta\mu$: 23 ms; 95% C.I.: [5 ms, 44 ms], $p < 0.05$; Region 21: $\Delta\mu$: 78 ms; 95% C.I.: [57 ms, 100 ms], $p < 0.01$). No differences were observed among intervener conditions in Region 20. However, in Region 21, the clausal complement conditions ([Spec,TP], Expletive-Associate) were read more slowly than the PP complement conditions ($\Delta\mu$: 27 ms; 95% C.I.: [5 ms, 51 ms], $p < 0.01$). However, there was no reliable difference between the two clausal complement conditions ($\Delta\mu$: 3 ms; 95% C.I.: [-24 ms, 34 ms], n.s.).

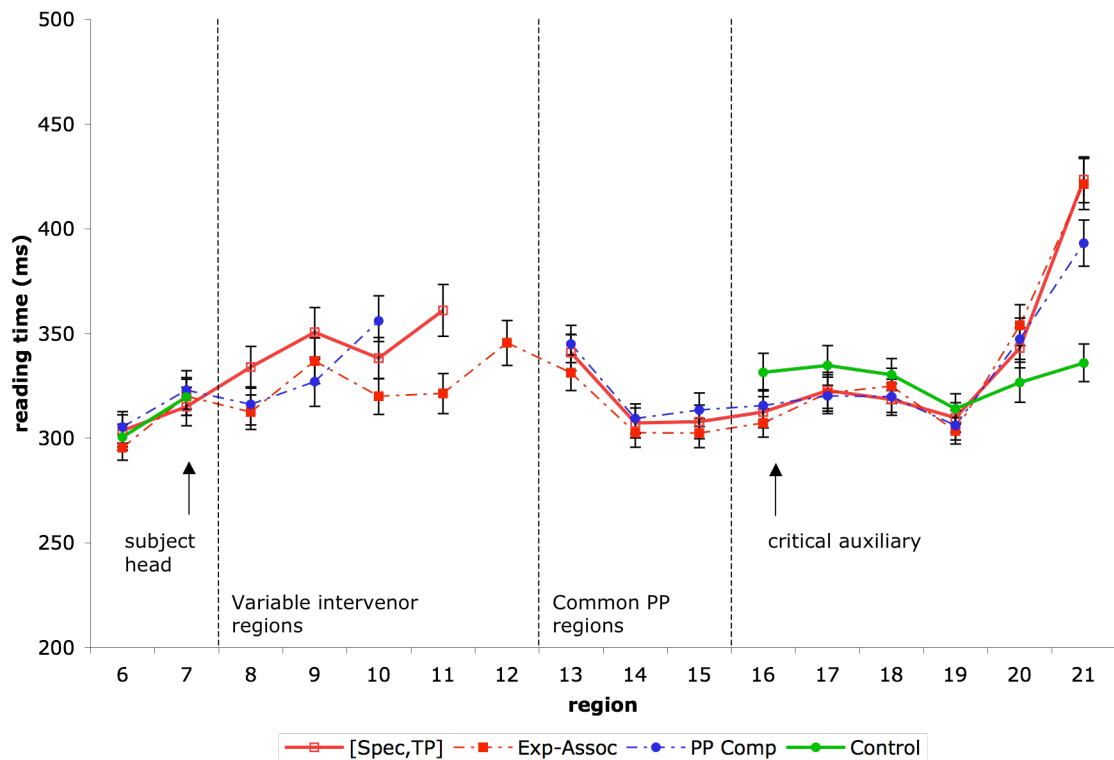


Figure 3-13 Experiment 6 Self-paced reading results

3.4.3.4 Discussion

The results of this experiment confirm that attaching material to a subject head affects the processing difficulty of a following VP. The online and offline results pull in two directions. The offline results are similar to Van Dyke's in the sense that the condition containing the most subject-like DP, the [Spec,TP] condition, led to the greatest decrements in performance. Consistent with Van Dyke's claim that a grammatically-inaccessible subject is retrieved and *interpreted* as the subject, we found that [Spec,TP] conditions lead to the worst performance on comprehension questions that require the embedded subject and predicate to be correctly paired (68%). For questions that required accurate processing of the intervener region itself, there were no differences among conditions. Therefore it seems that there were no differences in success at comprehending the intervener itself, which would correlate with the other deficits. [Spec,TP] conditions were processed the most slowly, as the online record reveals. In the critical region, the online results showed differences among the intervener conditions, which were all read more quickly than the non-intervener control. However, differences among the intervener conditions show up in the sentence-final regions, regions 20 and 21. There the clausal complement interveners were read most slowly, followed by the PP complement condition, and then the non-intervener condition. The difference with the non-intervener control can be attributed to the fact that the non-intervener condition is not sentence-final in those regions: recall the sentential complement of the embedded subject head was extra-posed to control for similar interpretations across materials.

In the online results, the only contrast we can claim is one having to do with whether or not the intervener is dominated by CP or not. These results support our contention that the differences Van Dyke & Lewis (2003) and Van Dyke (2007) obtained

were confounded by the extra clause in the high interference condition. Nonetheless the robust pattern obtained in the offline measures, in both Van Dyke's experiments, and our own remains to be explained. We can offer one explanation with Van Dyke's general viewpoint, which is that representations that interfere during a retrieval operation need not lead to online difficulty, measured in reaction times. We can suppose that the grammatically-inaccessible subject is a nearly good match for whatever retrieval structure the embedded verb provides. Some proportion of the time it selects that subject. If no information contained within the inaccessible subject signals that the subject is grammatically illicit, then the processing system proceeds as if it has constructed the correct representation. Consequently, there would be no observed reaction time difference, but only a difference in measures of interpretation. This explanation may account for why Van Dyke (2007)'s earliest online measures only showed an effect of syntactic interference for subjects that were LOSEM, if semantic fit counts as a signal for a good subject or not. However, it is important to recognize that while evidence for difficulty in "High interference" conditions seems clear, the evidence that actual mis-binding occurring in "High interference" is scant.

The results of Van Dyke's experiments, and our own, remains a somewhat mixed bag. The processing cost induced by clausal interveners seems clear enough, though its localization remains imprecise. What is least clear is that similarity-based interference is to blame for decrements in performance. One proposal for future research is to use a clearer case contrast, to see whether that yields more distinctive results. Given the effect of case in the agreement attraction research, this contrast seems promising. For example, the use of exceptional case marking constructions (ECM) could be used to create

interveners that have a DP in a subject position, yet bear inappropriate case for an agreeing subject. In the following paradigm, the intervening DP, in bold font, is increasing less subject-like based on whether case is Nominative or not, and whether it is explicitly marked in the input.

- (79) (a) *Subject position of a finite clause, Nominative, Lexically Case ambiguous*
The man [who believes **John** is foolish] was unsurprised by his behavior.
- (b) *Subject position of a non-finite clause, Accusative, Lexically Case ambiguous*
The man [who believes **John** to be foolish] was unsurprised by his behavior.
- (c) *Subject position of a non-finite clause, Accusative, Lexically Case unambiguous*
The man [who believes **him** to be foolish] was unsurprised by his behavior.

Similarity-based interference predicts a cline of difficulty, with (a) being the most difficult and (c) the least.

3.4.4 NPI Licensing v. Reflexive Anaphora

One problem with interpreting the data on complex subject attachment is that, unlike in subject-verb agreement, the link between the theory and the measures provided by reading times or comprehension accuracy is not as tight. In the case of subject-verb agreement, the patterns of difficulty could clearly be attributed to a discrete property of the stimulus: whether or not the verb matched the head of the subject. In the case of the complex subject attachment, the properties of the interference manipulation are more indirectly related to whatever processes take place at the auxiliary. Moreover in the case of agreement attraction, a putative interference manipulation could also predictably improve processing complexity grammaticality ratings. That is, in the agreement attraction experiments we reported, the manipulation did not change performance only by

making the comprehension tasks more difficult, but also by making comprehension easier.

Other claims of interference can be found for complex subjects. Most notably Negative Polarity Item (NPI) licensing has been found to be sensitive to a grammatically-inaccessible licenser found inside a complex subject (Drenhaus et al., 2005, Xiang, Dillon, & Phillips, submitted). NPIs are lexical items or constructions like ‘any’, ‘ever’, or ‘lift a finger’ that can only occur in certain semantic contexts: for example, under negation, in conditionals, or in rhetorical questions. The contrast in (80) illustrates the basic phenomenon with ‘ever.’ Only in (80a), when ‘ever’ occurs in the domain of negation, is the sentence acceptable.

- (80) (a) No candidate will ever apologize for slander
(b) *A candidate will ever apologize for slander.

It is not sufficient for negation to be present in the sentence. In (81), the negation-bearing element is embedded inside a relative clause, where it does not c-command ‘ever,’ and the resulting sentence is unacceptable.

- (81) *The candidate that no pundit likes will ever apologize for slander.

However, Drenhaus and colleagues (Drenhaus et al., 2005) have shown that the mere presence of an NPI licenser, even if it not in the right structural relationship with the NPI, can improve perception of acceptability in German sentences. For example, in a speeded acceptability task, they tested sentences with the German NPI ‘jemals’ (‘ever’). A negative licenser (‘kein’) was either present in the c-commanding subject position (82a), present but embedded in a subject-attached relative clause (82b), or absent (82c). In the sentences below, the NPI is in bold font, and the negative element is underlined. The accuracy rates in the acceptability task are given to the right.

- | | | | |
|------|-----|--|-----|
| (82) | (a) | <u>Kein</u> Mann, der einen Bart hatte, war jemals glücklich.
“No man who had a beard was every happy.” | 85% |
| | (b) | *Ein Mann, der <u>keinen Bart</u> hatte, war jemals glücklich.
“*A man who had no beard was ever happy.” | 70% |
| | (c) | *Ein Mann, der einen Bart hatte, war jemals glücklich.
“*A man who had a beard was every happy.” | 83% |

Drenhaus et al. (2005) found that participants were equally accurate in classifying (82a) and (82c), but were slower and less accurate in judging (82b). Crucially (82b) is the case in which the licenser is present but in a grammatically-inaccessible position. The basic finding seems quite robust, and has been replicated in ERP measures and reading times, both in German and in English (Vasishth et al., in press; Xiang, Dillon & Phillips, 2006, Xiang; Dillon & Phillips, submitted). Vasishth and colleagues (Vasishth, Drenhaus, Saddy, & Lewis, 2005; Lewis, Vasishth, & van Dyke, 2006) have argued that the intrusion of the grammatically-inaccessible licenser proceeds under a regime of cue-based retrieval, just as the inaccessible subject is claimed to have had its intrusive effect in Van Dyke’s studies. They have hypothesized that encountering the NPI initiates a search controlled by two kinds of cues: a semantic cue for negation and a syntactic cue for a c-commanding position. In grammatical sentences like (82a), there is a full match; in ungrammatical sentences with no negative element, like (82b), no constituents match the cues. However in sentences like (82c), there is a negative element, just not in a c-commanding position. The cues partially match with a constituent in this case, and so in some portion of cases, the NPI is mistakenly licensed³⁵.

³⁵ As we discussed in section 3.2.1.3, there is no way to encode the c-command property on an individual node, since c-command is a relational property. Therefore, for Vasishth et al. to include c-command in the retrieval structure, they must resort to a heuristic cue like [+Nom] or [+Subj]. In the German examples, the intrusive negative element occurred in an embedded object position, and so it would not be contacted by this cue, only the negation cue.

The logic of this explanation for NPI licensing illusions is identical to our own for agreement attraction. Two cues are provided by the constituent that needs to be licensed: one points to a grammatically accessible constituent, but one which does not have the licensing features, and the other points to a grammatically inaccessible constituent, but one which does have the licensing feature. However there are crucial differences that break the analogy. Xiang, Dillon & Phillips (submitted; henceforth XDP) have argued that the cue-based retrieval account is undermined by the fact that NPI licensing is not an item-to-item dependency. As they point out, there is a consensus that NPI licensing reflects the interaction between specific lexical properties of the NPI and the semantics and pragmatics of propositions (Chierchia, 2006; Fauconnier, 1975; Israel, 1998; Kadmon & Landman, 1993, Krifka, 1995; Ladusaw, 1992, *inter alia*). It is not a relationship between two items in a c-command configuration. Two sets of their examples illustrate this point nicely. In (83) the NPI ‘ever’ is licensed, despite there being

For NPIs occurring inside a VP, the subject is one prominent position in which a licenser could occur. However a cue for subjects will not be generally applicable. For example, a c-commanding licenser can occur as a VP-internal arguments, as in (i), as VP negation (ii), or as the verb itself in the case of adversative predicates:

- (i) It occurred to nobody that Mary would **ever** write a sonnet.
- (ii) Mary did not wish to **ever** write a haiku.
- (iii) Mary outright refused to **ever** write a villanelle.

The point is moot for the case of NPIs since, as we discuss in the text, NPI licensing does not have a direct c-command requirement. The c-command relationship is a by-product of the requirement that NPIs occur in a downward entailing environment (Ladusaw, 1979).

More generally though the use of a Subject cue illustrates the problem with addressing constituents in a structure-sensitive manner. It is in principle impossible to enforce a c-command requirement directly. In the case of agreement attraction, this problem is less dire, since the licensing of agreement truly occurs between a verb and the nominative-marked element in subject position, and not any possible c-commander.

no obvious lexical licenser in the sentence. In (84) ‘ever’ is licensed in sentence in which there is a negative item, ‘nobody,’ but one which does not c-command it.

- (83) (a) Has John **ever** cleaned his own dishes?
(b) The reason one **ever** bothers to decant a wine is to leave the sediment [...] behind in the bottle [*SouthWest Airlines Spirit* August 1994: 47; reported in Israel, 1998].

(84) Nobody’s mother has **ever** complained about his grades

NPIs are not directly licensed by c-commanding negation. One prominent analysis, Ladusaw (1979), holds that NPIs are instead licensed in downward entailing environments (though this still does not capture licensing in polar questions). What seems like a c-command requirement for many cases is not itself a licensing condition, but rather a by-product of being in particular downward entailing environments. Xiang and colleagues propose a theory of illusory licensing in which the contrastive function of the restricted relative clauses (which have been used in all NPI licensing experiments to date) can lead to the generation a pragmatically-sensible negative inference. It is this inference that comprehenders may use to license the NPI³⁶. What is relevant for our purposes is the comparison they go on to draw between NPI licensing and reflexive anaphora.

Reflexive anaphora is more plausibly licensed in comprehension as an item-to-item dependency. In English, reflexive anaphors like ‘himself’ or ‘herself,’ must occur in the presence of a c-commanding antecedent, as the comparison between (85a) and (85b) shows; and in the same local domain, as the comparison between (86a) and (86b) shows.

³⁶ XDP propose that, given a subject like (i), a comprehender infers that the set of individuals denoted by (i) has some property P (not yet expressed). However, because of the use of the restrictive relative, they might also consider a contrast set, expressed by the string in (ii), of which the speaker may have intended to assert that P does not hold, i.e. $\neg P$. If comprehenders generate the inference $\neg P$ for the contrast set, then that may have been used to license the NPI (cf. Israel, 2004)

- (i) The bills that no democratic senators have voted for ...
(ii) The bills that some democratic senators have voted for ...

- (85) (a) *John_i's mother bought a book for himself_i
 (b) John_i bought a book for himself_i
 (86) (a) *John_i wished that Mary would buy a book for himself_i
 (b) John wished that Mary_i would buy a book for herself_i

This restriction can be expressed syntactically (e.g., as Principle A, Chomsky, 1981) or semantically (e.g., Jackendoff, 1992). Unlike NPI licensing, it truly seems to depend on a structurally-conditioned relation between the two constituents that occur in the sentence³⁷. The question arises whether licensing reflexive anaphora would be subject to intrusion in just the same configurations that NPI licensing is. XDP reasoned that since an anaphoric dependency is more plausibly item-to-item in the case of reflexives, it would constitute a stronger test of cue-based retrieval.

The resolution of anaphora in comprehension can be tested by gender or number feature matches between an anaphoric element and candidate antecedents. A common way of assessing which antecedents the comprehender is entertaining is to use a stereotypical gender manipulation. Certain nouns, like ‘soldier,’ are associated with a stereotypical gender and this association is strong enough to lead to a processing disruption when they are identified as the antecedent of a reflexive which does not match that gender (Osterhout et al., 1997; Sturt, 2003). For example, in (87b), the anaphor ‘herself’ must be coreferent with ‘the tough soldier.’ However there is a processing disruption in resolving the anaphora, compared to (87a), because the stereotypical gender of ‘soldier’ is male.

- (87) (a) The tough soldier introduced **himself** to all the nurses.
 (b) The tough soldier introduced **herself** to all the nurses.

³⁷ There is another class of reflexives, the logophors, which is sensitive to discourse relations (e.g., Pollard & Sag, 1992; Reinhart & Reuland, 1993). We do not discuss these.

Sturt (2003) tested configurations containing two potential antecedents, one of which was grammatically accessible, and one of which was not. In two eye-tracking experiments he crossed grammatical accessibility with the gender match. (88) illustrates the design from his Experiment 2, which is most comparable to the NPI experiments.

- (88) (a) *Grammatically Accessible Match / Gram. Inaccessible Match*
The surgeon who treated Jonathan had pricked himself with a used needle.
(b) *Grammatically Accessible Match / Gramm. Inaccessible Mismatch*
The surgeon who treated Jennifer had pricked himself with a used needle.
(c) *Grammatically Accessible Mismatch / Gramm. Inaccessible Match*
The surgeon who treated Jennifer had pricked herself with a used needle.
(d) *Grammatically Accessible Mismatch / Gramm. Inaccessible Mismatch*
The surgeon who treated Jonathan had pricked herself with a used needle.

In this design the inaccessible antecedent is the name inside the subject-attached relative clause. If the reflexive anaphor initiates a search for an antecedent with both structural cues (i.e., c-commands) and morphological cues (i.e., masculine or feminine), then the grammatically inaccessible constituent should exhibit a partial match in conditions (a) and (c). Moreover, in Condition (c), in which the grammatically accessible constituent mismatches, the inaccessible constituent should be the only match. Consequently, there would be intrusion, as in the NPI cases. However Sturt found that in early eyetracking measures there was only an effect of match in grammatically accessible conditions. The feature match of the inaccessible constituent had no impact. He did however find an effect of the inaccessible constituent in second-pass and later region measures.

Nonetheless it would seem that the initial resolution of anaphora is faithful to the structural constraints on binding. This finding is inconsistent with the prediction that an inaccessible constituent is sometimes considered because of a partial match. In a separate experiment he replicated these results for a configuration in which the inaccessible constituent is a higher clause:

(89) He/she remembered that the surgeon had pricked himself/herself.

The findings thus generalize across both requirements for reflexive antecedence: that it c-command, and be in the local domain.

Xiang, Dillon, & Phillips (submitted) conducted a similar study to Sturt's, in which they measured evoked response potentials (ERPs). The interesting innovation in their design is that they also had participants read NPI sentences in the same experiment. They could therefore directly compare the relative response to intrusive NPI conditions and the relative response to intrusive antecedents within the same participants. The experimental design for anaphora conditions is illustrated in (90). In congruent conditions, a gender-matching antecedent occurred in subject position. In intrusive conditions, a gender-matching antecedent occurred in a grammatically-inaccessible embedded position. Finally, in incongruent conditions, there was no gender-matching antecedent.

(90) **XDP's reflexive anaphora sentences**

(a) *Congruent*

The tough soldier that Fred treated in the military hospital introduced **himself** to all the nurses.

(b) *Intrusive*

The tough soldier that Katie treated in the military hospital introduced **herself** to all the nurses.

(c) *Incongruent*

The tough soldier that Fred treated in the military hospital introduced **herself** to all the nurses.

The design of the NPI sentences mirrored these conditions (and the previous experiments discussed above) and is given in (91).

(91) **XDP's NPI sentences**

(a) *Grammatical*

No restaurants that the local newspapers have recommended in their dining reviews have **ever** gone out of business.

- (b) *Ungrammatical/Intrusive*
The restaurants that no local newspapers have recommended in their dining reviews have **ever** gone out of business.
- (c) *Ungrammatical/No licensor*
Most restaurants that the local newspapers have recommended in their dining reviews have **ever** gone out of business.

XDP had participants read these sentences in RSVP presentation while they recorded scalp voltage. They then analyzed the ERPs measured from the onset of the NPI or the reflexive. For both NPIs and reflexives, the ungrammatical conditions exhibited a posterior positivity resembling the P600, an ERP component characteristically evoked in response to syntactic or morphological violations (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown & Groothusen, 1993). In NPI intrusive conditions, the P600 was either absent or greatly attenuated (across electrodes). This is consistent with the idea that intrusive NPI licensors lead the comprehender into an illusion of grammaticality. However, in contrast to the NPIs, the intrusive reflexive conditions were not distinct from the incongruent reflexive ones. In other words, for NPIs the ERP response to intrusive conditions either groups with the good conditions or is in between good and bad conditions. For reflexive anaphora, however, the intrusive condition groups unequivocally with the bad conditions. It therefore seems that whatever mechanism is responsible for intrusion in the case of NPIs is not likewise operating in the resolution of reflexive anaphora.

As we have discussed, there are good formal reasons for suspecting that NPIs and reflexives should be distinct in their licensing procedures. However reflexives seemed like a stronger candidate for a cue-based retrieval resolution mechanism. We would further contend that the reflexive cases are stronger tests than either ours or Van Dyke's complex subject attachment experiments because the link between the measure and the

properties of the materials is clearer. Therefore the real point of comparison seems to be between reflexives and the agreement attraction cases. In agreement attraction cases the verb needs to agree with the subject, yet comprehenders seem unable to reliably target the subject for comparison. In the reflexive cases the anaphor needs to find its antecedent in the a same-clause c-commanding position, also corresponding to the subject. Here comprehenders never seemed led astray. What accounts for the distinction in grammatical accuracy in the two cases?

One potential distinction between the two cases is that the intruder in agreement attraction sentences came from a constituent in the same immediate clause. In the reflexive anaphora cases the potential intruder was in a different (embedded) clause. We have argued that immediate clausal context may be an important cue in the agreement attraction cases (to account for the pattern of judgments in RC attraction; see section 3.3.4). If clause context is used as a restrictor in the search for reflexive anaphors, then the RC-embedded intruder becomes a less-good partial match, becoming even less good the less that cue is weighted. There are good reasons for supposing the clause cue would be important in resolving reflexive anaphora. One is that there is no good way of identifying a c-commander, except by using a heuristic feature like +Subj (see fn. 35). The clause cue may be thus be the only structural licensing cue available to the system. Thus a better comparison with the agreement attraction case would be obtained if a potentially intrusive licenser occurred inside a PP modification, as in (92).

(92) The surgeon for Mary pricked himself/herself with a needle.

The structure of this example is more directly analogous to the agreement attraction cases. Crucially the potential licenser belongs to the same immediate clause in both cases.

Another possibility is that a reflexive does not pick up its antecedent by means of searching for a c-commanding constituent. Instead it could acquire it when it is integrated with the verb. In some frameworks, like Combinatory Categorical Grammar (Steedman, 1997), the reflexive is not treated as independent argument of the verb but as a device that reduces the verb's valency, changing the verb meaning to that of the corresponding reflexive predicate. For example, "The surgeon pricked himself" would essentially mean "The surgeon self-pricked." The procedure for matching the morphological features of the reflexive to the subject may thus be controlled by information contained in the encoding of the verb phrase, instead of by initiating an independent retrieval. Let us suppose that the verb has successfully integrated the subject as an argument, and that this is reflected by a pointer to the actual encoding of the subject constituent. The reflexive could check its morphological features by retrieving the subject via this unique pointer and not on the basis of abstract properties. If this account is on the right path, then we could construct a strong test for intrusion by using reflexives in non-argument positions of the verb, as only argument anaphors can reflexivize the verb³⁸.

3.5 Conclusions

In this Chapter we introduced some key properties of a content-addressable memory and the pitfalls of such an architecture for linguistically-structured representation. We illustrated these properties by examining a number of phenomena involving complex subjects.

³⁸ Thanks to Jeff Lidz for this observation.

We proposed a simple model of agreement attraction based on retrieval-based similarity interference. Using a small set of linguistically-motivated features we were able to achieve consistency across PP and RC attraction phenomena and crucially to capture the asymmetry between grammatical and ungrammatical cases, which the feature percolation model failed at. We tested and confirmed a novel prediction of this model in Experiment 5.

We examined Van Dyke's claims that complex subjects containing subject-like constituents are more difficult to attach. We refined her design in our own Experiment 6. On the whole we found that the online evidence for interference from inside complex subjects was equivocal. In our own data online measures indicated sensitivity to the number of clauses in an complex subject, but not to interference from a subject-like constituent. Offline measures consistently provided a clear indication of greater difficulty in the interference conditions, but they were not so clearly indicative that this difficulty stemmed from the wrong subject being integrated at the attachment site.

We then turned to NPI licensing and the resolution of reflexive anaphora. These phenomena have been tested across the same sentence structures as in Van Dyke's experiments. Evaluation of the phenomena and examination of existing experimental results provides little solid evidence for the occurrence of similarity-based interference in licensing these constructions. However, we also argued that these cases did not provide the strongest test for such effects. In the case of NPI licensing, this was because the licensing procedure did not involve an item-to-item dependency. In the case of reflexive anaphora, we speculated that clause-boundedness or a non-retrieval based mechanism could strongly restrict identification of candidates, and suggested some follow-ups.

In Chapter 4 we turn to the processing of *wh*-dependencies. While granting similarity-based interference a role in online comprehension, we develop a broader account in which predictability is a major determinant of fallibility. We briefly reconsider the attraction data in terms of predictability. However, we present novel evidence on the processing of unbounded *wh*-dependencies, which allows us to defend a link between predictability and interference based on how retrieval structures are composed.

4 Active dependency formation and mechanisms for the accurate recognition of grammatical dependencies

4.1 Introduction

The content-addressable memory architecture outlined in Chapter 3 introduces a potentially significant source of fallibility into online structure building. In the process of recognizing and licensing dependencies the comprehender is liable to establish grammatically-inaccurate relationships. The likelihood of doing so is jointly determined by the extent to which forming a dependency involves cue-directed retrieval of one constituent in the dependency and whether there are other similar constituents in the structure. The reasons that information retrieval cannot be tightly grammatically regulated inhere in the content-addressable memory architecture: search of memory identifies and ranks candidates effectively in parallel on the basis of the similarity between encoded representations and the retrieval structure. Grammatical restrictions that are stated over the relative hierarchical order of constituents cannot be encoded as a feature of item representations, and thus cannot be reflected in the match operation. In two sets of linguistic processes, agreement licensing and complex subject integration, grammatically-irrelevant constituents appear to impact grammatical accuracy, as reflected in online complexity, patterns of judgment, or comprehension accuracy. However, as we will review in greater detail in this chapter, though, linguistic comprehension is highly grammatically sensitive and accurate in many domains, notably

in the processing of *wh*-dependencies. The question we wish to address more broadly is why some processes are characteristically accurate and why some are not.

We have already articulated a mechanism in which inaccuracy seems to come for free. However, we also discussed several possibilities for introducing relative hierarchical order into the search of memory. In the first case, the generation of candidate matches unavoidably identifies grammatically inaccessible candidates; however more controlled processing verifies the structural relationships – for example, by carefully and serially following the inter-item associations – and ultimately delivers a licit constituent. This possibility aligns with debates in psycholinguistics over whether grammatical constraints act as early versus late filters on structure building (e.g., Badecker & Straub, 2002; Sturt, 2003). In a second case, candidates are ordered not only by their similarity to the retrieval structure, but by another analogue value, like activation, which may implicitly recapitulate hierarchical order. This possibility is related to accounts of (non-linguistic) serial order in the list learning literature (e.g. Page & Norris, 1998). We argued that this mechanism may have heuristic value in restricted scenarios, but that global hierarchical order cannot be recovered in any simple way. In both of these alternatives, searching a structured representation overgenerates, and a subsequent selection process must prune. We mentioned a third possibility: if dependency formation is anticipated, then some licensing can occur in the absence of complete information about the head of the dependency and the contents of the retrieval structure can be controlled to identify just the grammatically licit constituent. It is this possibility we will expand upon in this chapter. There are two aspects to this idea: one is that by simply restricting how much retrospective search the system performs, then the system can avoid the intrusion of

grammatically-inaccessible information. Information that can be carried forward in time can license dependencies without having to rely on retrieved information. The second aspect is that, if the system preserves enough distinctive features of the actual constituent-to-be-retrieved, then it can minimize the impact of similarity-based interference when it retrieves the head of the dependency. Our argument advances in four parts:

- (I) first, we glean the basic generalization from the processing of anaphora and comprehension of *wh*-movement dependencies;
- (II) next, we present some experimental evidence from *wh*-processing that constraints on global well-formedness motivate parsing decisions, independently of the local retrieval environment (Experiments 7-8);
- (III) then, we show that the mere presence of formally-similar but irrelevant *wh*-phrases in a structure does not interfere with the retrieval of a target *wh*-phrase (Experiments 9-10).
- (IV) thirdly, we provide more detailed experimental evidence from *wh*-processing that some information about the head of the dependency is preserved across increasing dependency lengths to guide decision making (Experiments 11-13);

4.2 The role of predictability

4.2.1 Forwards v. backward anaphora

In Chapter 3 we introduced one kind of referential dependency, that between a reflexive anaphor and its antecedent. Reflexive anaphora must be resolved in a tightly-constrained syntactic domain and all measures indicate that initial dependency construction is highly grammatically accurate (Sturt, 2003; Xiang, Dillon, & Phillips, submitted). In the studies we reported the reflexive anaphor was never related to a subject

that was not in the same immediate clause. The grammatical accuracy in forming this dependency contrasts with the licensing of NPIs and verbal agreement. Both of these processes seem liable to illusions of grammaticality caused by the presence of grammatically-inaccessible constituent inside of a complex subject. We defended a retrieval-based account of agreement attraction and rejected such an account for NPIs. The formal properties of reflexive anaphora seem to lend themselves to retrieval-based processing but, based on the experimental results, they do not seem to be liable to the same kind of fallibility as agreement is. We considered two conjectures: one, that the boundedness of reflexive anaphora is effective at excluding constituents from all but the immediately dominating clause; two, that anaphora is resolved through the verb, perhaps through a subject-pointer contained in the VP's encoding. Both of these required further experimental support. However we might also consider whether the referential nature of anaphora leads to more careful processing. Putatively nothing goes awry if an agreement violation fails to be noticed³⁹. However fixing the reference of anaphors is crucial for interpreting the sentence. Therefore perhaps referential dependencies are resolved with more deliberate care. We therefore turn to pronominal anaphora to see whether it uniformly grammatically resolved in real-time.

³⁹ Lau, Wagers, Stroud & Phillips (2008) have recently provided reading time evidence that agreement attraction violations do not have interpretive consequences for establishing grammatical roles. In an attraction sentence like “The phone by the toilets were what Yolanda used,” participants never mis-identified ‘toilets’ as the matrix sentence subject (and thus theme of the pseudocleft) despite that fact it agreed with the verb. One domain in which getting agreement right seems to matter is in the resolution of attachment ambiguities. The strings in (i) and (ii) are disambiguated entirely by the number on the verb in the relative clause

- (i) The actors [in the play [that always impresses critics]]...
- (ii) The actors [in the play] [that always impress critics] ...

When we look at pronouns, like ‘him’ or ‘her,’ we see that they have, as an approximation, the inverse licensing requirement of reflexive anaphors. If a pronoun gets its reference from an antecedent in the same sentence, then its antecedent must not c-command it in the same clause, a restriction known as Principle B (Chomsky, 1981). The contrast in (93) illustrates this restriction: ‘him’ in (a) cannot refer to the name ‘Phil,’ because ‘Phil’ c-commands it in the same clause; but ‘her’ in (b) can refer to ‘Laura’ because ‘Laura’ is outside of the immediate clause dominating ‘her.’

- (93) (a) *Laura remembered that Phil_i liked to buy him_i drinks.
(b) Laura_i remembered that Phil liked to buy her_i drinks.

As with reflexive anaphora we can ask whether the comprehender is immediately sensitive to the restrictions on pronominal anaphora: does the comprehender exclude within-clause, c-commanding constituents as candidate antecedents. The studies to address this question have given mixed answers. Using cross-modal lexical priming, Nicol & Swinney (1989) found that candidates excluded by Principle B were not considered during dependency formation. Clifton, Kennison, & Albrecht (1997) reached a similar conclusion in self-paced reading. However more recently both Badecker & Straub (2002) and Kennison (2003) have found that Principle B-excluded antecedents are identified nonetheless as candidates in anaphora resolution. There is therefore not the same kind of unanimity about pronouns that there seems to be for reflexive anaphors.

Interestingly, evidence that the resolution of pronominal anaphors is not grammatically accurate holds only in case of forwards anaphora, when a pronoun is encountered and must search the syntactic context for its antecedent. However pronouns may also occur in backwards anaphora configurations, in which case its antecedent occurs later in the sentence, as example (94) illustrates.

(94) While he_i was at the bar, Phil_i bought a drink for Laura.

In backwards anaphora candidate antecedents can be evaluated left-to-right. Van Gompel & Liversedge (2003) provided reading time evidence that comprehenders seek to resolve the referent of the pronoun quickly. However, there is a structural restriction on antecedents in backwards anaphora, and that is Principle C (Chomsky, 1981). Principle C restricts a noun phrase from co-referring with a pronoun that c-commands it. The examples in (95) illustrate this restriction both for simple and embedded clauses.

- (95) (a) *He_i bought Phil_i a beer.
(b) *He_i said that Phil_i should get a beer.

Given an incentive to resolve the reference of pronouns, do comprehenders ever consider noun phrases banned by Principle C? The two studies to examine this question (Cowan & Cairns, 1987; Kazanina et al., 2007) both show that Principle C is respected in processing backwards anaphora. Kazanina et al. (2007) used a gender match manipulation to make their case. They considered two kinds of sentences. In ‘Principle C’ sentences, e.g. (96), a pronoun (in bold) and a name (underlined) were on the same c-command path.

Consequently co-reference between the two was illicit. In ‘No constraint’ sentences, e.g. (97), the pronoun and name were not on the same c-command path. Co-reference was thus allowed.

(96) *Principle C condition*
Because last semester **she**_i was taking classes full-time while Kathryn/Russell was working two jobs to pay the bills, Erica_i felt guilty.

(97) *No constraint condition*
Because last semester while **she**_i was taking classes full-time Kathryn/Russell was working two jobs to pay the bills, Russell/Erica_i never got to see her.

For ‘No constraint’ sentences, Kazanina and colleagues found that when the name mismatched the pronoun in gender, there was a large reading time slowdown on the

name. Crucially, the same manipulation had no effect in Principle C sentences, in which co-reference could never be possible. Therefore the process of identifying candidates in backwards anaphora is constrained by structural restrictions provided by the grammar.

An important difference between forward and backwards anaphora is how the search process for candidate antecedents is triggered. When participants read the sentence-initial pronoun in Kazanina et al. (2007)'s study, they could not assign it reference unless they began searching for a name or description later in the sentence. In the Principle B studies, however, the pronoun occurred much later in the sentence, when there already a number of names and descriptions mentioned. There was no advance warning that a pronoun would occur and consequently participants had to search memory for candidate antecedents triggered entirely by a bottom-up signal. If that search occurred in parallel, as would happen in a content-addressable memory, then illicit candidates could intrude upon the process of dependency resolution. In the case of reflexive anaphora, it is possible to constrain potential referents by which immediate clause they belonged to (at least, in many cases; ECM constructions provide a counterexample). In the case of pronouns, it would have to be possible to constrain potential referents by excluding a certain clause and structural relation. It seems likely, as we argued in Chapter 3, that constituent representations are encoded with features related to their grammatical domain. However it seems implausible to encode constituents with information about which domains they do not belong to. If there is only a matching mechanism, and no mechanism for otherwise inhibiting certain kinds of constituents from being returned as candidates, then it is not surprising that the retrospective search required of Principle B is prone to grammatical inaccuracy.

Let us suppose that, in general, achieving grammatical accuracy is harder when a retrospective search is invoked. Retrieval in a content-addressable memory provides one explanation for reduced accuracy in these circumstances: the search cannot be ordered by important grammatical relations and as a consequence grammatically illicit candidates are sometimes contacted. If it were then true that forward, or prospective, searches were better at keeping track of grammatical relations, then we would expect to find greater grammatical accuracy when the need to form a dependency is announced early. Under these circumstances there would be a great premium in exploiting predictive information.

In the remainder of this chapter we will argue that *wh*-dependency formation exemplifies the virtues of prospective searching. We review the standard evidence that the parser recognizes *wh*-dependencies early and attempts to complete them as soon as possible. We provide new experimental evidence that the recognition process is highly sensitive to global well-formedness constraints, and that the decision to complete dependencies is based principally on top-down information. Indeed we conclude that initial dependency formation of *wh*-dependencies may be retrieval free.

4.2.2 *Reconsidering agreement attraction*

Before moving on to *wh*-dependencies we want to mention that the grammatical-ungrammatical asymmetry in agreement attraction may submit to an explanation in terms of prospective processing. Recall the basic data: grammatical attractor sentences, like (98a), do not induce illusions of ungrammaticality (pace the conclusions of Experiment 5). However ungrammatical attractor sentences, like (98b), induce illusions of grammaticality.

- (98) (a) The path to the monuments is littered with bottles.
(b) *The path to the monuments are littered with bottles.

One way of explaining this contrast is exclusively in terms of the match between cues provided by the verb and the constituent encodings of the subject, as we did in Chapter 3. The other possibility is that the attractor effect in comprehension is specifically a reanalysis effect. On this view, agreement computation is always carried out correctly on the first-pass. However when this computation fails, a reanalysis mechanism can check back to see if an error was made. The initial computation could be instantiated as a predictive process: when the head noun is encountered, a verb marked with the correct number can immediately be predicted. When the verb is encountered, its number features can be checked against the predicted features, and if they match, nothing more needs to be done. However, if the bottom-up features mismatch the top-down predicted features, a cue-based-retrieval is deployed to see if the correct feature was somehow missed in the context the first time. It is in this ‘rechecking’ stage that the attractor NPs might sometimes be mistakenly retrieved. The fact that attractor effects are mainly seen in ungrammatical cases thus naturally falls out from this view, as does the fact that attractor effects can be seen even when the subject and verb were adjacent in the first place. The mismatch in the adjacent case would set in motion the same content-based retrieval process, subject to the same errors. This view also has the advantage that it makes retrieval of the number feature unnecessary in the normal case, although there may be other costs involved in making a prediction⁴⁰. Furthermore, given that English agreement paradigms for lexical verbs are largely syncretic, it may be necessary to use

⁴⁰ If there were clear evidence that prediction plays a role in explaining the grammatical/ungrammatical asymmetry in agreement attraction, then the finding in Chapter 3 that there are small decrements in the accuracy of some grammatical sentences (in the subject-attached RC constructions) must be reconsidered. In this case, then the decrement might indicate how often prediction occurs; or it would bolster the alternative judgment/regeneration explanation account sketched in section 3.3.3.5.

top-down information, like the number of the subject head, to identify the number features of the verb in the first place.

4.3 Processing *Wh*-dependency constructions

4.3.1 Active dependency formation

In overt movement languages like English, displaced *wh*-phrases are found in clause-edge positions where they establish scope properties, but they are assigned a thematic role within the clause in grammatically licensed positions. When the language processor encounters a *wh*-phrase, it must have a way of deciding when and where to link that phrase with its thematic role assigner. Making this decision is complicated by two aspects of unbounded dependencies. Firstly, the tail of a *wh*-dependency is (usually) only indirectly signaled by the input, for example, by the absence of a verb's subcategorized constituent. Secondly, there are numerous and diverse island constraints that restrict where *wh*-dependencies can terminate (Ross 1967, see Szabolcsi & den Dikken 2002 for a review).

In the past two decades, it has been established that the sentence processor attempts to form *wh*-dependencies before direct evidence in the input signals the position of missing constituents (Crain & Fodor 1985; Stowe 1986; inter alia), a phenomenon commonly referred to as 'active filling' (Frazier & Flores D'Arcais 1989). Once the parser identifies a displaced element (henceforth, a *filler*), gaps are posited at each available position that would allow the dependency to be completed⁴¹. For example, in (99), a *wh*-dependency exists between the underlined filler phrase and the verb 'play':

⁴¹ We use the term 'gap' and the expression 'posit a gap' in a theory neutral way, as is standard in the psycholinguistics literature. Our discussion and results do not depend on whether or not the tail of the dependency is a trace or not. This question has been the

(99) Which CD does the toddler like her mother to play ___ before naptime?

A wide range of experimental findings indicates that, in the course of comprehending this sentence, speakers posit a direct object gap upon encountering the verb ‘like.’ This gap assignment must then be revised upon reach the pronoun ‘her’, and the ultimate gap assignment is made upon reaching the verb ‘play’.

This sequence of events has been established using several experimental paradigms. For example, previous work has compared strings like (99) to nearly identical ones that lack a *wh*-dependency, for example (100a) versus (100b):

- (100) (a) The babysitter forgot which CD the toddler **likes** her mother to play ___
(b) The babysitter forgot whether the toddler **likes** her mother to play a CD ...

Word-by-word reading times have shown that a disruption begins at the constituent ‘her mother’ in (100a) (Crain & Fodor 1985; Stowe 1986; Lee 2004). This processing disruption, called the Filled Gap Effect, suggests comprehenders posit a gap in advance of an overt constituent. Convergent evidence comes from measures in which the semantic fit between a filler and potential-gap host (in bold) is manipulated, as in the following pair from Traxler & Pickering (1996):

- (101) (a) That’s the pistol with which the heartless killer **shot** the hapless man ...
(b) That’s the garage with which the heartless killer **shot** the hapless man ...

These semantic fit or plausibility manipulations show that readers detect when a filler is an implausible argument of a verb while they are reading it, or very shortly thereafter. A disruption is reflected either by a slowdown in reading times (Traxler & Pickering 1996;

subject of periodic debate in psycholinguistic circles (McElree & Bever 1989; Nicol, Fodor & Swinney 1989; Pickering & Barry 1991; Gorrell 1993; Gibson & Hickok 1993). Current experimental techniques are most informative about the timing of dependency formation, but the timing facts in this case are orthogonal to the representational hypotheses (Phillips & Wagers 2007).

Phillips 2006), a deflection of a lexical-semantic ERP component, the N400 (Garnsey, Tanenhaus & Chapman 1989), or a sharply increased tendency to report that the sentence stops making sense (Tanenhaus, Stowe & Carlson 1985; Boland et al. 1995). These results show that comprehenders not only posit a dependency between a filler and potential gap site, but also evaluate the semantic impact of their decisions as soon as possible. They are strengthened by a wide array of findings using related methodologies and sampling a number of different languages (see footnote ⁴²).

ERP studies provide a different index of long-distance dependency completion. Processing of the verb that allows completion of a wh-dependency elicits a posterior positivity relative to the same verb in a sentence without a wh-dependency (102ab: Kaan, Harris, Gibson, & Holcomb, 2000).

- (102) (a) NO WH-DEPENDENCY: Emily wondered whether the performer in the concert had imitated a pop star for the audience's amusement.
(b) WH-DEPENDENCY: Emily wondered which pop star the performer in the concert had imitated for the audience's amusement.

Kaan and colleagues argue that this is the P600 evoked response typically associated with syntactic anomaly detection and reanalysis processes, and use this finding to suggest that the P600 is an index of 'syntactic integration difficulty' in general. Although the interpretation of this effect remains uncertain (cf. Fiebach, Schlesewsky, & Friederici,

⁴² CROSS-LINGUISTICALLY: Dutch (Frazier, 1987; Frazier & Flores D'Arcais, 1989; Kaan 1997); German (Schlesewsky, Fanselow, Kliegl, & Krems, 2000); Hungarian (Radó, 1999), Italian (de Vincenzi, 1991), Japanese (Aoshima, Phillips, & Weinberg, 2004); Russian (Sekerina, 2003)

CROSS-METHODOLOGICALLY: electrophysiology using EEG (Garnsey, Tanenhaus & Chapman 1989; Kaan, Harris, Gibson & Holcomb, 2000; Phillips, Kazanina, & Abada 2005) and MEG (Lau, Yeung, Hashimoto, Braun, & Phillips 2006); the "stops making sense" task (Tanenhaus, Stowe, & Carlson 1985; Boland, et al. 1995); eye-tracking (Traxler & Pickering, 1996); cross-modal lexical priming (Nicol & Swinney, 1989; Nicol, Fodor, & Swinney 1994); anticipatory eye movements (Sussman & Sedivy, 2003).

2002; Kazanina, Phillips, & Abada, 2005), its timing reinforces the generalization that *wh*-dependencies are formed as soon as a syntactically-appropriate thematic role assigner is encountered and before any direct evidence that a constituent is missing.

4.3.2 *Mechanisms of active dependency formation*

The computational problems that processing a filler-gap dependency, actively or otherwise, poses in English is schematized below:

- (103) (a) DEPENDENCY RECOGNITION
The parser must recognize long-distance dependencies, at least by:
(i) identifying the filler;
(ii) identifying the gap.
- (b) RETAINING A FILLER IN MEMORY
The parser enter into memory an appropriate *syntactic* representation of the filler for later re-integration. In languages where gaps may proceed fillers, there is an analogous problem of retaining the integration environment in memory when the filler is ultimately encountered.
- (c) FILLER REACTIVATION & INTEGRATION
The filler memory must be contacted, possibly reactivated, and integrated at the foot of the dependency.

For present purposes we will ignore the problems of recognizing the head of a *wh*-dependency. We assume that the use of morphological and positional cues are enough to start a search in English, though this may not be true cross-linguistically (cf. Yoshida, 2006). However, consider the problems of maintenance and retrieval. Dependency completion is active, in the sense of the Active Filler Strategy (AFS; Frazier 1987; Frazier & Flores d'Arcais 1989), because it is driven by a signal related to the filler, and not the gap (Wanner & Maratsos 1978; Fodor 1978): gaps are postulated by a top-down signal, and not on the basis of bottom-up evidence. What is the nature of this signal?

Wanner & Maratsos (1978), in their formulation of an Augmented Transition Network (ATN; Woods 1973) as a processing model, had proposed that when an NP is

identified as the head of a relative clause, that the NP is placed on a HOLD list⁴³. The HOLD list allows the assignment of grammatical function to be put off until an appropriate subsequent context, and, in this sense, enables the ATN to operationalize a transformational grammar. For example, in an object-extracted relative clause, after the verb has been processed, the ATN attempts to analyze subsequent input as an NP. When it is unsuccessful, it checks to see whether or not the HOLD list is empty; if it isn't, it retrieves its contents, treating them as input. This parsing sequence does not qualify as an active strategy, because a gap is not postulated unless the parser fails to recognize the verb's lexical argument. Frazier (1987) suggested that Wanner & Maratsos's HOLD model could itself be a 'decision principle': the identification of a filler – or the non-emptiness of the HOLD cell – could serve as a signal to postulate a gap. Both the identification of a filler and the status of the HOLD list are closely related candidate signals, but they are logically separable cues to active gap postulation. In Frazier & Flores d'Arcais's (1989) formulation of the AFS, it is the identification of a filler that "immediately predisposes" the parser to rank gaps more highly than lexical arguments. Later, though, they suggest that this implicates an active filler (one that is not 'inert'): at predicates below the head of a dependency, gap analyses are considered before lexical argument ones because the active filler is effectively always an input, until it is successfully incorporated into the structure. What it means to be not 'inert' is open to a few different interpretations, but, functionally, the view seems to be that a filler, while a dependency is incomplete, has the same representational status as bottom-up inputs. Considered this way, dependency

⁴³ HOLD doesn't contain the constituent NP, but the constituents of NP. Only when these categories are retrieved at a gap are they assigned to the category NP ('ASSEMBLED') (Wanner & Maratos 1978 p. 136).

completion is active because the filler is in the workspace before subsequent inputs, and it will effectively out-compete incoming categories for attachment⁴⁴.

What is the representation of the unintegrated filler like? One view is that fillers are separately and actively maintained in a distinguished buffer, like the HOLD list, and that the HOLD list is consulted just like an input buffer at each step in the processing of a sentence. Let us focus on the first proposal: Wanner & Maratsos hypothesized that, if memory were like this in sentence processing, we would expect some cost to keep the elements on the HOLD list active, until they were all discharged. This maintenance cost could derive either from devoting resources to clamping HOLD list items in an active state. Alternately (though perhaps equivalently), it could be due to the consumption of a fixed pool of memory resources, or decrementing a fixed number of open buffers. Wanner & Maratsos (1978) provided evidence that memory costs were higher when a dependency was incomplete, by comparing subject and object relatives in a dual-task paradigm: participants had to both comprehend a sentence, and recall a list of names that at some point intruded upon the word-by-word reading of the sentence.

Ford (1983) criticized Wanner & Maratsos's result, and claimed that the differential difficulty in subject- and object-relatives could be localized to the verb, where reading times are elevated for object relatives (King & Just 1991). Nicol & Swinney (1989; also Nicol, Fodor & Swinney 1994) later showed that, once the filler is introduced, semantic associates are primed in lexical decision, but that this priming does

⁴⁴ A similar logic has been applied to derive a MOVE-over-MERGE preference in the root-first generation of sentences (Richards 1999). Retrieving new items from the lexicon is asserted to be more costly than using what is already merged into the existing phrase structure. If the resource cost in Richards' account is taken to be processing time, then his view is fundamentally the same as Frazier & Flores d'Arcais (1989): what is already in the workspace wins because it's faster.

not persist. Only at the verb, when the filler must be contacted and thematically integrated, does the priming effect re-emerge. These results were interpreted in favor of a re-activation model, in which the filler is not fully active, and must be back into a state suitable for integration. However, it is important to recognize that, firstly, the reading time results (Ford, 1983; King & Just, 1991) do not exclude an active filler. Processing complexity, or whatever slows readers down on object-relative verbs, may arise for reasons having to do with memory, ambiguity, particular parsing operations, or some combination of these. Moreover the cross-modal paradigm used by Swinney and colleagues has been extensively criticized (McKoon & Ratcliff, 1994) for not uniquely admitting the interpretation they draw. It is not clear, moreover, how informative a demonstrated *lack* of priming is in dependency-medial positions, when other constituents are being active constructed and interpreted in those positions.

Electrophysiological evidence has lately been brought to bear on the question of memory load during an open dependency. Both King & Kutas (1995) and Fiebach, Schlesewsky & Friederici (2002) showed that in object-extracted filler-gap dependencies, the averaged EEG record reveals a sustained anterior negativity. Because the effect is modulated by performance and participants' memory span, and has been implicated in more explicit memory load tasks (e.g. Ruchkin et al., 1990), the presence of a sustained anterior negativity in open filler-gap dependencies has been interpreted as a direct reflection of the memory load consumed by actively maintaining the filler. Fiebach and colleagues are cautious to point out that the electrophysiological effect does not choose between alternative accounts of what precisely is being maintained. It could be a full semantic or syntactic representation of the filler, or perhaps just a few features. Or, it may

not be the content of the filler that's represented, but rather that prediction for a category that allows completion of the dependency, as in Dependency Locality Theory (Gibson 1998). Moreover, recent work by Phillips and colleagues (2005) has provided evidence that the sustained anterior negativity does not reflect a cumulative effect that accrues at each word, but derives mainly from the first few words of the dependency. Such a finding raises doubt that the effect reflects active maintenance of the filler representation. Taken together, the behavioral and electrophysiological evidence on the representational state of the filler during dependency completion is mixed.

The chief idea behind active maintenance is that the filler is in a privileged representational state, which renders it highly accessible while the dependency is open, and prevents it from decaying. This perspective has been argued to conflict with the memory architecture we have been exploring. McElree, Foraker, & Dyer (2003) are explicit in their rejection of postulating distinguished representational states, like stacks or buffers. The resources available for maintaining information in an immediately accessible state are limited, they argue (cf. McElree, 2006; and Chapter 3). Consequently information that does not participate in the computation at hand is effectively shunted out of immediate attention, and must be re-accessed by retrieval. In such a memory architecture the representation of a filler cannot generally be in an active state for very long, since the input that intervenes between filler and gap will displace it from the focus of attention. However, the evidence is overwhelming that dependency completion proceeds in advance of information in the input that could directly signal a retrieval. Therefore some internal signal must direct dependency completion. The active maintenance account of active filling has the virtue that, by effectively granting fillers

and inputs the same status, the signal, though internal, is virtually bottom-up. When a category of the same type as the filler is subcategorized in the structure, fillers have the right of first refusal, so to speak. On the one hand clear evidence in favor of the active maintenance hypothesis is lacking and some powerful architectural desiderata militate against it. On the other hand, there is the logical problem of knowing when to attempt dependency completion on the basis of local information alone.

One reasonable solution is to more explicitly embed active filling into the parsing routines: the filler representation itself would not compete for attachment, but the system would attempt retrieval of the filler at every grammatically-legal subcategorizer until the filler is integrated. In a series of studies we report below (section 4.6), we show that most lexically-anchored information is lost over increasing dependency lengths in the active portion of dependency completion. Before the gap site, the parser seems only robustly sensitive to coarse-grained categorial information about the filler. The lack of sensitivity to specific lexical features seems consistent with the explicit incorporation of active filling into parsing routines as a rule operating over syntactic categories. Below we argue that licensing syntactic dependencies works in this way, and that information contained in the full filler encoding may not be retrieved until much later in the timecourse of comprehension. We suggest, however, that some item-specific, partial information about the actual filler is truly carried forward in time. We conjecture that a highly restricted subset of the features of the episodic representation of the current filler encoding is maintained in a state immediately accessible to processing. This set of features can serve both as an internal signal to complete the *wh*-dependency as soon as possible but more

crucially may serve as a component of the retrieval structure used to target the grammatical filler.

4.3.3 *Similarity-based interference and wh-dependency completion*

In recent years several lines of evidence have been collected to support the idea that *wh*-dependency completion is subject to similarity-based interference. The force of this claim is to further undermine the position that the full filler encoding is preserved in a distinguished state. The first source of evidence comes from a series of studies by Peter Gordon and colleagues (2001, 2002, 2004) on the processing of relative clauses and clefts. A very basic and robust finding in sentence comprehension is the subject/object asymmetry in relative clause processing (Wanner & Maratsos, 1978; King & Just, 1991, *inter alia*). When the head of the relative clause is relativized from object position (104a), the relative clause is harder to process than when it is relativized from subject position (104b).

- (104) (a) The salesman [_{RC} that the accountant contacted] spoke very quickly >
(b) The salesman [_{RC} that contacted the accountant] spoke very quickly

For example, in two sets of baseline data collected by Gordon et al. (2001) in self-paced reading, object relatives raised comprehension error rates 6-9%, and RC verb reading times 50-85%. However Gordon et al. (2001) also showed that difficulty of object relatives depends on the relative referential types of the filler DP and the subject DP. In examples like (104) both filler and subject are definite descriptions. However, if the subject DP was a pronoun or a name, the asymmetry in verb processing was essentially eliminated. The same effect held true if the *wh*-dependency is in a cleft (though the amelioration is not as complete). Gordon et al. (2001) argued that the distinctiveness of

the representations was an important factor in determining how easily the constituents could be integrated at the verb. When the two DPs are of the same referential “type” (either both descriptions, or both names), then integration is more difficult, than if they are of mixed types. Gordon et al. (2002) used a memory load paradigm to further support the idea that dissimilar constituents were easier to integrate. In this experiment, participants were presented with a list of 3 names or 3 descriptions. They then had to read a sentence, answer a T/F comprehension question about it, and recall the memory list. The design of the experiment is summarized in the example materials set in (105). The clefts were either subject or object clefts, and the memory list either did or did not match the referential type of the DPs in the sentence.

(105) (DESCRIPTION) LOAD POET—CARTOONIST—VOTER

(a) SUBJECT CLEFTS

MATCH: It was the dancer that liked the fireman before the argument began.

MISMATCH: It was Tony that liked Joey before the argument began.

(b) OBJECT CLEFTS

MATCH: It was the dancer that the fireman liked before the argument began.

MISMATCH: It was Tony that Joey liked before the argument began.

Comprehension error rates and reading times both showed strong effects of subject v. object extraction, and match v. mismatch. These effects interacted such that a matching memory list made object clefts much harder to understand, compared to mismatching lists; the impact on subject clefts was smaller. In later work, Gordon et al. (2004) showed that when pronouns, names and quantifiers mismatched with definite descriptions, the Subject/Object asymmetry could be alleviated (but not by contrast with indefinites, generics, or superordinal category nouns). This effect can be intuited quite strongly if the level of embedding is increased, as Bever (1974) observed:

(106) (a) The reporter the politician the commentator met trusts said the president won't resign.

(b) The reporter everyone I met trusts said the president won't resign.

(106a), in which the three pre-verbal DPs are all descriptions, is much harder to comprehend than (106b), in which the three pre-verbal DPs are a description, the universal quantifier, and an indexical pronoun (cf. Gibson, 1998).

The interpretation that Gordon and colleagues gave their data is that, before the verb is encountered in object extractions, there are two unintegrated arguments: the filler DP and the subject DP. In their unintegrated state, these arguments can interfere with one another to the extent they are similar⁴⁵. Once integrated with the verb, the relative distinctiveness of the DP arguments does not matter (cf. Gordon et al. 2006 for further data and argument). In the present context it is worth noting that Gordon et al. are explicitly not committed to a retrieval-based account of their data. A working memory account in which multiple items are actively maintained is, in their view, a permissible account of their data. However their data can be seen to provide a further argument that the filler encoding is not in a distinguished state insulated from other constituents. If the filler occupied its own buffer or slot, it should not matter how distinctive it is with respect to the incoming subject DP. We find that this argument is undermined somewhat, if there is a pressure to structurally analyze the filler – i.e., to discharge the contents of such a buffer. Suppose an object-extracted filler is initially analyzed as the subject but the input forces reanalysis. The relative distinctiveness of the two representations may make the

⁴⁵ In Gordon et al.'s experiments, the relevant dimension of similarity was referential type, or how the expressions can function in a discourse. It is unclear whether there is a plausible feature-based encoding of the different expression types that can distinguish names, definite descriptions, quantifiers and pronouns. It may be that the syntactic category of the expression or the abstract internal structure of the phrase is actually controlling similarity.

reanalysis process more difficult, and thus be uninformative about how the filler is maintained in its unanalyzed state.

Van Dyke & McElree (2006) provided more direct evidence that the filler must be retrieved at the site of integration, and that this retrieval is subject to interference. They also examined the processing of clefts under a memory load manipulation. In half of the experimental conditions, participants were presented with a list of three nouns at the start of the trial, which would have to be recalled after the sentence comprehension task (Load conditions). For example:

(107) TABLE-SINK-TRUCK

Participants then read sentences like that following:

(108) It was the boat that the guy who lived by the sea sailed / fixed in two sunny days.

Two possible critical verbs are given in this sentence (underlined). For half of the conditions, exemplified by ‘sailed’, the critical verb was not a good fit for these nouns (Low interference conditions): it is not plausible to sail a table, sink or truck. For the other half of the conditions the critical verb was a good fit (High interference conditions): it is plausible to fix a table, sink or truck⁴⁶. Results for the critical region are reported in Figure 4-1. In the critical verb region they reported an interaction of interference and load conditions. Reading times were identical when there was no memory list; however, under load conditions, high interference conditions were read much more slowly.

⁴⁶ It is unclear to us why the subject was made complex; perhaps this induced slower processing, giving effects a greater chance of manifesting themselves.

Van Dyke & McElree (2006)

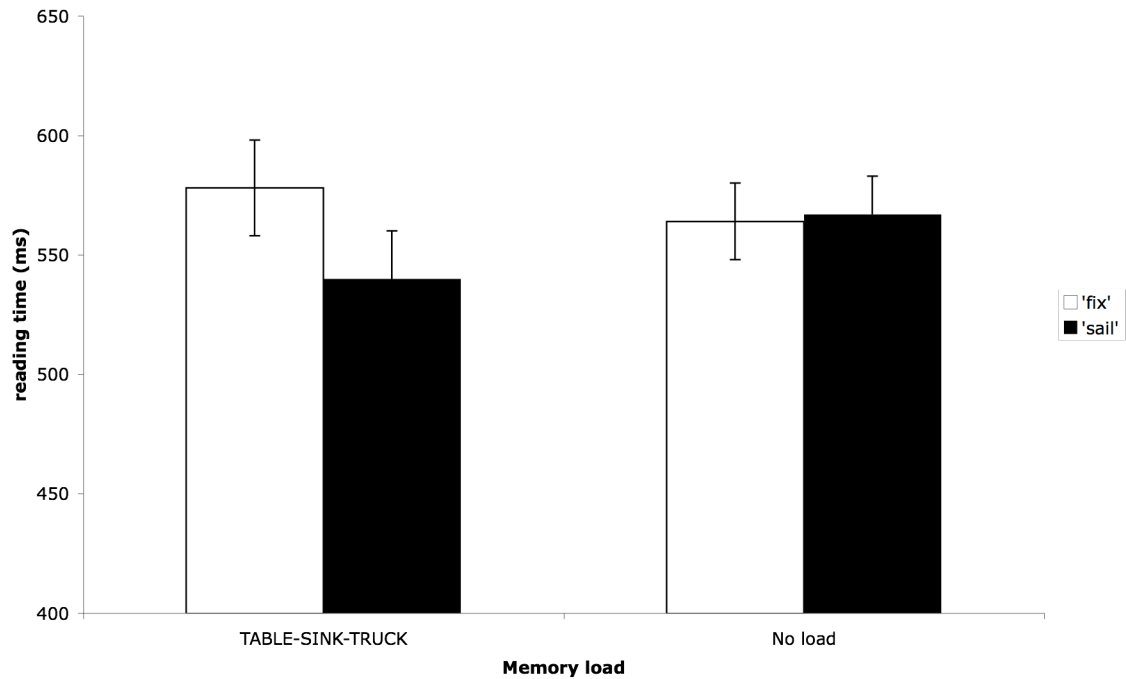


Figure 4-1 Van Dyke & McElree (2006) Critical verb region

The interpretation of this pattern is that the memory load items interfere with filler-gap completion. Similarity-based interference in this case, Van Dyke & McElree argue, arises during retrieval of the filler, as cues compete in the high-interference conditions. The memory list items (like ‘table-sink-truck’) have some common feature specified in the retrieval structure of the verb ‘fix’ not present in ‘sail,’ so that more memory items are activated in high-interference conditions than in low-interference conditions. If this explanation of the data is correct, then it is further support for a filler-gap processing regime supported principally by retrieval at the verb, and not by a maintained or distinguished representation of the filler.

The key finding in Van Dyke & McElree (2006)’s data was the interaction of interference and load at the critical verb. However, some further aspects of the data bear

consideration in relation to other experiments. In Van Dyke & Lewis (2003), Van Dyke (2007) and our Experiments 6, reported in Chapter 3, the purported high-interference manipulations most reliably led to decrements in offline measures of comprehension, even when online differences were not reliable. In Van Dyke & McElree (2006)'s data, there were main effects in comprehension accuracy: high interference conditions were harder than low interference conditions (83% v. 87% correct), and, interestingly, no-load conditions were harder than load conditions (also 83% v. 87% correct). The authors reported no difference between interference conditions specific to the load condition, however, which is exactly what other experiments lead us to expect. We might suppose that differences across experiments reflect how comprehenders trade speed and accuracy: where interference leads to errors in comprehension, particularly those that can be attributed to misbinding the constituent affected by the interference manipulation, we may see no difference in the online record. For example in Van Dyke (2007), the condition with the greatest interference (HiSYN/HiSEM) led to the worst comprehension accuracy, yet it was the intermediate interference condition (HiSYN/LOSEM) that most robustly disrupted online processing. We can imagine two kind of interfering constituents, corresponding to Van Dyke (2007)'s manipulation: just-partial matches, in which there is enough information in the constituent encoding to signal a potential error; and near-total matches, in which the only way to detect an error would be a process of chained retrievals to verify the structural relationship. In just-partial matches, the error signal leads to increased processing times online, since the system is triggered to select another constituent or otherwise attempt repair. In near-total matches, the system is sufficiently fooled by the item information in the interfering constituent and registers no error. In

just-partial match scenarios participants more often succeed in binding the correct constituent, so comprehension accuracy is less impacted; whereas in near-total match scenarios participants more often misbind the interfering constituent, which would be reflected as a decrement in comprehension accuracy. Unfortunately, there is no diagnostic measure of which constituent the comprehender took as the embedded object; a cloze task, as in Van Dyke (2007), could be helpful in that regard.

Van Dyke & McElree (2006)'s interference manipulation must surely count as a just-partial match, since the interfering elements are extra-syntactic: they are bare lexemes in a list. The item encodings should therefore carry no grammatical information (perhaps beyond lexical category), and are moreover encoded during a distinct task. Thus if one were misretrieved, it would be clear that it could not be a legal participant in any grammatical dependency. Given just how distinct one would expect the word list encodings to be, and given the presence of a fully matching in the sentence, it is surprising that the online effect is so strong. In our discussion of agreement attraction, one point emphasized was how the retrieval mechanism privileges full matches through non-linear cue combination. The partial matches in this experiment at best match on two dimensions: they are nominal (maybe) and have lexical features that match with the verb⁴⁷. The filler in [Spec,CP] would match on several others: e.g., its grammatical +wh

⁴⁷ It constitutes a further interesting puzzle whether we can give the right lexical specifications to admit selective semantic cues. In the example discussed, the word list is “table-sink-truck,” the grammatically-licensed constituent is “the boat” and the verbs are ‘fix’ and ‘sail.’ All items are ‘fixable’, but only one item is (prototypically) ‘sailable.’ There are therefore two factors relevant to the retrieval structure provided by the two verbs: how unselective ‘fix’ is, and whether ‘sail’ is selective. Is there a feature relevant to ‘fix’ that all representations share (e.g., +concrete; though surely [?] not +reparable)? Is there then a feature that ‘sail’ picks out (e.g., +navigable, +marine)? A further question is whether such features in the retrieval structure arise directly from the lexical

feature, a shared same-sentence/same-clause context encoding, its dominating category. Moreover one expects that a comprehension system that is even modestly grammatically constrained would assign greater attention to features like +wh, since structural relations ultimately determine the interpretation.

A priori assumptions about the retrieval structure could be misleading, however. And it is important to keep in mind that the effect of each partial match accrues because there are three, and it is the presence of three that competes with the grammatical constituent. Nonetheless there are sufficient concerns about specifying the retrieval structure to raise doubts about a purely retrieval-based account of these data⁴⁸. An important question is what effect the memory list has on retrieval of the verb itself from the lexicon (prior to any structure building), and whether shared lexical features with items on the memory list could slow the selection of the verb in high interference conditions. It should be straightforward to address whether it is filler-gap dependency completion per se, or another effect of the memory load conditions, that accounts for the interaction by testing sentences in which the verb does not participate in a filler-gap dependency. For example, in the following sentence, there is no filler-gap dependency to

specification of the verbs, or whether they are contextually provided by earlier processing.

⁴⁸ There is a question about whether the unergative thematic role assignment that ‘sail’ permits (but ‘fix’ does not) affects the retrieval structure. It is unknown whether this is systematic in Van Dyke & McElree (2006)’s materials, but it is worth considering in its own right. Independently of its occurrence in a filler-gap dependency, the string “the guy who lived by the sea fixed” has outstanding thematic requirements: the verb ‘fix’ must discharge its THEME role; on the other hand, in the string “the guy who lived by the sea sailed” is well-formed without further constituents. Inclusion of PPs like ‘in two days’ forces the transitive interpretation of ‘sail’ ultimately, though this information is not available in the critical region. The thematic structure of the verb is important because there is evidence for argument anticipation (Altmann & Kamide, 1999) and even some indication that it may occur semi-independently of filler-gap dependency completion (but cf. our discussion in section 4.4 below).

complete at ‘sail’/‘fix’ and consequently no reason to retrieve for these arguments (but see fn. 47 for a discussion of why it is important to match the thematic properties of the test verbs closely).

(109) Load: TABLE-SINK-TRUCK

Test: John heard that the guy who lived by sea sailed / fixed his boat in two sunny days. Assuming the load items are all equally bad subjects, there should be no interaction of verb class with load if that interaction is caused by cue competition in a retrieval. If, on the other hand, the presence of the memory list interacts in other ways (e.g., through lexical retrieval), the same interaction should be present in (109).

The recent sets of studies by Gordon and colleagues, and Van Dyke and McElree, give us grounds to believe that completing a filler-gap dependency does not depend on a special representation of the filler, at least one that can always be used easily and reliably. The presence of items other than the filler affects the ease and accuracy with which filler-gap dependencies are processed. These facts do not lead inevitably to the conclusion that the mechanism of interference is exclusively retrieval-based, nor that no information about the filler persists to guide initial dependency completion. In the following three sets of studies we argue that top-down properties of the dependency environment play a major role in dependency construction. We argued that the comprehender uses global context to determine when retrieval is attempted and what information it is based upon. In support of this mechanism we argue that some information is maintained and carried forward in time. Consistent with our conjecture about this kind of prospective processing, *wh*-dependency formation is found to be highly grammatically accurate.

4.3.4 *Three studies*

The first two sets of studies demonstrate that *wh*-dependency formation is highly grammatically accurate in two respects: first, in deciding where to attempt dependency completion; second, in targeting the head of the dependency to incorporate at a gap site (contra Van Dyke & McElree, 2006). Crucially, in the context of a highly retrieval-dependent architecture, this accuracy must hang on the parser being able to make accurate, grammar-driven predictions about where a dependency could terminate, and thus where to attempt retrieval. When it does retrieve, it must form a retrieval structure that is capable of (nearly uniquely) targeting the head of the dependency. In the third set of studies we provide evidence that category-level information is available to make a syntactic decision in active dependency formation, consistent with our hypothesis that some information is maintained.

In the first set of studies, we focus on island constraints (section 4.4, Experiments 7-8). There has been considerable attention devoted to the question of whether *wh*-dependency completion respects island constraints (Ross, 1967) online or whether the parser sometimes attempts to form a dependency that must ultimately be grammatically illicit. We test whether the online structure building respects the Coordinate Structure Constraint (CSC; Ross, 1967), a condition on extraction that forbids gaps inside coordinate structure, unless the same subconstituent is gapped in each coordinated phrase. The CSC is unique among island constraints in this regard: it does not ban extraction outright, but requires multiple extractions. Consequently we can test whether the comprehender can use the grammar to predict future retrieval sites.

In the second set of studies, we present a refined version of McElree & Van Dyke (2006). We conducted two experiments in which competition at retrieval comes from two

wh-dependency chains in the same sentence, not from an extra-syntactic list (section 4.5, Experiments 9-10). Our test for interference during filler-gap dependency completion simply concerned the presence or absence of a syntactically similar dependency head, and thus we could set aside potential lexical effects.

Finally, in the third set of studies, we test how well three different kinds of dependency formation probes survive increasing dependency lengths: verb-object plausibility, verb-PP selectional restriction, and a DP/PP filled gap effect. Each of these probes requires different kinds of information about the filler to generate a signal during active dependency completion. The specificity of the information required (minimally) for each probe is gradually decreased. The verb-object plausibility test requires information about the filler's lexical head features; the verb-PP selectional restriction requires just the lexical identity of the filler's functional head; and the DP/PP filled gap test only requires filler category. By varying dependency length, we are able to test what kinds of information is effective long after the filler was first encoded, and consequently what kinds of information may be guiding the parser's initial decision, before it attempts to recover full information about the filler.

4.4 The grammar's role in triggering wh-dependency formation

4.4.1 The motivation for active dependency formation and island constraints

Before we said that if it were true that no information about the filler encoding was maintained throughout processing, then we must be concerned with another mechanism for signalling that a dependency must be completed. The most obvious mechanism is simply a parsing rule: if a filler has been encountered, attempt to retrieve and integrate that filler at every licit subcategorizer. But why does this rule have high

priority? As Fodor (1978) discussed, it is not an inevitable element of the parser – it has to be motivated. Most discussions of why filler-gap dependency formation is so active identify a pressure imposed by the unintegrated filler. One kind of pressure sees the unintegrated filler as imposing a tax on processing resources, as it competes for maintenance resources in working memory. Wanner & Maratsos (1978) were early advocates of this kind of architectural approach to keeping dependencies as short as possible. In a retrieval regime, and one in which representations are not maintained in working memory, that pressure cannot come from the presence of the filler representation itself. If the filler is not consuming limiting memory space resources then there pressure to complete the dependency cannot stem from that burden⁴⁹.

Here we want to focus on another alternative, which is that the pressure to complete dependencies actively stems from either the well-formedness of the syntactic or semantic representation. This class of explanation is closely related to the how previously encountered information guides parsing, since the well-formedness of the representation can be gauged either top-down or bottom-up. A top-down indicator of well-formedness could be something like a list of outstanding requirements to license a structure. For example, one class of explanations, associated with principle-based parsing (e.g., Pritchett 1992; Weinberg 1992), identifies a pressure to satisfy grammatical licensing requirements as rapidly as possible. In the case of filler-gap dependencies, it has been

⁴⁹ By virtue of cue competition, retrieval itself is a limited capacity process, so the pressure could arise as an adaptation to maximize retrieval success. On the assumption that interpolating more material leads to a decline in retrieval accuracy of *wh*-phrases, then a parser that retrieves and integrates the *wh*-phrase sooner rather than later might on average be more successful in arriving at an interpretation. We have advocated this view elsewhere (cf. Wagers, 2006). We do not pursue it further here, though it raises interesting questions, concerning how the parser could adapt itself to this pressure.

suggested that there is a pressure to satisfy the Theta Criterion as soon as possible (e.g., Pritchett 1992; Aoshima et al. 2004). Until a structural relation can be established with the filler and a thematic role assigner, the parse is partially unlicensed, which is an undesirable state of affairs. This explanation can be framed in less grammaticized terms, under the assumption that the parser attempts to interpret as much of the sentence as soon as possible (e.g. Altmann & Kamide 1999; Sedivy et al., 1999). Under this view the active strategy is simply one manifestation of the parser's efforts to derive an interpretation from only partial information. By actively completing a *wh*-dependency the parser can yield a more informative interpretation from the limited input available. Altmann & Kamide (1999) in particular focus their explanation more on the outstanding properties of the verb. Verbs have licensing requirements as well; for example, they must discharge their theta roles. A single verb thus provides a bottom-up signal that there is an outstanding licensing requirement. Based on its inherent properties a retrieval structure could be assembled to search for arguments. The question arises, whether active dependency formation simply reflects the verb casting about for (thematically-unmarked) arguments, and finding it in the head of a *wh*-chain in the case of an incomplete filler-gap dependency⁵⁰.

Island constraints provide a natural way of testing whether active dependency formation is merely the result of a verb seeking to saturate its argument structure. Island constraints restrict the kinds of dependencies that can be formed, in ways that are potentially independent of constraints on interpretation or processing. For instance, the example sentence in (110) contains a complex noun phrase in subject position. It is

⁵⁰ Already we must suspect such an explanation, given the presence of filled-gap effects in subject position (Lee, 2004).

impossible for the filler phrase in the main clause to terminate in the NP-contained clause:

(110) * Which babysitter did [_{NP} the revelation [_S that the toddler tormented ____]]
frighten her mother ?

If the context with which the parser deals is really restricted by architectural constraints on the focus of attention, as has been suggested, then the question arises whether the verb ‘torment’ would search and identify ‘which babysitter’ as a argument, irrespective of the island boundary that separates them. The notion expressed by linking the filler ‘which babysitter’ with the verb ‘tormented,’ as in (110), is plausible and perhaps a likely state of affairs. Furthermore, as the discussion of the filled-gap effect in section 4.3.1 illustrates, in non-island domains the parser seems willing to make some mistakes and revise temporary commitments. However, linking ‘which babysitter’ with the verb ‘tormented’ can never turn out to be the right analysis, because of the constraint on constructing dependencies inside complex NPs. Showing that parsers do not engage in active dependency formation inside island domains would in principle constitute strong evidence that grammatical knowledge guides the parser’s decisions about when to recover filler information. In particular, it would indicate that the predicted well-formedness of a candidate analysis influences where the parser decides to recover information about the head of the dependency.

A number of studies have shown that measures of active dependency formation are not observed in island domains (Stowe, 1986; Bourdages, 1992; Pickering et al., 1994; Traxler & Pickering, 1996; cf. Freedman & Forster, 1985; Kurtzman & Crawford, 1991), and many share the consensus that island constraints are respected in incremental processing (Phillips, 2006). However this position is vulnerable, exactly because the

typical empirical consequence of respecting an island constraint is the absence of evidence that a dependency was ever entertained in that island.⁵¹ The strength of conclusions that can be drawn from a lack of evidence has raised concerns, particularly because some island domains have been argued to be themselves complex processing environments, whether a dependency is present or not (e.g. Deane, 1991; Kluender, 2005). Therefore null findings in island processing are liable to alternative interpretations that are unrelated to the parser's interaction with the grammar.

In the first series of experiments, we present a new argument that grammatical knowledge plays a definite role in the active formation of *wh*-dependencies, and one that does not suffer from the null-effect logic of previous studies on islands in language processing. Instead of considering island constraints that absolutely restrict the formation of a *wh*-dependency inside a certain domain, we consider a related constraint on *wh*-dependency formation, the Coordinate Structure Constraint, in which extractions are permitted in certain cases. The Coordinate Structure Constraint (CSC; Ross, 1967) rules out gaps within coordinate structures (111a-b), except in the case of across-the-board extraction, when one gap must occur in each coordinated phrase (111c)⁵².

⁵¹ Three EEG studies have demonstrated a processing disruption when the search for a gap encounters the boundary of island domain. This disruption is reflected in a particular evoked response potential (ERP; P600: McKinnon & Osterhout, 1996; LAN: Kluender & Kutas, 1993; Neville et al., 1991). However these ERPs are observed at island boundaries. Therefore they are not informative about whether the parser attempts to construct a dependency, only whether an island domain is noticed. Moreover, the observed ERPs are also sensitive to processing difficulty. Consequently, while these results may reflect calculation of ill-formedness in a formal account of island constraints, but they may equally well reflect increased complexity.

⁵² There are several well-known classes of exceptions to this generalization (Goldsmith 1985; Lakoff 1986), as, for example, in:

- (i) What did you go to the store and buy ___?
- (ii) How much can Josh drink ___ and still stay sober?

- (111) Phil generally dislikes the poetry ...
- (a) * that *The New Yorker* reviews authors or publishes ____
 - (b) * that *The New Yorker* reviews ____ or publishes interviews
 - (c) that *The New Yorker* reviews ____ or publishes ____

If the parser is guided by the Coordinate Structure Constraint, then there should be evidence that a second gap is actively posited in the second coordinate. Since this evidence would be positive, and not a null effect, then it could avoid the concerns raised by previous island studies. The presence of continued active dependency formation in coordinate structures could be explained by the real-time application of the CSC, but it could also be explained by a bottom-up retrieval mechanism in which the verb initiates a search for arguments. If a subsequent verb in a coordinate structure can take the filler as its argument, then it can satisfy its interpretive needs earlier than by waiting for an argument phrase. Therefore we consider a second kind of multiple dependency construction, a parasitic gap inside post-verbal adjunct clauses (Engdahl, 1983). Single extractions from a post-verbal adjunct clause are generally unacceptable (112a). In the presence of an extraction from direct or indirect object, the post-verbal adjunct clause can support an additional gap (112b), but, crucially, it is optional (112c).

- (112) Phil generally dislikes the poetry ...
- (a) * that *The New Yorker* reviews authors without publishing ____
 - (b) that *The New Yorker* reviews ____ without publishing ____
 - (c) that *The New Yorker* reviews ____ without publishing too much detail

A comparison of *wh*-dependency formation in coordinate structures and parasitic gap environments, like post-verbal adjunct clauses, is therefore informative: if dependencies are actively formed in both domains, then we would fail to isolate the role that

These exceptions occur in specific circumstances when certain narrative relationships hold between the coordinates. All materials used in our studies were designed so as to avoid these contexts. See Postal (1998) for further discussion of these environments.

grammatical principles play, since the results would not reflect any grammatical distinctions. If, on the other hand, CSC environments exhibit active dependency formation, but parasitic gap environments do not, then it would indicate that information in the syntactic context guides when the dependencies are formed, and when filler information is recovered. If this is the case, then it must be that an important grammatical constraint is reflected in parsing routines. Table 4-1 outlines the three candidate patterns of active dependency formation that might be observed in multiple dependency constructions, and the conclusion that could be drawn from each.

Firstly, it is entirely possible that active dependency formation ceases once a single, verified dependency is constructed. In this case, active dependency formation should not be observed either in second coordinates or in post-verbal adjunct clauses. We call this prediction “ENTIRELY FILLER DRIVEN,” since it corresponds to a parser that is driven solely by the requirements of the filler. We assume in this case that once the parser establishes a gap site or grammatical role for the filler, then active dependency formation is terminated. This pattern of results would also contradict an aggressively bottom-up account of *wh*-dependency formation. Secondly, active dependency formation might be observed in both the context of a second coordinate and a post-verbal adjunct clause. This prediction is called “VERB DRIVEN,” since it is expected if active dependency formation reflects the parser’s drive to saturate the verb’s licensing requirements bottom-up. Finally, active dependency formation might be observed only in second coordinates, and not in post-verbal adjunct clauses. This prediction is called “CONTEXT DRIVEN”, since it suggests that knowledge about the distinction between coordinate gaps and parasitic gaps affects parsing decisions.

Active dependency formation expected? <i>Active dependency formation principle</i>	<i>Verb Position</i>	
	Second coordinate	Adjunct clause
<u>Entirely filler driven:</u> Interpret or license displaced filler.	✗	✗
<u>Verb driven:</u> Identify arguments of the verb	✓	✓
<u>Context driven:</u> Satisfy grammatical constraints	✓	✗

Table 4-1 Predictions for active dependency formation in multiple dependency constructions.

A fourth logical possibility is that active dependency formation only persists in parasitic gap environments. This outcome seems unlikely, and it would be puzzling, as it would imply that the parser undertakes an effortful decision to construct an optional dependency, but not an obligatory dependency.

One previous study addresses the real-time status of the Coordinate Structure Constraint. Pickering, Barton, & Shillcock (1994) used a Filled Gap Effect design to compare sentences like:

- (113) (a) I know what you *hit* the cupboard and *broke* the mirror **with** ____
 (b) I know that you *hit* the cupboard and *broke* the mirror with a ball

In sentence (a), the filler *what* is the argument of the preposition *with*, but there are two predicates that intervene between filler and gap: *hit* and *broke*. In self-paced reading, Pickering, Barton, & Shillcock (1994) found that reading times were elevated at the determiner following *hit* in (a), compared to a control sentence (b) with no *wh*-dependency. But no such effect was observed following *broke*. One interpretation of these data is that the parser attempted to form a dependency with *hit*, but then had to retract this analysis because there was an overt direct object. However, recognizing that it was inside a coordinate VP, the parser did not then attempt to form a dependency with *broke*, since doing so would have violated the CSC. This study thus suggests that the

parser can exhibit real-time sensitivity to the CSC, but by restricting certain analyses. However, the signal to apply the grammatical constraint in this experiment comes in the form of a parsing failure followed by the coordinator ‘and’. Once the initial object gap site has failed, the reader might not be expected to figure out where it is possible to resolve the *wh*-dependency. For these reasons, the present study seeks to find evidence for application of the CSC that comes from a positive measure and to use a design where the signal to apply the constraint follows a successfully constructed representation.

4.4.2 The Coordinate Structure Constraint and Active Dependency Formation I (Experiment 7)

The goal of Experiment 7 was to test whether or not active dependency formation is operative in second coordinate phrases and parasitic gap environments. To do so, we created sentences containing object extractions from an initial VP, where active dependency formation is uncontroversial. Beyond the first gap, sentences had two possible continuations: (i) a coordinate VP, in which case a second gap is obligatory; or (ii) an adjunct clause that could host a parasitic gap, in which case a second gap is optional. As our index of active dependency formation, we manipulated the semantic fit of the filler with the second verb.

4.4.3 Materials and Methods

Participants

The participants were thirty-seven native speakers of American English from the University community, who were paid \$10.

Materials

Experimental materials consisted of 24 sets of 4 conditions organized in a 2 × 2 factorial design that independently manipulated the conditions of VP structure and plausibility. An example materials set is given in Table 4-2. The second verb is the critical region, where active dependency formation is tested. It is in bold font; the relative clause head is underlined.

<i>VP Structure</i>	<i>Plausibility</i>	
<i>Coordinated VP</i>	<i>Plausible</i>	The <u>wines</u> which the gourmets were energetically discussing ___ or slowly sipping ___ during the banquet were rare imports.
	<i>Implausible</i>	The <u>cheeses</u> which the gourmets were energetically discussing ___ or slowly sipping ___ during the banquet were rare imports.
<i>Adjunct Clause VP</i>	<i>Plausible</i>	The <u>wines</u> which the gourmets were energetically discussing ___ before slowly sipping the samples during the banquet were rare imports.
	<i>Implausible</i>	The <u>cheeses</u> which the gourmets were energetically discussing ___ before slowly sipping the samples during the banquet were rare imports.

Table 4-2 Sample materials set for Experiment 7

The semantic fit of the filler with the first verb was plausible across all conditions, so that processing would not be disrupted before the critical region. In the examples given, one can equally discuss wines or cheeses. The plausibility factor manipulated the semantic fit of the filler with the second verb only on two levels. This manipulation provides a measure of dependency formation, as a slowdown is expected for implausible verb-argument combinations (e.g., Traxler & Pickering, 1996). If this slowdown occurs at the verb, before any signal in the input that there is a missing constituent, then we can conclude that dependency formation is active.

The VP structure factor manipulated the structure that contained the second verb, on two levels: Coordinated VP or Adjunct Clause VP. This manipulation allowed us to compare evidence for active dependency formation in contexts that require second gaps, and those that merely allow second gaps, as outlined in Table 4-1. Coordinate VP sentences always contained direct object gaps in the second VP, as required by the Coordinate Structure Constraint. The Adjunct Clause sentences provided environments where a gap might be anticipated, but they did not actually include parasitic gaps (p-gaps). Thus when the initial VP was followed by an adverbial clause, the parser could be lured to a p-gap analysis, but such an analysis was never confirmed in our materials. This design permitted the identification of effects due to active dependency formation, rather than bottom-up, gap-driven processing. No p-gaps were present in our experimental target materials, as they might be potentially highly noticeable constructions that participants could use to strategically identify the target conditions.

As discussed in footnote 52, there are several classes of exceptions to the CSC, e.g. “What did you go to the store and buy ___?”, all involving expressions of purpose, outcome, and temporal contiguity. Since these environments seem most felicitous under conjunction with *and*, three coordinators were used – *and*, *but*, and *or* – equally balanced across the materials to mitigate against the potential confound that comprehenders might believe that they are in one of the CSC-exempt environments. Additionally, *and* conjuncts were constructed to avoid CSC-exempt construals by coordinating events that seemed equally felicitous in either order of mention. An auxiliary ratings study, reported below, shows that participants were not treating the materials as CSC exempt. We fully balanced closed-class lexical items in two other ways: in the adjunct conditions, four

prepositions were used: *while*, *without*, *before*, and *after*. Across materials, both the relative pronouns *who* and *which* were used. Analyses of the results showed that these various lexical differences had no effect on reading times, and therefore we do not discuss these manipulations further.

There were two additional constraints on the materials. First, both VPs contained adverbial modifiers before the verbs. These were included to provide a strong cue for the upcoming verb and thus allow participants sufficient time to recognize the CSC environment. Secondly, adjunct clause verbs were in the past progressive form, since simple past forms less readily host parasitic gaps. We matched both these features across all conditions, for comparability, and across both VPs, for parallelism.

Seventy-two filler sentences were included. In order to prevent recognition of target structures, the fillers included syntactic features characteristic of the target items, such as progressive morphology, coordinate structures, filler-gap dependencies, and anomalous predicate-argument combinations. Since there were no parasitic gaps in the experimental design, parasitic gaps were included in some fillers. There would therefore be no implicit cue in the experiment to only expect gaps in non-p-gap positions.

Acceptability rating study

In order to verify the generalization that multiple dependencies are necessary in coordinate structures but optional in parasitic gap environments we conducted an off-line rating study. 32 participants who did not take part in the on-line study completed an acceptability questionnaire using a 5-point scale. These participants either received extra credit in an introductory linguistics class, or payment as part of another set of experiments.

Half of the target item sets from the on-line study were included in this study. The sentences were minimally-modified versions of the on-line items from the plausible conditions, crossing the structural role of the second VP, as a *Coordinate VP* or as an *Adjunct Clause VP*, with the presence of a gap in the second VP.

- (114) The wines which the gourmets were energetically discussing ____ ...
- (a) *Coordinate VP, Gap*
... or slowly sipping ____ during the banquet were rare imports.
 - (b) *Coordinate VP, No Gap*
... or slowly sipping the samples during the banquet were rare imports.
 - (c) *Adjunct Clause VP, Gap:*
... before slowly sipping ____ during the banquet were rare imports.
 - (d) *Adjunct Clause VP, No Gap:*
... before slowly sipping the samples during the banquet were rare imports.

The 12 item sets of these conditions were distributed by a Latin Square across four lists and combined with 12 filler items of similar length and complexity. Six fillers were uncontroversially acceptable sentences and six were highly unacceptable sentences. Each list was permuted in two pseudo-randomized versions.

Results for the experimental items are given in Table 4-3. The average rating for uncontroversially acceptable filler items was 4.3 ± 0.09 (standard error), and 1.8 ± 0.07 for unacceptable filler items.

	GAP IN SECOND VP?		<i>C.I.</i>
	YES	NO	
SECOND VP			<i>C.I.</i>
COORDINATE VP	4.2 ± 0.1	2.9 ± 0.1	0.49
ADJUNCT CLAUSE	4.2 ± 0.1	3.7 ± 0.1	0.35
<i>C.I.</i>	0.30	0.80	

Table 4-3 Experiment 7 Acceptability Ratings Summary

Average ratings are given for each sentence type with standard error. 95% confidence intervals on mean ratings differences, across participants and items, are reported in margins. N = 32.

A repeated measures ANOVA showed a main effect of structure ($F_1(1,31) = 9.8$; MSE: 12.8; $p < 0.01$), a main effect of the presence of a second gap ($F_1(1,31) = 49.5$; MSE: 77.0; $p < 0.0001$), and, crucially, an interaction of the two factors ($F_1(1,31)$: 11.4; MSE: 14.3; $p < 0.01$). In the coordinate VP condition, ratings were substantially lower if there was no gap in the second coordinate ($p < 0.001$). This pattern confirms that participants were sensitive to the CSC. P-gap conditions were highly rated, and there was no difference between the two conditions with multiple gaps. This result mirrors an earlier finding for subject p-gap constructions (Phillips 2006), and neutralizes the potential concern that p-gaps are somehow marginal structures. Surprisingly, there was a moderate decline in ratings if no gap was present in the adjunct clause ($p < 0.05$). P-gap and non-p-gap materials were predicted to be equally highly rated, so it was unexpected that the gapless adjunct conditions were rated slightly lower than their p-gap analogues. It is worth emphasizing the small size of this effect relative to the drop in ratings observed for CSC-violations. The mean difference, normalized against variance, between p-gap and non-p-gap continuations (Cohen's d) was 0.4, much smaller than the corresponding difference

in coordinate structures (*d*: 1.2). We suspect that this difference may reflect a bias in the construction of materials. P-gaps seem most felicitous when there is a close relation between the events or states expressed in the main and adjunct clauses. Creating a non-p-gap analogue in the adjunct clause meant inserting a theme argument that was necessarily closely related to the displaced theme of the main clause. It may have seemed to experimental participants, therefore, an awkward way to express an idea more naturally expressed by a p-gap or even a coordination. Irrespective of the cause of this small difference, this result strengthens the logic of the online study. If speakers fail to actively construct a second gap in the adjunct conditions, despite the high acceptability of a p-gap in this structure, then this would show that active dependency formation is not merely motivated to derive a natural interpretation from the input, but interacts strongly with grammatical principles.

Plausibility Rating Study

At the conclusion of the on-line study, 24 of the 37 participants were presented with a questionnaire containing 24 sentences, and asked to rate each on a five-point scale for plausibility, with ‘5’ being the most plausible and ‘1’ the least plausible. These data were collected to confirm the effectiveness of the plausibility manipulation. The sentences were simple SVO clauses derived from the critical VPs from the on-line study (e.g. *The gourmets discussed the {wine/cheese}*, *The gourmets drank the {wine/cheese}*). The rating study used 4 conditions in a 2×2 design, crossing the factors filler type (*Plausible*, *Implausible*) and verb type (first vs. second VP in the target items). The four resulting conditions for the 24 item sets were distributed by a Latin Square across four lists, each of which was then permuted into two pseudo-randomized versions.

On-line materials had been designed such that all fillers should be plausible at the first verb position, but should differ in plausibility at the second verb position. The rating study results confirmed this manipulation, as there was a strong interaction between filler class and verb class ($F_1(1,23): 70.6$; $MSE: 160.4$; $p < 0.0001$). Sentences containing first-position verbs were rated equally highly, regardless of whether the object corresponded to a *Plausible* or an *Implausible*-class filler (mean: 3.9, difference n.s.). Whereas the average rating for *Plausible*-class objects, as objects of the second-position verb, remained high and consistent with the first-verb ratings (mean: 3.9), the average rating for *Implausible*-class direct objects was much lower (mean: 1.8, $p < 0.001$). We conclude that the filler class by verb position manipulation met the desired specifications.

Procedure and Analysis

The procedure followed was identical to Experiment 7. Self-paced reading times for experimental sentences were examined region-by-region. Sentences were aligned word-for-word up to the second verb, such that each ordinal word position corresponded to a separate region. Evaluation of statistical reliability was carried out by repeated measures analysis of variance⁵³. Both participants and items analysis is presented in appendix tables. In the text, however, only the participants analysis is given (which has been argued to be the correct, sufficient test statistic for counterbalanced designs such as our own: Raaijmakers, et al. 1999).

4.4.3.1 Results

Comprehension question accuracy for the target sentences was high (average: 88.8%) and did not differ reliably across conditions.

Figure 4-2 presents the region-by-region condition means for regions 7-19. The omnibus repeated measures ANOVA report is given in Appendix C. Region-by-region condition means and test results for Regions 1-6 are not reported, as materials did not differ across the structural manipulation. Materials did differ in the plausibility manipulation, as the filler was introduced in Region 2. However, no effect of filler type was observed in that region, or in any region before Region 8.

⁵³ In the rest of the text, we also used linear mixed effects models (LMEM), which are in many respects superior to repeated measures analysis of variance (see Baayen, Davidson, & Bates, submitted). However RMANOVA reports for Experiments 7, 8 and 11a have been submitted for publication, so we present those analyses here. LMEM, with participant and item random effects, generally give convergent test results with RMANOVA.

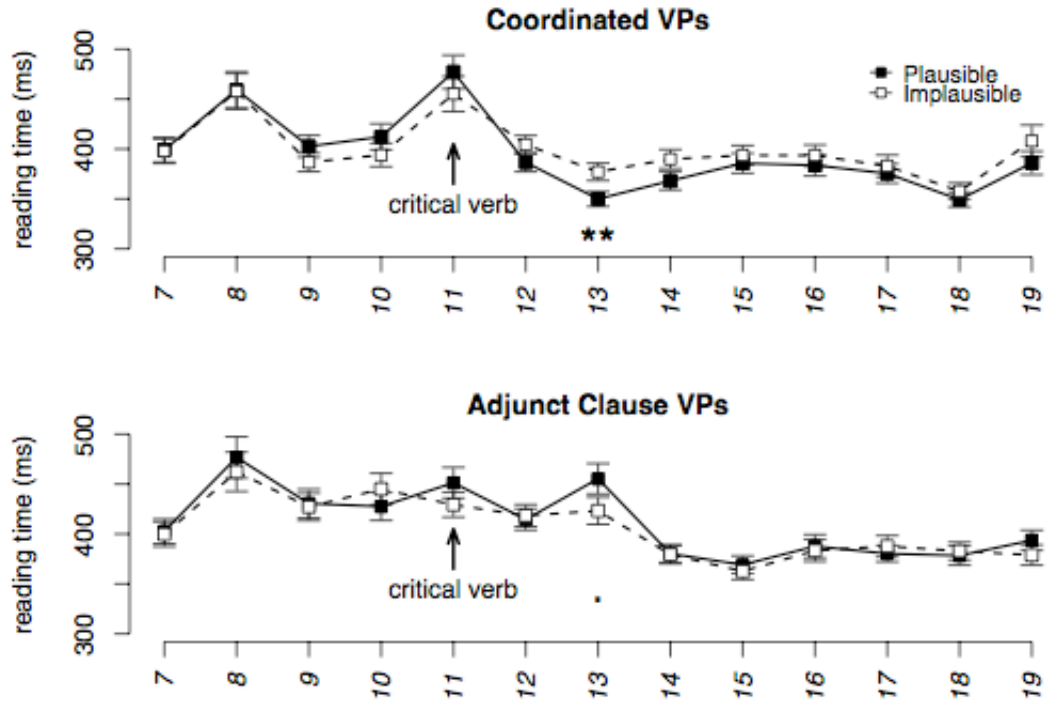


Figure 4-2 Experiment 7 Region-by-Region Reading Times

(The wines/cheese which the gourmets were)₁₋₆ energetically₇ discussing₈
...

Coordinate: or₉ slowly₁₀ sipping₁₁ during₁₃ the₁₄ banquet₁₅ were₁₆ rare₁₇
imports₁₈ from₁₇ ...

Adjunct: before₉ slowly₁₀ sipping₁₁ (the samples)/(some wine)₁₂₋₁₃
during₁₄ the₁₅ banquet₁₆ ...

Region-by-region reading times from the onset of the second VP to 8 regions beyond the critical verb. Punctuation indicates the result of a pairwise by-participants RMANOVA: p: ** < 0.01 * < 0.05 • < 0.10.

Preceding the second verb. Unremarkably, there were no main effects or interactions at Regions 7-8, the adverb and verb of the first VP. Materials were constructed to be structurally identical in these regions; and the plausibility norming survey reported in the Materials section confirmed that both types of fillers were equally plausible as the direct object of the verb in Region 8. Materials diverged structurally in Region 9, which consisted in a coordinator for Coordinated VPs and a preposition for

Adjunct Clauses, and there was a reliable increase in reading times for Adjunct Clauses in this region and in Region 10, the second VP adverb.

The second verb. No reliable effect of VP STRUCTURE, PLAUSIBILITY, or their interaction was observed at the critical second-VP verb in the participants analysis. A reliable main effect of PLAUSIBILITY was observed in the items analysis, due to plausible sentences being read slightly more slowly, although this was not reliable in pairwise comparisons.

Second VP post-verbal region. In the Coordinated VP conditions, implausible sentences were read more slowly in all regions subsequent to the critical verb, reaching significance two words downstream of the critical verb. In Adjunct Clause sentences, no consistent effect of plausibility was observed. In the region two words beyond the critical verb, there was a reliable interaction of plausibility and VP structure, due to slower reading times for implausible-filler sentences in coordinate VP conditions, and an opposite tendency in adjunct VP conditions. Planned pairwise comparisons in this region revealed a highly reliable slowdown due to implausibility for Coordinated VPs ($F_1(1,35)$: 7.87; MSE: 49793; $p < 0.01$). The opposite reading time pattern was observed in Adjunct Clauses, but it was only marginally significant ($F_1(1,35)$: 2.92; MSE: 106682; $p < 0.10$).

4.4.3.2 Discussion

Two conclusions follow from the reading time results in Experiment 7. Firstly, in Adjunct Clause VPs the lack of a slowdown due to implausibility suggests that dependency completion does not proceed actively in those environments. As no gap occurs in these conditions there is no bottom-up evidence to prompt dependency completion. Therefore the plausibility comparison within this condition constitutes a fair

test of purely active dependency completion both at the verb and in the following regions. Second, for Coordinated VPs, the presence of a reliable reading time slowdown for implausible fillers in the immediate post-verbal regions shows that filler-gap dependencies are constructed in the second VP. However this finding does not provide definitive evidence for active dependency formation, because of the timing of the effect. By the time the effect becomes reliable, there is bottom-up evidence for a gap in the form of missing constituents, and therefore dependency completion could have been cued from the input. Nonetheless, since spill-over effects are commonly observed in self-paced reading, this slowdown could reflect anomaly detection fed by active dependency completion.

If the effect in Coordinate VPs were indicative of active dependency completion, then we could conclude that the parser is sensitive to the Coordinate Structure Constraint, such that it recognizes when a filler that has already been successfully integrated with the first verb must participate in subsequent dependencies. This conclusion would be consistent with a grammatical licensing parser, in which active dependency formation is driven by the need to satisfy grammatical requirements. On the other hand, if the observed plausibility effect in Coordinate VPs reflects non-active bottom-up processes, then the difference between the plausibility contrasts in Coordinated VPs and the Adjunct Clause may have a more mundane explanation: there is a gap in one structure, but not the other. In order to determine whether active dependency completion in fact persists in coordinate VP environments, and thus tease apart the two possible interpretations of Experiment 7, Experiment 8 was designed such that effects of spill-over and of bottom-up gap-detection would be well separated in the time course of reading.

4.4.4 *The Coordinate Structure Constraint and Active Dependency Formation II (Experiment 8)*

In Experiment 7, the closeness of the critical verb and bottom-up evidence for a gap led to an ambiguous result. In order to separate the verb from the gap, ditransitive verbs were used in Experiment 8 as illustrated in (115). Consider the following example:

(115) The adhesive coating that the engineer **sprayed** the special test surfaces with ____
in his _____ new laboratory ...

In this example, the verb ‘spray’ subcategorizes for two internal arguments. When the second argument is relativized, the regions immediately following the verb do not provide evidence for a gap site. In a semantic fit manipulation, a slowdown due to implausibility could be observed at or beyond the verb but before bottom-up evidence for a missing constituent. It is thus possible to avoid the confound seen in Experiment 7. If the slowdown occurs at the verb or in the direct object regions, we can conclude that the parser actively completed the dependency, since it had to project the gap site before the input unambiguously signaled its location.

In this experiment we used coordinate VPs in which the second verb participates in *spray/load*-type locative constructions. As there was no implausibility effect in the Experiment 7 Adjunct Clause conditions, there were no such conditions in this experiment. Instead, the coordinate VPs were compared with length-matched conditions with a single filler-gap dependency, in order to compare patterns of dependency formation in a second coordinate VP with dependency completion in a single dependency, which was expected to be active uncontroversially.

4.4.4.1 Materials and methods

Participants

The participants were thirty-two native speakers of American English from the University community, who were paid \$10 to participate.

Materials and Procedure

Experimental materials consisted of 24 sets of 4 conditions organized in a 2 × 2 factorial design that independently manipulated the factors VP structure and plausibility. An example materials set is given in Table 4-4.

<i>VP Structure</i>	<i>Plausibility</i>	
<i>Coordinated VP</i>	<i>Plausible</i>	The <u>adhesive coating</u> that the talented engineer designed ___ or his boss and methodically sprayed the special test surfaces with ___ in his new laboratory could make the company lots of money.
	<i>Implausible</i>	The <u>computer program</u> that the talented engineer designed ___ or his boss and methodically sprayed the special test surfaces with ___ in his new laboratory could make the company lots of money.
<i>Single VP</i>	<i>Plausible</i>	The <u>adhesive coating</u> that the talented engineer from the high-tech aerospace firm methodically sprayed the special test surfaces with ___ in his new laboratory could make the company lots of money.
	<i>Implausible</i>	The <u>computer program</u> that the talented engineer from the high-tech aerospace firm methodically sprayed the special test surfaces with ___ in his new laboratory could make the company lots of money.

Table 4-4 Sample materials set for Experiment 8

The VP structure factor manipulated the VP structure containing the critical verb. Coordinate VP sentences contained two coordinated VPs, as in Experiment 7. The critical verb was the second verb in the coordinate. *Single VP* sentences contained only a single verb. A five-word PP modifier was attached to the relative clause subject in this condition, so that the ordinal position of the critical verb matched the position of the second verb in the Coordinate VP conditions.

The plausibility factor manipulated the semantic fit of the filler with respect to the critical verb by creating two classes of fillers. Plausible fillers were plausible as the direct object of both the first and second verb. Implausible fillers were plausible as the direct object of the first verb, but implausible as the direct object of the second verb. As in Experiment 7, the plausibility manipulation provided a measure of dependency formation.

Syntactically alternating locative verbs in the *spray/load* class were selected for the critical verb in each item set (Anderson, 1971; Fraser, 1971; Pinker, 1989; Rappaport & Levin, 1986 inter alia). Verbs from this class tend to impose greater semantic restrictions on both of their arguments than do simple datives, like *give*, or benefactives, like *buy*. This feature made it feasible to design a large number of items with a semantic fit manipulation that applies to the filler, regardless of the internal position it occupies. It is important that the filler be implausible both as direct or oblique object. Previous research has suggested that if a verb has multiple syntactic frames or argument positions, then fillers that are solely implausible as a direct object do not elicit a slowdown in filler-gap constructions (Boland et al., 1995; cf. Pickering & Traxler, 2003). Consider the verbs below, with arguments in the specified configuration:

(116) splash/spray/sprinkle/spread NP₁ with NP₂

For these verbs, NP₁, referred to as the *ground* argument, must typically be a concrete entity, while NP₂, referred to as the *figure* argument, must typically be either a liquid, plastic or particulate substance. There is nothing crucial about what the semantic selectional restrictions are, just that they tend to exist for both arguments in *spray/load* verbs. Twelve *spray/load* verbs were chosen from Levin (1993) as the critical verbs.

Each verb was used in two item sets. The first argument of the critical verb was always four words long, providing a large region between the verb position and the first direct evidence for a gap position. Any slow-down due to implausibility observed within this spill-over region could be attributed to active dependency formation, since direct evidence for the gap does not occur until after the subcategorized preposition *with*. The bottom-up cue for the gap was very strong, as two prepositions occurred in sequence (“sprayed the special test surfaces with in his new ...”).

Three further design constraints applied. As in Experiment 7, pre-verbal adverbs were used in all VPs. Unlike Experiment 7, all verbs appeared in simple past tense form. The progressive verb forms used in Experiment 7 were required by the Adjunct Clause conditions included in that study, which were not present in Experiment 8. Finally, the complementizer *that* was used to signal the onset of the relative clause, instead of the pronouns *who/which*. Both the complementizer and the relative pronoun serve as effective signals to the parser for a relative clause, and plausibility effects are obtained in both environments (First author, unpublished pilot results). In the context of these materials, the complementizer was judged to be more natural. Seventy-two fillers were adapted from the fillers in Experiment 7, so that the distribution of sentence lengths in fillers matched the distribution of target items.

Procedure and Analysis

Procedure and analysis was identical to Experiment 7. One participant who failed to perform the task as instructed was removed from further analysis.

Plausibility Rating Study

At the conclusion of the on-line study, 16 of the 32 participants were presented with a questionnaire containing 24 sentences, and were asked to rate each on a five-point scale for plausibility. The sentences were simple SVO versions of the VPs from the on-line study, and the design was the same as Experiment 7.

Sentences containing first-position verbs were rated equally highly, regardless of whether the object corresponded to Plausible or Implausible fillers (mean for both 4.2; difference n.s.). The average rating for Plausible-class objects as objects of the second-position verb remained high (mean: 3.7), but the average rating for Implausible-class direct objects was considerably lower (mean: 1.5, $p < 0.0001$). Thus, the filler class by verb position manipulation met the desired specifications.

4.4.4.2 Results

Question-answering accuracy was uniformly high. For the 24 experimental targets, accuracy was 92.3% overall. There were modest and reliable differences due to the VP STRUCTURE and PLAUSIBILITY manipulations. For coordinate structures accuracy was 95.7% for plausible conditions and 90.3% for implausible conditions; for single VP controls accuracy was 89.2% for plausible conditions and 94.1% for implausible conditions. A logistic mixed-effect model estimated that all factors had an odds ratio significantly different from zero. In contrast to Experiment 7, this model was significantly better than a null model that attributed all variation to participants, with no effects for each condition ($\chi^2(12): 2765.0; p \sim 0$).

Figure 4-3 presents the region-by-region condition means, segregated into two pair-wise comparisons for regions 6 to 20. These regions extend from the lexical offset of

the embedded subject head noun to 3 words beyond the preposition heading the critical verb's figure argument. The omnibus repeated measures ANOVA report is given in Appendix C for Regions 6-20. Materials did not differ in Regions 1-5, up to the lexical offset of the subject head noun, apart from the filler manipulation. No RT differences were observed in those regions.

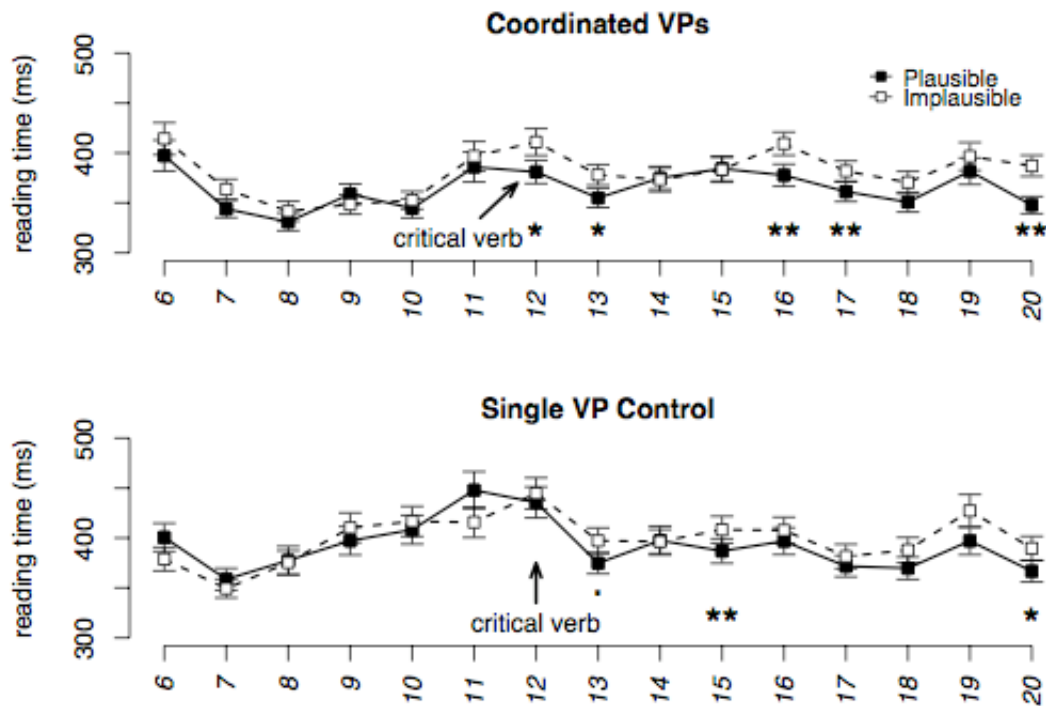


Figure 4-3 Experiment 8 Region-by-Region Reading Times

(The adhesive coating/computer program that the talented engineer)₁₋₅ ...

Coordinate: designed₆ for₇ his₈ boss₉ and₁₀ methodically₁₁ sprayed₁₂

Single VP: from₆ the₇ high-tech₈ aerospace₉ firm₁₀ methodically₁₁
sprayed₁₂

the₁₃ special₁₄ test₁₅ surfaces₁₆ with₁₇ in₁₈ his₁₉ laboratory₂₀ ...

Region-by-region reading times from the offset of the subject head noun to 8 regions beyond the critical verb, with example text for each region. Punctuation indicates the result of a pairwise by-participants RMANOVA: p: ** < 0.01 < * < 0.05 < • < 0.10.

Preceding the second verb. Materials diverged at the offset of the subject head

noun: for Coordinate VP sentences, a verb followed the subject noun, and for Single VP

sentences, a PP followed the subject noun. Accordingly, reading times differed as a function of VP structure, beginning three words downstream from the subject head noun, in Region 8, and persisting until three words downstream, in Region 14.

The second verb. At the second verb, in addition the effect of VP structure described above, there was a clear effect of plausibility, due to slower reading times for Implausible filler conditions. However, Coordinate VP sentences showed this contrast most strongly, with a variance-normalized mean difference of 0.18, compared to 0.04 for Single VP sentences. Moreover, as Figure 4-3 shows, the plausibility contrast was robust in pairwise comparisons for Coordinate VPs ($F_1(1,30)$: 5.63; MSE: 71964; $p < 0.05$) but not for Single VPs ($F_1(1,30) < 1$). However the interaction of plausibility and VP structure was not significant in this region ($F_1(1,30) < 1$)

The ground argument region. In the ground argument regions following the critical verb (Regions 13-16), we observed persistent effects of plausibility, especially at the determiner in Region 13 and at the noun in Region 16. The slow-down due to implausibility was of comparable size for both coordinate VPs and long single VPs in Region 13 (in raw ms). The effect was reliable in pairwise comparisons for Coordinate VPs ($F_1(1,30)$: 7.19; MSE: 45776; $p < 0.05$), but only marginally so for Single VPs ($F_1(1,30)$: 3.68; MSE: 56898; $p < 0.10$). Region 14 showed no effect of plausibility. Region 15 showed an effect only for Single VPs ($F_1(1,30)$: 7.59; MSE: 59747; $p < 0.01$). Region 16 showed a strong effect of implausibility for Coordinate VPs ($F_1(1,30)$: 10.36; MSE: 75008; $p < 0.01$), and a much weaker, and unreliable, effect for Single VPs ($F_1(1,30)$: 2.50; MSE: 29611; $p=0.12$).

The gap region. An effect of plausibility was present in Regions 17-20, which corresponded to the regions containing the preposition that selects the gap site and subsequent regions. Since Regions 17-18 were two prepositions in sequence (e.g. “with before”) they provided clear evidence of a missing constituent. Once again, the size and location of specific effects differed slightly across structural conditions, with Coordinate VP sentences showing contrasts in more regions, and displaying the largest contrast in Region 20.

4.4.4.3 Discussion

The reading time results from Experiment 8 showed that an implausible filler led to a slowdown at the second verb in a Coordinate VP structure and in subsequent regions in its argument field. The timing of this slowdown provides evidence that the second gap in coordinate structures is constructed actively, because it occurs unambiguously before the direct evidence of a gap position in the second coordinate. The use of *spray/load* verbs made it possible to put sufficient distance between the verb and the gap position, such that we can confidently interpret the slowdown as an effect of active dependency formation.

In Experiment 8, in comparison to Experiment 7, the effect of filler-verb plausibility appeared on the verb itself, and not one or two words downstream. There were differences in the experimental materials that could explain why the effect emerged earlier in Experiment 8. One important difference is that in Experiment 7 the second verb occurred only two words after the first verb, whereas in the Experiment 8 materials a short 3-word PP occurred in the first VP. To see why this could make a difference, consider that in order to detect implausibility comprehenders must not only posit a gap

location, but must also retrieve and integrate the filler syntactically and semantically. Even when there is no disruption in dependency formation many processes must take place. It is more likely that these processes could have persisted to the second verb in Experiment 7 than in Experiment 8, and thus could have delayed the emergence of an implausibility effect by one or two words.

A surprising finding in this experiment was that implausible fillers had a more disruptive effect on processing in a second coordinate than in the single dependency control conditions. In Experiment 11 (section 4.6.1), we delve into the reasons for this in greater detail.

4.4.5 General discussion of Experiments 7 & 8

Experiments 7 & 8 tested whether the parser persists in actively and incrementally constructing gaps in multiple-gap dependencies, even after the first filler-gap relationship has been successfully constructed. The goal of the study was to assess whether grammatical constraints actively direct the formation of *wh*-dependencies, controlling potential retrieval events in a top-down fashion. Two kinds of multiple-gap dependencies were compared: across-the-board extraction from coordinate VP structures, and parasitic gaps inside post-verbal adjunct clauses. Crucially, multiple gaps are obligatory in coordinate structures, but parasitic gaps are always optional. In Experiment 7 these generalizations were confirmed in an off-line rating study. The self-paced reading studies in Experiments 7 and 8 showed a strong effect of the semantic fit between the *wh*-phrase and the verb in the second coordinate of a coordinated VP. The use of the ditransitive *spray/load*-type verbs in Experiment 8 confirmed that this effect emerged before direct

evidence for the gap position. These results suggest that comprehenders re-engage in active dependency completion when they detect a coordinate structure containing a gap.

Comprehenders are sensitive in real-time to the grammatical implications of building multiple gap constructions. The experiments reported here add to previous findings that comprehenders are highly accurate in locating the tail of a *wh*-dependency. The advantage of testing the Coordinate Structure Constraint was that, because the index of dependency formation was a positive signal, we were able to avoid the specter of null effects stemming from an overload in complexity which other island studies are liable to. The grammatical generalizations about across-the-board extraction and parasitic gaps actively guide parsing decisions. The contrast between obligatory and optional gap environments strongly suggests that fillers are retrieved at the prompting of a parsing mechanism that keeps track of outstanding requirements in the syntactic context. Encountering a potential parasitic gap environment after completing a filler-gap dependency does not re-engage dependency completion mechanisms, despite the fact that an additional gap is fully acceptable and that doing so would saturate the argument structure of the verb sooner than waiting for the input.

4.5 The fidelity of retrieval in wh-dependency formation

4.5.1 Introduction

Experiments 7 and 8 concerned grammatical accuracy in locating the tail of a dependency. That the comprehender is accurate in locating the tail of a *wh*-dependency is consistent with the generalization offered in section 4.2 that predictive dependencies are grammatically most faithful. The next question is whether that fidelity extends not only to postulating a gap, but also to recovering the filler itself. Even if the parser anticipates

legal gap sites or carries forward a small amount of information to license them, it must still recover the full filler encoding at some point in comprehension.

Van Dyke & McElree (2006) argued that multiple candidates compete for the retrieval cues provided by the verb in filler-gap dependency completion (section 4.3.3). This is a natural consequence of the memory architecture, if there is overlap between the encoding of the filler representation and other constituents in memory. We raised a few concerns over their findings, however. Firstly we asked whether their memory load manipulation specifically influenced filler-gap dependency formation, or whether the effect arose elsewhere like in lexical access. Secondly we questioned whether an appropriate, linguistically-motivated feature structure could be devised in which the memory load items were serious competitors for a full match. In Experiments 9 and 10 we examine syntactic configurations that offer a tighter test of the hypothesis that similar constituents induce similarity-based interference during filler-gap dependency completion.

We conjectured that for a verb retrieving a filler, the strongest competitors for that filler would be constituents that are fillers in other dependencies, or, specifically, other *wh*-elements in clause-edge positions. For example, consider the sentence in (117). The critical verb is ‘revealed’ which hosts the gap for the extracted *wh*-phrase ‘what’ in the embedded question.

(117) The biographer asked what the idea that the professor often defended to his colleagues potentially **revealed** ____ about his character.

By hypothesis, at ‘revealed,’ the parser initiates a retrieval for the filler. However, a relative clause was recently processed (underlined), which also contained a filler-gap dependency. The question posed is whether ‘reveal’ would encounter difficulty retrieving

‘what’ because there is also a filler in the edge of the relative clause (like an unpronounced copy of *idea* or its coindexed operator). This filler, which corresponds to the relative clause head ‘idea’, is a reasonable semantic match for ‘reveal,’ but more importantly it has most of the syntactic properties that would license it as theme for ‘reveal.’ Crucially it occupies the same local syntactic position [Spec,CP] as ‘what,’ or putatively has a shared feature like [+*wh*]. What it lacks is relational: it is not in a c-command relationship with ‘reveal’ (or its gap) or in the edge of the immediate clause that dominates ‘reveal.’ It is therefore plausibly a full match in terms of the inherent features it bears. Alternatively, it is a near full match, if the retrieval structure includes a Clause/Context cue (as we proposed for agreement attraction in Chapter 3).

4.5.2 Experiment 9

In Experiment 9 we test for difficulty that can be attributed to similarity-based interference by crossing two experimental factors: whether the embedded question clause contains an extracted *wh*-phrase (*what* v. *if*), and whether the intervening complex subject is a relative clause or a sentential complement. If retrieval identifies multiple candidate filler phrases based on feature overlap, then resolving the filler-gap dependency in an embedded *wh*-question should be harder when the intervening subject contains a relative clause than when it contains a sentential complement. If difficulty at the verb is truly due to resolving a filler-gap dependency, then this difficulty should be selective to embedded *wh*-questions, and not observed in embedded *if*-clauses.

4.5.2.1 Materials and methods

Participants

Participants were 31 native speakers of English from the University of Maryland community with no history of language disorders. Participants were paid \$5 for their participation.

Materials

24 item sets were created with sentences containing an embedded clause whose subject was complex. Two experimental factors were crossed: embedded clause type (*wh*, *if*) and interference load of the complex subject. Interference load was defined as whether or not the complex subject contained a filler-gap dependency: either an object relative clause (high interference) or a sentential complement (low interference).

Sample set of experimental items for Experiment 9		
<i>Embedded clause type</i>	<i>Interference</i>	The biographer asked ...
<i>Wh-question</i>	High	... what the idea that the professor often defended ___ to his colleagues potentially revealed about his character.
	Low	... what the idea that the professor often deferred to his colleagues potentially revealed about his character.
<i>If question</i>	High	... if the idea that the professor often defended ___ to his colleagues potentially revealed anything about his character.
	Low	... if the idea that the professor often deferred to his colleagues potentially revealed anything about his character.

Table 4-5 Sample materials set for Experiment 9

The materials were distributed across 4 lists by a Latin Square. Each participant would therefore see six items per condition.

Procedures and Analysis

Presentation and analysis details were as described for Experiment 6.

4.5.2.2 Results

Comprehension accuracy is reported in Table 4-6.

Embedded clause type	Interference		
	<i>High</i>	<i>Low</i>	
<i>Wh-question</i>	73%	87%	80%
<i>If-question</i>	82%	81%	81%
	77%	84%	81%

Table 4-6 Comprehension question accuracy for Experiment 9

In comprehension accuracy there was a main effect of embedded clause type (β : 0.59 ± 0.51 ; $p < 0.05$), such that *if*-questions were overall answered more accurately; a main effect of interference load (β : 1.1 ± 0.56 ; $p < 0.001$), such that low interference conditions were overall answered more accurately; and an interaction of the two conditions (β : -1.2 ± 0.78 ; $p < 0.005$). This interaction reflects the fact that the interference factor only affected responses in embedded *wh*-clause conditions.

Reading time data is reported in Figure 4-4 for *wh*-clause conditions and Figure 4-5 for *if*-clause conditions. Effects are reported only for reliable differences and for the critical verb.

Region 11: Complex subject clause verb. Reading times at the verb inside the intervener clause showed a main effect of embedded clause type: reading times were slower at the intervener verb inside *wh*-clauses ($\Delta\mu$: 35 ms; 95% C.I. [2 ms, 66 ms], $p < 0.05$).

Although there is a sizeable numerical slowdown associated with the intervener verb in High interference conditions ($\Delta\mu$: 21 ms), it is not reliable (95% C.I. [-22, 80], $p < 0.3$). At the intervener verb in High interference conditions, the intervener clause

cannot unambiguously be identified as a relative clause, though the transitive verb is a strong cue (cf. Solomon & Mendelsohn, 2004).

Region 16: Critical verb. At the critical verb, there were no reliable effects. A slowdown was observed for low interference conditions that was nearly marginal ($\Delta\mu$: 25 ms; 95% C.I. [-20 ms, 60 ms], $p \sim 0.10$). Restricting comparison to *wh*-clauses, i.e. those conditions in which a filler-gap dependency is resolved at the verb, the difference is also numerically slower for the low interference condition, though also not reliable ($\Delta\mu$: 24 ms; 95% C.I. [-28 ms, 58 ms]).

Region 19-20: Sentence-final regions. Region 19 shows a main effect of clause type, such that *if*-clauses are read much faster than *wh*-clauses ($\Delta\mu$: 125 ms; 95% C.I. [96 ms, 157 ms], $p < 0.001$). This difference likely reflects the lexical differences between the two conditions: in the *if*-clause, the Region 19 lexical item is a determiner or possessive pronoun whereas in *wh*-clauses it is a noun. Conditions were lexically unmatched following the critical verb, since *if*-clauses by necessity had an overt argument whereas *wh*-clauses had a gap.

Region 20 also shows a main effect of clause type, with *if*-clauses being read much faster than *wh*-clauses ($\Delta\mu$: 75 ms; 95% C.I. [24 ms, 162 ms], $p < 0.05$). As in Region 19, the two levels of the clause are not matched lexically.

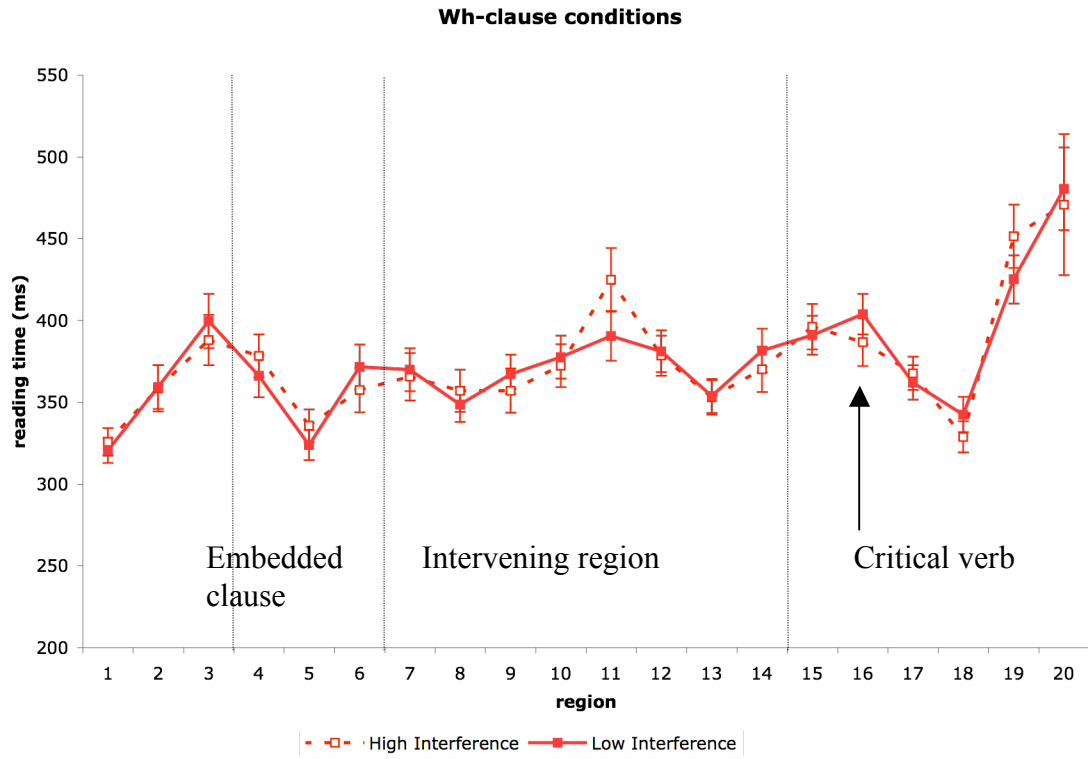


Figure 4-4 Experiment 9 Reading time results: Wh-clause Conditions

The₁ biographer₂ asked₃ what₄ the₅ idea₆
 [that₇ the₈ professor₉ often₁₀ defended₁₁/deferred₁₁ to₁₂ his₁₃ colleagues₁₄]
 potentially₁₅ revealed₁₆ about₁₇ his₁₈ character₁₉ ...₂₀

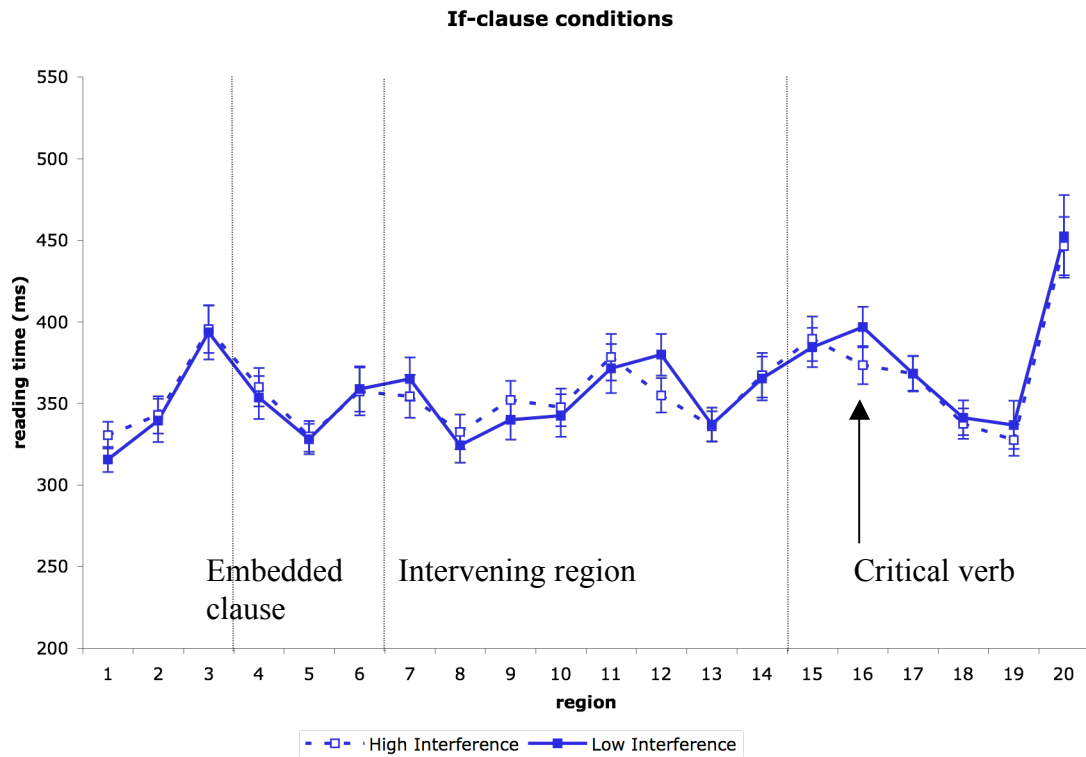


Figure 4-5 Experiment 9 Reading time results: If-clause Conditions

The₁ biographer₂ asked₃ if₄ the₅ idea₆
 [that₇ the₈ professor₉ often₁₀ defended₁₁/deferred₁₁ to₁₂ his₁₃ colleagues₁₄]
 potentially₁₅ revealed₁₆ anything₁₇ about₁₈ his₁₉ character₂₀

4.5.2.3 Discussion

In this experiment we tested whether or not the presence of an additional, but irrelevant filler-gap dependency, would make resolving a target filler-gap dependency more difficult. On the hypothesis that the fillers at the edge of both the embedded clause and relative clauses were substantially similar, a retrieval-based account of filler-gap completion predicts an interaction between clause type and interference load. The reading time data did not bear out this prediction at the critical verb. There was no indication that completing an embedded *wh*-clause filler-gap dependency led to greater processing times when a relative clause had recently been processed. As we noted in Chapter 3, though,

participants may trade speed for accuracy, or simply accept the wrong constituent and never verify the c-command relationship. If so, then the interference should be reflected in comprehension accuracy. Indeed, there we saw a selective effect of interference load on *wh*-clause conditions only, in the expected directions.

Interpreting the discrepancy in on-line and off-line data is complicated somewhat, though. The on-line data are localized to specific sentence regions, whereas the off-line data reflect the outcome of numerous processing events. This concern is relevant, because we observed selective difficulty in embedded *wh*-clause/relative clause conditions prior to the critical verb. We observed increased difficulty in the subject-attached clause for relative clauses. Moreover this difficulty was greatest for the verb inside a relative clause contained in an embedded *wh*-clause (though this effect was near-marginal). At the point of processing the verb, the relative clause and sentential complement analyses cannot be disambiguated; however the use of transitive verbs in the relative clause condition could have been a strong cue to differentiate the two analyses, either generally, or particularly if picked up in the experimental context. Pearlmutter & Mendelsohn (2000) have argued that in relative clause/sentential complement ambiguous strings, the relative clause analysis is at least as a strong competitor as the sentential complement. This ambiguity in analysis, combined with the search for licit retrieval sites, could explain why comprehenders experienced selective difficulty at the RC-contained transitive verbs inside embedded *wh*-clauses. If we grant this much, then the difficulty observed in the relative clause region could account for the interaction in the comprehension accuracy data.

Offline and online data were at odds in this experiment. The online data suggest that the mere presence of other filler-like constituents did not make resolving the filler-gap dependency at the critical verb more difficult. The offline data, however are consistent with this prediction, though somewhat equivocal because they do not localize the source of the difficulty. In Experiment 10, we attempt the same basic manipulation, though without having nested dependencies.

4.5.3 *Experiment 10*

In this experiment, we created a configuration in which the interference load region was not nested inside the target filler-gap dependency. To do so, we embedded the critical filler gap dependency in a relative clause attached to the sentence object. The interference region was defined as the relative clause attached to the sentence subject. The configuration is illustrated schematically below:

(118) Subj [_{RC} ---*Interference region* ---] V Obj [_{RC} ... V_{Critical} ____]

The critical filler-gap dependency was a object extraction from inside a full relative clause. The sentence subjects were made high interference regions by attaching a full relative clause, as in (119a). Consequently in high interference structures there were (putatively) two relative clause operators occupying identical structural positions. To create a low interference region, we used subject infinitival relatives, as in (119b), which have been argued to be reduced relative clauses (Kjellmer, 1975; Bhatt, 1999; cf. Kayne, 1994)⁵⁴.

(119) (a) *High Interference / Object Relative Clause*
 The brightest student_i [_{CP} Op_i that _____i took the test] wrote an essay_j [_{CP} Op_j that the instructor praised _____j for its mature style]

⁵⁴ We thank Alan Munn for suggesting this comparison.

(b) *Low Interference / Object Relative Clause*

The brightest student_i [_{PreDP} _____i to take the test] wrote an essay_j [_{CP} Op_j that the instructor praised _____j for its mature style]

There are clear distributional differences between subject infinitival relatives and subject full relatives that lead us to assume the filler in a full relative clause is more similar to the filler in the target region, than the head of the subject infinitival is. Bhatt (1999) argues that the subject infinitival relatives do not involve A' movement to the Spec of a [+wh] C. This analysis is supported by a pattern of observations that subject infinitivals lack crucial properties characteristic of A' dependencies in full relatives. Comparison with non-subject infinitival relatives reveals this is not a property of the infinitive per se. In comparison to both full relatives and non-subject infinitival relatives, subject infinitival relatives do not allow a complementizer, a relative pronoun, or long-distance movement. As well, unlike full relatives and non-subject infinitival relatives, there is no way for an [Op, t] chain to receive case in subject infinitivals. Whether Bhatt's analysis is necessary to capture these differences is not crucial: for our purposes what is crucial is that the filler's encoding in subject infinitivals be distinct from the filler's encoding in full relatives. Let us assume for discussion the relevant distinction is that full relative clause fillers are marked [+wh], that subject infinitival fillers lack this feature, and that this feature is highly weighted in the retrieval structure used in filler-gap dependency completion.

As a control condition we replaced the object-attached relative clause with a coordinate clause continuation containing the same subject and verb. A pronoun was inserted in object position, so that the control sentences would express the same thematic relations:

(120) The brightest student_i [that took the test]/[to take the test] wrote an essay and the instructor praised it for its mature style.

Crucially ‘praised’ in the control conditions does not require a retrieval operation, as it is not within a filler-gap dependency. Therefore we can test for an interaction between interference load and filler-gap dependency completion. If the system uses a set of generic structural cues to retrieve the filler during active dependency completion, the grammatically unavailable filler in the high interference sentences should compete with the actual filler for the dependency.

4.5.3.1 Materials and methods

Participants

Participants were 32 native speakers of English from the University of Maryland community with no history of language disorders. Each received partial course credit in an introductory linguistics class.

Materials

24 item sets were created with sentences containing an embedded clause whose subject was complex. The experimental factors were critical verb clause (*object relative*, *coordinated clause*) and interference load of the complex subject. Interference load was defined as whether or not the complex subject contained a full relative clause dependency: either there was a full relative (high interference) or an infinitival relative (low interference).

Sample set of experimental items for Experiment 10		
<i>Critical verb clause</i>	<i>Interference</i>	The brightest student
<i>Object relative</i>	<i>High</i>	... that took the test wrote an essay that the instructor praised ___ for its mature style.

	<i>Low</i>	... to take the test wrote an essay that the instructor praised _____ for its mature style.
<i>Coordinate clause</i>	<i>High</i>	... that took the test wrote an essay and the instructor praised it for its mature style.
	<i>Low</i>	... to take the test wrote an essay and the instructor praised it for its mature style.

Table 4-7 Sample materials set for Experiment 10

Superlatives and ordinals like “brightest” or “first” were used as prenominal modifiers of the subject across all conditions. We wanted to make interpretations as similar as possible across conditions. However, subject infinitival relatives permit modal interpretations (i.e., “the man to answer your question is Bill” ≈ “the man who should/can answer your question is Bill”). These modal interpretations can be quashed under ordinals, superlatives and *only* (i.e. “the first man to answer your question was Bill” ≈ “Bill was the first man who answered your question”).

Finally critical verbs were chosen that were semantically compatible with either the grammatical or interfering filler (e.g., it is fine to praise either students or essays).

The materials were distributed across 4 lists by a Latin Square. Each participant would therefore see six items per condition.

Procedures and Analysis

Presentation and analysis details were as described for Experiment 6.

4.5.3.2 Results

Comprehension accuracy is reported in Table 4-6. Overall accuracy was 93%. There were no significant differences between conditions. A model of the data with a single fixed coefficient could not be distinguished from the full model ($\chi^2 : 2.4$, d.f.: 3, n.s.).

Critical verb	Interference		
	High	Low	
<i>Object relative clause</i>	93%	95%	94%
<i>Coordinated clause</i>	91%	92%	92%
	92%	93%	93%

Table 4-8 Comprehension question accuracy for Experiment 10

Standard error of the cell means is 2% for all conditions.

Reading time data is reported in Figure 4-6 for object relative clause conditions and Figure 4-7 for the coordinated clause control conditions. Effects are reported only for reliable differences and for the critical verb.

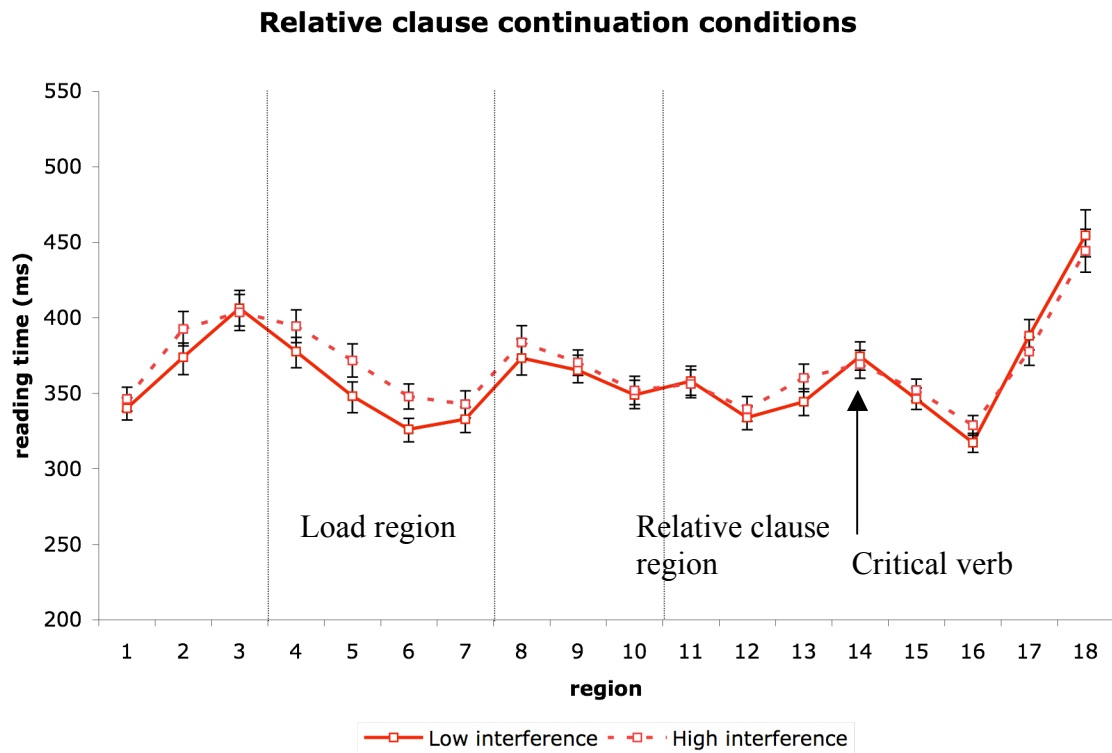


Figure 4-6 Experiment 10 Reading time results: Relative clause conditions

The₁ brightest₂ student₃ [to₄/that₄ take₅/took₅ the₆ test₇] wrote₈ an₉ essay₁₀ that₁₁ the₁₂ instructor₁₃ praised₁₄ for₁₅ its₁₆ mature₁₇ style₁₈.

Coordinated clause continuation conditions

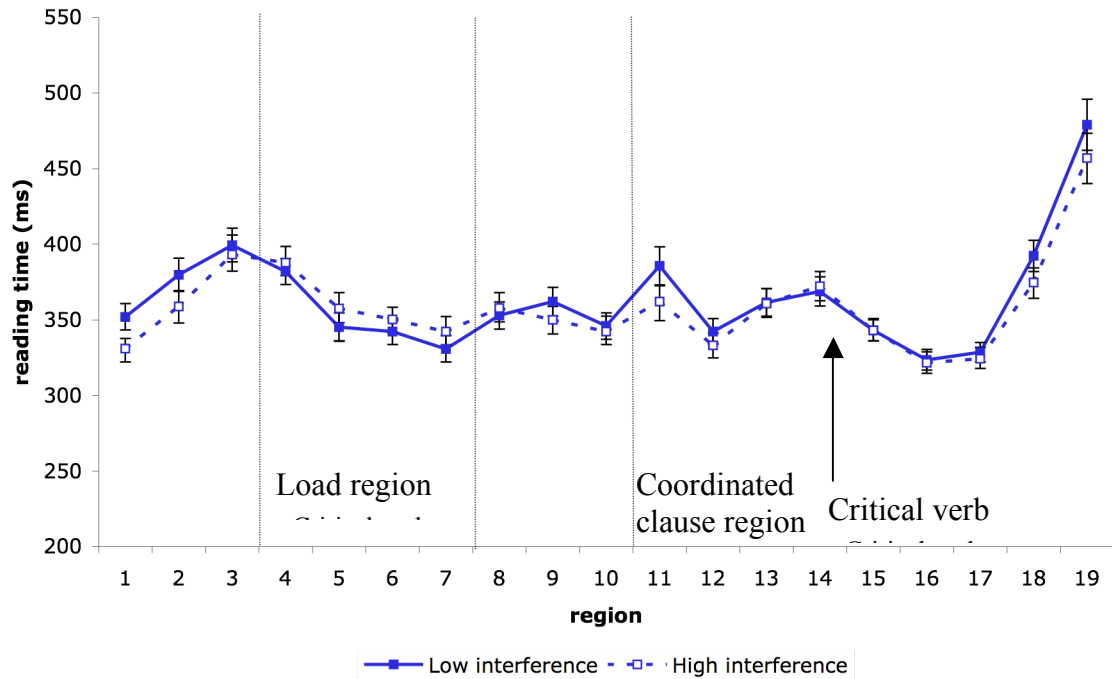


Figure 4-7 Experiment 10 Reading time results: Coordinated clause conditions

The₁ brightest₂ student₃ [to₄/that₄ took₅ the₆ test₇] wrote₈ an₉ essay₁₀
and₁₁ the₁₂ instructor₁₃ praised₁₄ it₁₅ for₁₆ its₁₇ mature₁₈ style₁₉.

Regions 1-2: Subject determiner and adjective. In Region 1 there was a reliable interaction between interference load and structural continuation factors (*high:conj* $\Delta\mu$: -25 ms; 95% C.I. [-47,-1], $p < 0.05$). This interaction reflects a pairwise difference between interference load conditions for coordinate clause sentences, but not relative clause sentences. Because all conditions are exactly matched in the first region, this difference must be spurious. The difference persists into Region 2, where the same kind of interaction was observed (*high:conj* $\Delta\mu$: -35 ms; 95% C.I. [-67,-7], $p < 0.05$). However the RTs are not different in Region 3, corresponding to the subject head noun.

Region 5: Subject relative clause verb. In Region 5 there was a reliable main effect of interference load, such that full relative clauses were read more slowly than infinitival relative clauses ($\Delta\mu$: 22 ms; 95% C.I. [1, 21], $p < 0.05$). Note that full relative clause verbs were longer and morphologically more complex in this region (e.g., V+ed v. V).

Region 6: Direct object determiner. In Region 6 there was also a reliable main effect of interference load, such that full relative clauses were read more slowly ($\Delta\mu$: 21 ms; 95% C.I. [4, 37], $p < 0.05$).

Region 11: Conjunction/relative clause complementizer. In Region 11 there was a reliable main effect of clause continuation type, such that coordinated clause conditions were read more slowly ($\Delta\mu$: 24 ms; 95% C.I. [4, 49], $p < 0.05$). This region corresponds to a conjunction, in coordinated clause conditions, or the object-attached relative clause complementizer, in relative clause conditions.

Region 14: Critical verb region. In Region 14, the critical verb region, there were no reliable effects. The crucial comparison is between interference load conditions, in an object relative clause (where the filler-gap dependency can first be resolved); and there was no difference between the two load conditions ($\Delta\mu$: 5 ms; 95% C.I. [-24, 20], n.s.).

Sentence-final regions. Region 18 shows a main effect of clause continuation type, such that coordinated clause continuations were read more quickly than relative clause continuations ($\Delta\mu$: 75 ms; 95% C.I. [-47, 107], $p < 0.001$). Because this region occurs after the critical verb, the conditions are lexically not aligned. This comparison involves an adjective in coordinate clause continuations and the sentence final noun in the relative clause continuations, so a difference is not surprising.

If we align the last four regions of the sentence, corresponding to the category sequence, P-D-A-N, so that regions are lexically matched, then the only region in which the two clause continuation conditions differ is in the P region: Region 15 for the relative clause condition, and Region 16 for the coordinated clause condition. Coordinated clause conditions were read more quickly ($\Delta\mu$: 24 ms; 95% C.I. [9, 39], $p < 0.005$). Note that in relative clause sentences, the preposition occurs immediately after the verb, since there is a gap; in coordinate clause sentences, it follows an overt pronoun.

4.5.3.3 Discussion

In Experiment 10 we found no effects, online or offline, that a structurally similar filler constituent interferes in dependency construction. Constructing a nearly identical dependency upstream (modulo the gap's case position) does not make filler-gap resolution more difficult, as reflected in reading times or comprehension accuracy. These data suggest that identifying the head of a filler-gap dependency is in fact a grammatically accurate process. We outline two mechanisms below.

4.5.4 Accurately identifying the head of a dependency

We propose two mechanisms below for accurately identifying the head of a filler-gap dependency. The first is specific to explaining our experimental data, and hinges on the idea that filler representations are re-encoded when they are successfully integrated at the gap. The second is more general, and supposes that some distinctive features of the actual filler representation has been carried forward to target it later.

First let us return to the observation in Experiment 9 that there is increased complexity in nested filler-gap dependency constructions. The pattern of difficulty inside the RC, but no difficulty outside, suggests that whether or not the dependency is complete

matters as to whether or not other fillers could interfere. Inside the RC, there are two open filler-gap dependencies, but outside the RC only one remains open. This observation is not surprising, and is of course familiar as characteristic of the difficulty of too many center self-embeddings (Miller & Chomsky, 1963; Cowper, 1976; De Roeck et al. 1982; Lewis, 1996). One response to this observation has been to assume that the architecture must so constrained such that unintegrated dependency constituents tax processing (Gibson 1998; Gordon et al. 2006). If there is a capacity-limited memory space, then this constraint can be cashed out without much difficulty. The difficult cases run up against the bounds of available memory. However, in the architecture we are considering, where there is virtually no capacity limitation, then the constraint must be cashed out in terms of cue competition. Either the cues must be different at the two verbs; or the encodings of the fillers must differ with respect to their status as open or closed; or both. We asserted in the introduction that when the critical verb was processed, the encodings of the two fillers in the structure highly overlap in the relevant features, and thus the grammatically-inaccessible head should compete with the grammatical one. However our data have raised the possibility that the encodings of the fillers change over time, such that the head of a complete dependency is dissimilar from that of an open one. One linguistically motivated way of accounting for this re-encoding is to suppose that unintegrated fillers have a feature indicating that they lack a thematic role, i.e.:

THETAROLE:Unmarked.⁵⁵ When the dependency is completed, this feature is substituted

⁵⁵ This encoding scheme would also be consistent with the principle-based parsing accounts of filler-gap processing, and in particular, Pritchett (1992), who argued active dependency formation was driven to license thematic structure.

with a THETAROLE:Theme feature (for example, or THETAROLE:Agent or even THETAROLE:Marked). Suppose, then, that retrieval inside a filler-gap dependency contains a (highly weighted) cue for something like THETAROLE:Unmarked. Under this encoding scheme, fillers that have been successfully integrated will be relatively more dissimilar than unintegrated fillers. It is worth worrying about this kind of single feature distinction, since cue combination is non-linear. As more such distinctions accrue, the likelihood of interference declines very quickly. Methodologically, the more differences we can plausibly impute to what seems like an interfering constituent, then the smaller and smaller the competition effects that are predicted, and the less confident we can be in null results like we report. Theoretically, though, it suggests that potential linguistic structure building systems could dampen the impact of the memory architecture by maximizing distinctiveness in encodings. An encoding system that can instantiate licensing requirements as unvalued features could thus minimize the impact of having to resolve multiple such dependencies in succession.

The second possibility for accurately targeting the heads of (predicted) dependencies is to carry forward ‘just a little information’ about the head of the dependency. If this information were not abstract, but rather, functioned like a tag or unique marker for the specific filler, then it could be used to target retrieval of just that filler. In effect, this proposal is a hybridized version of Wanner & Maratsos (1978)’s HOLD cell approach. In abstract terms it simply means recording a pointer to the filler’s location in memory, instead of preserving all of its contents. However the debate, as it

Badecker & Lewis (2007) also use unvalued attribute-feature pairs as retrieval cues. In their system this accounts for the plural markedness effect in agreement attraction. See Chapter 3.

has been cast in the psycholinguistics literature, has established a tension between maintenance and retrieval that is rather more extreme than the intermediate positions permitted by the architecture. It seems a reasonable trade-off to preserve a small amount of information for gains in the licensing process. When we consider the implications of the island results, discussed in the previous section, we see that the potential payoff of preserving some filler-specific information is great. The islands results suggests that the search for dependency completion sites obeys constraints on extraction: some configurational licensing of the dependency occurs left-to-right. If retrieving the filler then returns multiple candidates, then the system could potentially integrate a non-licensed filler, even though it has already (putatively) expended the effort to verify the path requirement from filler to gap site. To preserve accuracy in this scenario, the system would have to select the filler in the right configuration with the gap site. If our argument from Chapter 3 is sound, there is no way to make this selection on the basis of inherent features, and thus the selection process would essentially have to re-capitulate the processes that traced the path from filler to potential gap site in the first place. If, however, characteristically only one candidate encoding were returned in search, then it would be unnecessary to do any (right-to-left) licensing, and the remainder of verifying the dependency could occur locally (i.e., do features match between subcategorizer and filler? is there actually a gap?).

The motivation to ‘localize’ as much decision making as possible is familiar and strongly echoes Berwick & Weinberg (1984)’s account of subjacency. They argued that subjacency, the requirement that syntactic movement rules be bounded, reflects an adaptation of the grammar to a deterministic parsing mechanism (Marcus, 1980). To

consistent with our conclusion from section 4.5.4. We test how well three different kinds of dependency formation probes survive increasing dependency lengths: verb-object plausibility, verb-PP selectional restriction, and a DP/PP filled gap effect. Each of these probes requires different kinds of information about the filler to generate a signal during active dependency completion. The specificity of the information required (minimally) for each probe is gradually decreased. The verb-object plausibility test requires information about the filler's lexical head features; the verb-PP selectional restriction requires just the lexical identity of the filler's functional head; and the DP/PP filled gap test only requires filler category. By varying dependency length, we are able to test what kinds of information is effective long after the filler was first encoded, and consequently what kinds of information may be guiding the parser's initial decision, before it attempts to recover full information about the filler.

4.6.1 Lexically-specific features (Experiments 11a, 11b)

The first experiment in this series tests how well lexical features of the *wh*-phrase survive different dependency lengths. The probe for active dependency formation is the plausibility of the filler phrase as an internal argument of the verb. Plausibility was crossed with dependency length either by modifying the subject with a PP, a serial length manipulation (in Experiment 11a) or embedding the subcategorizing the verb in a further clause, a hierarchical length manipulation (in Experiment 11b). If lexical features of the filler phrase are maintained until the filler is licensed, then plausibility should be an effective probe of active dependency formation regardless of dependency length.

4.6.1.1 Experiment 11a: Plausibility and increased serial length

Participants

Twenty-four native speakers of American English from the University community were paid \$10 to participate in an experimental session lasting 50 minutes. All were naive to the purpose of the experiment and gave informed consent.

Materials, Procedure and Analysis

Experimental materials consisted of 24 sets of 4 conditions organized in a 2×2 factorial design that independently manipulated the factors dependency length and plausibility. Experimental materials followed the scheme in (122) and (123). As in Experiment 8, the filler was extracted from the oblique object position of an alternating locative verb. This manipulation provides a multi-word region between the subcategorizing verb and the evidence of a moved constituent. Therefore effects can spill-over from the verb and still be interpreted as active, if they occur before the gap region. Since we are interested in information that is maintained to aid the parser make decisions about *wh*-dependency formation, the active effects provide the crucial evidence.

(122) *Short, Plausible*

- (a) The adhesive coating that the talented engineer methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

Short, Implausible

- (b) The computer program that the talented engineer methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

(123) *Long, Plausible*

(a) The adhesive coating that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

Long, Implausible

(b) The computer program that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

The dependency length factor manipulated the serial distance between filler and gap. The long dependency sentences were identical to the corresponding conditions in Experiment 8 (testing the CSC). Short dependency sentences were derived from long dependency sentences by removing the subject-adjoined PP. Dependency length here is operationally defined as the number of words between the introduction of the filler and the verb. Seventy-two fillers were adapted from the fillers in Experiment 8 so that the distribution of lengths of fillers matched the distribution of lengths of target items.

The experimental procedure was identical to Experiment 8. Data treatment and analysis was identical. No participants were excluded.

Results

Question-answering accuracy was uniformly high. For the 24 experimental targets, accuracy was 89%. There were no reliable differences among conditions; a model with no fixed effect coefficients was indistinguishable from one with all fixed coefficients ($\chi^2(7): 4.40; p = 0.49$).

In both long and short dependencies, a slow-down due to implausibility appeared in the first word of the ground argument. As this effect occurred well before direct evidence for the gap, we interpret it as an effect of active dependency formation. In short dependencies, slower reading times persisted for implausible filler sentences throughout

the ground argument region and into the gap region, whereas for long dependencies the effect of implausibility was observed only once in the ground argument regions and then not again until the gap region. The results in these conditions were thus similar to the results observed in the Experiment 8 Single VP condition.

Figure 4-8 presents the region-by-region condition means from the beginning of the sentence until region 20, divided into two pair-wise comparisons. The main finding in the reading time data is the large attenuation of the plausibility effect in long conditions compared to the robust and large-lasting sensitivity exhibited by the short conditions. The omnibus repeated measures ANOVA report is given in Appendix C. Region-by-region condition means and test results for preceding regions did not differ (when the unmatched PP regions of the long condition were excluded).

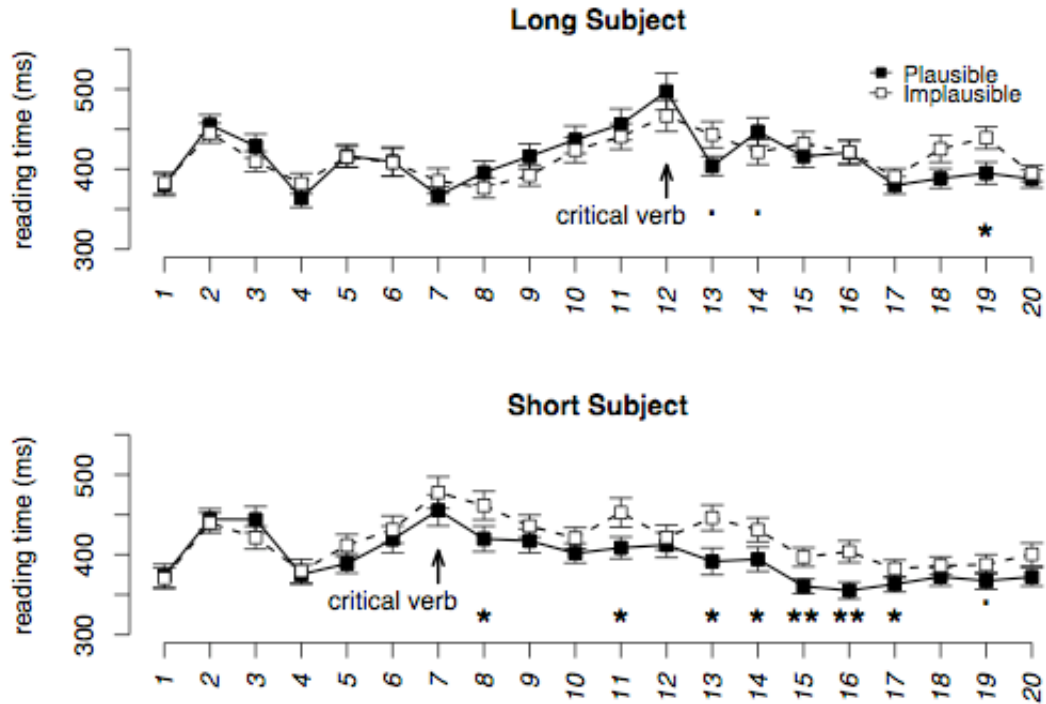


Figure 4-8 Experiment 11a Region-by-region reading times

(The coating/program that the engineer)₁₋₅

Long Subject: from₆ the₇ high-tech₈ aerospace₉ firm₁₀ methodically₁₁
 sprayed₁₂ the₁₃ special₁₄ test₁₄ surfaces₁₅ with₁₆ in₁₇ his₁₈
 laboratory₁₉ (...) ₂₀...

Short Control: methodically₆ sprayed₇ the₈ special₉ test₁₀ surfaces₁₁ with₁₂
 in₁₃ his₁₄ laboratory₁₅ (...) ₁₆...

Region-by-region reading times. Arrows indicate the critical verb.
 Punctuation indicates the result of a pairwise by-participants RMANOVA: p:
 ** < 0.01 < * < 0.05 < • < 0.10.

At the adverb preceding the verb there were no reliable main effects or interactions, although there was a non-significant tendency for long dependency conditions to be read more slowly. On the critical verb itself, there were also no reliable effects or interactions.

In the ground argument regions following the critical verb, plausibility effects were found in short and long conditions alike. However, the effect of plausibility was more long-lasting in the short dependencies. In the ground argument determiner region,

implausible sentences were read more slowly across both levels of the length factor. The size of the effect in long and short dependency conditions alike was consistent, but it was reliable in pair-wise comparisons only for the short dependency sentences. Neither levels of the length factor displayed an implausibility effect in the second and third words of the ground argument (regions 14-15), although short dependency sentences consistently displayed slower reading times for implausible sentences. In the final word of the ground argument region (region 16), corresponding to the head noun, short dependencies reliably showed a large slowdown due to implausibility, whereas long dependencies did not.

Short dependency conditions displayed sensitivity to the plausibility of the filler at the figure argument preposition, whereas long dependency conditions did not. At the two immediately following regions, however, there was a reliable slowdown due to implausibility for both long and short dependencies. In the third region subsequent to the preposition short dependency conditions again showed a reliable slowdown whereas Long dependencies did not.

Interim Discussion

The results of Experiment 11a showed the slowdown due to implausibility was attenuated when the distance between filler and gap was lengthened. This can be seen in the comparison of short and long conditions in Experiment 11a; it can also be seen in the comparison of coordinate and single VP conditions in Experiment 8. The largest and most sustained responses to an implausible filler were obtained when the filler-gap distance was short, or in a second coordinated VP.

4.6.1.2 Experiment 11b: Plausibility and increased hierarchical length

We performed a follow-up to Experiment 11a to test a different kind of length manipulation. In Experiment 11a, the long conditions were created by adjoining a PP to the subject; thus the model increased length was serial. In Experiment 11b, we also lengthened the dependency by inserting a whole clause between the filler and the gap-containing clause. Thus the experimental design was identical, crossing filler plausibility and dependency length. In this experiment, dependency length had three levels: short, long:clause, and long:PP. Experimental materials thus consisted of 24 sets of 6 conditions organized in a 2×3 factorial design that independently manipulated the factors dependency length and plausibility. A sample set is given below.

(124) *Short, Plausible*

- (a) It pleased the analyst that the coating that the talented engineer methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

Short, Implausible

- (b) It pleased the analyst that the program that the talented engineer methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

(125) *Long:Clause, Plausible*

- (a) The coating which the impressed analyst said that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

Long, Implausible

- (b) The program which the impressed analyst said that that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

(126) *Long:PP, Plausible*

(a) The coating that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

Long, Implausible

(b) The program that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

The short conditions were adapted from Experiment 11a; in this experiment, we matched the ordinal position of the critical verb in the short conditions, by embedding those sentences under ‘psych’ predicates. The long:clause conditions were adapted from the short conditions of Experiment 11a as well, by inserting a clause whose verb was heavily biased to subcategorize a CP. The long:PP level was identical to the long conditions in Experiment 11a, to test whether the results would replicate. In both long conditions, the number of words intervening between the filler and the critical verb was identical. In adapting the Experiment 11a materials, some lexical items were simplified; all fillers were reduced to single word phrase. The relative pronoun ‘which’ was used in place of the complementizer in long:clause conditions, to avoid the repetition of ‘that’ in a short span of time (which sounded unnatural to the experimenter and several informants).

All procedures were identical. There were 36 paid participants in this experiment, and each was paid \$10. Analysis of reading times was carried out via linear mixed-effects models (as for experiments in Chapters 2 and 3; see fn. 53 regarding the RMANOVA analysis for Experiments 7-8, 11a).

Results

Question-answering accuracy was uniformly high. For the 24 experimental targets, accuracy was 90%. There were no reliable differences among conditions; a with no fixed effect coefficients was indistinguishable from one with all of the fixed effect coefficients ($\chi^2(5): 6.09; p = 0.30$).

Figure 4-9 reports the reading times for just the long:clause condition, the new condition in Experiment 11b. The short and long:PP results replicated Experiment 11a; they are not reported in the text (but may be found in Appendix D). The main finding in the reading time data is a lack of sensitivity to filler plausibility in the active dependency completion regions in all but the short conditions.

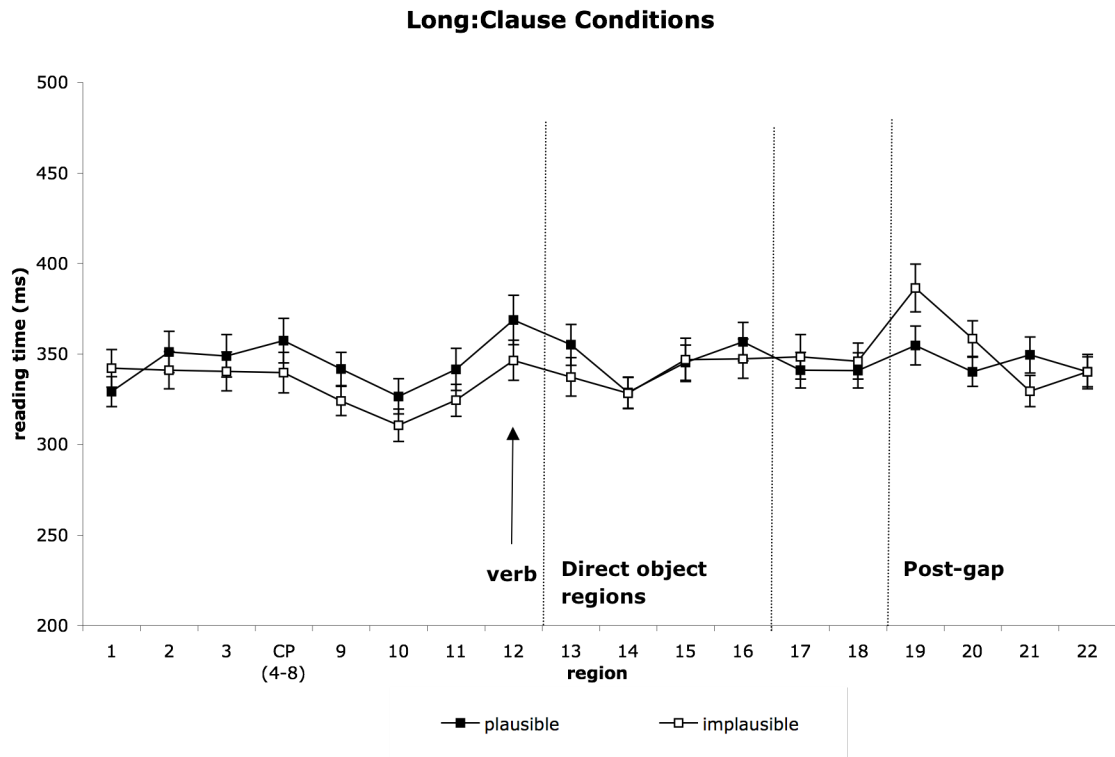


Figure 4-9 Experiment 11b Region-by-region reading times (Long:clause)

The₁ (coating / program)₂ that₃ (the impressed analyst said that]₄₋₈ the₉ engineer₁₀ sprayed₁₂ the₁₃ special₁₄ test₁₄ surfaces₁₅ with₁₆ in₁₇ his₁₈ laboratory₁₉ (...)₂₀₋₂₂

Pre-critical regions. We do not report tests for the structure comparison prior to the critical region, since the conditions are lexically unmatched. There were some reliable spurious differences in plausibility comparisons, observed between the following cells: plausible conditions slower in Region 4, short condition ($\Delta\mu$: 33 ms, 95% C.I. [12 ms, 56 ms], $p < 0.1$), plausible conditions faster in Region 5, long:clause conditions ($\Delta\mu$: 22 ms, 95% C.I. [0 ms, 40 ms], $p < 0.05$), Region 6, long:clause conditions ($\Delta\mu$: 29 ms, 95% C.I. [2 ms, 57 ms], $p < 0.05$).

Adverb region. There were no reliable differences across structure or plausibility in the adverb region.

Critical verb region. There were no reliable differences across structure or plausibility in the critical verb region.

Argument regions. A slow-down due to implausibility appeared in the first and second words of the ground argument region (Regions 13-14) (Region 13: $\Delta\mu$: 23 ms, 95% C.I. [-1 ms, 45 ms], $p < 0.10$; Region 14: $\Delta\mu$: 26 ms, 95% C.I. [1 ms, 50 ms], $p < 0.05$). This effect was significantly reduced in the long:clause conditions (Region 13: $\Delta\mu$: -39 ms, 95% C.I. [-4 ms, -69 ms], $p < 0.05$; Region 14: $\Delta\mu$: -27 ms, 95% C.I. [3, -64 ms], $p < 0.15$), though it was not different in the long:PP conditions, in the full model of the data. There was no reliable slowdown due to implausibility in long:clause conditions. Focusing on the comparison between short and long:PP conditions, the results are nearly identical to Experiment 11a. Pairwise comparisons over long:PP conditions revealed a reliable effect of plausibility in Region 13 ($\Delta\mu$: 18 ms, 95% C.I. [1 ms, 38 ms], $p < 0.05$), which did not persist into Region 14. The same comparisons over short conditions showed effects in both Regions 13 and 14. Despite being numerically larger, the effect of

plausibility was however only marginal in those regions (Region 13: $\Delta\mu$: 22 ms, 95% C.I. [-2 ms, 51 ms], $p < 0.10$; Region 14: $\Delta\mu$: 25 ms, 95% C.I. [-5 ms, 55 ms], $p < 0.10$).

What is notable, however, is that, like in Experiment 11a, the slowdown persists from the verb to the post-gap regions in short conditions, but not in long:PP conditions. If we collapse across the entire first argument region (Regions 13-16), then there is a reliable slowdown for short conditions ($\Delta\mu$: 19 ms, 95% C.I. [6 ms, 32 ms], $p < 0.01$), but not so for long:PP or long:clause conditions.

As a slowdown due to plausibility occurred well before direct evidence for the gap, we interpret it as an effect of active dependency formation. Interestingly, both long dependencies showed a strong effect of plausibility in the post-gap region. This effect replicates Experiment 11a as well. What seems to be characteristic of the short dependencies in both experiments is a long-lasting effect, slowing down processing from the verb onwards. This effect is “bi-phasic” in long:PP dependencies, showing a small slowdown in the region immediately following the verb, no difference in the intervening regions, and a larger response in the post-gap region. In long:clause dependencies, we only observed a slowdown in the post-gap regions.

It is worth considering whether baseline processing is more difficult in the long conditions, which might help to mask a plausibility effect. However, comparisons at the adverb, verb, and entire post-verb region – for just the plausible conditions – show no reliable variation in reading times due to structure. The comprehension accuracy results support this conclusion: though participants were overall 2 points lower in accuracy on long:clause conditions than short conditions, this difference is not reliable; moreover, they were numerically better on long:PP conditions (by 3%).

Discussion

The results of Experiment 11b conformed to those of Experiment 11a. In regions preceding the gap site, where the plausibility effect indexes active dependency completion, plausibility effects were longer lasting and more numerous in short conditions than in long:PP conditions, and they were absent in long:CP conditions. Both long conditions showed a strong effect of plausibility after the gap site had been signaled, indicating that participants were ultimately aware of the implausible verb-filler combination. However, only in the short conditions are they consistently aware in the active regions. What is striking about the present results is that only a modest amount of distance, a 5-word PP, is necessary to substantially disrupt sensitivity to an implausible filler. This disruption is selective: it is only in the pre-gap regions, where dependency completion is active, that sensitivity is lost. It suggests, however, that when the parser completes the dependency actively it may not have reliable access to the lexical features that allow evaluation of plausibility.

With regards to a retrieval model, the very presence of a plausibility effect indicates that the system is able to complete the dependency in spite of the lexical feature cues the verb might provide. We might suppose that categorial or positional cues are given priority (e.g., +wh) to retrieve the filler at the verb. Why then, should, the plausibility effect attenuate as dependency length increases? This attenuation with length might be explained if the interposed constituents provide sufficient interference, through spurious matches or shared features. Completing the filler-gap dependency is simply less probable as dependency length increases, because the likelihood of retrieving the filler declines. The fact that information is ultimately and robustly recovered in the post-gap

region casts doubt on this explanation. By the time the post-gap region is reached, the number of intervening constituents is even greater. There is perhaps more strongly constraining information, such as co-arguments, by the time the gap region has been reached in this experiment, which could help pinpoint the search for the filler. However it is difficult to see what highly distinctive features of the filler, instantiated at the time of encoding, could enter the retrieval structure only in the post-gap region. One conjecture is that the exact sense of the verb has been selected with more processing, and this yields a more informative retrieval structure. However notice that we must then interpret the implausibility effect in the post-gap region as a signal that the parser has failed to find the filler constituent. If additional processing allows the retrieval structure to accrue more cues, and in particular there are more lexically specific cues, then the implausible filler should be harder and harder to retrieve. It is not possible to rule this explanation out conclusively. However if the filler constituent ultimately failed to be retrieved, leading to an unlicensed dependency, it is surprising that comprehension accuracy is not impacted.

An alternative explanation finds greater traction with these data. We have conjectured that whatever information can be carried forward in time is used to complete the dependency actively. Active dependency formation, under this view, occurs largely independently of retrieval operations. Only after the dependency is constructed or licensed is the full filler encoding recovered to proceed with interpretation. Highly specific lexical information, by hypothesis, requires more space to maintain and is thus likely to be quickly displaced from focal attention as more relations have to be parsed. Consequently, in long dependencies, the plausibility effect only shows up reliably much later, when the filler representation is actually recovered. The plausibility effect in short

dependencies reflects the survival of some lexical information over the relatively short span from filler encoding to the occurrence of the verb. The retrievability of the filler does not vary substantially with the length of the dependency, which is consistent with the relatively small shifts in asymptote McElree, Foraker, & Dyer (2003) observed for 1 and 2 clause interpolations in filler-gap dependencies (see Chapter 3).

Finally it is worth emphasizing our finding that long:clause dependencies are worse than long:PP dependencies. Under successive cyclic movement, the *wh*-phrase is actually structurally equidistant from the verb in both short and long:clause dependencies (Chomsky, 1977). Therefore it is not unreasonable to expect long:clause dependencies to show the plausibility effect more robustly than the long:PP dependencies. The fact that they did not does not count as evidence against successive cyclicity (or any number of similar devices: Kayne, 1984; Gazdar, Pullum, Klein & Sag, 1985, etc.): it merely suggests that full filler information is not recovered successive cyclically. If that were true, it is also consistent with our alternative explanation. Coarse grained category information could be used to establish the cyclic trace or copy in the structure and no retrieval would be necessary⁵⁶.

4.6.2 Lexical identity (Experiment 12)

In Experiment 11, we demonstrated that lexical features appear to be ineffective at longer dependency lengths during active dependency formation. In this experiment, we asked whether slightly more coarsely-grained lexical information is preserved: the lexical

⁵⁶ In an earlier presentation of this work (Wagers & Phillips, 2006), we claimed that the filler was retrieved at clause boundaries (which we interpreted as consistent with Gibson & Warren (1999)'s observations). This conclusion, however, was premature and based on a partial data set. The addition of more participants has confirmed that there is no implausibility effect in multi-clause dependencies, prior to the gap site.

identity of a function word. To answer this question, we took advantage of the fact that some verbs select for prepositional phrases with a certain head. For example, in (127a), the verb ‘entrust’ requires a goal argument, which in English is headed by the preposition ‘to’. Correspondingly, in (127b), the verb ‘inherit’ requires a source argument, which is headed by the preposition ‘from’.

- (127) (a) The secretary entrusted the correspondence to/*from the courier.
(b) The orphan inherited a fortune from/*to the millionaire.

The head of the PP can be pied-piped in a *wh*-dependency.

- (128) (a) The courier to whom the secretary entrusted the correspondence ...
(b) The millionaire from whom the orphan inherited a fortune

In this experiment, we conjectured that comprehenders would be sensitive online to the selectional requirement of the verb for its PP argument. By means of the pied-piping manipulation, we could ask whether comprehenders would be sensitive to this requirement over short and long dependency lengths. We conjectured that recognizing whether or not the PP is of the right type in active dependency formation would require preserving the lexical identity of the head.

4.6.2.1 Materials and Methods

Participants

Participants were 18 native speakers of English from the University of Maryland community with no history of language disorders. Participants received \$10.

Materials

A sample item set is given in Table 4-9. The experimental factors were the match between verb and PP, and dependency length between filler and gap host. Gap position is marked with an underscore and the interposed PP is bracketed.

Dependency length	Match	“ The courier ... unfortunately wrecked his bike in traffic.”
<i>Short</i>	<i>Match</i>	... to whom the secretary recently entrusted the confidential business correspondence _____ after some hesitation ...
	<i>Mismatch</i>	... from whom the secretary recently entrusted the confidential business correspondence _____ after some hesitation ...
<i>Long</i>	<i>Match</i>	... to whom the secretary [for the high-powered defense attorney] recently entrusted the confidential business correspondence _____ after some hesitation ...
	<i>Mismatch</i>	... from whom the secretary [for the high-powered defense attorney] recently entrusted the confidential business correspondence _____ after some hesitation ...

Table 4-9 Sample materials set for Experiment 12

Items were balanced so that the *match* preposition was ‘to’ in 12 sets and ‘from’ in 12 sets⁵⁷. Materials were distributed according to a Latin Square across four lists, and accordingly each participant read six sentences from each condition. 72 filler sentences were included, adapted largely from Experiment 11; however some new distractors were devised that included pied-piping in other contexts beside the head of a subject-relative clause. The purpose of this manipulation was to ensure that participants could not strategically identify the experimental targets.

Procedures and Analysis

Presentation and analysis details were as described for Experiment 11b. Regions were structurally aligned, so that the VP regions common to all item sets were analyzed word-for-word across conditions.

4.6.2.2 Results

Comprehension accuracy is reported in Table 4-10.

⁵⁷ In preliminary analysis, we split the data set into the 12 items for which *from* was the mismatching prepositions and the 12 items for which *to* was the mismatching prepositions. The results were identical and conform to the patterns reported below.

		<i>Dependency length</i>		
		<i>Short</i>	<i>Long</i>	
<i>Verb-PP Selection</i>	<i>Match</i>	88%	84%	86%
	<i>Mismatch</i>	84%	80%	82%
		86%	82%	84%

Table 4-10 Comprehension accuracy for Experiment 12

Average percentage correct over participants, with row, column and grand means. Standard error of the cell means is 4% for all conditions, except the *mismatch:long* condition, in which it is 3%. N = 18.

There was a 4% decrement observed for long dependencies, and a 4% decrement for verbs with mismatching PP arguments. However none of these effects reached significance in the mixed-effects logit models.

Reading time results are reported in Figure 4-10. The main finding in the reading time data is that there is only sensitivity to the verb-PP match in short dependency conditions.

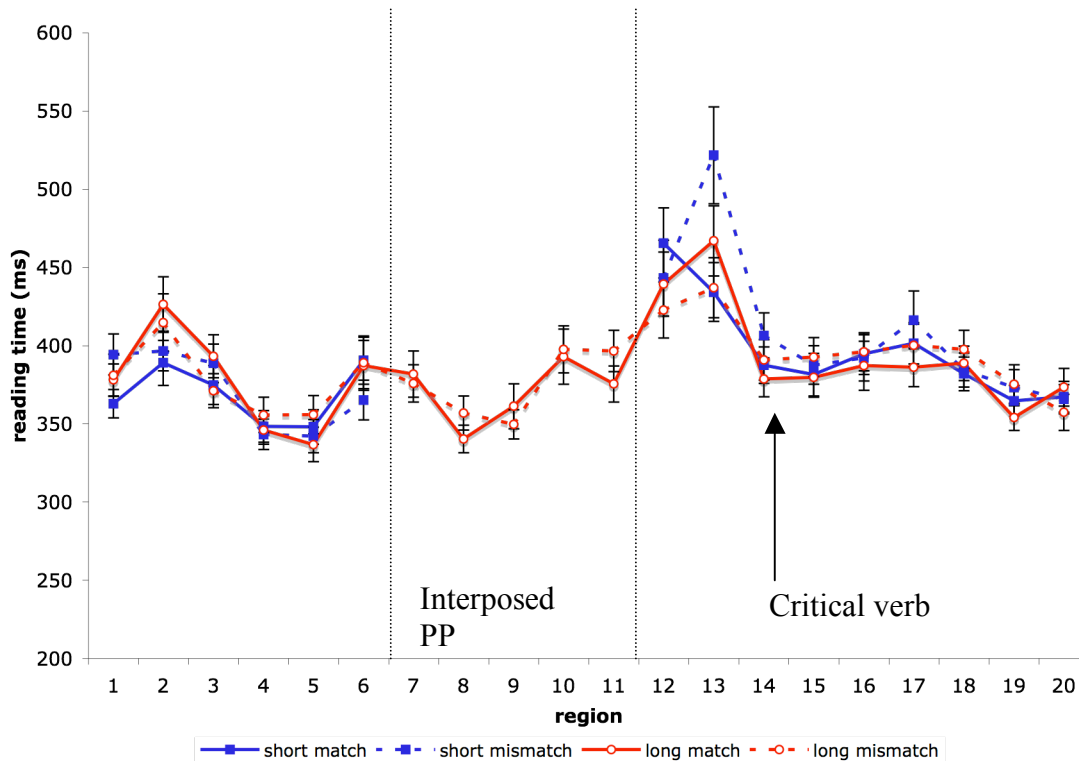


Figure 4-10 Experiment 12 Reading time results

Example sentence, with region subscripts:

The₁ courier₂ to₃/from₃ whom₄ the₅ secretary₆
 [for₇ the₈ high-powered₉ defense₁₀ attorney₁₁]
 recently₁₂ **entrusted**₁₃ the₁₄ confidential₁₅ business₁₆ correspondence₁₇
 after₁₈ some₁₉ hesitation₂₀ (unfortunately wrecked his bike in traffic).

Only significant effects are reported.

Region 1-3: Sentence-initial regions. Despite being identical, there are some baseline differences observed in Regions 1-3, corresponding to the sentence-initial determiner, subject head noun and pied-piped preposition. In Region 1, there is a main effect of match ($\Delta\mu$: 35 ms; 95% C.I. [10 ms, 60 ms], $p < 0.01$) and an interaction of match and length (*long:mismatch* $\Delta\mu$: -40 ms; 95% C.I. [-2 ms, -74 ms], $p < 0.05$). This interaction reflects the fact that a pairwise difference in match is only observed in the two long dependency conditions. In Region 2, there is a reliable effect of length ($\Delta\mu$: 45 ms;

95% C.I. [9 ms, 83 ms], $p < 0.05$). In Region 3, there is a reliable interaction of the two experimental factors (*long:mismatch* $\Delta\mu$: -43 ms; 95% C.I. [-4, -84], $p < 0.05$).

Because these differences occurred before any experimental manipulations, they are interpreted as spurious. In Regions 4-6, corresponding to the *wh* phrase and the subject of the relative clause, there are no differences among conditions. Since all the early differences involve comparisons among the *long* conditions, it is important to note that in the PP region, Regions 7-11, there are no reliable differences. Therefore across all conditions, the baselines are stabilized well in advance of the critical verb.

Region 13: Critical verb. In Region 13, the critical verb region, there is an effect of match ($\Delta\mu$: 91 ms; 95% C.I. [33, 148], $p < 0.01$), and an interaction with distance (*long:mismatch* $\Delta\mu$: -125 ms; 95% C.I. [-45, -215], $p < 0.01$). Pairwise comparisons show that there is only a slowdown for mismatch conditions when the dependency is short ($\Delta\mu$: 90 ms; 95% C.I. [5, 166], $p < 0.05$), and no reliable difference between long dependency conditions. Indeed, match conditions are (numerically) slower in long dependencies ($\Delta\mu$: 34 ms; 95% C.I. [-15, 85], n.s.)

No differences are observed in subsequent regions of the sentences. In an attempt to detect a match difference in long conditions, we pooled Regions 14-19; results were not significant.

4.6.2.3 Speeded grammaticality follow-up

The results of the reading time data suggest that comprehenders are only sensitive to the selectional properties of the verb with respect to a pied-piped PP in short dependencies. Unlike plausibility detection, reported in Experiment 11 above, participants showed no ultimate sensitivity to mismatch in online measures.

Comprehension accuracy indicated a decrement due to match, regardless of length, but those results were not reliable. However, even if comprehenders failed to notice the ill-formedness of the sentences, it is likely they could determine the correct interpretation, by knowing the verb's semantic properties (i.e. 'inherit' needs a source argument; even if the preposition in the pied-piped oblique argument is forgotten, it can be taken a source). In a follow-up speeded grammaticality experiment, we asked a different group of participants to judge the materials used in the reading-time experiment as acceptable or not. We report these results below.

Identical item sets from the current experiment were included in the speeded grammaticality experiments reported in Chapters 3. In results reported below, 16 participants came from Experiment 3 and 16 from Experiment 6. Analysis revealed no between group differences, so we consolidate them below.

Results of the speeded grammaticality task are reported in Figure 4-11.

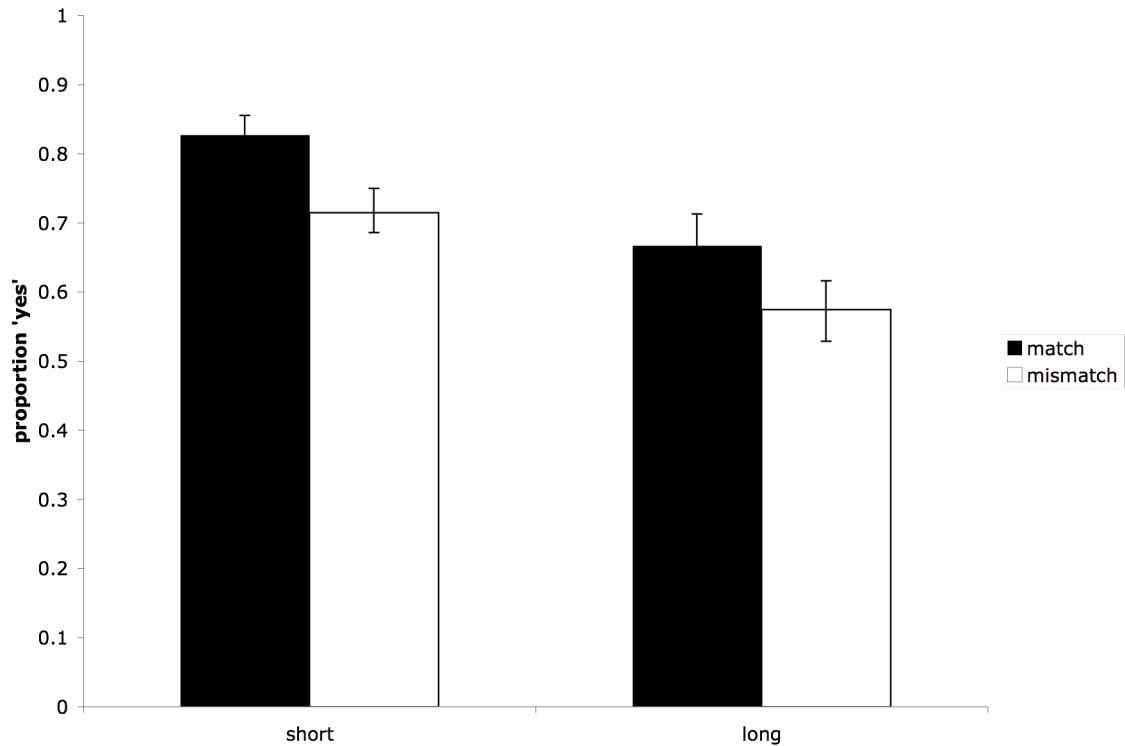


Figure 4-11 Experiment 12 Follow-up results
Speeded grammaticality, proportion ‘yes’ responses

Error bars are standard error of the mean proportion across participants.

There are two patterns to note in these data. Firstly, judgments of acceptability were sensitive both to the selection mismatch ($\beta: -.71 \pm 0.49, p < 0.001$) and dependency length ($\beta: -.95 \pm 0.49, p < 0.001$). There was no interaction of the two: participants were equally less likely to say ‘no’ to pied-piped verb-PP mismatches in both short and long dependency conditions. Secondly, however, overall the differences between match and mismatch conditions were small: for example, in short dependencies participants accepted match conditions 83% of the time and mismatch conditions 71% of the time. This cannot be attributed to an overall ‘yes’ bias in the experiment, since we see from data in Experiments 3 and 6, run simultaneously, that participants were willing to say ‘no’ 70-80% of the time for some agreement violations.

4.6.2.4 Discussion

The data in Experiment 12, and the speeded grammaticality follow-up, indicate that as dependency length increases, sensitivity to the lexical identity of the pied-piped preposition declines in active dependency formation. In short dependencies, a mismatch between the verb and the PP leads to a strong reading time disruption. This disruption is entirely absent in long dependencies. Thus we conclude that lexical identity cannot be well maintained in an immediately accessible state to guide active dependency formation. The data in the speeded grammaticality follow-up demonstrate that the violation is not incapable of being noticed. Consequently it is possible, regardless of length, to recover a representation over which the selectional restriction between verb and PP can be evaluated. However it seems that information about the identity of the head of the PP is not reliably available to the parser during initial dependency formation.

4.6.3 *FG (Pied-piping) (Experiment 13)*

In this experiment we turn to less lexically anchored properties of the filler. It is possible in English to extract both *wh*-phrases and the PPs that contain them. By using a modified filled gap effect design, we can test whether the categorial identity of the extracted phrase is well maintained over long dependencies. We used comparisons like the following:

- (129) (a) The website which the blogger recently designed | the flashy graphics for ...
(b) The website for which the blogger recently designed | the flash graphics ...

In (129a), by the time comprehender encounters the verb (indicated by the '|'), the extracted phrase will be analyzed as the direct object. As the direct object is recognized, the comprehender must reanalyze the site of extraction. However, in (129b), the pied-piped extraction does not lead to a direct object analysis. Comparison across the direct

object regions in (129) is thus expected to show a slow-down for simple DP extraction, reflecting the cost of reanalysis. Comparing regular and pied-pied extractions is a modified version of the filled-gap logic (Stowe, 1986) and has been demonstrated to be effective before (Lee, 2004). In an important respect, it is preferable to the standard filled gap experimental design because it compares two sets of conditions that both involve an extraction dependency. In contrast, the standard design (Stowe, 1986) compares the direct object region inside a movement dependency with one inside an *if of whether* clause.

What would happen in long dependencies? If the categorial identity of the extracted phrase is preserved across longer dependencies, then we expect a filled gap effect in both short and long conditions. If, however, categorial identity is lost, then long conditions should be identical in processing complexity, either because both regular and pied-piped extraction leads to a direct object analysis or because neither do.

4.6.3.1 Materials and methods

Participants

Participants were 42 native speakers of English from the University of Maryland community with no history of language disorders. Participants received \$10.

Materials

To achieve the desired contrast, we constructed sentences that involved extraction from a benefactive PP. The experimental design crossed extraction type, either ‘simple’ or ‘pied-piped,’ with dependency length. As in Experiment 11b, the length factor had three levels: short, long:PP and long:clause. Conditions were combined in a 2×3 factorial design creating six conditions for each item set. There were 24 items sets. Thus

each participant would see four examples of each condition. The ordinal position of the critical verb was matched across conditions by embedding the short dependency under a psych predicate. A sample set of materials is given in Table 4-11. The length manipulation in long:* conditions is bracketed.

<i>Dependency length</i>	<i>Extraction</i>	
<i>Short</i>		The CEO was worried that the website ... would soon be obsolete.
	<i>Simple</i>	... which the blogger recently designed the flashy graphics for ___ after user demand grew ...
	<i>Pied-piped</i>	... for which the blogger recently designed the flashy graphics ___ after user demand grew ...
<i>Long:PP</i>		The website ... would soon be obsolete.
	<i>Simple</i>	... which the blogger [from the local art school] recently designed the flashy graphics for ___ after user demand grew ...
	<i>Pied-piped</i>	... for which the blogger [from the local art school] recently designed the flashy graphics ___ after user demand grew ...
<i>Long:Clause</i>	<i>Simple</i>	... which [the company CEO said that] the blogger recently designed the flashy graphics for ___ after user demand grew ...
	<i>Pied-piped</i>	... for which [the company CEO said that] the blogger recently designed the flashy graphics ___ after user demand grew ...

Table 4-11 Sample materials set for Experiment 13

Procedures and Analysis

Presentation and analysis details were as described for Experiment 11b. Regions were structurally aligned, so that the VP regions common to all item sets were analyzed word-for-word across conditions.

4.6.3.2 Results

Comprehension accuracy is reported in Table 4-12.

		<i>Dependency length</i>			
		<i>Short</i>	<i>Long:PP</i>	<i>Long:Clause</i>	
<i>Extraction type</i>	<i>Simple</i>	88% ± 2%	89% ± 2%	76% ± 3%	84%
	<i>Pied-piped</i>	85% ± 3%	85% ± 3%	74% ± 3%	81%
		87%	87%	75%	83%

Table 4-12 Comprehension accuracy for Experiment 13

Average percentage correct over participants, with row, column and grand means. Standard error of the cell means is reported in the table.

There was a significant (negative) effect of long:clause conditions (β : -1.0 ± 0.6 , $p < 0.005$). No other comparisons were reliable.

Reading time data are reported in Figure 4-12 for short conditions, Figure 4-13 for long:PP conditions, and Figure 4-14 for long:clause conditions. The main finding in the reading time data is the appearance of a filled-gap effect across all dependency length conditions, occurring in the critical direct object regions.

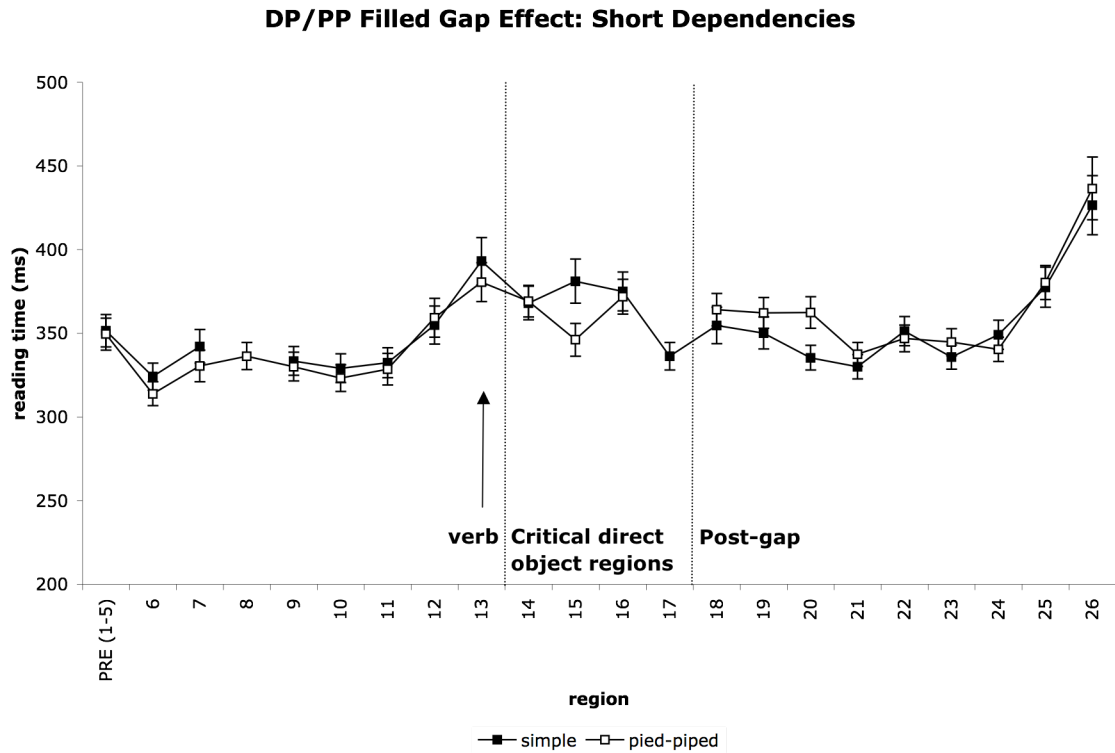


Figure 4-12 Experiment 13 Region-by-region reading times: Short conditions

(The CEO was worried that)₁₋₅ the₆ website₇

simple: which₉ the₁₀ blogger₁₁ recently₁₂ designed₁₃ the₁₄ flashy₁₅
graphics₁₆ for₁₇

pied-piped: for₈ which₉ the₁₀ blogger₁₁ recently₁₂ designed₁₃ the₁₄ flashy₁₅
graphics₁₆

after₁₈ user₁₉ demand₂₀ grew₂₁ would₂₂ soon₂₃ be₂₄ obsolete₂₅ (...)₂₆

DP/PP Filled Gap Effect: Long/PP Dependencies

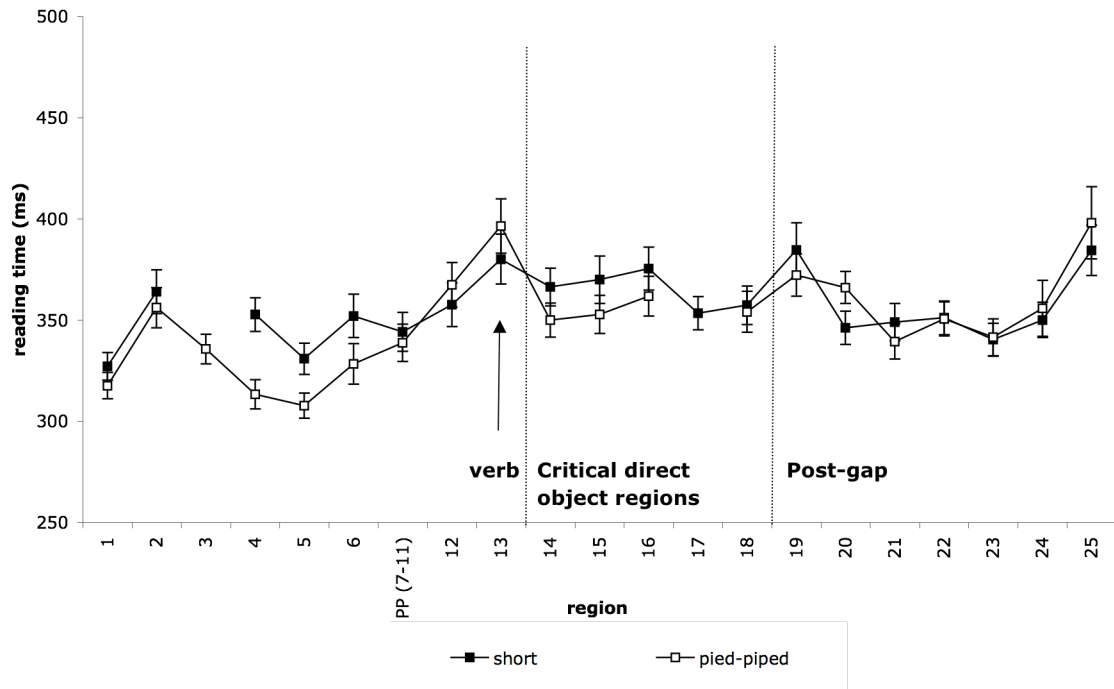


Figure 4-13 Experiment 13 Region-by-region reading times: Long/PP conditions

The₁ website₂

simple: which₄ the₅ blogger₆ (from the local art school)₇₋₁₁ recently₁₂
designed₁₃ the₁₄ flashy₁₅ graphics₁₆ for₁₇

pied-piped: for₃ which₄ the₅ blogger₆ (from the local art school)₇₋₁₁
recently₁₂ designed₁₃ the₁₄ flashy₁₅ graphics₁₆

after₁₈ user₁₉ demand₂₀ grew₂₁ would₂₂ soon₂₃ be₂₄ obsolete₂₅ (...)₂₆

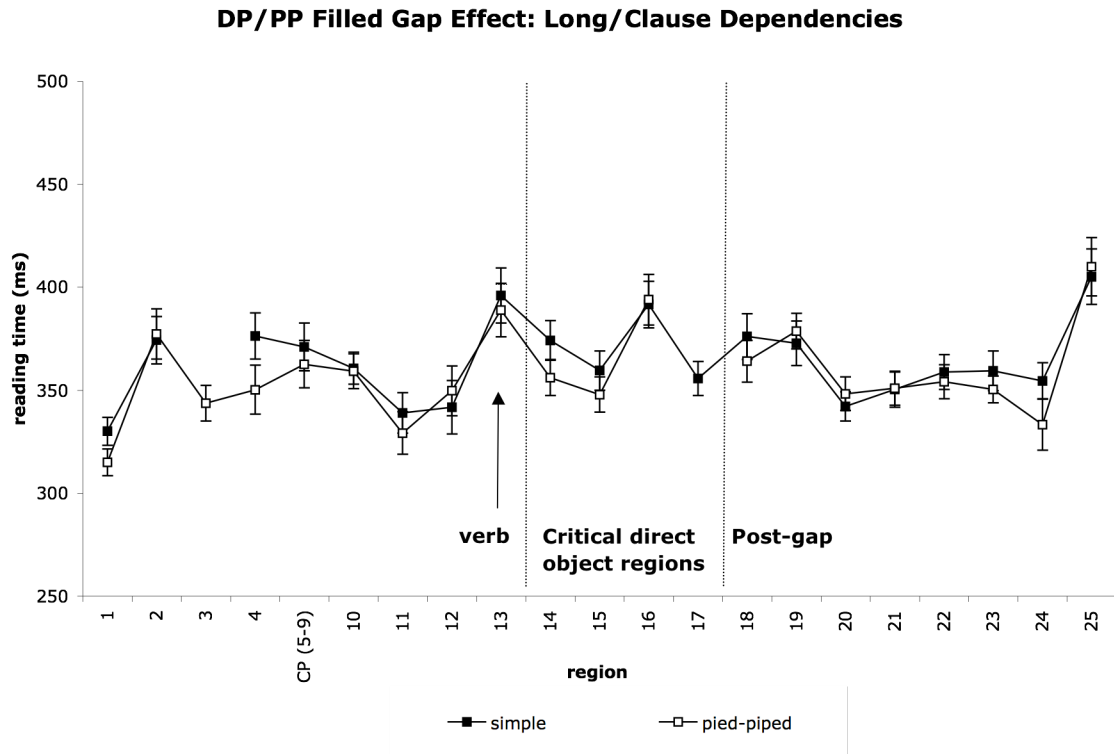


Figure 4-14 Region-by-region reading times: Long/Clause conditions

The₁ website₂

simple: which₄ (the company CEO said that)₅₋₉ the₁₀ blogger₁₁ recently₁₂
designed₁₃ the₁₄ flashy₁₅ graphics₁₆ for₁₇

pied-piped: for₃ which₄ (the company CEO said that)₅₋₉ the₁₀ blogger₁₁
recently₁₂ designed₁₃ the₁₄ flashy₁₅ graphics₁₆

after₁₈ user₁₉ demand₂₀ grew₂₁ would₂₂ soon₂₃ be₂₄ obsolete₂₅ (...)₂₆

Pre-verbal regions. There were no reliable differences among conditions in

Regions 1 – 12, with the exception of Regions 4 – 6 in the long conditions. These regions correspond to the sequence ‘which – Det – N’ at the edge of the relative clause. Simple extractions were read more slowly in these regions ($\Delta\mu$: 23 ms, 95% C.I. [15 ms, 31 ms], $p < 0.005$). We speculate that the initial ambiguity of parsing ‘for’ in pied-piped extraction may have contributed to the discrepancy between extraction type conditions. This slowdown was restricted to the long conditions, where it occurred sentence initially, which could magnify such effects. Because this difference occurs so-far upstream of the

critical regions (Region 14 and subsequently), and there are no differences in Regions 7-12, we are not concerned that this difference would impact tests in the critical regions.

Verb. Reading times at the verb, Region 13, showed no effect of length or extraction type.

Critical direct object region. Regions 14 – 16 corresponded to the critical direct object regions where a filler-gap effect was expected. All length conditions showed a slowdown for simple extractions in either Region 14 or Region 15. The full mixed effects model of the data revealed a reliable effect of extraction type in Region 15, such that simple conditions are read more slowly across all length conditions. For short conditions it was only in Region 15 that a slowdown was observed. However, for long:PP conditions, a smaller slowdown was observed in Region 14, 15 and 16. This slowdown was marginally significant in Region 14 alone ($\Delta\mu$: 16 ms, 95% C.I. [-1 ms, 32 ms], $p < 0.10$), and reliable in Region 15 ($\Delta\mu$: 16 ms, 95% C.I. [3 ms, 26 ms], $p < 0.01$). When Regions 14 – 16 were combined in a pooled analysis of long:PP conditions, the result was also reliable ($\Delta\mu$: 16 ms, 95% C.I. [4 ms, 27 ms], $p < 0.01$). For long:clause conditions, the slowdown was reliable only in the Region 14 pairwise comparison ($\Delta\mu$: 19 ms, 95% C.I. [0 ms, 36 ms], $p < 0.05$).

The mean difference due to extraction type was numerically smaller in both long conditions compared to the short conditions. This difference was not reliable in the full model for the Region 15 data. However, as a stronger test we combined the Region 14 data for long:PP conditions with the Region 15 data for long:clause conditions to perform a single long v. short comparison, with the Region 15 data for short conditions. The combination of the two regions to create the pooled long condition reflects the fact the

differences were not time-locked across conditions; however we wanted to compare the largest and most reliable conditions for each condition. This analysis returned a significant effect of extraction type ($\Delta\mu$: 35 ms, 95% C.I. [14 ms, 55 ms], $p < 0.005$), but the interaction with length is not reliable ($\Delta\mu$: 17 ms, 95% C.I. [7 ms, 42 ms], $p < 0.20$). We therefore cannot conclude that the filler-gap effect was smaller for long conditions. It is possible our design simply does not give us enough statistical power to detect the interaction. The raw mean differences in effect sizes are somewhat misleading: normalizing each extraction-type comparison against the pooled standard deviation in that condition shows that the effect of extraction type is less discrepant across length conditions (*Cohen's d* – short: 0.24, long:PP: 0.14, long:clause: 0.15)⁵⁸. Moreover, while one larger effect is observed for short conditions, several smaller effects are observed for the long conditions.

Gap and post-gap regions. Region 18 is the first post-gap region, and constitutes the signal that a constituent is missing (as in previous experiments, with a sequence of two preposition). In this region long:CP conditions were read significantly more slowly ($\Delta\mu$: 21 ms, 95% C.I. [3 ms, 42 ms], $p < 0.05$). In Region 19, long:PP conditions were read significantly more slowly ($\Delta\mu$: 38 ms, 95% C.I. [17 ms, 60 ms], $p < 0.001$). There was a marginal slow-down for long:CP conditions ($\Delta\mu$: 21 ms, 95% C.I. [-1 ms, 42 ms], $p < 0.10$); there was also a marginal interaction of long:PP conditions with pied-piped extraction, such that pied-piped conditions were read faster ($\Delta\mu$: 26 ms, 95% C.I. [-6 ms,

⁵⁸ Establishing a reliable interaction between extraction type and dependency length would require power to detect an effect size of roughly $d = 0.09$ (or $R^2 = 0.002$; a very small effect). Given our design and the between condition correlations observed within participants in the actual experiment, we estimate nearly 120 participants would be necessary to achieve power ($1 - \beta$) of 0.80 for that effect size (calculations carried out in G*Power 3, Faul, et al., 2007).

53 ms], $p < 0.10$). In Region 20, pied-piped extractions were read more slowly ($\Delta\mu$: 23 ms, 95% C.I. [5 ms, 42 ms], $p < 0.05$).

4.6.3.3 Discussion

The modified filled gap paradigm used in Experiment 13 gave us our first indication that some effects of active dependency formation can robustly survive longer dependency lengths. We detected a filled-gap effect in both short and long conditions. It was numerically larger in the short condition, but restricted to one region only; in long conditions, it was smaller but spread over several regions. The numerical differences were not reliable. We conclude that while the longer dependency lengths may introduce variability in when the reanalysis is triggered, comprehenders robustly detect that they have misparsed the simple extractions. In order to do so, they must have maintained the distinction between the category extracted: DP or PP. Consequently, basic category information seems to survive longer dependency lengths better than either lexical features or lexical identity. Interestingly, comprehenders are slower at showing the filled-gap effect in short dependencies, but faster in long dependencies. We believe this may be indirectly related to our hypothesis: in the long dependencies, comprehenders only have filler category information to rely upon, and can so quickly assess (and reject) the direct object analysis of the filler in simple extractions. In short dependencies, comprehenders may be evaluating the direct object analysis using a broader array of information about the filler, since more lexically-anchored information survives at shorter dependencies.

4.6.4 *Conclusions*

In Experiments 11 – 13 we examined how different measures of active dependency formation respond to different dependency lengths. In Experiments 11a and

11b, we used a plausibility manipulation, which has been successfully used before in a variety of tasks (Garnsey, Tanenhaus & Chapman 1989; Tanenhaus, Stowe & Carlson 1985; Boland et al. 1995; Traxler & Pickering 1996; Phillips, 2006; Lau, Yeung, Hashimoto, Braun & Phillips 2006) and has thus generally been considered a robust index of dependency formation. In Experiment 12, we used a selectional restriction between a verb and its PP argument, which we devised for the study. In Experiment 13, we used a modified version of the filled gap paradigm (Stowe, 1986; Lee, 2004) comparing simple extractions with pied-pied extractions. Only in the latter case did we observe evidence for active dependency formation in long dependencies. It is worth emphasizing that most of our length manipulations involved adjoining a five-word PP to the subject. Based on word-by-word reading times, this corresponded to a 1-1.5 second increment in the time elapsed from filler to verb. The mismatches in Experiment 11a were especially strong, so it seems counter-intuitive that this additional processing time, spent as it was parsing a structurally unambiguous substring, should nearly extinguish sensitivity.

Crucially in all experiments we observed ultimate sensitivity to the dependency formation manipulation. In Experiments 11a & 11b, sensitivity to an implausible verb-argument pair was manifested immediately following unambiguous evidence for the gap. In Experiment 12, we never observed sensitivity to the verb-PP mismatch in the online record, but an off-line followup revealed (length-independent) sensitivity. If we restricted our attention to just the results in Experiments 11-12, then the data would suggest that the impact of length is to switch into a non-active mode of dependency formation. The pattern in Experiments 11a & 11b support this conclusion most strongly: in short

dependencies, the reading time disruption immediately follows the verb, whereas in long dependencies, it immediately follows evidence for the gap site. Such conclusion would constitute a serious challenge to the generality of active dependency formation, suggesting that only in very short dependencies does the parser posit gaps before direct evidence for a missing constituent. However, in Experiment 13, the filled gap effect survived both PP and clausal extension of the dependency. The active positing of gap sites thus persists across the same dependency lengths that extinguish sensitivity to specific lexical information. These data support the view that dependency formation precedes independently of access to the detailed contents of the filler.

4.7 Conclusions

In this chapter we considered how the parser could adapt to a content-addressable memory to facilitate the accurate recognition of grammatical dependencies. Based primarily on previous results in the processing of anaphora, we concluded that the predictable dependencies were most likely to be constructed and licensed without considering grammatically accurate constituents. Experimental studies on the formation of *wh*-dependencies elaborated this viewpoint in several ways:

- 1) **Top-down, grammar driven dependency formation.** The Coordinate Structure Constraint experiments (7-8) provide strong evidence that the parser uses its knowledge of island constraints to prompt the construction of filler-gap dependencies. The contrast with potential p-gap environments affirms that the active dependency formation strategy proceeds top-down with reference to outstanding licensing requirements, and not on the basis of the bottom-up compatibility of an analysis.

- 2) **Accurately targeted dependency heads.** The Interference experiments (9-10) attempted to follow-up and refine Van Dyke & McElree (2006)'s demonstration that filler-gap dependency completion is liable to similarity-based interference. On balance they support the idea that (predicted) dependency formation is free from interference, when the comparison set is other dependency heads (and not extra-syntactic lexemes). We presented two mechanisms to account for accurate performance: in the first, the encoding system can, in a grammatically motivated way, render an incomplete dependency head featurally distinct from a complete one; in the second, a small amount of filler-specific information is carried forward that is used to retrieve it at licit retrieval sites. The data do not choose between the two accounts. From the standpoint of structurally licensing the dependency, the first strategy is heuristic; the second one requires some maintenance, but allows licensing to occur entirely left-to-right.
- 3) **Robust availability of filler category information.** The Dependency Length experiments (11-13) asked whether it was plausible that any information was carried forward in time to guide dependency formation. Lexically-anchored information was lost quickly, even in monoclauses. Categorical information survived across longer dependencies. These results are consistent with the top-down nature of active dependency formation and decision-making process that is most robustly supported by coarse-grained information.

We propose that these results taken together are characteristic of a processing system in which the initial licensing of predicted dependencies could be largely retrieval-free. The synthesis of the latter two set of experiments supports this point most strongly.

Experiments 11-12 suggest that most of the information about a filler is not recovered during the active formation phase of a long dependencies, but that it is eventually recovered. Given a direct access retrieval mechanism, one explanation for the near complete insensitivity observed in active regions of Experiments 11-12 is the competition of many similar constituents. Longer dependencies mean more encodings in memory that could be activated by the retrieval structure at the verb. However, Experiments 9-10 pull in the opposite direction. Even when dependencies are short, the most similar constituents, fillers in other dependencies, do not interfere. It seems unlikely, therefore, that the extra length in Experiments 11-12 introduced strongly interfering constituents.

5 Conclusions

5.1 *Specific conclusions*

The major empirical target of this dissertation is how accurately on-line comprehension reflects grammatical principles and constraints. This dissertation reported the results of several reading time and speeded grammaticality judgment experiments on two kinds of dependency completion processes: subject-verb agreement licensing and *wh*-dependency formation. The goal of these experiments was to test under which conditions real-time dependency formation is faithful to grammatical principles and constraints. For the same reasons, we also examined the existing literature on processing complex subjects. Assessing those conditions was part of a broader effort to determine how structure-sensitivity could be achieved in a memory architecture that is inherently not well-suited to verifying hierarchical relations between constituents. We first report the specific experimental conclusions in this section. In the next section we report the broader conclusions.

5.1.1 *Agreement attraction*

5.1.1.1 Experimental results (Experiments 1 – 4)

Agreement attraction occurs in English when a DP other than the subject matches the verb in number features. The canonical example of agreement attraction, discussed most heavily in the production literature, occurs in complex subjects modified by a PP (Bock & Miller, 1991):

(130) The path to the monuments were littered with bottles.

Here we provided the first online complexity data for a relatively understudied species of agreement attraction: attraction between a relative clause head and a relative clause verb, first reported by Kimball & Aissen (1971):

(131) The runners who the driver wave to ...

We found that comprehenders processed agreement attraction highly similarly in both RC and complex subject attraction. The occurrence of a plural attractor eased the RT disruption normally associated with subject-verb mismatches. Crucially a plural attractor only reliable eased the subject-verb mismatch disruption; it did not increase difficulty for grammatical sentences. This pattern, which we describe as eliciting illusions of grammaticality, but never illusions of ungrammaticality, was mirrored in the judgment results.

We concluded that agreement attraction is selectively fallible in comprehension: it is liable to intrusion of a ungrammatical analysis only when a grammatical analysis is not available. The results of our studies, across sentence types and experimental measures, failed to confirm the Symmetry Prediction of feature percolation accounts of attraction (Vigliocco & Nicol, 1998; Eberhard, Cutting, & Bock, 2005).

5.1.1.2 Modeling results (and Experiment 5)

We presented a cue-based retrieval account of our data which we formalized using Shiffrin and colleagues' Search of Associative Memory (Gillund & Shiffrin, 1984). In this model, attractors intrude in online comprehension when a constituent that fully matches the verb's retrieval cues is not found. In ungrammatical cases, a subject cannot be found with plural number, but a plural, non-subject is partially activated. In

grammatical cases, there is no plural feature to contact the non-subject. This account was developed initially for complex subject attraction, and then extended to RC attraction. There it was necessary to introduce a clause context cue, since the RC attractor is subject-like. Finally the use of case cues predicted that there should be a small amount of intrusion from RC heads in subject position even in grammatical sentences, but not in object position. This was confirmed in an speeded grammaticality study: attraction was found in ungrammatical sentences when the RC was either subject- or object-attached. However a slight but reliable effect was found in grammatical sentences as well when the RC was subject-attached.

5.1.2 Wh-dependency formation

5.1.2.1 The Coordinate Structure Constraint (Experiments 7-8)

We tested whether the Coordinate Structure Constraint (CSC) was respected in real-time processing. The CSC forbids extraction out of coordinated phrases, unless the same subconstituent is extracted out of each coordinate (Ross, 1967). We found that when a gap was detected inside a coordinate environment, the parser actively completed a second filler-gap dependency in the second coordinate. Experiment 8 confirmed that the parser did so without any evidence of a missing constituent. In a very similar environment, i.e., a post-verbal adjunct clause which could support a parasitic gap, we found no evidence of active dependency completion. Consequently we concluded the comprehender was immediately aware of the implications of completing a *wh*-dependency in a coordinate phrase. Grammatical knowledge was directing dependency completion.

5.1.2.2 Locating the head of a *wh*-dependency (Experiments 9-10)

We tested whether the parser was as accurate at locating the head of a dependency as previous experimental results, including our own Experiments 7-8, indicate it is for the tail of a dependency. Specifically we tested for whether the presence of additional *wh*-dependency heads in a sentence, in similar structural positions, would interfere with the dependency formation process. We found no online evidence at the critical verb that an additional candidate *wh*-phrase interfered in completing a target dependency. There was evidence in Experiment 9 that completing two *wh*-dependencies led to decreased accuracy. In that experiment, the distractor *wh*-phrase was nested inside the target dependency. It is possible that there was interference at the medial verb; but there may have also been an interaction between the gap search and the ambiguity resolution present in those materials. In Experiment 10, which consider same-sentence dependencies which were not nested, there was no observed decrement in accuracy. We presented two candidate mechanisms to explain these results in a content-addressable memory: in the first mechanism, *wh*-phrases are re-encoded when they have been successfully integrated and thus can be restricted from a verb-triggered search; in the second mechanism, a small amount of idiosyncratic information about the *wh*-phrase is carried forward that allows targeted retrieval. Our findings contrast somewhat with Van Dyke & McElree (2006) who found that words from a extra-syntactic memory load list could interfere in *wh*-dependency formation. We argued, however, that they provided no evidence that it was the dependency resolution process which the memory list interfered with, and that our manipulation constitutes a stronger test.

5.1.2.3 Carrying forward information in time (Experiments 11-13)

In the final set of experiments, we tested how well different measures of active dependency formation survive increasing serial and hierarchical dependency length. These measures included a plausibility manipulation (Traxler & Pickering, 1996), a verb-PP selectional restriction, and a pied-piping filled gap effect (Lee, 2004). We found that only the filled gap effect survived the longer dependencies. The increase in dependency length was relatively modest – in the serial length conditions, it only involved interpolating a 5-word PP region, corresponding to roughly 1 second of elapsed processing time – so it was surprising that some measures of active dependency formation were so effectively attenuated. Both the plausibility and verb-PP selectional restriction manipulations required lexically-anchored information, while the filled-gap effect only required information about the category of the filler. We concluded that active dependency formation is most effectively guided by coarse-grained categorial information.

5.2 *Broader Conclusions*

Achieving structure sensitivity in a content-addressable memory is inherently difficult. Properties that constituents have by virtue of their hierarchical relation to other constituents, like c-command, cannot be encoded in constituent representations. These kind of restrictions therefore cannot be enforced in a content-addressable, direct access search for constituents that license or participate in a dependency. Consequently, grammatically inaccessible constituents will be generated as candidates in the search process. This fact about embedding linguistic representations in a content-addressable

memory architecture is a major determinant of grammatical inaccuracy in real-time processing.

There are ways of countering this inaccuracy. Our own experiments on *wh*-dependency processing, and a review of the existing literature, revealed relatively little inaccuracy in active dependency formation. Islands are respected or enforced and the head of the dependency is located without interference. Likewise the processing of backwards anaphora is highly grammatically accurate. In both *wh*-dependencies and backwards anaphora, the need to construct a dependency is announced by the first (temporally-occurring) element in the relationship; consequently the parser can search left-to-right. A predictive or prospective search for candidate constituents thus appears to be more accurate than a retrospective one. Some information about the syntactic context must be carried forward in these cases to allow local evaluation of the dependency. We proposed that a small amount of information carried forward could aid in licensing the dependency and in targeting the retrieval of the right dependency head.

The selective fallibility observed in ungrammatical agreement attraction sentences and in the resolution of pronominal anaphora is consistent with a content-addressable memory architecture. Previous researchers have also proposed that complex subject attachment and NPI licensing are liable to partially matching candidates (Vasishth et al., 2005; Van Dyke & Lewis, 2003). Pursuing Xiang, Dillon & Phillips (submitted)'s line of argumentation, we dismissed NPI licensing as relevant for triggering a constituent retrieval and thus reflective of retrieval-based interference. It is an interesting question, though, whether NPI licensing would be more accurate if there were some kind of early signal that a NPI would appear in the sentence. Xiang, Dillon, & Phillips' explained the

spurious licensing of NPIs by appealing to a pragmatically-supported negative interference generated by contrastive relative clauses. If, however, an NPI was expected, the comprehender may be more careful about verifying whether or not an appropriate downward entailing environment is available. As for complex subject attachment, we argued that previous experimental work has confounded clause number with the presence of a subject-like constituent (Van Dyke & Lewis, 2003; Van Dyke, 2007). In an experiment that deconfounded the two, we provided new evidence that complex subject attachment is not more difficult when there is a subject-like constituent in the context. We still found offline evidence, however, of increased difficulty in subject-interference conditions. In the previous literature, offline measures were also clearly most affected. On balance, it seems that that subjects embedded within subjects do impact the interpretation of sentences, but they do not robustly lead to increases in online complexity. This pattern of results is compatible with retrieval-based interference of the inaccessible subject, if the presence of a partially-matching constituent does not lead to RT increases in the selection process. However, it is also compatible with the influence of later comprehension processes or task-specific processes, such as sentence regeneration.

The final domain we discussed concerned reflexive anaphora. Even though reflexive anaphora resolution cannot generally be anticipated by the comprehender, it is highly grammatically accurate. Unlike agreement attraction, feature-matching constituents inside a complex subject do not intrude in processing. One possibility is that the use of clause context cues can be made salient enough to restrict the retrieval of candidate constituents to the immediate clause. This account gains greater traction if there really is no syntactic interference in complex subject attachment: we could apply

the same mechanism in both cases. The other possibility is that reflexive anaphora resolution does not rely purely on abstract cues: that is, it does not assemble its retrieval structure based on general grammatical cues, but based on cues specific to the present episode of sentence encoding. By hypothesis the verb or VP encoding contains a specific pointer to the actual subject, which it could pass to the anaphor. This conjecture is similar to one we offered to explain accuracy in locating the head of a *wh*-dependency: unique information can be used to retrieve specific constituents in a sentence that contains many abstractly similar constituents.

If the second line of explanation is on the right track, then we might expect more broadly that certain grammatical devices, like feature-passing, could serve the function of encoding constituents both with abstract features but also pointers that may prove useful later. The memory architecture we have considered not only restricts the amount of information available to make parsing decisions, but also grants access to syntactic context in a structure-insensitive fashion. While it is undeniable that there is a cost to passing information forward in time, such a cost would often be justified by the benefit of rendering information about non-local constituents as effectively local.

6 Appendices

A The Symmetry Prediction of Feature Percolation and RTs

The Symmetry Prediction of feature percolation accounts of agreement attraction states that the proportion of illusions of grammaticality should equal the proportion of illusions of ungrammaticality. If the perception of grammaticality, as a distribution of binary responses, shifts symmetrically in both grammatical and ungrammatical sentences, will reaction times also shift symmetrically? That is, would reading times in grammatical sentences be slowed down as much as reading times in ungrammatical sentences be sped up?

Response times have a characteristically right-skewed distribution (Luce, 1986) and this is true of reading times in a self-paced reading task. The distribution of responses can be modeled as an ex-Gaussian (Hohle, 1965), a distribution formed by convolving the normal and exponential distributions. By means of simulation we confirm that mixing two ex-Gaussian distributions, corresponding to ‘perceived grammatical’ and ‘perceived ungrammatical’ internal responses, leads to linear shifts in the mean of the composite distribution. Therefore the Symmetry Prediction leads us to expect a symmetrical interaction in attractor sentence RTs.

We us assume a simple model, where the reaction time response at the region of interest in attraction sentences is determined by mixing the reaction time distribution for grammatical responses with the reaction time distribution for ungrammatical responses. The mixing proportion is determined by the rate of percolation. The equations in (132) and (133) reflect this assumption. In (132), RT_{AU} is the composite reaction time

distribution for a ungrammatical sentence containing an attractor. Reaction times in this distribution are sampled with probability p from the grammatical distribution RT_G and probability $(1 - p)$ from the ungrammatical distribution RT_U . In (133), the composite reaction time distribution for a grammatical sentence containing an attractor, $RT_{A/G}$, is given in the same terms.

(132) *Ungrammatical attraction distribution*

$$RT_{A/U} = p \cdot RT_G + (1 - p) \cdot RT_U \quad p < 1$$

(133) *Grammatical attraction distribution*

$$RT_{A/G} = (1 - p) \cdot RT_G + p \cdot RT_U \quad p < 1$$

It is not difficult to show that given this model, the means of composite distributions are linearly related to the mixing proportion. We do so by simulation. The logic of the analysis and its results are given step-by-step.

1. An ex-Gaussian distribution is generated by adding a normal distribution (which determines the leading edge of the RT distribution) to an exponential (which gives the long tail). ex-Gaussian distributions have three parameters: μ - the normal mean; σ - the normal variance; τ - the exponential mean. The mean of the ex-Gaussian is simply $\mu + \tau$; and its variance $\sigma^2 + \tau^2$. See Hohle (1965) for further details.
2. The ex-Gaussian parameters for RT_G were estimated from the Sg [Sg] grammatical conditions in Wagers, Lau & Phillips (2008) Experiment 4, Region 8. Parameters for RT_U were estimated from the Sg [Sg] ungrammatical conditions. Distributions were fit via maximum likelihood

3. The ex-Gaussian parameters are fit via maximum likelihood. See the following web site for useful discussion on how to do this in the R language: <http://users.fmg.uva.nl/rgrasman/rpages/2007/07/ex-gaussian-distribution-for-reaction.html>.
4. Figures I & II below show the observed Sg [Sg] grammatical and ungrammatical distributions⁵⁹. The continuous ex-Gaussian distribution generated by the estimated parameters is superimposed. Details of the observed and fit distribution are given in the figure captions.

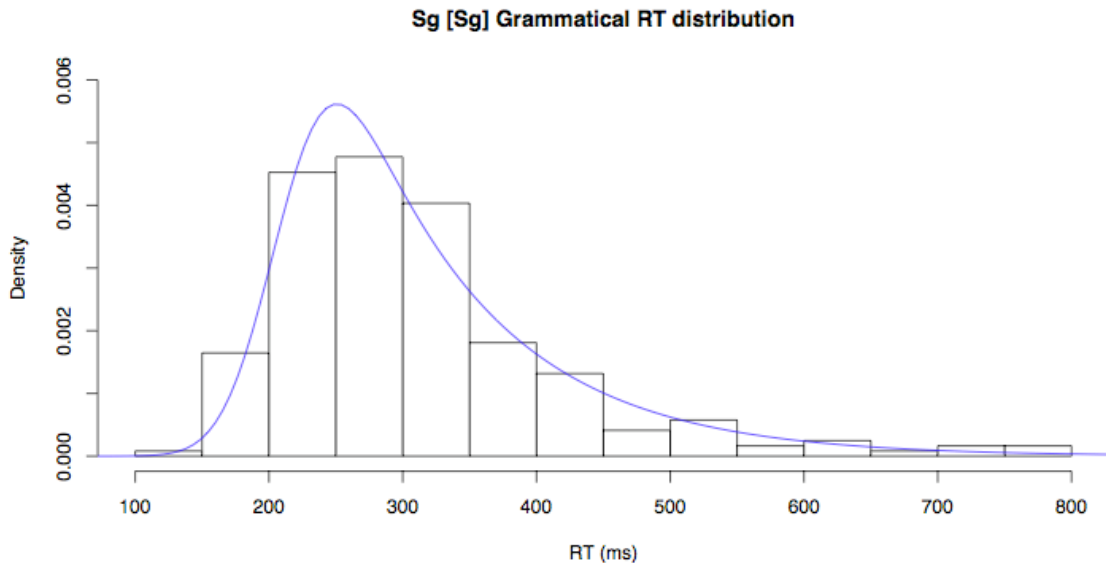


Figure 6-1 Sg [Sg] Grammatical RT Distribution: estimated RT_G

Observed parameters: $\mu = 313 \text{ ms}$; $\sigma^2 = 12010 \text{ ms}^2$

Estimated ex-Gaussian parameters: $\mu = 210 \text{ ms}$; $\sigma = 32.8 \text{ ms}$; $\tau = 104 \text{ ms}$

⁵⁹ These distributions represent all RTs in the condition/region collapsed across participants. Please note that we are assuming that the composite distributions are generated at the participant level, who samples from the pure distributions from trial to trial. There was not enough data to estimate parameters per participant and average. For the purposes of the simulation, that is irrelevant since we are only interested in how any ex-Gaussian behaves (and not interpreting the parameters).

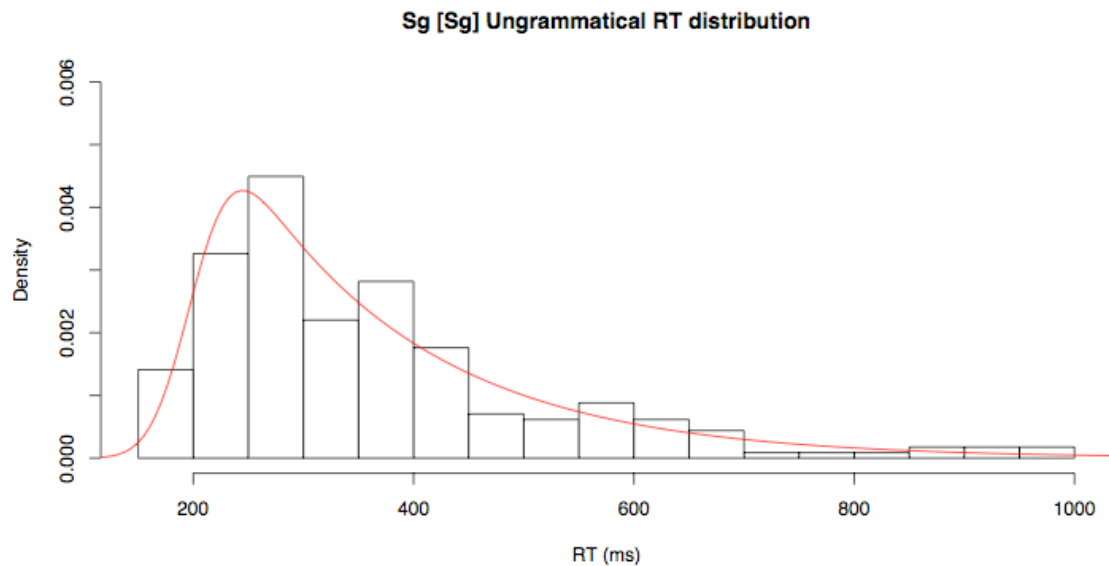


Figure 6-2 Sg [Sg] Ungrammatical RT Distribution: estimated RT_U

Observed parameters: $\mu = 365$ ms; $\sigma^2 = 25891$ ms²

Estimated ex-Gaussian parameters: $\mu = 200$ ms; $\sigma = 29.4$ ms; $\tau = 165$ ms

5. Inspection of the parameters shows that the difference seems to be carried largely in the mean of the exponential component⁶⁰.
6. Based on the parameter estimates, it is possible to generate ‘mixed’ populations according to equations (132)-(133). For example, if $p = 0.15$, then $RT_{A/G}$ can be generated by sampling 15% of its values from RT_U and 85% from RT_G .
7. 50 experiments were simulated, in which attractor RT distributions were generated for $n = 225$ trials. Mean differences between the baseline distribution and the mixed distribution computed. Figure 6-3 reports the results as follows:

⁶⁰ This is a casual observation. Note, the distribution ex-Gaussian parameters is not known analytically, so the only way to do statistical inference of the parameters would be something like bootstrapping.

- a. The x-axis corresponds to the proportion of trials drawn from the grammatical distribution. The y-axis corresponds to the mean difference between the attractor and non-attractor distributions. Error bars correspond to the standard deviation of mean differences.
 - b. Blue symbols correspond to grammatical attractor condition, and indicate how much one would slow down in the presence of an attractor, assuming percolation.
 - c. Red symbols correspond to ungrammatical attractor conditions, and indicate how much one would speed up in the presence of an attractor, assuming percolation
 - d. To work an example: Assume percolation happens 30% of the time. $p = 0.3$.
 - i. Grammatical slow-down is given by $(1 - p)$ on the x-axis: 0.7.
 - ii. Ungrammatical speed-up is given by p : 0.3.
8. Inspection reveals that speed-ups and slow-downs are symmetrical. That is, the relationship between mixing proportion and RT difference is linear.

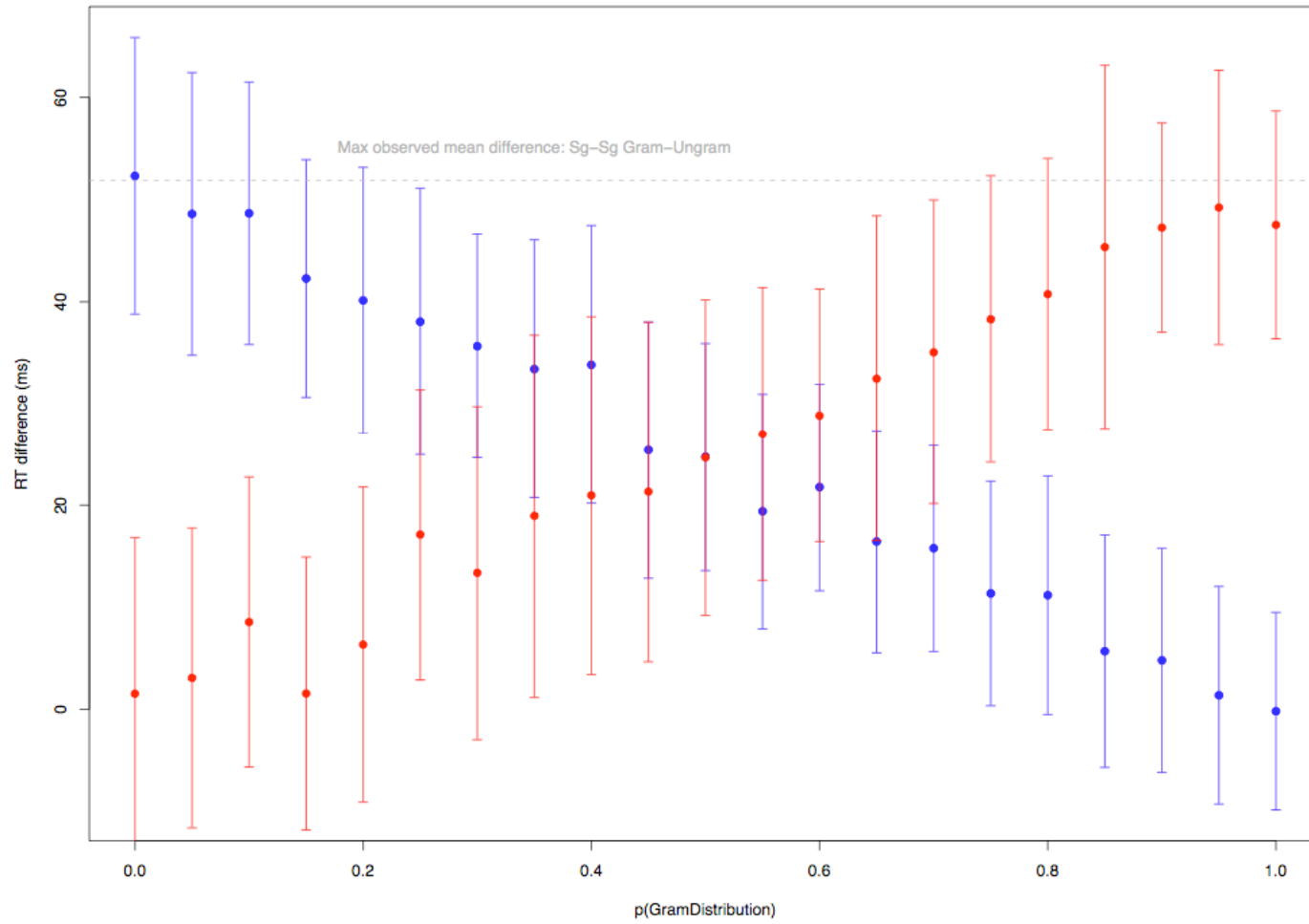


Figure 6-3 Simulation results: $RT_{A/G}/RT_{U/G}$ means shift symmetrically

B Experiments 1 – 2 Omnibus RM-ANOVA Tables

Experiment 1

ANOVA Tests reliable at $\alpha = 0.05$ in bold.

MSE: MS_{effect}

	By participants				By items				MinF ²		
	Df	MSE	F ₁	p	df	MSE	F ₂	p	df	minF ²	P
<i>Region 2 (RC head)</i>											
Attractor number	1,27	3642	5.7	0.02	1,47	7715	2.84	0.09	1,74	1.89	0.17
Grammaticality	1,27	1866	1.68	0.21	1,47	3506	2.15	0.14	1,63	0.94	0.34
Number x gram.	1,27	307	0.5	0.49	1,47	4547	2.42	0.12	1,38	0.41	0.53
<i>Region 3 ('who')</i>											
Attractor number	1,27	4109	6.49	0.02	1,47	5004	1.97	0.16	1,69	1.51	0.22
Grammaticality	1,27	704.1	1.23	0.28	1,47	397	0.25	0.62	1,64	0.21	0.65
Number x gram.	1,27	1471	1.56	0.22	1,47	2914	1.87	0.17	1,65	0.85	0.36
<i>Region 4 ('the')</i>											
Attractor number	1,27	400	1.75	0.20	1,47	1058	0.79	0.38	1,73	0.54	0.46
Grammaticality	1,27	977	2.31	0.14	1,47	284	0.29	0.59	1,58	0.25	0.62
Number x gram.	1,27	187	0.22	0.64	1,47	243	0.13	0.72	1,74	< 0.1	0.78
<i>Region 5 (RC subject)</i>											
Attractor number	1,27	4001	6.26	0.02	1,47	10474	4.8	0.03	1,73	2.71	0.1
Grammaticality	1,27	282	0.61	0.44	1,47	11	< 0.1	0.94	1,48	< 0.1	0.92
Number x gram.	1,27	51	< 0.1	0.81	1,47	1020	0.51	0.47	1,33	< 0.1	0.82
<i>Region 6 (verb)</i>											
Attractor number	1,27	3051	2.06	0.16	1,47	7341	2.04	0.15	1,69	1.02	0.32
Grammaticality	1,27	270	0.33	0.57	1,47	335	< 0.1	0.78	1,66	< 0.1	0.79
Number x gram.	1,27	2722	0.32	0.58	1,47	321	0.12	0.73	1,71	< 0.1	0.77
<i>Region 7 (verb+1)</i>											
Attractor number	1,27	28444	8.24	0.01	1,47	49733	12.2	<0.01	1,60	4.91	0.03
Grammaticality	1,27	38706	5.31	0.03	1,47	56498	9.87	<0.01	1,55	3.45	0.07
Number x gram.	1,27	17078	6.92	0.01	1,47	31842	10.6	<0.01	1,59	4.18	0.05
<i>Region 8 (verb+2)</i>											
Attractor number	1,27	2846	2.37	0.14	1,47	4518	1.09	0.30	1,73	0.75	0.39
Grammaticality	1,27	3312	1.97	0.17	1,47	4587	1.32	0.25	1,74	0.79	0.38
Number x gram.	1,27	869	1.27	0.27	1,47	1102	0.3	0.58	1,66	0.24	0.63
<i>Region 9 (verb+3)</i>											
Attractor number	1,27	3845	4.26	0.05	1,47	7502	4.24	0.04	1,69	2.12	0.15
Grammaticality	1,27	0.6	< 0.1	0.98	1,47	490	0.1	0.75	1,27	< 0.1	1
Number x gram.	1,27	590	0.88	0.36	1,47	1264	0.91	0.34	1,68	0.45	0.5
<i>Region 10 (verb+4)</i>											
Attractor number	1,27	4451	3.75	0.06	1,47	7104	3.8	0.05	1,68	1.89	0.17
Grammaticality	1,27	5062	3.3	0.08	1,47	11269	5.91	0.02	1,56	2.12	0.15
Number x gram.	1,27	188	0.227	0.64	1,47	1198	0.503	0.48	1,51	0.16	0.69

Experiment 2: 2 x 2 x 2 ANOVA

ANOVA Tests reliable at $\alpha = 0.05$ in bold.

MSE: MS_{Effect}

	By participants				By items				MinF'		
	df	MS_{effect}	F_1	p	df	MS_{effect}	F_2	p	df	minF'	P
<i>Region 2 (RChad)</i>											
Grammaticality	1,55	643	0.35	0.56	1,47	2416	2.08	0.16	1,73	0.3	0.59
Attractor number	1,55	20084	8.18	0.01	1,47	14098	5.89	0.02	1,96	3.42	0.07
Subject number	1,55	2643	1.43	0.24	1,47	4792	2.05	0.16	1,101	0.84	0.36
Att-Num x gram.	1,55	9320	2.52	0.12	1,47	3402	1.05	0.31	1,82	0.74	0.39
Sub-Num x gram	1,55	711	0.39	0.59	1,47	580	0.26	0.62	1,100	0.14	0.71
A-num x S-num	1,55	9320	2.52	0.12	1,47	6379	2.36	0.13	1,101	1.22	0.27
3-way interaction	1,55	1042	0.53	0.47	1,47	353	0.15	0.70	1,72	0.12	0.73
<i>Region 3 (who')</i>											
Grammaticality	1,55	314	0.29	0.59	1,47	75	< 0.1	0.84	1,102	0.91	0.34
Attractor number	1,55	8909	5.52	0.02	1,47	5052	3.6	0.06	1,94	2.18	0.14
Subject number	1,55	2212	1.26	0.27	1,47	7988	4.22	0.05	1,84	0.97	0.33
Att-Num x gram.	1,55	2	< 0.1	0.97	1,47	18	< 0.1	0.93	1,78	< 0.1	1
Sub-Num x gram	1,55	152	< 0.1	0.76	1,47	570	0.47	0.5	1,75	< 0.1	0.78
A-num x S-num	1,55	1876	1.01	0.32	1,47	1854	1.9	0.18	1,97	0.66	0.42
3-way interaction	1,55	4427	4.03	0.05	1,47	3437	1.93	0.17	1,86	1.3	0.26
<i>Region 5 (RC subj)</i>											
Grammaticality	1,55	703	0.27	0.60	1,47	2749	1	0.32	1,82	0.21	0.65
Attractor number	1,55	15895	7.03	0.01	1,47	10596	4.75	0.03	1,95	2.83	0.1
Subject number	1,55	8661	4.34	0.04	1,47	9244	5.35	0.03	1,102	2.4	0.12
Att-Num x gram.	1,55	828	0.39	0.54	1,47	3723	1.02	0.32	1,90	0.28	0.6
Sub-Num x gram	1,55	1945	0.76	0.39	1,47	2940	1.36	0.25	1,98	0.49	0.49
A-num x S-num	1,55	2941	1.21	0.28	1,47	1948	0.76	0.39	1,93	0.46	0.5
3-way interaction	1,55	3871	1.67	0.20	1,47	3319	1.21	0.28	1,96	0.70	0.4
<i>Region 6 (verb)</i>											
Grammaticality	1,55	942	0.3	0.59	1,47	1831	0.72	0.40	1,92	0.21	0.65
Attractor number	1,55	1443	0.61	0.44	1,47	261	< 0.1	0.80	1,57	< 0.1	0.81
Subject number	1,55	35005	11.6	0.001	1,47	31055	11.4	0.001	1,101	5.76	0.02
Att-Num x gram.	1,55	540	0.22	0.64	1,47	61	0.02	0.88	1,56	< 0.1	0.89
Sub-Num x gram	1,55	717	0.25	0.62	1,47	1477	0.45	0.50	1,98	0.16	0.69
A-num x S-num	1,55	1	< 0.1	0.99	1,47	12	< 0.1	0.96	1,64	< 0.1	0.99
3-way interaction	1,55	7927	2.39	0.13	1,47	6128	1.79	0.19	1,7	1.02	0.32
<i>Region 7 (verb+1)</i>											
Grammaticality	1,55	363776	28.6	<0.001	1,47	251224	42.1	<0.001	1,101	17	<0.001
Attractor number	1,55	11147	1.81	0.18	1,47	4380	1.12	0.29	1,93	0.69	0.41
Subject number	1,55	4289	0.45	0.50	1,47	1817	0.45	0.51	1,101	0.22	0.64
Att-Num x gram.	1,55	19253	3.48	0.07	1,47	13675	1.89	0.18	1,89	1.23	0.27
Sub-Num x gram	1,55	2216	0.54	0.47	1,47	141	< 0.1	0.86	1,52	< 0.1	0.86
A-num x S-num	1,55	130	< 0.1	0.89	1,47	1517	0.29	0.59	1,62	< 0.1	0.89
3-way interaction	1,55	1044	0.30	0.58	1,47	5546	1.37	0.25	1,78	0.25	0.62
<i>Region 8 (verb+2)</i>											
Grammaticality	1,55	5540	1.39	0.24	1,47	1566	0.47	0.50	1,77	0.35	0.56
Attractor number	1,55	2723	0.92	0.34	1,47	2785	1.04	0.31	1,102	0.49	0.49
Subject number	1,55	45	< 0.1	0.90	1,47	49	0.02	0.90	1,102	< 0.1	0.9
Att-Num x gram.	1,55	7739	2.65	0.11	1,47	9174	2.85	0.10	1,102	1.37	0.24
Sub-Num x gram	1,55	4338	1.30	0.26	1,47	2174	0.90	0.35	1,95	0.53	0.47
A-num x S-num	1,55	1501	0.61	0.44	1,47	545	0.25	0.62	1,82	0.18	0.67
3-way interaction	1,55	27327	6.03	0.02	1,47	31886	7.68	0.01	1,102	3.38	0.07

Experiment 2: 2 x 2 ANOVAs

First number is 2x2 for RC Subject=singular, second number is 2x2 for RC Subject=plural. ANOVA Tests reliable at $\alpha = 0.05$ in bold. MSE: MS_{Effect}

	By participants				By items				MinF'		
	df	MS _{effect}	F ₁	p	df	MS _{effect}	F ₂	p	df	minF'	p
<i>Region 2 (RC head)</i>											
Grammaticality	1,55	1354 1	0.65 < 0.1	0.43 0.98	1,47	2682 314	1.79 0.16	0.19 0.69	1,88 1,55	0.47 < 0.1	0.49 0.98
Attractor numb.	1,55	28383 1021	9.4 0.35	< 0.01 0.56	1,47	19721 755	7.39 0.31	0.01 0.58	1,98 1,100	4.14 0.17	0.04 0.68
Number x gram.	1,55	7950 1894	4.19 0.84	0.05 0.36	1,47	2974 781	0.95 0.31	0.33 0.58	1,68 1,79	0.78 0.23	0.38 0.63
<i>Region 3 ('who')</i>											
Grammaticality	1,55	451 15	0.35 < 0.1	.56 .92	1,47	115 530	< 0.1 0.34	0.79 0.56	1,66 1,56	< 0.1 < 0.1	0.81 0.92
Attractor number	1,55	1305 9480	0.54 9.1	0.47 < 0.01	1,47	393 6514	0.29 6.35	0.59 0.02	1,89 1,96	0.19 3.74	0.66 0.06
Number x gram.	1,55	2319 2110	0.35 1.87	0.56 0.18	1,47	1478 1978	0.84 0.86	0.36 0.36	1,92 1,85	0.25 0.59	0.62 0.44
<i>Region 5 (RC subj.)</i>											
Grammaticality	1,55	155 2494	< 0.1 0.78	0.78 0.38	1,47	2 5687	< 0.1 2.2	0.98 0.14	1,48 1,85	< 0.1 0.57	0.98 0.45
Attractor number	1,55	2581 16255	0.98 7.85	0.33 0.01	1,47	1729 10815	0.36 3.86	0.36 .06	1,100 1,87	0.46 2.59	0.5 0.11
Number x gram.	1,55	559 4140	0.27 1.76	0.61 0.19	1,47	6 7036	< 0.1 1.78	0.96 0.19	1,48 1,101	< 0.1 0.88	0.96 0.35
<i>Region 6 (verb)</i>											
Grammaticality	1,55	8 1651	< 0.1 0.42	0.95 0.52	1,47	10 3299	< 0.1 1.1	0.95 0.3	1,101 1,87	< 0.1 0.31	0.97 0.58
Attractor number	1,55	677 767	0.27 0.2	0.6 0.66	1,47	194 80	< 0.1 < 0.1	0.84 0.89	1,61 1,55	< 0.1 < 0.1	0.84 0.89
Number x gram.	1,55	6302 2165	2.14 0.74	0.15 0.39	1,47	3708 2482	1.13 0.82	0.29 0.37	1,89 1,102	0.74 0.39	0.39 0.53
<i>Region 7 (verb+1)</i>											
Grammaticality	1,55	154603 211389	18.2 25.5	<.001 <.001	1,47	131642 119723	17.8 38.1	<.001 <.001	1,101 1,98	25.5 38.1	<.001 <.001
Attractor number	1,55	4436 6841	0.93 0.87	0.34 0.36	1,47	5527 371	1.11 < 0.1	0.3 0.77	1,102 1,57	0.51 < 0.1	0.48 0.36
Number x gram.	1,55	14631 5666	3.40 1.22	0.07 0.27	1,47	18319 902	3.84 0.14	0.06 0.71	1,102 1,58	1.8 0.12	0.18 0.73
<i>Region 8 (verb+2)</i>											
Grammaticality	1,55	37 9841	< 0.1 2.12	0.91 0.15	1,47	25 3715	< 0.1 1.37	0.78 0.25	1,73 1,93	< 0.1 0.83	0.92 0.36
Attractor number	1,55	4135 90	1.44 < 0.1	0.24 0.85	1,47	2896 433	0.93 0.26	0.34 0.62	1,94 1,70	0.56 < 0.1	0.46 0.86
Number x gram.	1,55	32075 2991	8.52 0.81	< 0.01 0.37	1,47	37633 3427	8.63 1.14	0.01 0.29	1,101 1,101	4.29 0.47	0.04 0.49

C Experiments 7, 8, & 11A Omnibus RM-ANOVA Tables

NB: Significance tests: p: 0 *** 0.001 ** 0.01 * 0.05 • 0.10

MSE: MS_{Effect}

Experiment 7

	VP STRUCTURE	FILLER PLAUSIBILITY	STRUCTURE × PLAUSIBILITY
<i>Region 7</i> VP₁ adverb	F ₁ : 0.4; MSE: 8177 F ₂ : 0.0; MSE: 298	F ₁ : 0.0; MSE: 169 F ₂ : 0.1; MSE: 1811	F ₁ : 0.0; MSE: 0.0 F ₂ : 0.0; MSE: 13
<i>Region 8</i> VP₁ verb	F ₁ : 0.7; MSE: 41882 F ₁ : 0.4; MSE: 25226	F ₁ : 0.3; MSE: 12982 F ₁ : 0.5; MSE: 19470	F ₁ : 0.0; MSE: 177 F ₂ : 0.2; MSE: 11257
<i>Region 9</i> Coordinator/ Preposition	**F ₁ : 11.1; MSE: 295434 *F ₂ : 5.2; MSE: 231453	F ₁ : 1.4; MSE: 29350 F ₂ : 1.3; MSE: 27215	F ₁ : 0.3; MSE: 5091 F ₂ : 0.3; MSE: 7181
<i>Region 10</i> VP₂ adverb	**F ₁ : 8.7; MSE: 250432 *F ₂ : 5.0; MSE: 182765	F ₁ : 0.0; MSE: 493 F ₂ : 0.0; MSE: 114	F ₁ : 2.8; MSE: 83234 F ₂ : 1.3; MSE: 65292
<i>Region 11</i> VP₂ verb	F ₁ : 1.6; MSE: 84783 F ₂ : 2.0; MSE: 155712	F ₁ : 2.4; MSE: 74711 *F ₂ : 5.5; MSE: 76929	F ₁ : 0.0; MSE: 208 F ₂ : 0.0; MSE: 89
<i>Region 12</i> VP₂ verb + 1	•F ₁ : 3.0; MSE: 98836 •F ₂ : 3.0; MSE: 94310	F ₁ : 0.9; MSE: 15066 F ₂ : 1.1; MSE: 15248	F ₁ : 0.6; MSE: 7934 F ₂ : 0.4; MSE: 8257
<i>Region 13</i> VP₂ verb + 2	***F ₁ : 31.6; MSE: 1194454 ***F ₂ : 19.9; MSE: 1077296	F ₁ : 0.20; MSE: 5083 F ₂ : 0.07; MSE: 2119	**F ₁ : 7.9; MSE: 147273 *F ₂ : 4.5; MSE: 172297
<i>Region 14</i> VP₂ verb + 3	F ₁ : 0.3; MSE: 3391 F ₂ : ~0; MSE: 54	F ₁ : 1.7; MSE: 18320 F ₂ : 0.9; MSE: 20788	F ₁ : 0.8; MSE: 16320 F ₂ : 0.8; MSE: 21034
<i>Region 15</i> VP₂ verb + 4	•F ₁ : 3.9; MSE: 80104 F ₂ : 2.7; MSE: 108783	F ₁ : 0.1; MSE: 539 F ₂ : ~0; MSE: 52	F ₁ : 0.4; MSE: 4331 F ₂ : 0.3; MSE: 6060

Numerator *df* in each manipulated factor: 1. Subject *n*: 36. Item *n*: 24

Experiment 8

	VP STRUCTURE	FILLER PLAUSIBILITY	STRUCTURE × PLAUSIBILITY
<i>Region 6</i> VP₁ verb or PP prep.	F ₁ : 2.0; MSE: 47908 F ₂ : 0.8; MSE: 48478	F ₁ : ~0; MSE: 65 F ₂ : 0.1; MSE: 6555	F ₁ : 2.5; MSE: 48105 F ₂ : 1.0; MSE: 60876
<i>Region 7</i> VP₁ verb + 1 PP prep. + 1	F ₁ : 0.3; MSE: 393 F ₂ : 0.0; MSE: 58	F ₁ : 1.0; MSE: 6225 F ₂ : 0.2; MSE: 4768	*F ₁ : 4.8; MSE: 33625 F ₂ : 1.8; MSE: 31967
<i>Region 8</i> VP₁ verb + 2 PP prep + 2	**F ₁ : 12.6; MSE:261161 **F ₂ : 8.5; MSE: 281351	F ₁ : 0.3; MSE: 2585 F ₂ : 0.0; MSE: 630	F ₁ : 0.3; MSE: 3288 F ₂ : 0.4; MSE: 11544
<i>Region 9</i> VP₁ verb + 3 PP prep + 4	***F ₁ : 21.9; MSE:497323 **F ₂ : 9.5; MSE: 42453	F ₁ : 0.1; MSE: 2716 F ₂ : 0.0; MSE: 138	F ₁ : 1.7; MSE: 20166 F ₂ : 0.5; MSE: 8951
<i>Region 10</i> PP prep + 5	***F ₁ : 27.2; MSE:654984 ***F ₂ : 16.7; MSE:720729	F ₁ : 0.6; MSE: 9607 F ₂ : 0.2; MSE: 5576	F ₁ : 0.1; MSE: 2037 F ₂ : 0.1; MSE: 3089
<i>Region 11</i> Adverb	***F ₁ : 15.7; MSE:290247 F ₂ : 2.7; MSE: 277167	F ₁ : 0.5; MSE: 17590 F ₂ : 0.4; MSE: 22401	F ₁ : 2.4; MSE: 66569 F ₂ : 2.3; MSE: 88303
<i>Region 12</i> VP₂ verb	***F ₁ : 17.7; MSE:330954 *F ₂ : 5.9; MSE: 372924	*F ₁ : 4.9; MSE: 76503 F ₂ : 1.5; MSE: 41584	F ₁ : 0.4; MSE: 10272 F ₂ : 0.3; MSE: 9308
<i>Region 13</i> Ground Arg 1	*F ₁ : 4.8; MSE: 64446 *F ₂ : 6.4; MSE: 75865	*F ₁ : 6.8; MSE: 102345 •F ₂ : 3.0; MSE: 88161	F ₁ : 0.1; MSE: 386 F ₂ : 0.0; MSE: 574
<i>Region 14</i> Ground Arg 2	**F ₁ : 8.9; MSE: 77807 •F ₂ : 4.2; MSE: 87912	F ₁ : ~0; MSE: 60 F ₂ : 0.0; MSE: 359	F ₁ : 0.3; MSE: 1680 F ₁ : 0.1; MSE: 1640
<i>Region 15</i> Ground Arg 3	F ₁ : 3.8; MSE: 48399 F ₂ : 2.3; MSE: 44052	•F ₁ : 3.5; MSE: 26226 F ₂ : 0.6; MSE: 14936	*F ₁ : 4.6; MSE: 33288 F ₂ : 0.6; MSE: 17751
<i>Region 16</i> Ground Arg 4	F ₁ : 0.9; MSE: 14312 F ₂ : 1.3; MSE: 21848	**F ₁ : 9.1; MSE: 99796 F ₂ : 2.6; MSE: 74079	F ₁ : 0.5; MSE: 4012 F ₂ : 1.3; MSE: 2012
<i>Region 17</i> Figure prep.	F ₁ : 0.9; MSE: 6877 F ₂ : 1.0; MSE: 7083	**F ₁ : 7.8; MSE: 37100 F ₂ : 1.5; MSE: 41409	F ₁ : 0.3; MSE: 1497 F ₂ : 0.3; MSE: 3683
<i>Region 18</i> AdvP Word 1	*F ₁ : 6.7; MSE: 53301 F ₂ : 3.5; MSE: 59611	*F ₁ : 8.0; MSE: 71070 F ₂ : 2.2; MSE: 59426	F ₁ : 0.0; MSE: 308 F ₂ : ~0; MSE: 50
<i>Region 19</i> AdvP Word 2	*F ₁ : 4.8; MSE: 75894 •F ₂ : 4.1; MSE: 120830	*F ₁ : 5.1; MSE: 137215 •F ₂ : 3.0; MSE: 97036	F ₁ : 0.1; MSE: 860 F ₂ : 0.4; MSE: 14768
<i>Region 20</i> AdvP Word 3	F ₁ : 1.9; MSE: 24955 F ₂ : 1.0; MSE: 23005	***F ₁ : 22.3; MSE: 78628 **F ₂ : 10.4; MSE: 189923	F ₁ : 1.1; MSE: 7324 F ₂ : 0.4; MSE: 7361

Numerator *df* in each manipulated factor: 1. Subject *n*: 31. Item *n*: 24

Experiment 11A

	DEPENDENCY LENGTH	FILLER PLAUSIBILITY	LENGTH × PLAUSIBILITY
<i>Region 11</i> Adverb	F ₁ : 2.4; MSE: 84672 F ₂ : 2.6; MSE: 67148	F ₁ : 0.5; MSE: 10728 F ₂ : 0.0; MSE: 1078	F ₁ : 0.7; MSE: 19646 F ₂ : 1.1; MSE: 34595
<i>Region 12</i> Verb	F ₁ : 0.8; MSE: 29538 F ₂ : 0.6; MSE: 26439	F ₁ : 0.2; MSE: 7270 F ₂ : 0.0; MSE: 1085	F ₁ : 2.2; MSE: 95446 F ₂ : 1.8; MSE: 69534
<i>Region 13</i> Ground Arg 1	F ₁ : 1.2; MSE: 38526 F ₂ : 1.7; MSE: 45099	**F ₁ : 13.3; MSE: 201845 **F ₂ : 9.1; MSE: 277367	F ₁ : 0.2; MSE: 6148 F ₂ : 0.1; MSE: 2435
<i>Region 14</i> Ground Arg 2	F ₁ : 0.3; MSE: 7987 F ₂ : 0.3; MSE: 6818	F ₁ : 0.3; MSE: 5894 F ₂ : 0.1; MSE: 2768	•F ₁ : 3.5; MSE: 66349 F ₂ : 1.9; MSE: 46906
<i>Region 15</i> Ground Arg 3	F ₁ : 0.7; MSE: 20853 F ₂ : 1.1; MSE: 25423	F ₁ : 1.6; MSE: 27875 F ₂ : 1.3; MSE: 43107	F ₁ : 0.3; MSE: 5668 F ₂ : 0.0; MSE: 110
<i>Region 16</i> Ground Arg 4	F ₁ : 1.4; MSE: 13097 F ₂ : 0.5; MSE: 11494	•F ₁ : 3.4; MSE: 49719 F ₂ : 2.3; MSE: 71552	F ₁ : 1.8; MSE: 68998 F ₂ : 1.5; MSE: 34298
<i>Region 17</i> Figure Prep	*F ₁ : 6.3; MSE: 109005 **F ₂ : 9.4; MSE: 131393	F ₁ : 0.6; MSE: 8899 F ₂ : 1.1; MSE: 11860	F ₁ : ~0; MSE: 68 F ₂ : 0.1; MSE: 1104
<i>Region 18</i> AdvP Word 1	F ₁ : 0.7; MSE: 14124 F ₂ : 0.7; MSE: 20491	*F ₁ : 7.4; MSE: 227500 *F ₂ : 7.2; MSE: 205372	F ₁ : 0.5; MSE: 15231 F ₂ : 0.1; MSE: 2471F
<i>Region 19</i> AdvP Word 2	F ₁ : 0.3; MSE: 3636 F ₂ : 0.1; MSE: 2368	**F ₁ : 13.1; MSE: 197729 **F ₂ : 8.9; MSE: 243801	F ₁ : ~0; MSE: 182 F ₂ : 0.0; MSE: 1171
<i>Region 20</i> AdvP Word 3	F ₁ : 1.5; MSE: 17357 F ₂ : 1.9; MSE: 15909	*F ₁ : 5.4; MSE: 52683 **F ₂ : 9.4; MSE: 78446	•F ₁ : 4.1; MSE: 34742 F ₂ : 1.1; MSE: 22117

Numerator df in each manipulated factor: 1. Subject *n*: 24. Item *n*: 24
Significance tests: p: 0 *** 0.001 ** 0.01 * 0.05 • 0.10.

Notes: In short dependencies, the regions above correspond to ordinal word positions 6-15, but these have been renamed to facilitate comparison with long dependencies.

D Experiment 11B Reading times for Short and Long:PP Conditions

2.5 S.D. Trimmed Reading Times

Mean and standard error reported in ms

<i>LENGTH</i>	<i>Short</i>				<i>PP</i>			
	Plausible		Implausible		Plausible		Implausible	
PLAUSIBILITY Region	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
1	347	9	353	11	339	9	331	9
2	345	10	356	11	349	11	339	10
3	327	9	336	10	379	14	345	12
4	317	9	353	12	321	8	332	9
5	337	11	339	11	348	11	330	10
6	300	7	319	8	322	9	336	9
7	326	9	337	10	322	8	313	7
8	326	10	340	10	317	8	318	9
9	336	12	322	9	352	12	330	9
10	329	10	333	10	352	11	330	9
11	341	11	354	12	348	12	350	11
12	353	13	366	11	359	11	362	11
13	342	10	360	12	343	9	357	10
14	348	12	372	12	348	10	359	10
15	347	12	364	12	359	12	359	11
16	343	11	352	10	340	9	341	9
17	325	8	341	9	337	8	342	8
18	329	10	353	10	337	9	336	8
19	336	10	358	9	345	10	371	13
20	341	10	367	10	357	12	363	12
21	349	11	369	10	351	9	345	9
22	324	9	347	10	336	9	344	8

7 References

- Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247–64.
- Anderson, J.R. (1983). Retrieval of information from long-term memory. *Science*, 220(4592), 25-30.
- Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, 111(4), 1036-60.
- Anderson, M.C. & Neely, J.H. (1996). Interference and inhibition in memory retrieval. E.L. Bjork & R.A. Bjork (Eds.), *Memory. Handbook of Cognition and Perception* (2nd ed.), pp. 237-313. San Diego: Academic Press.
- Anderson, S.R. (1971). On the role of deep structure in semantic interpretation. *Foundations of Language*, 7, 387-396.
- Aoshima, S., Phillips, C., & Weinberg, A.S. (2004). Processing filler–gap dependencies in a head–final language. *Journal of Memory and Language*, 51, 23–54.
- Baddeley, A.D. (1986). *Working Memory*. Oxford: Clarendon Press.
- 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.
- Badecker, W., & Lewis, R. (2007). A new theory and computational model of working memory in sentence production: agreement errors as failures of cue-based retrieval. *Proceedings of the 20th Annual CUNY Conference on Human Sentence Processing*. La Jolla, CA: University of California, San Diego.
- 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.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (submitted). Mixed-effects modeling with crossed random effects for subjects and items.
- Barss, A., & Lasnik, H. (1986). A note on anaphora and double objects. *Linguistic Inquiry*, 17, 347-54.
- Bayer, S., & Johnson, M. (1996). Features and Agreement. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*. Morristown, NJ: Association for Computational Linguistics, 70-76.

- Bertram, R., Hyona, J. & Laine, M. (2000). Word morphology in context: Effects of base and surface frequency. *Language and Cognitive Processes*, 15, 367-388.
- Berwick, R. & Weinberg, A. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.
- Bever, T. (1974). The Ascent of the Specious or There's a lot we Don't Know about Mirrors. In Cohen, D. (ed.), *Explaining Linguistic Phenomena*. Washington, DC: Hemisphere Publishing Company.
- Bhatt, R. (1999). *Covert Modality in Non-finite Contexts*, Ph.D. Dissertation, University of Pennsylvania.
- Blouin, D. C., & Riopelle, A. J. (2005). On confidence intervals for within-subjects designs. *Psychological Methods*, 10, 397-412.
- Bock, J. K. & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99-127.
- Bock, J. K. & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, 8, 57-99.
- Bock, J. K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, 51, 251-278.
- Bock, J. K., & Miller, C. A. (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.
- Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., & Carlson, G.N. (1995). Verb argument structure in parsing and interpretation: evidence from wh-questions. *Journal of Memory and Language*, 34, 774-806.
- Bourdages, J.S. (1992). Parsing complex NPs in French. In Goodluck, H., & Rochemond, M.S. (eds.), *Island constraints: Theory, acquisition and processing*. Dordrecht: Kluwer.
- Bousfield, W.A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229-40.
- Bresnan, J. (2001). *Lexical-Functional Syntax*. Malden, MA: Blackwell.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Oxford University Press.

- Chomsky, N. (1977). On Wh-Movement. In Culicover, P.W., Wasow, T., & A. Akmajian, eds., *Formal syntax*. New York: Academic Press. 71-132.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1994). Bare phrase structure. *MIT Occasional Papers in Linguistics 5*. Cambridge, MA: MIT Working Papers in Linguistics.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clark, H. (1973). The language-as-a-fixed-effect fallacy: critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review*, 3, 37-60.
- Clifton, C., Kennison, S.M., & Albrecht, J.E. (1997). Reading the words *him* and *her*: implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36, 276-292.
- Clifton, C., Jr., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, 11-39.
- Cofer, C.N. Bruce, D.R., & Reicher, G.M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 71, 858 - 866.
- Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford UP.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87-185.
- Cowart, W., & Cairns, H.S. (1987). Evidence for an anaphoric mechanism within syntactic processing: some reference relations defy semantic and pragmatic constraints. *Memory & Cognition*, 15, 318-331.
- Cowper, E.A. (1976). *Constraints on sentence complexity: a model for syntactic processing*. Ph.D. Dissertation, Brown University.
- Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.

- Crain, S., & Fodor, J.D. (1985). How can grammars help parsers? In Dowty, D., Karttunen, L., & Zwicky, A. (eds.), *Natural language parsing*. Cambridge: Cambridge University Press.
- Deane, P. (1991). Limits to attention: A cognitive theory of island constraints. *Cognitive Linguistics*, 2, 1 – 63.
- Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- den Dikken, M. (2001). “Pluringulars”, pronouns and quirky agreement. *The Linguistic Review*, 18, 19-41.
- Dosher, B.A. (1976). The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, 8, 291-310.
- Dosher, B.A. (1981). The effects of delay and interference: A speed-accuracy study. *Cognitive Psychology*, 13, 551-582.
- Drenhaus, H., Saddy, D. & Frisch, St. (2005) Processing negative polarity items: When negation comes through the backdoor. In Kepser, S. & Reis, M. (Eds.), *Linguistic Evidence -- Empirical, Theoretical, and Computational Perspectives*. Berlin: Mouton de Gruyter, 145-165.
- Eberhard, K. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, 36, 147-164.
- Eberhard, K., (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language*, 36, 147-164.
- Eberhard, K. M., Cutting, J. C., & Bock, J. K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112, 531-559.
- Eich, J.E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, 8, 157-173.
- Eich, J.M. (1982). A composite holographic associative recall model. *Psychological Review*, 89, 627-61.
- Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, 6, 5–34.
- Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject-verb agreement errors. A study in French written language. *Quarterly Journal of Experimental Psychology A*, 47, 437-464.

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-91.
- Ferreira, F. & Henderson, J.M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*, 725-745.
- Fodor, J.A., Bever, T.G. & Garrett, M.F. (1974). *The psychology of language: an introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- Fiebach, C.J., Schlesewsky, M., & Friederici, A.D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*, 250-272.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*, 427-473.
- Ford, M. (1983). A method of obtaining measures of local parsing complexity throughout sentences. *Journal Verbal Learning & Verbal Behavior*, *22*, 203-18.
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Attraction in sentence production: The role of syntactic structure. *Language and Cognitive Processes*, *17*(4), 371-404.
- Francis, W. N. (1986). Proximity concord in English. *Journal of English Linguistics*, *19*, 309-17.
- Fraser, B. (1971). A note on the spray-paint cases. *Linguistic Inquiry*, *2*, 603-7.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*, 519-560.
- Frazier, L., & Flores D'Arcais, G.B. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory of Language*, *28*, 331-44.
- Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*, 343-370.
- Freedman, S.E., & Forster, K.I. (1985). The psychological status of overgenerated sentences. *Cognition*, *19*, 101-31.
- van Gelderen, E. (1997). *Verbal Agreement and the Grammar behind its Breakdown*. Tübingen: Max Niemeyer Verlag.
- Garnsey, S.M., Tanenhaus, M.K. & Chapman, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 51-60.

- Garrett, M. (1980). Levels of processing in speech production. In B. Butterworth (Ed.), *Language production, Vol. 1: Speech and talk* (pp. 177-220). New York, NY: Academic Press.
- Goldsmith, J. (1985). A principled exception to the coordinate structure constraint. In *Papers from the Twenty-First Annual Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411-1423.
- Gordon, P.C., Hendrick, R., & Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 13, 425-430.
- 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.
- Gorrell, P. (1993). Evaluating the direct association hypothesis: a reply to Pickering and Barry (1991). *Language and Cognitive Processes*, 8, 129-46.
- Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*, 8, 147-161.
- Gibson, E., & Warren, T. (2004). Reading time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55-78.
- Greenberg, J.H. (1966). *Universals of Language*. Cambridge, MA: MIT Press.
- Grossberg, S., (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17, 199-219.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-483.
- Harley, H. (1994). Hug a tree: deriving the morphosyntactic feature hierarchy. *MIT Working Papers in Linguistics*, 21, 289-320.
- Harley, H. (2002). Possession and the double object construction. *Linguistic Variation*

Yearbook, 2, 29-68.

Harrison, A.J. (2004). Three way attraction effects in Slovene. In *Proceedings of the Second CamLing Postgraduate Conference on Language Research*. Cambridge: University of Cambridge.

Hartsuiker, R. J., Antón-Méndez, I., & Van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language*, 45, 546-572.

Haussler, J. & Bader, M. (submitted). Agreement checking and number attraction in sentence comprehension: Insights from German relative clauses.

Hintzman, D.L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, 16, 96-101.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace model. *Psychological Review*, 95, 528-551.

Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 11, 741-749.

Howard, M.W. & Kahana, M.J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269-299.

Huang, C.T.J. (1982). Move wh in a language without wh movement. *The Linguistic Review*, 1, 369-416.

Jackendoff, R. S. (1977). *X'-Syntax: A Study of Phrase Structure*. Cambridge, MA: MIT Press.

Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Kaan, E. (1997). *Processing subject-object ambiguities in Dutch*. Ph.D. Dissertation, University of Groningen.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *Journal of Psycholinguistic Research*, 31(2), 165-193.

King, J., & Just, M.A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580-60.

Kjellmer, G. (1975), Are Relative Infinitives Modal? *Studia Neophilologica*, 47(2), 323-32.

- Jarvella, R.J. (1971). Syntactic Processing of Connected Speech. *Journal of Verbal Learning and Verbal Behavior*, 10, 409-16.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 3, 228-238.
- 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*, Oxford: Blackwell Publishers, pp. 439-479.
- 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, 15.1-15.32
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201.
- Kayne, R. (1984). *Connectedness and Binary Branching*. Dordrecht, Netherlands: Foris.
- Kayne, R. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kazanina, N., Lau, E.F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56, 384-409.
- Kennison, S.M. (2003). Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language*, 49, 335-352.
- Kluender, R. (2005). Are subject islands subject to a processing account? In Chand, V., Kelleher, A., Rodriguez, A.J., & Schmeiser, B. (eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press.
- Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573-633.
- Kimball, J., & Aissen, J. (1971). I think, you think, he think. *Linguistic Inquiry*, 2, 241-246.
- King, J.W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376-95.
- Knuth, D. (1965/1997). *The Art of Computer Programming. Volume 1: Fundamental Algorithms*. 3rd ed. Reading, MA: Addison-Wesley.
- Kurtzman, H.S., & Crawford, L.E. (1991). Processing parasitic gaps. In Sherer, T. (ed.), *Proceedings of the 21st Annual Meeting of the North Eastern Linguistics Society*.

Amherst, MA: LSA Publications.

Ladusaw, W. (1979). *Polarity sensitivity as inherent scope relations*. PhD thesis, University of Texas, Austin.

Lakoff, G. (1986). Frame semantic control of the coordinate structure constraint. In *Papers from the Twenty-Second Annual Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Larson, R.K. (1988). On the double object construction. *Linguistic Inquiry*, 19, 335-391.

Lau, E., Rozanova, K., & Phillips, C. (2007). Syntactic prediction and lexical surface frequency effects in sentence processing. *University of Maryland Working Papers in Linguistics*, 16, 163-200.

Lau, E., Wagers, M., Stroud, C., & Phillips, C. (2008). Agreement and the subject of confusion. *Proceedings of the 21st Annual CUNY Conference on Human Sentence Processing*. Chapel Hill, NC: University of North Carolina, Chapel Hill.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33, 51-73.

Lehtonen, M., Wande, E., Niska, H., Niemi, J. & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, 35(2), 121-146.

Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.

Lewis, R.L. (1996). Interference in short-term memory: the magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25, 193-215.

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375-419.

Marcus, M.P. (1980). *Theory of Syntactic Recognition for Natural Languages*. Cambridge, MA: MIT Press.

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.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition*, 24, 173-187.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, 38, 225-252.

- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory. *Journal of Psycholinguistic Research*, 29(2), 111-123.
- McElree, B. (2001). Working memory and focal attention. *J Exp Psychol: Learn. Mem. Cogn.* 27, 817-35.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation*, vol. 46. San Diego: Academic Press.
- McElree, B., & Bever, T. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research*, 18, 21–35.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General*, 118, 346-373.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory & Language*, 48(1), 67-91.
- McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 134-157.
- McGeoch, J.A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39, 352-70.
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognition*, 11, 495-523.
- McKoon, G., & Ratcliff, R. (1994). Sentential context and on-line lexical decision tasks. *Journal of Experimental Psychology: Language, Memory and Cognition*, 20, 1239-43.
- Miller, G.A., & Chomsky, N. (1963). Finitary models of language users. In Luce, R.D., Bush, R.R., & Galanter, E. (eds.), *Handbook of Mathematical Psychology, Volume II*. New York: John Wiley.
- Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.
- Murdock, B.B. (1983). A distributed memory model for serial-order information. *Psychological Review*, 90, 316-338.
- Murdock, B.B., & Walker, K.D. (1969). Modality effects in free recall. *Journal of Verb Learning and Verbal Behavior*, 8, pp. 665–676.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53-81.

Neville, H.J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically-based sentence processing classes: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *3*, 151-165.

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, *51*, 568-585.

Nicol, J.L., Fodor, J.D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *20*, 1229-38.

Nicol, J., Forster, K., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*, 569-587.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5-19.

Nilsson, L.-G. (1974). Further evidence for organization by modality in free recall. *Journal of Experimental Psychology*, *103*, 948-957.

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, *15*(4/5), 389-420.

Osterhout, L., & Mobley, L.A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *31*, 785-806.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, *25*, 273-285.

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, *57*, 126-149.

Page, M.P.A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761-781.

Pearlmutter, N. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research*, *29*, 89-98.

- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427-456.
- Pearlmutter, N.J., & Mendelsohn, A.A. (2000). Serial versus parallel sentence comprehension. Ms.
- Pesetsky, D.M. (1995). *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795–823.
- Phillips, C., Kazanina, N., & Abada, S.H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407–428.
- Phillips, C. & Wagers, M.W. (2007). Relating Structure and Time in Linguistics and Psycholinguistics. In Gaskell, G. (ed.), *Oxford Handbook of Psycholinguistics*, Oxford: Oxford University Press.
- Pickering, M.J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In Clifton, C., Frazier, L., & Rayner, K. (eds.), *Perspectives on Sentence Processing*. London: Lawrence Erlbaum.
- Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
- Postal, P. (1998). *Three Investigations of Extraction*. Cambridge, MA: MIT Press.
- Potter, M.C. (1988). Rapid serial visual presentation (RSVP): a method for study language processing. In Kieras, D.E. & Just, M.A. (eds.) *New methods in reading comprehension research*. Hillsdale, NJ: Erlbaum Press.
- Pickering, M.J., & Barry, G.D. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*, 229–259.
- Plate, T.A. (1994). *Distributed representations and nested compositional structure*. PhD Thesis, Department of Computer Science, University of Toronto.
- Pollard, C. & Sag, I. (1994). *Head-driven Phrase Structure Grammar*. Chicago: CSLI Publications.
- Potter, M.C., & Lombardi, L. (1992). The regeneration of syntax in short-term memory. *Journal of Memory and Language*, *31*, 713-33.
- Potter, M.C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265-82.

Pritchett, B.L. (1992). *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A comprehensive grammar of the English language*. London: Oxford UP.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “The language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416-426.

Rappaport, M., & Levin, B. (1986). What to do with Theta Roles. *Lexicon Project Working Papers* 11, Cambridge, MA: Center for Cognitive Science, MIT.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.

Radó, J. (1999). Discourse effects in gap-filling. *Proceedings of the Architecture and Mechanisms for Language Processing Conference 1999*. Edinburgh: University of Edinburgh.

De Roeck, A.N., Johnson, R.L., King, M., Rosner, M., Sampson, G., & Varile, N. (1982). A myth about centre-embedding. *Lingua*, 58, 327-40.

Rouveret, A., & Vergnaud, J.-R. (1980). Specifying Reference to the Subject: French Causatives and Conditions on Representations. *Linguistic Inquiry*, 11, 97-202.

Ross, J.R. (1967). *Constraints on variables in syntax*. Ph.D. Dissertation, MIT.

Ruchkin, D.S., Johnson, R., Jr., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76, 419-39.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-47.

Sekerina, I.A. (2003). Scrambling and processing: dependencies, complexity, and constraints. In Karimi, S. (ed.), *Word order and scrambling*. Malden, MA: Blackwell.

- Selman, B., & Hirst, G. (1985). A rule-based connectionist parsing system. In *Proceedings of the 7th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum. pp 212-221.
- Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In Hemforth, B., & Konieczny, L. (eds.), *German sentence processing*. Dordrecht: Kluwer Academic Publishers.
- Shallice, T., & Vallar, G. (1990). *Neuropsychological impairments of short-term memory*. New York: Cambridge University Press.
- Shieber, S. M. (1986). *An introduction to unification-based approaches to grammar*. CSLI Lecture Notes No. 4. Stanford: CSLI.
- Shiffrin, R. M., Murnane, K., Gronlund, S., & Roth, M. (1989). On units of storage and retrieval. In Izawa, C. (Ed.), *Current Issues in Cognitive Processes: The Tulane Flowerree Symposium on Cognition*. Hillsdale, NJ: Erlbaum.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4(2), 145-166.
- Solomon, E., & Pearlmutter, N. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, 49, 1-46.
- Stevenson, S. (1994). Competition and Recency in a Hybrid Network Model of Syntactic Disambiguation. *Journal of Psycholinguistic Research*, 23, 295-322.
- Strang, B. (1966). Some features of S-V concord in present day English. In I. Cellini & G. Melchiori (Eds.), *English Studies Today: Fourth Series* (pp. 73-87). Rome: Edizioni di Storia et Letteratura.
- Stowe, L. (1986). Parsing wh-constructions: evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-46.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3), 542-562.
- Sussman, R., & Sedivy, J. (2003). The time course of processing syntactic dependencies. *Language and Cognitive Processes*, 18, 143-63.
- Szabolcsi, A., & den Dikken, M. (2002). Islands. In Cheng, L., & Sybesma, R. (eds.), *The Second State-of-the-Article Book*. Mouton de Gruyter.
- Tabor, W., Galantucci, T., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355-370.

- Tanenhaus, M.K., Stowe, L.A., & Carlson, G. (1985). The interaction of lexical expectation and pragmatics in parsing filler-gap constructions. *Proceedings of the Seventh Annual Cognitive Science Society Conference*, Irving, CA.
- Traxler, M.J., & Pickering, M.J. (1996). Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language*, 35, 454-475.
- Trollope, A. (1883/1999). *An Autobiography*. New York: Oxford University Press.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.
- Uriagereka, J. (1999). Multiple Spell-out. In Epstein, S.D. & Hornstein, N. (eds.), *Working Minimalism*. Cambridge, MA: MIT Press.
- Usher, M., & Cohen, J.D. (1999). Short Term Memory and Selection Processes in a Frontal-Lobe Model. In *Connectionist Models in Cognitive Neuroscience*, London: Springer-Verlag.
- Van Dyke, J. (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., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285-316.
- Van Dyke, J.A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157-166.
- Van Gompel, R.P.G., & Liversedge, S.P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 128-139.
- Vasishth, S. (2006). On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories. Talk presented at the International Conference on Linguistic Evidence, Tübingen, Germany.
- Vasishth, S., Drenhaus, H., Saddy, D., & Lewis, R. (2005). Processing negative polarity. Talk presented at the CUNY Sentence Processing Conference, Arizona, USA.
- van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, 29, 37-70.

- Verhaegen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. *J. Exp. Psychol.: Learn. Mem. Cogn.* 30, 1322-37.
- Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, 34, 186-215.
- Vigliocco, G., Butterworth, B., & Garrett, M. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 61, 261-298.
- Vigliocco, G., Harstuiker, R., Jarema, G., & Kolk, H. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, 11, 407-442.
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production. Is proximity concord syntactic or linear? *Cognition*, 68, 13-29.
- de Vincenzi, M. (1991). *Syntactic Parsing Strategies in Italian*. Dordrecht: Kluwer Academic Publishers.
- Wagers, M.W. (2006). (Re)active dependency formation. College Park, MD: University of Maryland Department of Linguistics. Ms.
- Wagers, M.W., Lau, E., & Phillips, C. (2008; submitted). Agreement attraction in comprehension: representations and processes. Ms.
- Wagers, M.W., & Phillips, C. (2006). Re-active filling. *Proceedings of the 19th Annual CUNY Conference on Human Sentence Processing*. New York: City University of New York Graduate Center.
- Wagers, M.W., & Phillips, C. (2008; submitted). Multiple dependencies and the role of the grammar in real-time comprehension. Ms.
- Wanner, E., & Maratsos, M. (1978). An ATN Approach to Comprehension. In Halle, M., Bresnan, J., & Miller, G.A. (eds.), *Linguistic Theory and Psychological Reality*, Cambridge, MA: MIT Press.
- Weinberg, A. (1992). Parameters in the theory of sentence processing: minimal commitment theory goes east. *Journal of Psycholinguistic Research*, 3, 339-64.
- Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 3(2), 189-202.
- Wickelgren, W. A., Corbett, A. T., & Doshier, B. A. (1980). Priming and retrieval from

short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 387-404.

Woods, W. (1973). An experimental parsing system for transition network grammars. In Rustin, R., (ed.), *Natural Language Processing*, pp 111–154. New York: Algorithmics Press.

Xiang, M., Dillon, B. & Phillips, C. (2008; submitted). Illusory licensing across dependency types: ERP evidence.