Is cue-based memory retrieval 'good-enough'? Agreement, comprehension, and implicit prosody in native and bilingual speakers of English

Elizabeth Pratt
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Part of the Linguistics Commons

Recommended Citation
http://academicworks.cuny.edu/gc_etds/1099

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
IS CUE-BASED MEMORY RETRIEVAL ‘GOOD-ENOUGH’?: AGREEMENT, COMPREHENSION, AND IMPLICIT PROSODY IN NATIVE AND BILINGUAL SPEAKERS OF ENGLISH

by

ELIZABETH PRATT

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor in Philosophy,
The City University of New York.

2015
This manuscript has been read and accepted by the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date
________________
Professor Eva M. Fernández
Chair of Examining Committee

Date
________________
Professor Gita Martohardjono
Executive Officer

Professor Eva M. Fernández
Professor Marcel den Dikken
Professor Andrew Rosenberg
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

IS CUE-BASED MEMORY RETRIEVAL ‘GOOD-ENOUGH’?: AGREEMENT, COMPREHENSION, AND IMPLICIT PROSODY IN NATIVE AND BILINGUAL SPEAKERS OF ENGLISH

by

Elizabeth Pratt

Advisor: Professor Eva M. Fernández

This dissertation focuses on structural and prosodic effects during reading, examining their influence on agreement processing and comprehension in native English (L1) and Spanish-English bilingual (L2) speakers. I consolidate research from three distinct areas of inquiry—cognitive processing models, development of reading fluency, and L1/L2 processing strategies—and outline a cohesive and comprehensive processing model that can be applied to speakers regardless of language profile. This model is characterized by three critical components: a cognitive model of memory retrieval, a processing paradigm that outlines how resources may be deployed online, and the role of factors such as prosody in parsing decisions.

The general framework of this integrated ‘Good-enough Cue’ (GC) model assumes the ‘Good-Enough’ Hypothesis and cue-based memory retrieval as central aspects. The ‘Good-Enough’ Hypothesis states that all speakers have access to two processing routes: a complete syntactic route, and a ‘good enough’ heuristic route (Ferreira, Bailey, & Ferraro, 2002; Ferreira, 2003). In the interest of conserving resources, speakers tend to rely more on heuristics and templates whenever the task allows, and may be required to rely on this fallback route when task demand is high. In the proposed GC model, cue-based memory retrieval (CBMR) is the instantiation of the complete syntactic route for agreement and long-distance dependencies in
particular (Lewis & Vasishth, 2005; Wagers, Lau, & Phillips, 2009; Wagers, 2008). When retrieval fails using CBMR (due to cue overlap, memory trace decay, or some other factor), comprehenders may compensate by applying a ‘good-enough’ processing heuristic, which prioritizes general comprehension over detailed syntactic computation. Prosody (or implicit prosody) may reduce processing load by either facilitating syntactic processing or otherwise assisting memory retrieval, thus reducing reliance on the good-enough fallback route. This investigation explores how text presentation format interacts with these algorithmic versus heuristic processing strategies. Most specifically, measuring whether the presentation format of text affects readers’ comprehension and ability to detect subject-verb agreement errors in simple and complex relative clause constructions.

The experimental design manipulated text presentation to influence implicit prosody, using sentences designed to induce subject-verb agreement attraction errors. Materials included simple and embedded relative clauses with head nouns and verbs that were either matched or mismatched for number. Participants read items in one of three presentation formats: a) whole sentence, b) word-by-word, or b) phrase-by-phrase, and rated each item for grammaticality and responded to a comprehension probe.

Results indicate that while overall comprehension is typically prioritized over grammatical processing (following the ‘Good-Enough’ Hypothesis), the effects of presentation format are differentially influential based on group differences and processing measure. For the L1 participants, facilitating the projection of phrasal prosody (phrase-by-phrase presentation) onto text enhances performance in syntactic and grammatical processing, while disrupting it via a word-by-word presentation decreases comprehension accuracy. For the L2 participants however, phrase-by-phrase presentation is not significantly beneficial for grammatical
processing—even resulting in a decrease in comprehension accuracy. These differences provide insight into the interaction of cognitive task load, processing strategy selection, and the role of implicit prosody in reading fluency, building toward a comprehensive processing model for speakers of varying language profiles and proficiencies.
ACKNOWLEDGEMENTS

First, I would like to thank my dissertation advisor, Eva Fernández, for her tireless support and guidance in my research. From the initial stages of experimental design to data analysis to final edits, she has guided me in such a way that I have gained a tremendous amount of knowledge in the last few years, and even more, developed exponentially as an independent and inquisitive researcher. Her influence has extended beyond this work alone—helping me remember that there is life beyond the dissertation, and giving me advice on how to stay human until I get there. I am a stronger researcher, a stronger writer, and a stronger person because of her guidance.

I would also like to thank my committee members, Marcel den Dikken and Andrew Rosenberg. Marcel has been my faculty advisor from my first year in the program as a Master’s student, and I have benefitted immensely from his incredible depth of knowledge on anything and everything linguistic. Even as my research focus gradually shifted from syntax to psycholinguistics, he has enthusiastically guided me in the syntactic and interface aspects of my projects, and encouraged me to continue strengthening the bridge between these areas. I am also grateful to Andrew, in particular for his methodological advice and insightful questions, all of which led not only to better analyses, but also to my own deeper understanding of the processes involved.

I have had the privilege of working at the Graduate Center in various capacities throughout the course of my graduate studies, which has allowed me to be more closely involved with the program that I could have been otherwise. Gita Martohardjono has always had an open door and open heart – giving the best and wisest advice, both academic and otherwise. I’m not sure I would have made it through, semester after semester, without Nishi Bissoondial who has not only been a guiding resource, but a workmate and a friend. Without the help of Paula
Villacres, the crazed and complicated data collection process may have been my undoing. Her joy and meticulousness made a tedious process smooth and even enjoyable. I’ve also been fortunate to have studied under so many dedicated and distinguished faculty members: Janet Dean Fodor, Marcel den Dikken, Ricardo Otheguy, Gita Martohardjono, Di Bradley, Bill McClure, Christina Tortora, Martin Chodorow, and Elaine Klein. I’ve learned a tremendous amount from each, and am grateful for their input, guidance, and support throughout.

In many ways, a dissertation is a necessarily lonely process, which no one can completely enter into with you. So for my friends—those who stayed by me and those who came alongside, my gratitude goes beyond words. Everyone I met in the trenches of the program who made life bearable—Alex Funk, Alejandro Cuellar, Chris Warnasch, Christen Madsen, David Haase, Edmund O’Neill, Emily Zane, Erin Quirk, Ingrid Heidrick, Jeremy Rafal, Jesse Meyerson, Katie Hawkland, Marisa Nagano, Nazik Dinçtopal, and Teresa O’Neill—it was like the best bits of high school, mixed with later nights, nerdier discussions, and more drinks. My Astoria community group, my “Circle of Panic”, who have loved and lifted me out of many deep pits along the way. Giulia Bencini, for her beautiful spirit and life wisdom, and for long runs and deep talks. My longest and oldest friend, Kelly Crispell Kromidas, who has calmly talked me through a veritable fountain of dissertation and life panics, and distracted me with basement dance parties and cat therapy. Lucia Pozzan, my friend, linguistic muse, and unofficial committee member. There is no way to summarize the years of discussion, support, laughter and kindness—across all long and short distances. Having walked the same path, one step ahead the entire way, she has tamed the future unknown, and assured me that I will be all right. There is no one else to whom I am more indebted. Marisa Nagano, without whom I’m not sure I would have survived the last few years. Not everyone has the benefit of simultaneously writing a dissertation
with a friend, and I couldn’t have asked for a better person to help keep me sane. Priscilla Obeada, for her tireless compassion, who has only known me in school/dissertation mode, and has stayed my friend anyway—through kitchen table tears, Friday night dance parties, and yelling at cats. Vasiliki Liakos, who originally motivated me to go back to school in the first place (stop having a pity party and just apply already!), and who has now waited so patiently for the end. No one has been more understanding during these long years, or shared in the struggle and joy as if it were her own.

Last, but not least, I’d like to thank my family who has supported and come through for me in amazing ways during this process. My mother, for her unconditional love and patience, listening as I neurotically rehashed each day, and recounted my seemingly endless to-do lists. For her understanding of why I couldn’t come home that often, why some days I needed to talk for hours, and other days be a reclusive hermit. For her emails of encouragement that came on difficult days, and for being my role model in compassion and perseverance. And finally, I’d like to thank my father, for his calm and quiet confidence in me, and his reassurance that success lies in simply doing your best. He challenged me to work toward becoming the best version of myself, and although he’s not here to see this day, I hope I have made him proud so far.
TABLE OF CONTENTS

1. INTRODUCTION ........................................................................................................................................ 1

2. AGREEMENT PROCESSING ......................................................................................................................... 4
   2.1 Introduction ................................................................................................................................................. 4
   2.1.1 Fundamentals of agreement: Encoding language representations in real time ................................. 4
   2.2 Agreement processing in production ...................................................................................................... 7
      2.2.1 Factors affecting agreement computation ......................................................................................... 10
         2.2.1.1 Plural markedness ......................................................................................................................... 10
         2.2.1.2 Syntactic/structural constraints ................................................................................................. 10
         2.2.1.3 Semantic factors ......................................................................................................................... 12
         2.2.1.4 Morphophonological factors ...................................................................................................... 13
         2.2.1.5 Summary .................................................................................................................................... 14
   2.3 Agreement processing in comprehension ............................................................................................ 14
      2.3.1 Number mismatch effect .................................................................................................................. 15
   2.4 Theoretical accounts of agreement ....................................................................................................... 17
      2.4.1 Feature percolation .......................................................................................................................... 18
      2.4.2 Spreading Activation/Marking and Morphing .................................................................................. 19
      2.4.3 Production Syntax/Feature Selection & Copy .................................................................................. 20
      2.4.4 Constraint-based accounts ............................................................................................................. 20
      2.4.5 Semantic integration in production ................................................................................................ 21
      2.4.6 Semantic integration in comprehension .......................................................................................... 23
      2.4.7 Attraction and processing ................................................................................................................ 27
      2.4.8 Cue-based memory retrieval ......................................................................................................... 30
   2.5 ACT-R in sentence processing .................................................................................................................. 31
      2.5.1 Cue-based retrieval in production .................................................................................................. 33
2.5.2 Cue-based retrieval in comprehension ................................................. 34
2.5.3 Summary ................................................................................................. 36

3. PROSODY AND PROCESSING ........................................................................ 38

3.1 Prosodic components and hierarchical organization ...................................... 39

3.2 Prosody in comprehension ........................................................................... 41
  3.2.1 Parsing the speech stream ........................................................................ 41
  3.2.2 Cues to information structure and semantic interpretation ......................... 43

3.3 Syntax-prosody interface .............................................................................. 44
  3.3.1 Coordination ............................................................................................ 44
  3.3.2 Garden paths ............................................................................................ 45
  3.3.3 PP-attachment .......................................................................................... 47

3.4 Prosody and working memory ....................................................................... 49

3.5 Prosody and reading ....................................................................................... 50
  3.5.1 Oral reading fluency and development ...................................................... 52
  3.5.2 Prosodic phrasing and reading fluency ...................................................... 56

3.6 Prosody in silent reading ................................................................................. 58
  3.6.1 Implicit Prosody Hypothesis ...................................................................... 58
  3.6.2 Implicit reading and agreement ................................................................. 60
  3.6.3 Summary .................................................................................................... 61

4. L2 PROCESSING ............................................................................................... 63

4.1 L2 syntactic processing .................................................................................. 64
  4.1.1 Variation in ultimate attainment ................................................................. 64
    4.1.1.1 Relative clause attachment ................................................................. 65
    4.1.1.2 Filler gaps ............................................................................................ 66
    4.1.1.3 Syntactic ambiguities .......................................................................... 68
4.1.1.4 Morphology and agreement ................................................................. 70
4.1.1.5 Summary ................................................................................................ 74
4.1.2 Factors contributing to L2 ultimate attainment .......................................... 74
  4.1.2.1 L1 effects ............................................................................................. 75
  4.1.2.2 Proficiency ......................................................................................... 76
  4.1.2.3 Processing limitations ........................................................................ 77
  4.1.2.4 Summary ............................................................................................. 78
4.2 L2 processing models .................................................................................. 79
  4.2.1 Shallow Structure Hypothesis ................................................................. 79
  4.2.2 Declarative-Procedural Model ............................................................... 81
  4.2.3 ‘Good-Enough’ Processing Hypothesis .................................................. 82
  4.2.4 Summary ................................................................................................. 85
4.3 L2 prosody in processing ............................................................................. 86
4.4 L2 reading .................................................................................................. 88
  4.4.1 L1 transfer of reading skills .................................................................... 89
  4.4.2 L2 automaticity in reading ..................................................................... 90
  4.4.3 L2 prosody and reading ........................................................................ 92
  4.4.4 Text segmentation in L2 reading ............................................................ 95
  4.4.5 Summary ................................................................................................. 97
5. EMPIRICAL EVIDENCE FOR A ‘GOOD-ENOUGH CUE’ MODEL .............. 99
  5.1 Proposal and review of predictions ............................................................ 100
  5.2 Development of current experimental design ............................................ 101
  5.3 Methods .................................................................................................... 102
    5.3.1 Participants ........................................................................................... 102
    5.3.2 Materials ............................................................................................... 103
5.3.3 Procedure ........................................................................................................ 104

5.4 Results .................................................................................................................. 107

5.4.1 Analysis overview .......................................................................................... 107
5.4.2 Grammaticality ratings .................................................................................... 108
5.4.3 Comprehension accuracy ................................................................................ 118
5.4.4 Response times (RTs).................................................................................... 121
  5.4.4.1 Grammaticality rating RTs ...................................................................... 121
  5.4.4.2 Comprehension probe RTs ....................................................................... 122
  5.4.4.3 Phrase format: Segment 1 RTs ................................................................. 123
  5.4.4.4 Phrase format: Segment 2 RTs ................................................................. 125
  5.4.4.5 Phrase format: Segment 3 RTs ................................................................. 126
  5.4.4.6 Sentence format: Sentence RTs ............................................................... 126
5.4.5 Proficiency correlations ................................................................................... 127
5.4.6 Tradeoff effects ............................................................................................... 128
5.4.7 Summary of results ......................................................................................... 133
  5.4.7.1 Grammaticality ....................................................................................... 133
  5.4.7.2 Comprehension ....................................................................................... 133
  5.4.7.3 Response times (RTs) ............................................................................. 133
  5.4.7.4 Tradeoff effects ....................................................................................... 134

6. DISCUSSION .............................................................................................................. 135

6.1 Elaboration of the Good-enough Cue (GC) integrated model ......................... 135
  6.1.1 Overview ....................................................................................................... 135
  6.1.2 Cue-based retrieval mechanism ................................................................... 138
  6.1.3 Good-enough processing strategy .................................................................. 142
  6.1.4 The role of prosody ...................................................................................... 143
6.2 Discussion of results .............................................................................................................. 144
  6.2.1 Cue-based memory retrieval in agreement ................................................................. 145
  6.2.2 ‘Good-enough’ processing strategy ............................................................................. 148
  6.2.3 The role of prosody during reading ............................................................................. 150
    6.2.3.1 L2 proficiency and use of prosody ................................................................. 153
  6.3 Conclusions ...................................................................................................................... 155

APPENDIX A – PILOT STUDY .................................................................................................. 160
APPENDIX B – EXPERIMENTAL ITEMS .............................................................................. 168
APPENDIX C – FILLER ITEMS ............................................................................................. 174
REFERENCES ......................................................................................................................... 177
LIST OF TABLES

Table 1. Average proficiency measures for L1 and L2 participants ........................................... 103
Table 2. Sample set of materials ..................................................................................................... 104
Table 3. Overview of presentation paradigms .................................................................................. 105
Table 4. Response time (RT) data exclusions .................................................................................. 108
Table 5. Linear mixed-effects model of overall grammaticality ratings.  
Effects of format, group, grammaticality, and complexity.................................................... 109
Table 6. Linear mixed-effects model of L1 grammaticality ratings.  
Effects of format, grammaticality, complexity, and English reading proficiency.... 111
Table 7. Linear mixed-effects model of L2 grammaticality ratings.  
Effects of format, grammaticality, complexity, and English reading proficiency.... 113
Table 8. Linear mixed-effects model of L2 grammaticality ratings.  
Effects of format, grammaticality, complexity, and Spanish reading proficiency. .. 114
Table 9. Linear mixed-effects model of grammaticality ratings in the SENTENCE condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 117
Table 10. Linear mixed-effects model of grammaticality ratings in the WORD condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 117
Table 11. Linear mixed-effects model of grammaticality ratings in the PHRASE condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 118
Table 12. Generalized linear mixed-effects model of overall comprehension accuracy.  
Effects of format, group, complexity, and target. ................................................................. 119
Table 13. Linear mixed-effects model of overall grammaticality rating response times.  
Effects of format, group, grammaticality, and complexity................................................. 122
Table 14. Linear mixed-effects model of overall comprehension probe response times.  
Effects of format, group, complexity, and target. ................................................................. 123
Table 15. Linear mixed-effects model of Segment 1 reading times in the PHRASE condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 124
Table 16. Linear mixed-effects model of Segment 2 reading times in the PHRASE condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 125
Table 17. Linear mixed-effects model of Segment 3 reading times in the PHRASE condition.  
Effects of group, grammaticality, complexity, and English reading proficiency. .... 126
Table 18. Linear mixed-effects model of reading times in the SENTENCE condition. Effects of group, grammaticality, complexity, and English reading proficiency. 127

Table 19. Correlation table for Woodcock-Johnson/Woodcock-Muñoz passage comprehension and self-rated proficiency (English and Spanish) 128
LIST OF FIGURES

Figure 1. Speech processing models ........................................................................................................ 5

Figure 2. L1 grammaticality ratings by presentation format. ......................................................... 110

Figure 3. L2 grammaticality ratings by presentation format. ......................................................... 112

Figure 4. Grammaticality ratings by group and complexity, SENTENCE condition only. ........ 115

Figure 5. Grammaticality ratings by group and complexity, WORD condition only. .............. 116

Figure 6. Grammaticality ratings by group and complexity, PHRASE condition only. .......... 116

Figure 7. L1 comprehension accuracy by presentation format and complexity. ....................... 120

Figure 8. L2 comprehension accuracy by presentation format and complexity. ....................... 120

Figure 9. L1 comprehension accuracy by presentation format and comprehension target ......... 121

Figure 10. L2 comprehension accuracy by presentation format and comprehension target ....... 121

Figure 11. Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy, Simple experimental items, SENTENCE condition. .... 130

Figure 12. Average comprehension accuracy and grammatical rating accuracy by individual. Simple experimental items, SENTENCE condition......................................................... 130

Figure 13. Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy, Complex experimental items, SENTENCE condition. . 131

Figure 14. Average comprehension accuracy and grammatical rating accuracy by individual. Complex experimental items, SENTENCE condition......................................................... 131

Figure 15. Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy, Filler items, SENTENCE condition............................. 132

Figure 16. Average comprehension accuracy and grammatical rating accuracy by individual. Filler items, SENTENCE condition. ............................................................... 132

Figure 17. Language processing model (adapted from Friederici 2011)........................................ 138
1. Introduction

This dissertation focuses on structural and prosodic effects during reading, examining their influence on agreement processing and comprehension in native English and Spanish-English bilingual speakers. I consolidate research from three distinct areas of inquiry—cognitive processing models, development of reading fluency, and L1/L2 processing strategies—in order to construct a cohesive and comprehensive processing model that can be applied to all speakers of a language, regardless of language background.

Recent work has attempted to integrate syntactic structural features within a cognitive cue-based memory retrieval model (Lewis & Vasishth, 2005; Wagers et al., 2009; Wagers, 2008); however, due to the real-time constraints imposed by communicative needs, the use of such a detailed processing route is not always feasible. There is evidence that when cognitive demand is high, comprehenders often fall back on ‘good enough’ heuristics (Ferreira et al., 2002; Ferreira, 2003) that prioritize thematic templates, pragmatics, and plausibility over full morphological and syntactic processing. This investigation integrates both implicit prosody and cue-based memory retrieval into the ‘Good-Enough’ Hypothesis framework, investigating how text presentation format interacts with these algorithmic versus heuristic processing strategies. Specifically, measuring whether the presentation format of text affects readers’ comprehension, and the ability to detect subject-verb agreement errors in simple and complex relative clause constructions.

The computation of agreement has been extensively examined from both linguistic and psycholinguistics standpoints, and studies have identified numerous features that contribute to the processing of agreement in both production and comprehension. Despite the relatively
uninformative nature of subject-verb agreement in English, it is nonetheless calculated almost automatically, with errors triggering processing difficulty and rapid reanalysis effects (Pearlmutter, Garnsey, & Bock, 1999). Also, even in the case of English, where the agreement system is comparably simple, errors still occur during production, and may go unnoticed during comprehension. The circumstances under which these effects may occur provide compelling insight regarding the features relevant to the computation of agreement, and how it may proceed online. While agreement is often considered to be mainly a syntactic computation, it also integrates information from non-syntactic sources, making it an ideal testing ground for not only grammatical processing measures, but for general processing strategies as well.

The online use of prosody in agreement processing is relatively underexamined. This is particularly the case for agreement processing during silent reading, where there are no direct methods of measuring how prosody may be used, or whether it functions similarly as in oral production and comprehension. The effect of segmented text on comprehension has been explored within the pedagogical and reading subfields, but has not been explicitly linked with implicit prosody effects, or extensively adapted for psycholinguistic research. The experiments described here take a step in this direction, applying text segmentation as a method of tapping into the role of (implicit) prosody during reading and exploring how it interacts with comprehension and grammatical processing—specifically, agreement in complex relative clause constructions.

Agreement phenomena have been less extensively studied in L2 speakers, and there is conflicting evidence supporting both their sensitivity to agreement errors (Tanner, 2011), and insensitivity to errors (Jiang, 2004; Keating, 2009). Models of overall L2 processing also tend to diverge in their predictions of performance, and attributed sources of non-convergence with L1
speakers. This study, which tests agreement within constructions of varying complexity, can shed additional light on the question of whether L2 processing differences are primarily representational or more substantially dependent on performance differences.

This dissertation is laid out as follows. In Chapter 2, I discuss the factors influencing agreement in both production and comprehension, and review several proposals on how the computation of agreement may proceed. I then introduce the cue-based memory retrieval mechanism, proposing it as the instantiation of a complete parsing route, particularly in the case of dependency relations. In Chapter 3, I discuss the role of prosody in parsing, and in particular, how the projection of prosody onto text (implicit prosody) enters into the determination of parsing strategy. In Chapter 4, I describe several differences in L2 ultimate attainment, and relate strategies from the ‘Good-Enough’ Hypothesis to performance variations in both L1 and L2 populations. I also discuss the relationship between reading comprehension and reading fluency, and how the manipulation of text presentation may interact with both of these aspects, regardless of language background. In Chapter 5, I present empirical support for a comprehensive ‘Good-enough Cue’ processing model, in which I integrate the cue-based memory retrieval mechanism into the ‘Good-Enough’ Hypothesis framework, and describe how considerations of implicit prosody, as demonstrated by the effects of text segmentation, interact with deployment of parsing strategy. In Chapter 6, I develop the Good-enough Cue model in further detail, and discuss and interpret the results of the study in light of this paradigm.
2. Agreement Processing

2.1 Introduction

Over the last decades, subject-verb agreement has been widely studied as a window into the mechanisms of real-time sentence processing. Despite the morphological poverty of the agreement system in English, speakers nonetheless produce errors in both oral and written production (Bock & Cutting, 1992; Bock & Miller, 1991), with similar rates of error detection in comprehension (Pearlmutter et al., 1999; Wagers, 2008). It is clear that, regardless of its informativeness to a particular parse, agreement is computed in English, with violations of agreement resulting in processing difficulty (Pearlmutter et al., 1999), and triggering neural responses to subject-verb mismatches (Kaan, 2002). The study of agreement computation is thus an effective measure of online sentence processing, and allows for further inference into the general processing mechanisms of production and comprehension.

2.1.1 Fundamentals of agreement: Encoding language representations in real time

The study of agreement processing is backgrounded by models of linguistic processing in general. One persistent issue lies in acknowledging the complexity of language processing while addressing the cognitive limitations and time constraints of online linguistic encoding and decoding. Many processing models, such as those of Bock and Levelt (1994), and Levelt, Roelofs, and Meyer (1999) (see Figure 1), contain three levels of lexical encoding—conceptual, grammatical (lemmas, morphemes), and phonological (lexemes, forms)—as well as a self-monitoring mechanism. While the models illustrate a largely sequential procedure of lexical production, given the complexity of the production process, the question is then how this
encoding can be handled efficiently under the real time constraints imposed by communicative needs.

**Figure 1.** Speech processing models

a. Bock and Levelt’s lexical network (from Bock & Levelt, 1994, Fig. 2)

b. Levelt et al.’s model of speech production (from Levelt et al., 1999, Fig. 1)

The Minimalist program (Noam Chomsky, 2000, 2001, 2004) takes up the issue of linguistic competence and derivational economy, maintaining that all grammatical representations can be derived by the two functions of MERGE and AGREE. These operations are further constrained by scope (only occurring in response to specific triggers) and domain (only occurring with the current stage or ‘phase’ of the derivation). This approach attempts to reduce the processing burden implied by multiple levels or modules of grammatical operations.

Time constraints and limited cognitive resources suggest a similar limiting of processing operations, leading to proposals that grammatical limits can be perceived as parsing limits. The Garden Path Model (Frazier & Fodor, 1978; Frazier, 1987) utilizes a serial processing approach,
but assumes several principles of derivational economy which limit the drain on processing resources.

According to the Minimal Attachment principle (Frazier & Fodor, 1978; Frazier & Rayner, 1982; Frazier, 1987), the parser assumes the simplest structure possible for the given material. For example, it prefers to attach incoming material as a direct object of the current clause, rather than beginning a new subordinate clause. In a sentence beginning such as in (1), the parser initially analyzes the NP the mayor’s position as the object of argued, thus preferring continuation (1), which matches the projected structure, to continuation (1), which does not, leading to a garden-path effect.

(1) a. The city council argued the mayor’s position…
   b. forcefully.
   c. was incorrect.

The Late Closure principle (Frazier & Fodor, 1978; Frazier & Rayner, 1982; Frazier, 1987) calls for an incoming element to be attached to the current clause, rather than to previously processed structures. Violations of this principle can also result in garden-path effects, as in (2). The parser expects to attach yesterday as the modifier of call, but the conflicting temporal interpretation forces the parser to reanalyze and attach it to the verb said:

(2) Miranda said she will call yesterday.

The garden path model, as with minimalist models, seeks economical parsing based on syntactic structure and principles. Constraint-based theories, on the other hand, utilize a parallel processing model whereby features from multiple streams—such as semantic or frequency-based information—are processed simultaneously and contribute to structure formulation and parsing. On these accounts, a broad range of factors external to the syntax, such as transitivity (Garnsey et
al., 1997; Traxler et al., 2005; Trueswell, Tanenhaus, & Garnsey, 1994), frequency (Trueswell, Tanenhaus, & Kello, 1993; Trueswell, 1996), plausibility (Boland et al., 1995; Garnsey et al., 1997; Traxler et al., 2005), context (Spivey-Knowlton, Trueswell, & Tanenhaus, 1993), priming (Trueswell & Kim, 1998), and prosodic patterns (Snedeker & Yuan, 2008) play in role in parsing decisions.

In addition to non-syntactic factors, a comprehensive parsing theory must also be able to accommodate instances of what Phillips and colleagues term ‘grammatical illusions’ (Phillips, Wagers, & Lau, 2011), where an ungrammatical sentence as in (3) is judged to be grammatical.

(3) *The path to the monuments are littered with bottles.

While comprehenders often demonstrate a striking ability to detect errors in online processing (e.g., island constraints, Principle C violations), in some conditions they are susceptible to illusory grammaticality—judging a construction as grammatical when it is not. Comprehenders are particularly prone to these illusions in case-licensing scenarios (Bader, Meng, & Bayer, 2000), NPI-licensing (Vasishth et al., 2008), comparatives (Townsend & Bever, 2001) and subject-verb agreement (Bock & Miller, 1991). Variation in the relative strength of grammatical constraints suggests that while the representation parse is structurally rich, it is susceptible to feature cue- and memory-based effects when elements must be held in memory and retrieved during processing (Phillips et al., 2011).

2.2 Agreement processing in production

Subject-verb agreement errors as in (4) are well attested, both anecdotally and experimentally:

(4) *The time for fun and games are over. (Kimball & Aissen, 1971)
Because of the tendency in subject-verb agreement errors for the verb to agree with the closest noun, terms such as *proximity concord* (Quirk et al., 1972) and *attraction* (Zandvoort, 1961) have been used to describe the phenomenon. More recent investigations have attempted to identify the mechanisms behind agreement, and define the domain of agreement errors in both production and comprehension.

Early research attributed subject-verb agreement errors to linear proximity or distance between a local noun and the verb. There is evidence for this in that attraction typically does not occur unless an attractor noun intervenes linearly between the head noun and verb (Bock & Cutting, 1992; Bock & Miller, 1991, among others; but see also Wagers, 2008), as well as in disjunction data from Haskell and MacDonald (2005) that suggest linear effects do play a role. While agreement errors have been elicited in nearly all head-local noun number combinations, by far the most robust error patterns occur with singular head and local plural noun constructions (e.g., *the key to the cabinets*). In a meta-analysis by Eberhard, Cutting, and Bock (2005), production of singular-plural agreement errors (e.g., *the key to the cabinets were…*) averaged at about 13%, while those of plural-singular agreement errors (e.g., *the keys to the cabinet was…*) was only 3%.¹ This pattern, termed the ‘mismatch effect’ or ‘plural markedness effect’, is not restricted to agreement in English, but has also been observed in German (Hemforth & Konieczny, 2003), French (Negro et al., 2005), Dutch (Hartsuiker, Antón-Méndez, & van Zee, 2001), and Italian (Vigliocco, Butterworth, & Semenza, 1995), among other languages that display richer agreement morphology than English.

¹ Percentages refer to studies using count nouns; for more detailed data on other noun types and constructions, see Table 3 (p. 546) in Eberhard et al. (2005).
In addition to the plural markedness effect, structural (versus linear) distance effects contribute greatly to the occurrence of agreement errors. In many of the studies described below (e.g., Bock & Cutting, 1992; Solomon & Pearlmutter, 2004), the more hierarchically distant a local (plural) noun is from the head noun, the less likely it is to interfere with agreement marking on the verb. And, as suggested by Franck, Vigliocco, and Nicol (2002), this structural distance effect is strong enough to overpower potential linear distance effects—at least in the case of stacked PP-modifiers. Thus, despite the potential influence of linear distance, plural marking on the attractor and structural distance between the head and local nouns are stronger predictors of errors patterns.

Much of the current experimental research into the mechanism of subject-verb agreement has been inspired by Bock and Miller's (1991) seminal work on agreement attraction. In an elicited production task, Bock and Miller presented participants with an auditory preamble consisting of a subject phrase as in (5). Participants were instructed to repeat back the preamble and complete the sentence.

(5) a. The key to the cabinets…
   b. The keys to the cabinet…

Following their results, Bock and Miller proposed the following characteristics of subject-verb agreement errors: (i) verb number errors are more likely to occur when the head noun and local noun are mismatched for number—almost exclusively when a singular head noun is followed by a plural local noun (plural markedness effect); (ii) errors are more likely in complex subject constructions with a PP modifier than with a relative clause modifier (clause boundary effect); and (iii) conceptual number and animacy have little effect on the number of errors produced.
Bock and Cutting (1992) investigated the effect of modifier type on error rates. They found that local plurals in a PP-modifier were more likely to induce verb errors than those embedded in a relative clause, suggesting that a local noun attractor is more likely to interfere if it is found within the same clause as the head noun than if it is contained within an embedded clause. Following this, they proposed the Clause Bounding Hypothesis, where clause boundaries act as barriers (or at least obstacles) to agreement attraction effects.

This initial set of studies set the stage for subsequent investigations into the types of information that contribute to agreement processing.

2.2.1 Factors affecting agreement computation

2.2.1.1 Plural markedness

Bock and Eberhard (1993) investigated the influence of number representation on error production and concluded that it is not conceptual, phonological, or morphological properties that most influence agreement errors, but rather number representation within the syntax. They suggest that while plural is marked within the syntax, singular is a default feature, and thus unmarked. When the head noun is singular, there is no number feature with which the verb must agree, so a singular verb form is chosen by default. However, when the head noun is plural, the marked feature signals the verb that a plural form must be selected. In the case of a head and local number mismatch, a local noun with a plural marked feature is much more likely to overwrite an unmarked singular feature during verb selection than vice versa.

2.2.1.2 Syntactic/structural constraints

Vigliocco and Nicol (1998) and Nicol (1995) provided evidence that structural distance of the intervening noun from the verb plays a role in error rates; primarily in relation to the timing of
verb planning. Vigliocco and Nicol (1998) investigated agreement errors in the production of declarative sentences and in their inverted interrogative forms:

(6) The helicopter for the flights is safe.
(7) Is the helicopter for the flights safe?

While (6) and (7) differ in linear word order, both have the same hierarchical structure, at least prior to T-to-C movement. The results demonstrated that there was no difference in error rates between the declarative and interrogative forms, and thus, that the processing of agreement takes place during syntactic structuring, not during the later arrangement of words into their final string order.

Further evidence for hierarchical effects in attraction is that agreement errors increase when an attractor is hierarchically closer to the head noun than when it is more distant, even when linear distance is the same in both conditions. Hartsuiker et al. (2001) looked at SOV clauses in Dutch, comparing subject-modifier constructions (8) with direct object constructions (9):

(8) Karin zegt dat het meisje met de krans(en) heeft gewonnen. ‘Karen says that the girl with the garland(s) has won.’
(9) Karin zeg dat het meisje de krans(en) heeft gewonnen. ‘Karen says that the girl has won the garland(s).’

While in both constructions, the potential attractor, krans(en) ‘garland(s)’, is adjacent to the verb, it is hierarchically closer to the head noun in the subject-modifier condition (being part of the NP) than in the direct object condition. Hartsuiker et al. found that although agreement errors occurred in both conditions, they were significantly more likely in the subject-modifier condition (11.5% errors) than in the direct object condition (4.9% errors).
Building on this, Franck, Lassi, Frauenfelder, and Rizzi (2006) investigated the influence of several structural factors on subject-verb agreement in French and Italian. In a series of three experiments, they established a hierarchy of influence on agreement error patterns: while linear precedence alone is sufficient to trigger attraction, a c-command relation increases this likelihood. However, when the subject has moved to a Spec-head position (e.g., OSV order), the verb is more resistant to error than when the subject remains in a postverbal position. This last effect is attributed to the checking of agreement once in an OVS construction via \textsc{agree}, and twice in an OSV construction, via both \textsc{agree}, and local Spec-head agreement.

These findings support not only a hierarchically informed agreement mechanism, but demonstrate the influence of both intermediate and final structural representations.

2.2.1.3 Semantic factors

\textit{Distributivity and notional number}

Within agreement, the term \textit{distributivity} refers to how many ‘tokens’ are indicated by the head noun. For example, in (10), \textit{road} is notionally singular—there is only one road. In (11), \textit{label} is notionally plural—there are multiple instances of the label. Both nouns are grammatically singular but have different number construals.

(10) The road to the mountains…
(11) The label on the bottles…

Bock and Miller (1991) and Bock and Cutting (1992) found that notional number did not affect agreement attraction patterns in English. However, Vigliocco and colleagues (Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Hartsuiker, et al., 1996) demonstrated that this can vary
crosslinguistically depending on morphological richness: the richer the morphology in the language, the greater the sensitivity to distributive number.

*Plausibility*

Thornton & MacDonald (2003) found that agreement errors and processing difficulty increase when both the head and local nouns are plausible agents for the target verb, compared to when only the head noun is a plausible agent.

*Noun animacy and semantic overlap*

Barker, Nicol, & Garrett (2001) found that sentences with animate subjects are less susceptible to agreement errors than those with inanimate subjects. Semantic similarity (including animacy) between the head and local noun also leads to more agreement errors than unrelated noun pairs.

2.2.1.4 *Morphophonological factors*

Bock and Eberhard (1993) demonstrated that agreement processing is not reliant on phonological number features (in English, the plural marking allomorphs */-s/, */-z/, */əz/), but rather on the notional number of the elements. Thus, agreement errors were found with both regular and irregular plurals, (e.g., *boys, men*), but not with nouns that are homophonous with true plurals (e.g., *cruise-crews*).

Hartsuiker, Schriefers, Bock, and Kikstra (2003) found that attraction errors are more likely when the Case of the local noun is phonologically ambiguous (e.g., *die* – nominative or accusative vs. *den* - dative) (German), and when the determiner of the head NP is ambiguous with regard to number versus unambiguous (e.g., *de* – singular or plural vs. *het* – singular in Dutch; *die vs. das* in German)
2.2.1.5 Summary

While the mechanism of agreement is often viewed as primarily syntactic, these studies demonstrate that agreement computation utilizes and integrates information from multiple sources. Semantic, phonological, and pragmatic information\textsuperscript{2} contribute to how agreement and number features are calculated, in addition to syntactic and morphological information. Since agreement phenomena are not confined to syntactic factors, there must be processing constraints which allow efficient computation in real time, and which may shed light on how and why agreement errors occur. If errors are due to processing constraints, the processing of agreement during comprehension may show similar patterns of influences and errors, suggesting that similar mechanisms are deployed in both. I review some of this research below, then summarize several proposed accounts of agreement in both production and comprehension.

2.3 Agreement processing in comprehension

One consideration in studying agreement in English is the frequent lack of overt agreement morphology on the verb. Despite this concern, online measures of effects in comprehension have been successfully implemented in a number of studies. While the research on comprehension is not as extensive as on production, the findings to date indicate that a similar mechanism is active in both.

In the past several decades, agreement attraction in comprehension has been explored using sentence matching tasks (Freedman & Forster, 1985), grammaticality judgments (Clifton, \textsuperscript{2}Psycholinguistic studies on the contribution of certain pragmatic information to agreement processing are not as extensive. However, discourse-related features such as Topic and Focus can be assumed to be visible to syntactic computations such as agreement, particularly since they can be associated with functional projections in the left periphery (i.e., TopP, FocP), which are able to participate in syntactic operations (Aboh, 2010; Cruschina, 2012).
Frazier, & Deevy, 1999; Häussler & Bader, 2009), reading time measures (Nicol, Forster, & Veres, 1997), eye tracking (Pearlmutter et al., 1999), and event-related potentials (Osterhout & Mobley, 1995). An advantage of looking at agreement processing in comprehension is the ability to more freely manipulate the agreement attraction environment, and determine the contributing error factors with greater sensitivity.

2.3.1 **Number mismatch effect**

Building on production findings, Nicol et al. (1997) investigated whether agreement errors in comprehension would occur in similar environments, and whether syntactic distance between a head and local noun would create interference effects in agreement error detection.

Their first experiment used a self-paced reading ‘maze task’ in which, at each word, participants were given a choice of two continuations of the sentence—one grammatical, and one not. The task was designed to ensure that participants were interpreting the sentence online, and thereby pinpoint areas of processing difficulty. Sentences such as in (12), containing a singular or plural head, and a matching or mismatching local noun, were presented, and reading time at the verb was measured.

(12)  
  a. [SS] The author of the speech is here now.  
  b. [SP] The author of the speeches is here now.  
  c. [PP] The authors of the speeches are here now.  
  d. [PS] The authors of the speech are here now.

They found a main effect of NP-match, as well as an interaction between head number and NP-match: while the singular NP-match condition was faster than the singular mismatch condition (SS < SP), there was no reading time difference between the two plural head noun conditions.
To address a potential confound of covert production in this first task, a sentence classification task was performed in which participants read word strings presented on a screen, and were asked to judge whether the words appeared in the correct order. The results again showed a main effect of NP-match, as well as an interaction of head number and NP-match—as in Experiment 1, the singular NP-match condition was faster than its NP-mismatch counterpart, and there was no reading time difference found in the plural head noun conditions.

Nicol et al.’s Experiments 3 and 4 tested the effects of semantic plausibility and notional number on agreement processing, which supported that the results of the first two experiments were, in fact, due to subject-verb agreement processes, and not confounding effects of plausibility or notional number. In Experiment 5, Nicol et al. explored the effect of syntactic distance between the head and local nouns using relative clause constructions as in (13).

(13) High attachment
   a. [SS] The owner of the house who charmed the realtor was no longer willing to sell.
   b. [SP] The owner of the house who charmed the realtors was no longer willing to sell.

Low attachment
   c. [SS] The owner of the house which charmed the realtor was no longer willing to sell.
   d. [SP] The owner of the house which charmed the realtors was no longer willing to sell.

The relative clause either modified the head NP (High Attachment condition: (13a) and (13b)), or the second NP (Low Attachment condition: (13c) and (13d)). Although the linear distance between the head and local nouns was the same for both conditions, the syntactic distance is greater in the Low Attachment condition than in the High Attachment condition. If erroneous feature percolation is responsible for errors in agreement processing (see next section for summary of this view), there would be a greater likelihood of processing difficulty in the High Attachment condition, where the local NP is syntactically closer to the head.
The results showed a significant interaction between NP-match and attachment site. The NP-match condition was significantly faster than the NP-mismatch condition when the attachment site was high, but there was no reading time difference in the Low Attachment condition. Because of this interaction, it was concluded that syntactic distance between a head and local NP does factor into agreement processing.

Nicol et al. found parallels between production errors and comprehension effects (e.g., the mismatch effect) that suggest a similar mechanism is at work in both aspects of agreement processing. Although production and comprehension systems do differ, both require a representation of structure in working memory, and both are susceptible to factors that interfere with the maintenance of that structure. In production, agreement may be checked via a “forward-specifying” system, where the features of the head NP specify the attributes of the upcoming verb. In comprehension, agreement may be checked via a “backward-checking” system, in which an inflected verb would trigger feature checking of the head NP. Both systems would thus be vulnerable to erroneous feature percolation from the local NP to the head NP. Underspecification of the singular feature in both production and comprehension would contribute to this error, and align with results indicating plural asymmetry effects.

2.4 Theoretical accounts of agreement

A number of theoretical accounts have been proposed to accommodate the findings summarized above. I review several of the prominent accounts below, and briefly discuss the strengths and weaknesses of each.
2.4.1 Feature percolation

In the feature percolation account (Franck et al., 2002; Nicol et al., 1997; Vigliocco & Nicol, 1998), features of a noun phrase can percolate upward through the syntactic structure and erroneously value a noun higher in the structure. The greater the structural or hierarchical distance between the attractor and head noun, the lower the likelihood of feature percolation.

The model is able to account for both clause boundary effects and plural markedness effects: the greater the structural distance between the two nouns, the lower the likelihood of feature percolation; and the more marked a feature (such as plural), the greater the likelihood of it overwriting a less marked feature (such as singular).

However, agreement error patterns in comprehension raise a critical concern: the model predicts that percolation effects should appear regardless of the grammaticality of the sentence: ungrammatical sentences as in (14) may appear grammatical, and grammatical sentences as in (14) may appear ungrammatical.

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

Contrary to this prediction, results obtained by Wagers, Lau, and Phillips (2009) indicate that while ungrammatical constructions are susceptible to appearing grammatical, ‘illusions of ungrammaticality’ in sentences that are actually grammatical are exceedingly rare. This strongly suggests that agreement attraction is not due to the erroneous representation of the subject via percolation (Wagers et al., 2009; Wagers, 2008).
2.4.2 Spreading Activation/Marking and Morphing

One of the most enumerated accounts of agreement and attraction effects, the Marking and Morphing model (Eberhard et al., 2005), was originally proposed as an elaboration of the percolation model, and describes the agreement mechanism as two primary procedures, *marking* and *morphing*. Marking maps conceptual number information onto syntactic elements (e.g., the subject NP). Morphing is the structural integration of the elements, during which the verb inherits number from the subject. This step aligns the appropriate morphological information to particular points or nodes in the structure, assigning number features to the constituents and, in the case of complex subjects, combining them appropriately. The verb does not receive its number morphology directly from the semantics or lexicon, but instead via a copying procedure from the subject.

Agreement errors may occur at either the marking or the morphing stage. Conceptual or semantic-based errors may occur at the marking stage where, for example, a subject with a multiple token representation (e.g., *the label on the bottles*), may be erroneously marked as plural, even though the noun form itself is singular. The verb would then inherit this feature and be marked with plural morphology. An error during the morphing stage would be structurally based—the verb may erroneously inherit the features of a local NP, particularly if the local and head NPs are within the same clause.

Based on evidence that conceptual/notional features of the head noun influence agreement while those of the attractor noun do not (Bock & Middleton, 2011; Eberhard, 1999, among others), attraction is said to more plausibly occur during morphing, when number features are integrated within the noun phrase, and the resulting feature is conveyed to the verb. This model has been criticized however by Franck (2011) in its inability to account for finer aspects
of the syntactic structure (such as c-command and precedence distinctions) that modulate the strength of agreement attraction, as well as for not adequately explaining attraction across clause boundaries.

2.4.3 Production Syntax/Feature Selection & Copy

In partial response to the Marking and Morphing model, Franck, Vigliocco, Antón-Méndez, Collina, and Frauenfelder (2008) proposed integrating theoretical syntax with lexical production processes. The authors suggest that agreement is composed of two primary operations—feature selection and feature copy. Feature selection operates within the lexicon, selecting number features according to conceptual input, and even morphophonological influences. Feature copy operates within the syntax as an AGREE function that transmits number features to the verb. Under this model, agreement errors are found to be sensitive to c-command/precedence relations, both during and after the structural derivation process. As a sentence is built, certain configurations—such as precedence of the attractor or lack of Spec-head agreement—may increase the likelihood of a verb copying the features of an attractor rather than the subject. However, this model, while acknowledging semantic and conceptual input during feature selection, does not incorporate these non-syntactic sources into the error model, which occurs wholly within the syntax alone. Additionally, it does not explicitly address the singular-plural asymmetry found in the majority of attraction studies.³

2.4.4 Constraint-based accounts

Constraint-based accounts attribute agreement errors primarily to semantic factors rather than syntactic structure. The general premise, as put forward by Haskell and MacDonald (2005) and

³ Although this could be remedied by adding a markedness effect, where plurals, being marked, are more likely to interfere with agreement than singulars.
Thornton and MacDonald (2003), is that subject-verb agreement is essentially meaning-based. A number feature indicated in the semantics of a construction transmits this information to both the subject and verb in parallel. Thus, agreement errors are not based on faulty transmission of number from subject to verb, but on competing verb forms that have varying levels of support from the available nouns.

Taking this type of model even further are the Maximalist account (Vigliocco & Franck, 1999, 2001; Vigliocco & Hartsuiker, 2002, 2005), and the sign-based framework developed by Reid (1991, 2011). Both accounts emphasize the predominant role of meaning in subject-verb agreement, with the implication that agreement ‘errors’ are the result of a conceptual view rather than a failure to correctly match features. According to Reid, number features are selected independently by the subject and the verb. Thus, there is no direct agreement process between the subject and verb at all, but a semantic feature retrieval process that is responsible for morphological realization of number.

2.4.5 Semantic integration in production

Solomon and Pearlmutter (2004) focus on the conceptual and semantic relationship between nouns, which they refer to as “semantic integration”. They distinguish two potential models: a serial-processing model and a parallel-processing model. In a serial system, production is constrained to the current phrase and proceeds sequentially from one phrase to the next. Any operation that requires maintenance or recall of information from a previous phrase requires additional resources. In a parallel model, activation is not strictly sequential: words and structures can be processed simultaneously, and to varying degrees, may overlap in their activation during the mapping from conceptual structure to syntactic structure. This reduces the
strain on resources to coordinate non-adjacent elements, but then requires more effort to prevent interference from other active elements. Solomon and Pearlmutter predict that in a parallel processing model, elements with a stronger semantic relation are more likely to be planned, and thus activated, simultaneously.

For subject-verb agreement processing, they predicted that constructions containing more related/integrated elements would be more susceptible to error, since the subject noun and local noun would likely be active at the same time and may interfere with each other during grammatical mapping. In the case of a number mismatch between the head and local noun, the incorrect verb form may be chosen. They explored agreement patterns in argument versus adjunct PPs, and embedded clause structures, and performed correlational analyses to compare the semantic integration model to previously described factors contributing to agreement processing.

In contrast with earlier views on clause boundedness, Solomon and Pearlmutter found evidence that clause boundary locations are correlated with error rates only insofar as they reflect the degree of relation between the nouns. For example, plural nouns in an attributive (argument) PP, such as toppings in (15a), are more likely to induce agreement errors than those in an accompaniment (adjunct) PP, such as beverages in (15b). A clause-bounding theory cannot explain these results, since it does not differentiate between PP types.

(15) a. The pizza with the yummy toppings… [attribute PP: high relation]
   b. The pizza with the tasty beverages… [accompaniment PP: lower relation]

In addition, Solomon and Pearlmutter replicated the head-local mismatch effect, where plural local nouns induced more errors than singular local nouns. They also supported Bock and Cutting’s (1992) findings that error effects are smaller in embedded constructions: the greatest
effects were found in PP constructions (16a), then relative clauses (16b), then sentential complements (16c).

(16)  
   a. The report of the nasty auto accidents…
   b. The report that described the traffic accidents…
   c. The report that Megan described the accidents…

However, while error rate differences between relative clauses and sentential complements can be explained by the semantic integration effect within a parallel processing model, it does not predict the difference found between the *of*-PPs and relative clauses, which did not differ in degree of semantic integration.

In a series of correlational meta-analyses, Solomon and Pearlmutter compared the explanatory power of semantic integration with other factors previously described in agreement attraction. The results indicate that semantic integration is related to factors such as plausibility and distributivity, but is better able to handle the data than either of the two. However, it cannot accommodate the robust plural markedness effect, nor the *of*-PP and relative clause discrepancy, indicating that additional factors are required to account for how agreement is computed in production.

### 2.4.6 Semantic integration in comprehension

Looking at comprehension, Pearlmutter et al. (1999) considered two primary proposals for how featural information may be handled: (i) a hierarchical feature-passing system, where the features of heads percolate up through the syntactic tree, allowing the processor to identify and match agreeing elements; or (ii) a linear slot-based model, where the number feature of the subject is

---

4 Degree of semantic integration was determined on the basis of relatedness ratings collected from an additional set of participants.
kept in memory until the verb is reached, when it can then be compared to the verb’s number marking.

In the case of agreement errors in comprehension, in a hierarchical feature-passing model, errors are predicted to arise when features of non-subject NPs percolate too high in the tree and overwrite information in the subject NP. In a linear-slot-based model, errors would occur due to signal decay over time and or intervening distance between the subject and the verb.

Pearlmutter et al. tested complex subject constructions as in (17)—varying NP number match and grammaticality—in a series of experiments involving either word-by-word reading or eye tracking.

(17)  a.  [SS-S] The key to the cabinet was rusty from many years of disuse.
    b.  [SP-S] The key to the cabinets was rusty from many years of disuse.
    c.  [SS-P] *The key to the cabinet were rusty from many years of disuse.
    d.  [SP-P] *The key to the cabinets were rusty from many years of disuse.

Experiment 1 presented stimuli in a non-cumulative word-by-word self-paced moving window paradigm. At the verb, they found an effect of NP-match, where the NP-match conditions (SS-S, SS-P) were faster than the NP-mismatch conditions (SP-S, SP-P). A grammaticality effect, as well as an interaction with the NP-match were found immediately following the verb—grammatical conditions were read more quickly than ungrammatical conditions, and although the NP-match grammatical condition (SS-S) was the fastest, the NP-match ungrammatical condition (SS-P) was the slowest. Interestingly, while overall, grammatical sentences were read more quickly, both grammatical and ungrammatical verbs were read more quickly when they matched the local noun (*SP-P < *SS-P; SS-S < *SP-S).

Thus, at the verb, an NP-mismatch increased processing difficulty in both grammatical and ungrammatical conditions. After the verb, NP-mismatch only increased processing difficulty
for the grammatical conditions, while decreasing it for ungrammatical conditions, suggesting that early agreement processing is influenced by the markedness of local noun number (i.e., plural), after which the head noun number becomes relevant, and grammatical effects are seen.

In Experiment 2, Pearlmutter et al. homed in on the timing of the observed mismatch and grammaticality effects by using the same constructions in an eye-tracking paradigm. In this study, measures were made of first-pass reading time—the initial time spent reading each word, total reading time—first-pass reading time, as well as any time spent re-reading a word, and regressive saccade probability—the probability of eye movements returning to a previous word. First-pass reading time reflects initial or early processing, while both total reading time and saccade probability reflect later processing or integration.

Results indicated no significant difference in first-pass reading time of the verb region in any of the conditions. Total reading time at the verb showed significant main effects of grammaticality and NP-match as well as an interaction, with the longest total reading times occurring in the NP-match ungrammatical condition (*SS-P). While there was no difference in total reading time in the verb region (verb + following word) when the local noun was plural (SP-S = *SP-P), total reading time of the head noun was significantly greater for the NP-mismatch ungrammatical condition than for the NP-mismatch grammatical condition (*SP-P > SP-S). Probability of saccade at the verb did not differ for any conditions except for the NP-match ungrammatical (*SS-P) condition, however the target of the regressive saccades (when they occurred) was most often the local noun, not the head noun.

These results support those of Experiment 1, indicating an effect of grammaticality and of NP number mismatch, while further clarifying the time course and details of agreement processing.
In their final experiment, Pearlmutter et al. returned to a self-paced reading paradigm to further examine the effect of NP number. Using only grammatical sentences, they measured reading times of SS-S, SP-S, PP-P, and PS-P constructions as in (18).

(18)  
   a. [SS-S] The key to the cabinet was rusty from many years of disuse.  
   b. [SP-S] The key to the cabinets was rusty from many years of disuse.  
   c. [PP-P] The keys to the cabinets were rusty from many years of disuse.  
   d. [PS-P] The keys to the cabinet were rusty from many years of disuse.

At the verb, they found a main effect of NP-match, as well as an interaction between head number and NP-match. For singular head noun conditions, NP-match was faster than NP-mismatch (SS-S < SP-S), but there was no difference in the plural head noun conditions. At the word following the verb, there were no main effects of either head number or match, but there was an interaction between the two, and the plural head mismatch condition was faster than the plural match condition (PS-P < PP-P).

Pearlmutter et al. concluded that number mismatches between the nouns increase interference, and thus increase the likelihood of erroneous feature matching/retrieval of the subject NP at the verb. As suggested by the authors, an alternative model would allow number features to attain a level of activation via processing, which allows checking of those features when the verb is reached. However, the interference of intervening elements (particularly competing elements with number features of their own), increases the rate of NP-mismatch effects, as well as grammaticality effects, as the number of the verb does not match that of the retrieved head number feature.
2.4.7 Attraction and processing

Following key points in Solomon and Pearlmutter (2004), Wagers et al. (2009) argue against representation-based theories of attraction such as feature percolation, noting that this model in particular generates several testable predictions that are critical to its viability: (i) strong attraction effects should not occur in constructions where the local noun is not directly dominated by the head, and where it does not intervene between that head and the agreeing verb; (ii) attraction effects should occur regardless of whether the construction is grammatical or not. Wagers et al. examined these predictions via a series of self-paced reading experiments.

In Experiment 1, Wagers et al. established the baseline effects of ungrammaticality and plural processing. Ungrammatical materials were read more slowly than the grammatical materials at the verb region. In addition, reading times were longer immediately following a plural noun, indicating that plurals incur a greater processing cost, regardless of additional complexity factors.

Experiment 2 tested one of the implications of the feature percolation model—that an attractor must intervene hierarchically between the subject and the verb. Wagers et al. used object relative clause constructions as in (19), where the relative clause head does not intervene between the subject driver and the verb see:

(19) The runner(s) who the driver see(s) during the commute every morning always wave(s) to say hi.

While there was no main effect of grammaticality in the relative clause verb region, the region immediately following the verb showed a significant effect of both attractor (head noun) number and grammaticality, as well as an interaction between the two. Ungrammatical materials were read more slowly than grammatical materials, but only when the head noun was singular.
The results align with previous studies in which the attractor does intervene—linearly, hierarchically, or both—between the subject and the verb. The observed effect of reading time slowdown in the absence of some form of structural intervention argues against a feature percolation account of agreement attraction.

However, since all previous studies have found a singular subject-plural attractor effect, but no significant effect with plural subjects, Wagers et al.’s Experiment 3 included a plural subject condition, as in (20).

(20) The runner(s) who the drivers see(s) during the commute every morning always waves to say hi.

The results align with those of previous studies (i.e., attraction was found in the singular subject-plural attractor items, but not in the plural subject-singular attractor items), confirming that the mechanism and processes being measured are in fact the same. Additionally, both Experiments 2 and 3 demonstrate an attraction effect in ungrammatical, but not grammatical sentences (i.e., ungrammatical sentences were potentially perceived as grammatical, but grammatical sentences were not perceived as ungrammatical); this outcome again argues against a feature percolation account, where errors are predicted to occur in both conditions.

While Experiments 1–3 seemed to provide evidence against percolation, it may have been the case that in relative clauses, percolation is delayed until the main verb is reached, so no effect would be seen at the earlier critical region. Also, percolation may proceed relatively slowly, and thus grammatical sentences may be processed too quickly to show an effect.

Experiments 4 and 5 again used a self-paced reading paradigm to examine attraction effects in PP-modifier constructions as in (21). Experiment 4 included an adverb between the attractor noun and verb, while Experiment 5 did not.
(21) The slogan on the poster(s) (unsurprisingly) was/were designed to get attention.

   In both experiments, a significant effect of attractor number was found in both the attractor number region and the verb region. Following the verb, there was a main effect of grammaticality, as well as an interaction between grammaticality and attractor number. In the ungrammatical conditions, plural attractor sentences were read more quickly than singular attractor sentences, but there was no significant difference in the grammatical conditions.

   Finally, in Experiments 6 and 7, Wagers et al. further explored the effect of attraction in grammatical PP-modifier sentences. If attraction effects do occur in grammatical sentences (as predicted by the percolation model), then reading time should be slower in plural attractor conditions than in singular attractor conditions. In addition, attraction should either increase or decrease acceptability ratings.

   In Experiment 6, only the grammatical PP-modifier constructions were presented in a self-paced reading paradigm. Unlike the preceding experiments, which contained both grammatical and ungrammatical sentences, here there was no effect of attractor number on reading time at either the critical verb region, or the following regions.

   Experiment 7 examined the same constructions, this time using a rapid serial visual presentation (RSVP) paradigm, and acceptability judgments were recorded to measure processing difficulty. Results indicated a main effect of grammaticality as well as an interaction of grammaticality and attractor number: participants were more likely to accept an ungrammatical sentence when the attractor was plural, but were not more likely to reject a grammatical sentence under the same conditions.

   This series of experiments by Wagers and colleagues provides evidence against the percolation model of agreement attraction by debunking two of its key predictions: (i) that a
potential attractor must intervene between the subject and verb, and (ii) attraction effects will be seen for both grammatical and ungrammatical constructions.

Wagers et al. argue for an agreement system that is not based on erroneous number representation of the subject (either by percolation or other means), but on a cue-based retrieval mechanism triggered at the verb itself. This type of mechanism—originally developed by Lewis and Vasishth (2005) and McElree (2006)—is triggered at the verb after detection of a subject-verb mismatch, and initiates a search for the subject in memory.

2.4.8 Cue-based memory retrieval

The evidence gathered by Wagers and colleagues (2008, 2009) militates not only against a feature percolation effect, but also against any other accounts based on the misrepresentation of subject number. Wagers et al. (2009) suggest that the semantic integration model, while tapping into essential patterns in subject-verb agreement errors, is not sufficient to account for all of the effects found—particularly hierarchical features found in PP versus relative clause constructions. They argue for a cue-based memory retrieval mechanism that can accommodate both structural and semantic/planning effects. For both production and comprehension, a search process is initiated by the verb to either copy or check the feature of a controller (i.e., its subject). This search is not structurally ordered, but begins with a set of retrieval cues which are then compared to all representations in memory to identify the most appropriate controller candidate.

However, since agreement errors are rare overall, feature overlap cannot be the sole contributing factor in verb agreement computation. Lewis and Vasishth (2005) introduce a formula to capture the additional effects of structure, distance, and memory decay over time (381). In their model, an element held in memory begins to decay immediately. This decay
however, can be offset by reactivation of the element at critical nodes, meaning that the higher an element is in the structure, the greater the strength of its feature set will be, due to reactivation. Hierarchically more dominant categories are more likely to be retrieved, since they have undergone more processing (i.e., have been reactivated at a greater number of critical nodes) and thus have a higher level of activation (Badecker & Lewis, 2007; Lewis & Vasishth, 2005).

Recent literature has strengthened the approach of memory retrieval and interference to agreement and attraction errors. The cue-based retrieval model adopted by Wagers (2008) and Wagers et al. (2009) is based on proposals for both production (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007) and comprehension (Lewis & Vasishth, 2005; McElree, Foraker, & Dyer, 2003), which draw in part on a broader cognitive model, Adaptive Control of Thought-Rational (ACT-R). A brief introduction to the model—as it has been applied to sentence processing—is given below, followed by a closer look at how it handles the mechanism of agreement.

2.5 ACT-R in sentence processing

The ACT-R model (Anderson & Lebiere, 1998; Anderson et al., 2004), as applied to linguistic phenomena, incorporates sentence processing mechanisms with general principles of cognitive theory. It assumes that elements must be maintained in memory and retrieved at a relevant point for integration and interpretation (Abney & Johnson, 1991; Chomsky & Miller, 1963; MacWhinney, 1986; Miller & Chomsky, 1963; Wanner & Maratsos, 1978, among many others), and coordinates theories of working memory with features of linguistic processing. The background assumption is that language is not an isolated processing module, but rather is affected by general constraints on cognition. Working memory limitations observed during
linguistic tasks are artifacts of cognitive functions and limitations, such as activation levels, decay, and rehearsal (Lewis & Vasishth, 2005). A benefit of framing language processing in these terms is the ability to incorporate complex linguistic aspects within a model that is fundamentally concerned with the strict time constraints in all forms of cognitive processing.

The ACT-R model has been successfully applied to account for memory retrieval patterns in non-linguistic domains, and has accurately predicted the time course of various cognitive processes and tasks. Badecker and Lewis (2007) and Lewis & Vasishth (2005) have extended the model into language production and comprehension, respectively.

The ACT-R model also incorporates various independent cognitive modules that are able to interact with one another. A critical feature is that each module has a buffer that can only maintain a limited amount of information at a time. For language processing, the two relevant modules are declarative memory and procedural memory. Declarative memory holds chunks of information, or presentational “feature-value pairs” (Lewis & Vasishth, 2005), that can form associations with other chunks. Procedural memory stores the rules used to evaluate and act upon the information in other buffers. It is also responsible for the development of retrieval cues that will access back to information chunks held in declarative memory.

In sum, there is a two-part structure of memory store and focal attention system; the attentional focus consists of one to three items (McElree et al., 2003; McElree, 2001). As items come in, previous items are moved from attentional focus into the memory store. When an incoming item must reference or integrate with items outside of the current focus, they must be retrieved from the memory store. This retrieval takes place via a cue-based search system, whereby specific features or cues of an item are targeted for search.
2.5.1 Cue-based retrieval in production

In production, the verb looks for features in a controller—this search can be considered either a copying mechanism or a checking mechanism. The verb searches for features that indicate subjecthood (e.g., Case features, or structural features such as Spec,TP, etc.), but is also drawn to more marked features such as plural markers.

In many cases, the subject and verb are adjacent, and no direct retrieval is necessary since both can occupy the focus of attention simultaneously. However, in the case of complex subjects, relative clauses, and other constructions, the verb is constructed well after the head of the subject, so the subject must be retrieved using the system described above. If a sentence contains multiple NPs, there may be competition between them in selecting appropriate features for the verb. While the feature set for the head NP will typically correctly identify it as the appropriate subject, the feature set of another NP may partially overlap, and, under certain conditions, create interference in the retrieval process. The presence of a plural NP will increase the likelihood of its retrieval since the verb is drawn to more marked features. Hierarchically more dominant categories are also more likely to be retrieved, since they have undergone more processing and thus have a higher level of activation (Badecker & Lewis, 2007).

Semantic influences such as notional number can be handled by incorporating additional appropriate cues to the retrieval process. It is likely that retrieval does in fact take multiple levels of cues into account, so such a proposal is not arbitrary.

---

5 A control-based system would call for a copying mechanism, and a constraint-based system would call for a checking mechanism. See Badecker and Lewis (2007) for further discussion.
2.5.2 Cue-based retrieval in comprehension

In comprehension, the agreement mechanism is argued to be similar—a search for matching features initiated at the verb. As in production, the verb uses a set of retrieval cues to identify its correct controller. However, since the incoming verb is already marked, the set of cues includes not only licensing features such as Case or structural position, but other features such as number that will match its own marker (Wagers, 2008).

The model specifically does not operate within a structural or hierarchical framework, however, structural depth effects attested in previous studies of attraction can be accounted for in Lewis and Vasishth’s model. Following McElree, Lewis and Vasishth’s model assumes that items held in storage begin to decay rapidly after initial processing. However, as suggested by Badecker and Kuminiak (2007), processing of structures such as relative clauses leads to some reactivation of the relevant items held in storage. In the case of agreement, retrieval of the correct controller is dependent on two factors: feature/cue match with the controller, and activation level of the controller. These two factors account for both locality effects, as well as structural and clause bounding effects in attraction.

The cue-based memory retrieval system functions in two directions: both top-down and bottom-up. Based on previously processed input, the system predicts what features an agreeing element should have. For example, a subject marked with particular features will prime the system to expect a verb with matching features. As seen in (22), a singular subject may be marked with a set of morphosyntactic features such as number, Case (and others such as gender, if relevant), and perhaps semantic cues such as Animate. While processing, the comprehender reaches the subject, which is marked with features that allow prediction of the upcoming verb form. If the verb matches the prediction, no additional action is needed. If there is a mismatch, a
retrieval mechanism activates, resulting in a slowdown for ungrammatical sentences. If the search finds a partial match in the memory store, the ungrammatical detection effect may be mitigated or even eliminated.

(22) *The athlete who hired the agents run five miles every morning.

Complicating the process is the issue of decay over time. Once items are processed, they begin to rapidly decay in storage, making retrieval more difficult as more time passes between presentation of subject and verb. However, just as in other memory models, rehearsal improves retention and elements can be reactivated at particular nodes in the structure, strengthening their signal. In a relative clause, for example, the head noun may be reactivated at the original and the successive cyclic trace positions, strengthening the signal before decay and making attraction effects less likely than in a PP clause.

This interaction of decay and reactivation, along with the initial feature match strength, forms the basis of this retrieval model for agreement. The activation level is a function of base activation, number of retrievals, and elapsed time since last retrieval. Both syntactic and semantic/pragmatic cues can influence retrieval, although syntactic cues are stronger, and can override semantic/pragmatic cues (Van Dyke & McElree, 2011). There is a disconnect in much of the sentence processing literature, which states that certain structures are more difficult than others, but makes no statement regarding the architecture underlying the processing mechanism.\(^6\) The strength of this model is in its ability to incorporate cues from multiple sources and

---

\(^6\) Notable exceptions are McElree’s speed-accuracy tradeoff work and associative parallel retrieval parsing model, and Van Dyke and Lewis's (2003) theory of decay and interference, which is compatible with Lewis and Vashishth’s model presented here.
processing levels. This allows not only for similarity-based interference on a lexical-featural level, but also frequency effects of lexical items or structural forms, which may account for processing asymmetries not accounted for by activation and decay alone. Using such a mechanism to capture the effects of hierarchy on attraction can be combined with structural position markers and semantic cues, together forming a robust system for encoding structure into the retrieval process.

2.5.3 Summary

In this chapter, I have reviewed the existing literature on agreement processing and attraction, and introduced the cue-based memory retrieval (CBMR) mechanism I believe is best able to accommodate the relevant data within a generalizable cognitive model. As part of my proposal, I assume CBMR as an instantiation of complete syntactic processing, at least within the domain of dependency resolution.

Variations of this cue-based retrieval mechanism have been successfully applied to agreement and memory during processing (see Tanner, 2011; Wagers et al., 2009). Since prosodic phrasing and implicit prosody—the projection of phrasing rhythm onto text during silent reading—have been found to affect processing strategies and parsing preferences, how could prosody be incorporated into the model? The role of prosody may act to enhance memory via two routes: directly on the memory system, as well as indirectly via memory for structure. The prosodic contour may give additional cues to the memory system, or reinforce particular cues as relevant to the parse of a given structure. Disruption of the prosodic projection during processing (including reading) may further impair ungrammatical detection in complex sentences. In the sections below, I introduce some of the relevant research regarding prosody and
processing, and discuss how it bears on the current project, and how it may be integrated into the proposed model.
3. Prosody and Processing

The term *prosody* is a broad term that may collectively refer to a number of suprasegmental features such as pitch, intensity, duration, and the organization of words into groups or phrases. This dissertation will focus on this last aspect, the grouping of words into prosodic phrases. Since much of the available literature suggests that prosodic boundaries—indicated by pauses between phrases—play a considerable role in interpretation, the question arises as to the nature of that role. Is prosody primarily a byproduct of syntactic structure, or does it provide unique information to the parser?

In processing and comprehension, the question of the prosody-syntax relationship has been approached from multiple perspectives, given their close (although not exact) alignment. It has been argued that prosody is used by the listener to align the input into a representation that corresponds with the syntax (Cutler, Dahan, & van Donselaar, 1997). Under an assumption that prosody reflects syntax in some sense, to what degree does it do so, and how do listeners use and interpret prosody during parsing?

A productive area of research has examined the influence of prosody and prosodic boundaries on attachment preferences in ambiguous sentences, some results of which will be summarized below. Of equal interest is the view of prosody as a reflection of broader processing and working memory mechanisms. Such mechanisms may work in tandem with syntactic considerations on the parts of the speaker and the listener.

In production, it has often been claimed that prosodic patterns are based on syntactic and semantic attributes of the utterance (Selkirk, 1984). In studies of monolingual speakers, as well as of bilinguals and L2 learners, prosody has consistently been shown to influence syntactic
interpretations, e.g. ambiguous relative clause interpretations, as well as perform a wide range of other communicative functions (Cutler et al., 1997). Oral prosody influences the parsing strategies of readers, affecting their interpretation of ambiguous constructions such as (23), where *on the balcony* can be interpreted as modifying either *servant* (high attachment) or *actress* (low attachment) (% indicates a prosodic boundary).

(23)  
a. Someone shot the servant % of the actress who was on the balcony.  
b. Someone shot the servant of the actress % who was on the balcony.

While English speakers generally tend to prefer low attachment (Cuetos & Mitchell, 1988), presenting the sentence above with a prosodic phrase break after the first noun *servant* (as in (23a)) biases the interpretation toward low attachment, while placing a prosodic phrase break after the second noun *actress* (23b) biases toward high attachment (Fernández, 2007; Fodor, 1998; Maynell, 1999, 2000). Additionally, studies suggest that prosody facilitates comprehension by not only enhancing structural information, but also by helping to retain information in working memory while the message is processed (Kjelgaard & Speer, 1999; Koriat, Greenberg, & Kreiner, 2002; Kreiner, 2005). This prosodic representation may then provide the initial framework for syntactic and semantic parsing (Schafer, 1997; Speer, Shih, & Slowiaczek, 1989). This view typically assumes a hierarchical intonation structure as described within autosegmental theory, briefly summarized below.

3.1 **Prosodic components and hierarchical organization**

In its most general definition, prosody is the segmentation of elements within a sentence or utterance, and the relative prominence of those segments to each other. Within prosodic theory,
these segments or units can be organized hierarchically as in (24) (Hayes, 1989; Nespor & Vogel, 1986):

(24)

![Diagram of the prosodic hierarchy](image)

The prosodic hierarchy has been formalized by way of the Strict Layering Hypothesis, which states that a unit at each level is exclusively comprised of units from the next level down in the hierarchy (Nespor & Vogel, 1986; Selkirk, 1984, 2003). The purpose of the hierarchy is primarily to define the domain in which certain phonological and prosodic phenomena are observed to take place.

In this framework, an utterance is composed of at least one intonational phrase (IP). An IP is a prosodic unit associated with a unique intonational contour and is classically defined as consisting of at least one phrase-level stressed syllable and a boundary tone (Pierrehumbert & Hirschberg, 1990). IPs often align with syntactic units, but can be determined by semantic and discourse-related factors as well. Parenthetical clauses, nonrestrictive modifiers, and other forms

---

7 Exceptions to the Strict Layering Hypothesis have led to its reformulation as a set of strong, yet violable constraints (see Ito & Mester 1992, or Shattuck-Hufnagel & Turk 1996, for an overview).
are typically contained within their own IP (Nespor & Vogel, 1986). Each IP consists of one or more phonological phrases. Phonological phrases ($\phi P$) are roughly equivalent to syntactic maximal projections, and are characterized by at least one pitch accent, but not necessarily a final boundary tone.

A clitic group ($cg$) contains no more than one content word, and optionally any adjacent monosyllabic clitics, or function words. The prosodic word ($\omega$) refers to a morphosyntactic word, which may be defined as either a content or function word (Hayes, 1989; Nespor & Vogel, 1986), or may be restricted to content words only (Selkirk, 2003). Finally, foot ($F$) refers to a unit containing at most one stressed syllable, followed by any number of weak syllables dominated by the same node (Nespor & Vogel, 1986).

### 3.2 Prosody in comprehension

Prosody plays a role in comprehension at multiple points in the parsing process. Beginning with auditory input, it facilitates segmentation of the speech stream, guides word recognition, and is used to provide and interpret acoustic cues to information structure. It can also disambiguate between syntactic structures, and support memory during parsing. These functions are discussed in more detail below, as well as in the sections following.

#### 3.2.1 Parsing the speech stream

Young children and infants are sensitive to prosodic patterns in their language of exposure, and it has been argued that not only are they able to discriminate varying patterns at both the phrase

---

8 The hierarchical level between the intonational phrase and the prosodic word has also been termed a Major ($MaP$) or Minor Phrase ($MiP$) (Selkirk & Tateishi, 1988), Accentual Phrase ($AP$), or Intermediate Intonational Phrase ($ip$) (Beckman & Pierrehumbert, 1986). For more detailed descriptions, see the original works, or Shattuck-Hufnagel and Turk (1996) for an overview.
and word levels (Bull, Eilers, & Oller, 1984, 1985; Christophe, Guasti, & Nespor, 1997; Christophe, Mehler, & Sebastián-Gallés, 2001) but that they use that ability as part of their acquisition process.

The ‘prosodic bootstrapping hypothesis’ claims that prosody allows infants to segment fluent speech and identify the critical elements of a sentence and/or utterance (Nazzi & Ramus, 2003). Soderstrom, Seidl, Kemler Nelson, and Jusczyk (2003) found that infants as young as 6 months old are able to use prosodic cues to identify and segment phrasal units in connected speech. Other researchers have found that infants as young as 2 months old use prosodic grouping of speech into clauses to organize the input and encode information from the speech signal into memory (Hirsh-Pasek et al., 1987; Mandel, Jusczyk, & Kemler Nelson, 1994).

Prosodic information is also used by adults during auditory word recognition. Grosjean and Gee (1987) argue for increased attention to prosodic structure and features in lexical access models; they cite previous evidence that syllabic saliency in the speech stream guides word recognition (Cutler, 1976; Ladd, 1980), and that pausing patterns during oral reading align with prosodic structures, but not necessarily syntactic ones (Gee & Grosjean, 1983). Slowiaczek (1990) presented words with either a correct or incorrect stress pattern (noTORious/notorIous). She found that response times were faster for correct patterns than for incorrect patterns, and that word/non-word identification was faster for stimuli presented with correct lexical stress. However, response times were still faster for mis-stressed words than non-words. Slowiaczek (1991) additionally found that stress pattern information was often ignored in judging whether a word was acceptable in a given context, as long as it was semantically compatible.

Interestingly however, speakers of Dutch and German are affected by incorrect stress patterns, which more strongly interfere with word recognition (Friedrich, 2003; Koster & Cutler,
1997). This discrepancy may be due to lexical patterns in the languages, where potential
competitors during word recognition are more likely to be eliminated by attending to stress
patterns. Thus, the use of prosodic cues such as lexical stress is more advantageous in languages
with high competitor populations for certain initial syllable groups.

3.2.2 Cues to information structure and semantic interpretation

Selkirk (1984) proposed that IPs are subject to a semantic constraint termed the Sense Unit
Condition. A ‘sense unit’ is comprised of either a single or multiple constituents that, in the
semantic interpretation of the sentence, share a modifier or argument relation. The condition
does not directly dictate which elements must form an IP, but prohibits elements that do not form
a sense unit from occupying the same IP (Selkirk, 1984).

Schafer, Carter, Clifton, and Frazier (1996) make a compelling case for the use of
prosodic information during parsing, in particular, its interaction with information structure, and
the alignment of prosodic features with semantic interpretation. They report results from studies
testing whether pitch accent placement influences relative clause attachment preferences. In the
absence of strong intonational boundary cues, listeners were more likely to attach a relative
clause to the noun that received a pitch accent. Additionally, relative clauses were more likely to
be attached to contrastively accented NPs than to focally accented NPs. This supports arguments
that the information status of incoming elements is inferred online, and that listeners have access
to full prosodic representations during parsing.
3.3 Syntax-prosody interface

Beyond the use of prosodic acoustic features during parsing, the structure of the prosodic hierarchy allows for phrasal and grouping effects. These effects influence interpretation and have been shown to informatively interact with certain syntactic configurations.

3.3.1 Coordination

A basic function of prosody within syntactic disambiguation is coordination, or bracketing within a phrase. A sentence such as (25) is ambiguous between interpretation a) and interpretation b) (Lehiste, 1973):

(25) Old men and women sat on the bench.
   a. [[[Old men] and [women]]] sat on the bench.
   b. [Old [men and women]] sat on the bench.

In speech, the two forms can be disambiguated prosodically: in interpretation (25a), the word men tends to be longer in duration than in (25b), with characteristic F0 differences that allow hearers to differentiate between the two (Katz et al., 1996). Katz et al. (1996) examined this effect in production with both adults and children. Participants were presented with three colored blocks (Pink, Green, and White) that were grouped into one of three configurations as in (26):

(26) a. \[P G W\]
    b. \[P G W\]
    c. \[P G W\]

Participants were asked to describe the arrangement of blocks using the phrase “pink and green and white” in a way that a blindfolded or absent person could determine the grouping. The adults were able to reliably produce word and pause durational cues to indicate groupings, although the children were not. However, in a related study by Beach, Katz, and Skowronski (1996), both
children and adults were able to consistently use prosodic cues to disambiguate similar stimuli in comprehension.

### 3.3.2 Garden paths

Auditory processing studies have examined the role of prosody in both global and local ambiguities, and its ability to signal structural variations to the parser. Studies investigating globally ambiguous sentences indicate that prosody can be used to favor one syntactic interpretation over another (Nicol & Pickering, 1993; Schafer, Speer, & Warren, 2005; Schafer, 1997).

Studies involving locally ambiguous sentences have explored the online effect of prosody as the listener is processing incoming structure. Processing principles, such as Minimal Attachment and Late Closure (Frazier & Rayner, 1982; Frazier, 1987), served as a testing ground for much of the early work on prosody and locally ambiguous constructions, specifically looking at whether prosody could be manipulated to override such processing preferences.

Marslen-Wilson, Tyler, Warren, Grenier, and Lee (1992) used the ambiguity of a verb taking either a direct object or a clause complement to test the effect of prosody on the Minimal Attachment principle. According to this principle, in the absence of other cues, the parser prefers to attach incoming material as a direct object of the current clause, rather than beginning a new subordinate clause. Thus, in a sentence preamble as in (27), the parser initially analyzes the NP as the object of knew, thus preferring continuation (27a), which matches the projected structure, to continuation (27b), which does not match it, leading to a garden-path effect.
(27)  a. The pupils knew several solutions to the problem
    b. …would be quite possible.
    c. …in Physics 100.

Participants were presented the first part of the sentence auditorily, with prosody either
favoring clause complement continuation (sharp F₀ fall on verb and upstep on NP) or a direct
object continuation (continuing F₀ declination across verb and NP). Their reaction time was then
measured during oral reading of a visual probe that displayed the first word of the clause
complement version (would in (27a)). They found that reaction times were faster in the clause
prosody condition than the direct object prosody condition, and concluded that prosody is not
only used by the parser to resolve ambiguities while structure building, but is also able to
override parsing preferences such as Minimal Attachment.

Speer et al. (1996) examined the effect of prosody on the Late Closure principle (Frazier
& Rayner, 1982), which calls for an incoming element to be attached to the current clause. Speer
et al. presented sentences as in (28)⁹ with cooperating, neutral, or conflicting prosody in both a
comprehension task and a cross-modal naming task. In the cooperative condition, the prosody
and syntax matched in early or late cl
osure, and in the conflicting condition, they did not. The
neutral condition had no prosodic boundary in the critical region, and was considered equally
appropriate for either interpretation.

(28)  a. Cooperating, Late Closure: Whenever the guard checks the door % / it’s locked.
    b. Cooperating, Early Closure: Whenever the guard checks % / the door is locked.
    c. Conflicting, Late Closure: Whenever the guard checks % the door / it’s locked.
    d. Conflicting, Early Closure: Whenever the guard checks the door % is locked.
    e. Neutral, Late Closure: Whenever the guard checks the door it’s locked.
    f. Neutral, Early Closure: Whenever the guard checks the door is locked.

⁹ % indicates a prosodic boundary, and / indicates a syntactic boundary.
In the neutral condition of the naming task, there was an advantage for the (28e) interpretation, following the Late Closure principle. In the Cooperating condition, overall reaction times were faster, while in the Conflicting condition, reactions times were slower. Crucially, however, there was no advantage in the Cooperating Late Closure condition, suggesting that prosodic cues do not necessarily further facilitate processing of preferred parses.

These results indicate that the presence of a prosodic boundary at a point of ambiguity can influence the interpretation, and, as in the Marslen-Wilson et al. (1992) study, can override parsing preferences. Speer et al. (1996) propose that a prosodic boundary serves to close the current constituent, allowing the parser to assume a potentially structurally dispreferred interpretation. Kjelgaard and Speer (1999) further explored this interaction between prosody and closure strategies with equivalent results, concluding that the facilitation and interference effects of prosody indicate that listeners use prosodic information online, and at an early stage, during parsing.

3.3.3 PP-attachment

For PP-attachment constructions, Watson and Gibson (2004) proposed the Anti-Attachment Hypothesis, which states that a prosodic boundary after an attachment site increases processing difficulty because it indicates to the parser that there are no upcoming attachments (29a), while a boundary after a non-attachment site facilitates processing (29b).

(29) The artist gave a portrait of the president to the manager on Wednesday.
   a. The artist gave a portrait (1) of the president to the manager on Wednesday.
   b. The artist gave a portrait of the president (2) to the manager on Wednesday.

Pynte and Prieur (1996) attempted to validate these dissociative effects of prosodic boundaries in French using sentences with ambiguous prepositional phrase attachment. They measured
response times to detect a target word in a given utterance, varying the attachment site of the PP (VP vs. NP), and prosodic break position, as in (30).

(30) a. VP-attach, 1 break: The spies inform % the guards of the conspiracy.
    b. NP-attach, 1 break: The spies inform % the guards of the palace.
    c. VP-attach, 2 breaks: The spies inform % the guards % of the conspiracy.
    d. NP-attach, 2 breaks: The spies inform % the guards % of the palace.

According to the garden-path model (Frazier 1978), attaching a PP to the verb is the default preference, but a break inserted after the verb closes off the VP and facilitates attachment to the NP. The presence of an additional break following the NP ((30c) and (30d)) should counteract this effect and reset the VP-attachment preference. In line with this prediction, Pynte and Prieur found that a single break after the VP slowed response times in VP-attachment sentences (in line with an advantage for NP attachment), but the insertion of a break after the NP reestablished the VP-attachment preference.

A subsequent experiment followed the same paradigm, but with a prosodic break before the verb instead of after. While the two-break condition favored VP-attachment as before, they found that, contrary to the predicted dissociative effect, a break before the verb did not facilitate VP attachment, nor inhibit NP attachment.

These results fail to provide support for prosodic breaks as indicators of dissociation, and Pynte and Prieur interpret this as support for Perfetti’s (1990) processing model of an initial stage where low-level constituents are constructed before being mapped onto the argument structure of the verb. Specifically, they suggest that prosodic breaks are not used to close off a constituent, but instead group elements together into such low-level constituents. Following the trajectory of this argument, it may be that this process allows elements to be held in memory until they can be integrated into a larger structure.
3.4 Prosody and working memory

Early research on working memory maintained that items held in memory are stored and processed in specialized components based on information type—the visuospatial sketchpad for visual field items, and the phonological loop for spoken and written material (Baddeley & Hitch, 1974).

Related to the phonological loop is the concept of an auditory buffer, which forms part of the selective listening and memory model put forth by Broadbent (1958, 1971). As elaborated by Frankish (1989, 1995), the auditory buffer privileges memory for last items in a list, even if multiple lists are presented. This suggests that the group of spoken utterances into phrases may take advantage of this feature in the auditory buffer. The temporal phrasing of speech may increase the efficiency of auditory memory, and thus play a significant role in comprehension.

The rapid decay of the elements held in memory can be slowed by rehearsal, such as repeating a telephone number until it can be dialed. Slowiaczek and Clifton (1980) demonstrated that during silent reading, rehearsal takes the form of subvocalization, assisting readers in building a mental representation of the sentences being read. In their experiment, when subvocalization was disrupted by the readers performing a verbal task (e.g., repeating syllable strings silently), comprehension was impaired. It is hypothesized that prosody is a critical component of the subvocalization routine, and thus presents consequences for the storage and processing of material in the phonological loop.

Prosody has been shown to mediate the effect of syntactic and semantic disruption on recall (Stine & Wingfield, 1987), and interestingly, when prosody and syntax conflict, prosody ‘wins’, leading to potential errors in recall (Wingfield, 1975). The effect of prosody on memory is clearly demonstrated in studies showing that synthetic speech produced without prosodic cues
adversely affects memory and comprehension (Paris et al., 2000), and natural prosody in speech and oral reading is more effective in aiding memory than monotone reading (Koriat et al., 2002). Rosner, Grabe, Nicholson, Owen, and Keane (2004) further show that the facilitating effect of prosody is seen most prominently in contexts which are more difficult to process, in particular with long or complex utterances.

Beyond the effect of prosody on syntactic computations, there is a long history of research supporting its effect on memory for speech. Epstein (1961) and subsequent work by O’Connell, Turner, and Onuska (1968), Leonard (1973), and Harriman and Buxton (1979) demonstrated that memory for nonsense syllables was increased by adding morphosyntactic structure, and that the addition of sentence prosody alone could improve memory performance.

While prosody alone may not be a primary information source during parsing, it can provide a level of ancillary support from the input that aids processing. When prosody is disrupted or removed completely, greater strain is placed on memory and memory-based tasks, an effect seen even more robustly in complex and demanding parsing contexts.

Assuming that prosody plays a role in structural memory, it would likely interact with a cue-based memory retrieval mechanism such as in Lewis and Vasishth's (2005) model. The research presented here will shed light on that interaction, contributing data on an often overlooked, yet critical factor in processing and memory.

3.5 Prosody and reading

The complexity involved in how a listener interprets prosody shifts when looking at comprehension and the processing of text. In the relative absence of explicit prosodic cues and
boundaries, does prosody continue to contribute to processing? If so, how does it operate, and what does that contribution look like?

It is suggested that the ability to project appropriate prosody in reading requires that the reader correctly assign syntactic roles to the sentential elements, demonstrating a grasp not only of the structure, but of the general message of the sentence (Chafe, 1988). Studies show a strong correlation between phrasing ability and comprehension. It has been contended that the ability to appropriately phrase textual material into meaningful units is fundamental to fluency in both oral and silent reading. Appropriately chunking text into syntactically and semantically related groups reflects cognitive restructuring of the input and successful encoding into memory. Once this level of reading skill has been achieved, simultaneous improvements in comprehension are often observed (Rayner & Pollatsek, 1989).

Beyond the correlation between prosody, reading fluency, and comprehension, the precise nature of the relationship between prosody and reading comprehension is unclear (see Schwanenflugel et al., 2004), and it is difficult to discuss the relationship between prosody and reading comprehension without reference to overall oral reading fluency. Most experimental work on reading fluency and comprehension has concentrated on the developing reading skills of children. Early studies noted the correlation of fluency with comprehension, also noting that greater reading fluency is associated with more ‘appropriate’ prosodic phrasing and contours (Clay & Imlach, 1971; Dowhower, 1987). Moreover, studies have shown that chunking text in such a way as to preserve major syntactic and prosodic boundaries improves oral fluency (LeVasseur et al., 2006; LeVasseur, Macaruso, & Shankweiler, 2008; Rasinski et al., 1994). Prosody may be seen as an intermediary between fluency and comprehension, such that
individuals who demonstrate appropriate prosody are more likely to exhibit better comprehension as well (Paige et al., 2014).

3.5.1 Oral reading fluency and development

The act of reading is a complex skill that requires the implementation of a number of simultaneous and sequential tasks. The primary purpose of most written texts is to transmit information to readers. To achieve proficiency, a reader must be able to decode the orthographic material into a linguistic representation, extract its meaning, and integrate the textual information with previous knowledge and experience (Miller, 1988).

These processing steps involve the coordination of a number of subtasks, including basic decoding, word recognition and lexical access, and phrasing or “chunking” of material into units of meaning (Kuhn & Stahl, 2003). If these component processes occur sequentially as presumed, it requires that they be implemented both rapidly and accurately for successful interpretation of text.

To accommodate this, at least some of the lower-level processes (e.g., word recognition, lexical access) must be automatized, freeing up cognitive resources to handle higher-level processes such as phrasing and semantic extraction (LaBerge & Samuels, 1974). The motivation for routinization of basic reading processes is primarily the limitations of the working memory system. Unless some aspects of decoding and word recognition are performed in a rapid, attention-free manner, too much of working memory is recruited to analyze the input, and overall meaning extraction is impaired. Additionally, slow decoding results in a reduced reading rate, making it difficult to hold large enough portions of material in memory for accurate comprehension (Perfetti, 1985, 1988).
In early oral language acquisition, learners must build associations between acoustic input and stored linguistic knowledge. As experience with the input increases, this process of association occurs more rapidly and automatically, and fluency may be attained. In early reading acquisition, learners must similarly build associations between orthographic symbols and phonological representations (Hoover & Gough, 1990). However, unlike spoken language, acquisition of reading skill is not automatic, and the end state of reading fluency is not uniform across individuals. In both L1 and L2, the acquisition of reading skill follows a similar pattern; initially, reading is slow and laborious, with decoding proceeding on a sound-by-sound or word-by-word basis. At this level, comprehension is typically low. As proficiency increases, decoding occurs more rapidly, and comprehension of the material increases appreciably (Breznitz, 2006).

Achieving fluency requires not only lexical and grammatical knowledge of a language, but also sufficient processing experience with reading text. The study of reading fluency has expanded over recent years, due in part to a report by the National Reading Panel naming it as a critical factor in reading instruction (National Reading Panel, 2000).

Reading fluency is also correlated with comprehension measures. Fluent readers—as judged by oral reading rates, accuracy, and prosodic patterns—consistently perform better on reading comprehension tests (Fuchs et al., 2001; Fuchs, Fuchs, & Maxwell, 1988; Jenkins et al., 2000; Nathan & Stanovich, 1991). This suggests that at least some aspects of fluency align with the ability to extract meaning.

In the reading fluency literature, there is a consensus on its primary components: (i) accurate decoding, (ii) automatized word recognition, and (iii) appropriate use of prosodic stress and phrasing (Anema, 2008; Carol Chomsky, 1978; Dowhower, 1991; Kuhn & Stahl, 2003; National Reading Panel, 2000; Rasinski, 1990; Samuels, Schermer, & Reinking, 1992;
Schwanenflugel et al., 2004). The end goal of fluent reading—rapid, accurate comprehension of text—depends on the reader’s skill level in each of these components, and their influence corresponds to their development stage in reading acquisition (Wolf & Katzir-Cohen, 2001). Under this view, accurate decoding reflects the goal of the initial stage of reading acquisition. Once accuracy reaches a critical threshold, automatized word recognition can develop. Appropriate prosody and phrasing strategies are often considered the last to emerge, requiring sufficient proficiency in the other components.

Although there is little debate on the components constituting fluency, there are differing views as to which aspects must be automatized to achieve it. Two competing theories concerning this dispute are the Automaticity Theory, which emphasizes the contribution of automatization in basic processing, and the Verbal Efficiency Theory, which emphasizes the contribution of prosodic phrasing.

The Automaticity Theory of LaBerge and Samuels (1974) is premised on limited processing capacity or working memory available for reading. The number of subtasks required in successful reading can overwhelm this capacity, making reading and comprehension effortful. However, if a sufficient number of lower-level tasks can be automatized, more attention can be given to higher-level tasks, increasing both reading speed and comprehension. More specifically, lower-level lexical processes such are word recognition are likely to be automatized in skilled readers. Without this automatization, readers will use a larger portion of their processing capabilities on decoding and word recognition, leaving an insufficient amount left for the higher-level tasks which are linked to comprehension (Adams, 1990; LaBerge & Samuels, 1974; Stanovich, Cunningham, & Feeman, 1984; Stanovich, 1980). Word recognition is a likely
candidate for automatization, given that lexical units are relatively invariant compared to phrasal units, and may become more optimized via exposure (Fukkink, Hulstijn, & Simis, 2005).

Under the Automaticity approach, automatization in these lower-level processes predicts parallel improvements in reading comprehension. However, many studies have found that training and improvements in word recognition do not correlate with improvements in reading comprehension (Dowhower, 1987; Rashotte & Torgesen, 1985; Taguchi, Gorsuch, & Sasamoto, 2006; Young, Bowers, & MacKinnon, 1996).

On the other hand, according to the Verbal Efficiency Theory (VET) (Perfetti, 1985, 1988), automatization of lower-level processes is necessary, but insufficient for attaining a skilled reading level. Of more importance is the automatization of higher-level processes, e.g., phrasing or “chunking”. Accordingly, phrasing or “chunking” ability is a better predictor of fluency and comprehension. The VET is better able to address data from a number of studies investigating the role of automatization in reading. Proponents of VET suggest that, in such studies, lower-level lexical processes may indeed be automatized, but higher-level post-lexical processes are not sufficiently developed. In particular, it is the automatization of post-lexical processes such as phrasing that demonstrates the reader’s grasp of the syntactic-semantic organization of the language, and fundamentally determines reading fluency (Rasinski, 1999). From this perspective, processing efficiency, and thereby fluency, is primarily determined by the ability to parse input into appropriate phrases or meaning units, projecting grammatical knowledge onto a text.

Despite the contribution of appropriate prosody in reading fluency and comprehension, it is, relatively speaking, poorly understood, and many discussions do not explicitly address how it functions online during the reading process. The current study, which manipulates the
presentation of text, may help to determine which of these theories is more plausible. If text presentation taps into projection of prosody during reading, and if, following the VET, automatization of prosodic phrasing is integral to reading fluency, we may expect to see different effects of presentation depending on individual fluency differences. Building toward this, several hypotheses on how prosody may apply during reading are given below.

### 3.5.2 Prosodic phrasing and reading fluency

Much of the research on prosody and reading comprehension focuses on automaticity and the developing skill sets of children and other early readers, which may also be relevant to L2 reading development. However, what is the role of prosody for more advanced native readers? Is prosody simply transference of auditory processing techniques to reading?

The ability to group words into appropriate units is a key aspect of fluent reading (see Allington 2006, among others), and non-fluent readers tend to either read word-by-word or group words differently than typical oral speech patterns (Kuhn & Stahl, 2003). Prosody in oral language can be used as a cue to syntactic and semantic information. Children are particularly attentive to prosody in the auditory input (Schreiber, 1980) and must transition this skill to reading by using punctuation and grammatical cues to appropriately segment text. Even with advanced word decoding skills, fluency is not achieved unless this ability to segment text is developed (Schreiber, 1991).

Pedagogical research has investigated how manipulation of text presentation format may enhance reading skill and support the development of reading fluency. Pre-segmentation of text into meaningful phrasal chunks has been shown to improve reading performance in both children (LeVasseur et al., 2006; O’Shea & Sindelar, 1983) and less skilled adult readers (Cromer, 1970).
Skilled readers seem to be more resistant to imposed text segmentation, suggesting that their own phrasing skills override cues from the input.

In an early study, Cromer (1970) investigated the effect of text segmentation on the comprehension of good and poor reader groups. Sentences were presented in one of four formats: regular sentence (31), single word (32), phrase (33), or fragmented group (34).

(31) The cow jumped over the moon.
(32) The | cow | jumped | over | the | moon.
(33) The cow jumped | over the moon.
(34) The cow | jumped over the | moon.

Assuming that skilled readers chunk text into phrases, while less skilled readers read word-by-word, Cromer predicted that guiding the less skilled groups to read in phrases would improve their comprehension (i.e., make them look like more skilled readers), and guiding the skilled readers to read word-by-word would disrupt their comprehension (i.e., make them look like less skilled readers).

All participants were matched for IQ, but were grouped based on their performance on a comprehension measure and vocabulary score. Participants were first divided based on comprehension score into ‘good’ and ‘poor’ reader groups. The ‘poor’ readers were further divided into a ‘difference’ group, which had low comprehension scores but were matched with the good readers in vocabulary scores, and a ‘deficit’ group, which had low comprehension and vocabulary scores. Cromer found that comprehension patterns varied based on both reading skill level and presentation format. ‘Good’ readers were unaffected by presentation format, and comprehended equally well in all conditions. ‘Difference’ readers were disrupted in the word and
fragment conditions, but improved in the phrase condition. ‘Deficit’ readers comprehended best in the word condition and were not significantly affected by any of the other conditions.

Cromer concluded that while ‘good’ and ‘deficit’ readers may be more or less impervious to text manipulations, either due to the strength of their own phrasing skills (‘good’ readers), or deficit in their reading fluency (‘deficit’ readers), ‘difference’ readers may benefit from textual phrasing as an aid to reading fluency. This suggests that (i) reading comprehension involves additional components beyond general IQ and vocabulary knowledge, and (ii) as long as sufficient vocabulary skills have been acquired, facilitating text presentation may directly influence reading comprehension.

3.6 Prosody in silent reading

Explicit prosody has been shown make use of phrasing (breaks) and intonational cues (pitch accents) to disambiguate between syntactic representations (Schafer et al., 1996). In the absence of explicit prosodic cues, is there evidence that phonological features such as phrasing and intonation are projected, and can similarly influence interpretation? Early evidence from Baddeley and Hitch (1974) and Slowiaczek and Clifton (1980) suggests that if subvocalization is blocked, comprehension during silent reading is impaired; thus, it appears that rehearsal involves projection of phonological/prosodic information, and that this information contributes to the processing and comprehension. More recent inquiries into this domain of ‘implicit’ prosody are reviewed below.

3.6.1 Implicit Prosody Hypothesis

In the years following, there has been additional evidence that fluent readers are not only able to produce prosody during oral reading, but while reading silently as well (Bader, 1998; Fodor,
The experience of hearing an ‘inner voice’ during silent reading has long been anecdotally attested, however, more recent research also supports this experience theoretically and empirically. Because prosody is not consistently or always explicitly indicated in written language, it is often unclear what role it plays in reading and whether that role is critical to reading comprehension. As in production and comprehension of oral prosody, there is some controversy as to whether prosody is a direct reflection of syntactic processes, or whether it makes a unique contribution to syntactic (and other) analyses during parsing.

Koriat et al. (2002) proposed the Structural Precedence Hypothesis, which claims that during reading, readers establish an early structural frame for a phrase or sentence based on function words and morphosyntactic cues that may indicate general phrase structure. Prosody during oral reading reflects this early processing and may be used to help maintain a structure in memory while further processing takes place. While some semantic influences may play an early role, syntactic and prosodic processing precede complete semantic analysis, as evidenced by the ability to project appropriate prosody onto nonsense sentences as long as morphosyntactic cues remain intact (like ‘Jabberwocky’).

Other researchers have suggested that prosody more directly influences both early syntactic and reanalysis processes. Bader (1998) shows that prosody is able to affect the ease of reanalysis during reading of ambiguous sentences: he claims that during reading, both a prosodic and a syntactic structure are produced. If revision of syntactic structure is necessary, it is made more difficult if the prosodic structure must be revised as well (Prosodic Constraint on Reanalysis, Bader 1998: 8). The Implicit Prosody Hypothesis takes this claim further by stating that a default prosodic contour is projected onto text during silent reading, and this projection directly affects interpretations such as ambiguity resolution (Fodor, 1998, 2002).
Related claims have been made for lexical stress variations (Breen & Clifton, 2011), rhythmic stress patterns (Ashby & Clifton, 2005; Ashby & Martin, 2008; Kentner, 2012), and prosodic phrase lengths (Hirose, 2003; Hwang & Schafer, 2009; Hwang & Steinhauser, 2011), suggesting that multiple prosodic features contribute to processing during reading.

3.6.2 Implicit reading and agreement

Disruption of implicit prosody has been show to affect agreement processing during reading, where errors in subject-verb agreement are less likely to be detected if natural reading rhythm is impeded. Kreiner (2005) proposed that the processing difficulty resulting from a subject-verb mismatch reflects online syntactic integration. Natural reading prosody facilitates integration, and allows subject-verb mismatches to be more easily detected. However, if this integration is disrupted in some way, mismatches will be more difficult to detect, and less processing difficulty will be seen.

Testing this hypothesis, Kreiner compared eye movements of participants reading sentences using natural prosody, with those using a fixed grouped prosody that did not align with the phrasal structure of the experimental sentences. Kreiner used three-word groupings—a rhythm which, having been shown to improve recall in list reading (Ryan, 1969), would demonstrate that performance differences would be due to prosody, not general memory effects.

Additionally, to test the prediction that natural prosody helps to maintain a structural representation in memory during integration, Kreiner contrasted an ‘adjacent’ condition (35a), where the subject and verb were immediately adjacent, with a ‘distant’ condition (35b), where the subject and verb were separated by a relative clause.\(^\text{10}\):

---

\(^{10}\) English translations of the Hebrew sample sentences, as provided by Kreiner (2005).
Adjacent match/mismatch:
The audience who have been waiting excitedly to this fashion show watch how the model/s *step* proudly on the stage.

Distant match/mismatch:
The audience watch with a smile how the model/s dressed in a ridiculous hat *step* proudly on the stage.

Kreiner found an overall significant difference between matching and non-matching verb items in the natural prosody condition, but no significant difference in the grouped prosody condition. There was no significant effect of prosody when the subject and verb were adjacent; in the distant condition however, a mismatch effect was found in the natural prosody condition, but not when grouped prosody was applied. These results suggest that natural prosody does facilitate subject-verb agreement processing, particularly when processing and/or working memory load is greater.

Relatedly, it has been found that rapid serial visual presentation (RSVP) of text has the potential to accelerate reading rates, however, it may result in reduced comprehension rates (Bernard, Chaparro, & Russell, 2001; Kang & Muter, 1989). This effect is attributed to the fixed presentation rate of the materials, and it has been suggested that the invariable pace interferes with the projection of prosody onto the text (Castelhano & Muter, 2001; Fernández, 2007).

**3.6.3 Summary**

In this chapter, I have reviewed some of the ways in which prosody interacts with processing and comprehension, with a view to establish its relevance in both syntactic processing and memory functions. Prosody is often omitted from structurally based parsing models, a fact which may be partially due to difficulty in defining which features of prosody may enter into such models, and what the scope of application might be. However, the research overwhelmingly demonstrates
that prosody is used by the parser to clarify syntactic variations, identify clausal and phrasal boundaries, as well as facilitate online memory for structure and memory for recall. This effect extends even to silent reading, where implicit prosody is shown to affect ambiguity resolution and agreement processes, particularly in high processing or memory load contexts.

Given this evidence, it is reasonable to integrate it into the Good-enough Cue model, with the recognition that it is able to explain variations in performance that might otherwise remain unaccounted for in the predominant processing models. Generally speaking, I propose that prosody acts to reduce processing load by either facilitating syntactic processing or otherwise assisting memory retrieval. However, individual factors such as oral fluency or reading fluency may determine the degree to which prosodic features and phrasing are beneficial. Differences such as these are relevant to the current study, which includes participants from different language profiles, which may translate to differences in reading fluency in particular.

In Chapter 4, I summarize some characteristics and models of nonnative processing, and discuss L2 fluency and reading within these contexts.
4. L2 Processing

While a large body of work has been produced on subject-verb agreement errors in L1 speakers, relatively few studies have investigated the same phenomenon in L2 speakers, particularly in comprehension. Nonetheless, there is ongoing debate as to whether L2 speakers and bilinguals utilize the same processing strategies as L1 speakers, or whether there is a fundamental difference between the groups. In general, second language learners are highly variable in their ultimate L2 attainment, particularly if acquisition begins after puberty (Dörnyei & Skehan, 2003; Miyake & Friedman, 1998). They often may not achieve native-like performance in areas such as phonology and morphosyntax, particularly if the relevant features are not shared between the L1 and L2 (Frenck-Mestre et al., 2009; Hawkins & Franceschina, 2004). Since L2 speakers tend to have less difficulty with features of the L2 that are also present in their L1, subject-verb agreement may not necessarily be problematic as long as it is implemented in both languages. Nonetheless, performance differences may still be found when processing speed is critical, or when performing under other processing duress.

In this chapter, I discuss some of the characteristics of nonnative processing, including variation in attainment within syntactic domains, differences based on L1 effects, proficiency, and processing limitations. I then introduce several L2 processing models that address how ultimate attainment of a target language may differ for native and nonnative speakers.
4.1 L2 syntactic processing

4.1.1 Variation in ultimate attainment

Clahsen and Felser (2006) discuss four factors that may contribute to variation in how L2 speakers process a target language. First, nonnative speakers may have greater difficulty integrating information from multiple sources. They seem to better utilize semantic rather than syntactic cues in online processing (Felser et al., 2003), and are less efficient in mapping prosodic cues to semantic structures (Akker & Cutler, 2003). Second, nonnative speakers may process input more slowly than native speakers, as suggested by ERP studies indicating delayed responses to semantic and morphosyntactic violations (Hahne & Friederici, 2001; Hahne, 2001). Third, unlike native learners, L2 learners may transfer their L1 processing strategies to the L2. These strategies may not be optimal for parsing of the target language, and may constrain competence. Finally, maturational effects, or some form of critical period effect may prevent L2 learners from achieving native-like representation and processing of the target language.

Differences in L2 processing have been attributed to L1 transfer (Fernández, 2000), greater reliance on lexical rather than syntactic cues (Papadopoulou & Clahsen, 2003), and processing or working memory resources (Dussias, 2003). In the next section, I discuss studies of L2 processing in several domains that examine what types of strategies are recruited by nonnative speakers during online parsing.

Most recent neurophysiological evidence suggests that, at least for advanced L2 speakers, the neural regions recruited for processing are not qualitatively different—L2 speakers show activation in the same cortical areas as native speakers, although they may show activation in additional areas, suggesting that L2 processing is more effortful and/or less efficient (Abutalebi,
Online behavioral studies can more accurately measure these quantitative differences, and help to characterize the parsing strategies utilized by nonnative speakers. Processing domains in which crosslinguistic differences have been found are frequently used to compare L2 processing with native processing. Below, I review some of this research, looking at relative clause attachment, filler-gap dependencies, structural ambiguities, and morphosyntactic processing.

4.1.1.1 Relative clause attachment

Ambiguous relative clauses have been studied with L2 speakers because of the crosslinguistic differences in attachment preferences. In a sentence such as (36), the relative clause may refer to the first NP niece (high attachment), or the second NP teacher (low attachment).

(36) Andrew had dinner yesterday with the niece of the teacher who belonged to the Communist Party.

Languages such as Spanish (Cuetos & Mitchell, 1988), German (Hemforth, Konieczny, & Scheepers, 2000), Dutch (Brysbaert & Mitchell, 1996), Greek (Papadopoulou & Clahsen, 2003), French (Zagar, Pynte, & Rativeau, 1997) prefer high attachment, while languages such as English (Cuetos & Mitchell 1988), Brazilian Portuguese (Miyamoto, 1999), Swedish, Norwegian, and Romanian (Ehrlich et al., 1999) prefer low attachment. In L1 processing, factors such as input frequency (Cuetos & Mitchell, 1988), implicit prosody (Fodor, 1998), and activation level (Gibson, Schütze, & Salomon, 1996) have been proposed to account for these preferences, but no consensus has been reached as to the primary influences. In L2 research then, the goal has often been to determine whether preferences pattern with those of the L1 or those of the target L2.
Results from L2 attachment studies argue against both positions however, and instead support an L2 processing capacity profile. Dussias (2003) reports that while L1 Spanish-L2 English bilinguals demonstrate target-like preference for low attachment, L1 English-L2 Spanish bilinguals also prefer low attachment in the L2. This suggests a learner bias toward low attachment, irrespective of L1 preference. Papadopoulou and Clahsen (2003) took high-proficiency L2 learners of Greek, whose L1s (Spanish, German, and Russian) were also high attachment languages, and examined their preferences in sentences containing either a genitive (37a) or a preposition (27b) condition:

\[(37) \quad \begin{align*}
\text{a. } & \text{The student of the teacher who…} \\
\text{b. } & \text{The student with the teacher who…}
\end{align*}\]

While both native and nonnative speakers showed low attachment preference in the preposition condition (with), only native speakers consistently preferred high attachment in the genitive (of) condition.

4.1.1.2 **Filler gaps**

It is well documented in the sentence processing literature that maintaining displaced constituents in working memory increases processing load. Because this load increases with the distance between the element and its associated underlying position (Gibson, 1998; King & Just, 1991), the parser attempts to integrate a displaced element as soon as possible in the incoming structure. This preference in the processing of filler-gap dependencies, known as the Active Filler Hypothesis (Frazier & Clifton, 1989), has been found in both native and nonnative speakers (Juffs & Harrington, 1995, 1996; Williams, Möbius, & Kim, 2001). However, these studies did not distinguish between structurally based and thematic-based gap-filling strategies.
Modelling their study on Gibson and Warren's (2004) study of *wh*-dependencies, Marinis, Roberts, Felser, and Clahsen (2005) examined intermediate trace effects in native and nonnative English speakers from several L1s. In a self-paced reading experiment, they tested whether the presence of an intermediate gap (38) facilitated processing at the original gap position.

(38)  

a. The nurse *who* the doctor argued *e*1′ that the rude patient had angered *e*1 is refusing to work late.  
b. The nurse *who* the doctor’s argument about the rude patient had angered *e*1 is refusing to work late.

While both native and nonnative speakers showed integration effects (increased RT) at the original gap position, only native speakers were facilitated by the presence of an intermediate trace—their RTs were faster at the original trace position than when no intermediate trace was present. L2 speakers showed no difference. Marinis et al. interpret these results as evidence the L2 speakers do not use syntactic information in the same way (or to the same extent) as native speakers.

Contributing to the suggestion that L2 speakers use syntactic information differently from native speakers is a cross-modal priming study by Felser and Roberts (2007). Nonnative speakers listened to sentences such as (39), and were visually presented with a picture target to which they had to make a discriminatory judgment response. The probe was either an image of the displaced element (e.g., squirrel), or an unrelated item (e.g., toothbrush), and the probe position was either at the original gap (marked by ‘2’) or at a control position 500 ms earlier (marked by ‘1’).

(39) Fred chased the squirrel to which the nice monkey explained the game’s 1 difficult rules 2 in the class last Wednesday.

While the nonnative speakers’ RTs were facilitated when the picture probe was identical to the antecedent, in contrast to the native speakers, this effect was not greater at the actual gap
position. This suggests that while the nonnative speakers maintained the displaced elements in memory throughout the sentence, they did not necessarily reactivate them at gap positions.

While results from these studies do suggest that L2 speakers use structural information differently during online processing, there may also be processing-based effects at work, since trace-reactivation effects were not found in native speakers with low working memory spans (Roberts et al., 2007).

4.1.1.3 Syntactic ambiguities

Other studies have investigated how L2 speakers resolve syntactic ambiguities during reading. Frenck-Mestre and Pynte (1997) compared the parsing strategies of native and nonnative speakers in sentences containing local syntactic ambiguities and manipulating properties of the verb. In both (40) and (41), the PP is compatible with either VP (a) or NP (b) attachment. However, in (40), the verb is monotransitive, allowing only one argument following it, while in (41), the verb is ditransitive, allowing for two arguments.

(40) a. He rejected the manuscript on purpose because he hated its author.
   b. He rejected the manuscript on horses because he hated its author.

(41) a. They accused the ambassador of espionage but nothing came of it.
   b. They accused the ambassador of Indonesia but nothing came of it.

If an initial analysis is based on general syntactic principles, then the VP attachment versions (a) of both (40) and (41) should be easier to process than the NP attachment versions (b). However, if lexical or subcategorization information is also used during the initial parse, then the (a) version should be easier than (b) in the ditransitive condition, but not necessarily in the monotransitive condition.
Frenck-Mestre and Pynte found that there was no overall preference for VP attachment in either the native or nonnative group, and in fact, both groups appeared to rapidly process thematic information from the verb, which determined their PP-attachment preferences. Nonnative speakers did demonstrate more processing difficulty than native speakers, particularly in the VP-attachment condition.

In their second experiment, Frenck-Mestre and Pynte looked at whether nonnative speakers use thematic information from their L1 when processing the L2. They used verb pairs such as obey-obéir, which are optionally transitive in English, but strictly intransitive in French. While both native and nonnative speakers were affected by syntactic ambiguities, nonnative speakers showed an additional processing slowdown at the point of disambiguation, suggesting that they may have initially considered the thematic constraints of the verb in their L1 before proceeding with a target-like parse.

Juffs (1998) investigated the ability of L2 speakers from several L1s to process reduced relative clauses in English, specifically looking at their sensitivity to the morphology and argument structure of English, and whether their respective L1s influenced how they processed ambiguities. Using a moving window self-paced reading procedure, Juffs presented participants with reduced relative clauses, varying the argument structure of the verbs, and good or bad cues to the appropriate parse. The acceptability and RT data of the L2 speakers patterned with that of the native speakers, suggesting that both groups are similarly sensitive to argument structure differences and cues. However, participants whose L1 was typologically similar to English (Romance languages, as opposed to Chinese-Korean-Japanese) were advantaged in acceptability judgment measures (although not in RT). While there are concerns in conflating results from
multiple L1s, Juffs’ findings do lend support to a processing-based difference between native and nonnative parsing, rather than a qualitative one.

4.1.1.4 Morphology and agreement

Research into L2 morphological processing has also provided rich data on L2 processing profiles. Generally speaking, adult L2 learners are not always sensitive to morphological features during comprehension—e.g., tense, number, and agreement (Johnson et al., 1996; Lardiere, 1998). Difficulties seem to persist despite advanced proficiency, high L2 exposure, and equivalent features in the L1 (Hopp, 2007). Most L2 morphological processing research examines the extent to which L2 speakers are sensitive to morphological features during comprehension, and whether divergence from native speakers is due to representation (competence) or processing (performance) differences.

Foote (2011) investigated whether early and late English learners of Spanish were sensitive to agreement violations in both local (42) and nonlocal (43) dependencies in a self-paced reading study.

(42) Veo que tu padre es/*son de Texas.
    see1sg that your father3sg is3sg/are3pl from Texas
    ‘I see that your father is from Texas.’

(43) El reloj del hombre es/*son de Suiza.
    the watch3sg of the man3sg is3sg/are3pl from Switzerland
    ‘The man’s watch is from Switzerland.’

Results indicated that all groups were sensitive to errors, regardless of the distance between the subject and verb. However, both native and early bilingual groups showed increased RTs in both the target verb region and the following region, while the late bilinguals slowed down only at the region following the verb, potentially indicating a processing delay.
In an eye-tracking study, Keating (2009) investigated Spanish L2 learners’ sensitivity to noun-adjective gender agreement violations within local and nonlocal domains: the DP (44), the main verb clause (45), and subordinate VP clause (46).

(44) *Una casa blanco [vp no se vende rápido en esta zona de la ciudad.] 
A white house does not sell quickly in this part of the city.

(45) *Una casa [vp es bastante pequeño cuando tiene solo una habitación.] 
A house is quite small when it has only one bedroom.

(46) *Una casa [vp se vende bastante bien [cp cuando [vp es nuevo y muy grande.]]] 
A house sells quite well when it’s new and very large.

Keating found that advanced L2 learners were sensitive to violations when the noun and adjective were adjacent, but not when the noun and adjective were separated by intervening material. These results again point to a processing-based deficit rather than a competence deficit, since the learners were able to detect errors in the local condition.

McDonald (2006) further strengthened this argument for processing-based differences in L2 error detection. In her Experiment 1, McDonald compared native and L2 speakers’ performance on three processing measures: a verbal working memory task, a phonological decoding task, and a processing speed task, as well as an auditorily presented grammaticality task, which included sentences with tense, number, and agreement errors. Results indicated that the L2 learners performed significantly worse in all three processing measures as well as the grammaticality judgment task. Interestingly, as measured by grammaticality judgment accuracy and response latency, the L2 learners had less difficulty detecting less complex errors, such as word order, than more complex ones, such as past tense and subject-verb agreement.

In Experiment 2, McDonald tested the effect of processing burden on native speakers. She used the same paradigm as in Experiment 1, with the addition of one of five stressors: low
digit load, high digit load,\textsuperscript{11} white noise overlay, timed response, or compressed speech. While errors in the less complex constructions (word order) were still detected, the processing burden incurred by all but the low digit load stressor significantly reduced error detection the more complex types (e.g., agreement). These patterns align with those of the L2 learners in Experiment 1, supporting the conclusion that general cognitive or processing factors contribute to differences in L2 error detection performance.

Jiang (2004) tested Chinese L1 English learners in their sensitivity to morphological errors in a self-paced reading task. In one experiment, native and L2 speakers read sentences in three error conditions: agreement between a common noun-phrase subject and a finite verb (47), agreement between a pronominal subject and a finite verb (48), and subcategorization schema (49).

(47) The bridges/*bridge to the island were about ten miles away.
(48) I told you I/*she am a professor of psychology.
(49) The teacher encouraged/*insisted the children to mail the letter to the president.

While native speakers were sensitive to errors in all three conditions, the L2 speakers had slower reaction times in the pronom-verb and subcategorization conditions, but not the subject-verb condition.

Jiang frames these results in terms of L2 competence and automaticity, claiming that some grammatical knowledge may not be ‘integrated’, and so cannot be automatically activated during L2 processing. He argues that since the L2 speakers performed almost perfectly in offline grammar tests, but failed to detect errors online, the problem lies in their L2 competence—i.e., a

\textsuperscript{11} For the digit load conditions, participants were required to maintain a set of numbers in memory while performing the task(s).
faulty representation within their grammar. However, these results more parsimoniously align with a processing deficit approach: the L2 speakers demonstrate correct grammatical knowledge, but processing burden reduces the ability to detect errors online. There are additional methodological and analytical concerns with the study, in that all L2 participants were from the same L1 (Chinese), which does not exhibit morphological agreement, and were all considered to be of the same proficiency level, despite large variation in their TOEFL scores and other measures. Despite these concerns, the overall results suggest that the source of error is not representational, but is instead related to processing demand.

Tanner (2011) investigated agreement attraction effects in L2 learners of English to determine whether they demonstrate the same grammatical processing profiles as native speakers. He measured behavioral and ERP responses to constructions containing grammatical or ungrammatical PP subject modifiers (50) and relative clause subject modifiers (51).

(50) The winner of the big trophy/trophies has/have proud parents.
(51) The winner who received the trophy/trophies has/have proud parents.

Behavioral results showed that the L2 learners’ responses were similar to those of native speakers: they were more sensitive to agreement errors when both the head and local noun were singular (match condition), and in the relative clause condition over the PP condition. ERP response profiles aligned with these behavioral data, as well as with the profiles of native speakers, indicating stronger neural responses to agreement violations in the match condition and in the relative clause condition.

The difference between native speakers and L2 learners was found to be largely quantitative rather than qualitative, and Tanner found correlations particularly with L2
proficiency, age of arrival, and motivational factors. While both native and L2 speakers responded with either N400 or P600 responses to agreement violations, the L2 speakers were more likely to respond with an N400 than the native speakers. This effect was modulated by sentence complexity, where increased complexity resulted in increased P600 effect size. Tanner rejects a ‘shallow’ processing explanation for the L2 results, pointing to the native speaker data, which indicate that some native speakers also have an N400 response to agreement violations. I return to this in more detail later; however, this may be additional evidence for an effect of processing load and/or processing strategy on how L2 speakers respond to syntactic errors.

4.1.1.5 Summary

The findings of L2 syntactic processing studies indicate that divergence from native-like attainment lies more in processing constraints rather than representational or competence-based deficits. Learners may have native-like knowledge of the grammar, but have difficulty deploying that knowledge online, particularly when under high processing load. Evidence that even native speakers can look like learners when under sufficient processing burden further validates this view. In the next section, I consider some factors in L2 processing that may account for the increased load, and differences in ultimate attainment.

4.1.2 Factors contributing to L2 ultimate attainment

Most L2 processing studies compare L2 speakers with native speakers, however, factors unique to second language acquisition are critical to framing the L2 processing profile. Below, I review several of these factors, including L1 effects, proficiency, and online processing capacity.
4.1.2.1  **L1 effects**

While mapping of an L2 onto an existing L1 template may occur during the initial stages of language acquisition, it is less clear how the L1 may continue to guide L2 processing strategies and interpretations at high proficiency levels.

L1 transfer theory holds that shared features or mechanisms instantiated in the L1 can be transferred to the L2. This could mean, for example, that it is easier for an English L1 learner of Spanish to process subject-verb agreement, which is a feature of English, than noun-adjective agreement, which is not. Sabourin (2003) and Sabourin, Stowe, and de Haan (2006) support this view, finding in an ERP study of L2 Dutch gender agreement that the morphological similarity of grammatical gender marking in the L1 affects gender acquisition in the L2.

Also, speakers of languages such as Spanish seem to be more sensitive to notional number than speakers of English when processing agreement features. Research suggests that Spanish-dominant bilinguals are more likely to make agreement errors in English when a head noun is conceptually (although not grammatically) plural, while for native English speakers, there is no significant effect of notional number (Nicol, Teller, & Greth, 2001).

However, in terms of online syntactic processing, there is less conclusive evidence of transfer of L1 strategies, at least at high L2 proficiency levels. Marinis et al. (2005) found no evidence of trace reactivation in L2 processing of *wh*-constructions, regardless of whether *wh*-movement occurs in the L1. Papadopoulou and Clahsen (2003) and Felser et al. (2003) similarly found no effect of L1 on attachment preferences in PP-genitives, and Frenck-Mestre et al. (2009) found that even in cases of L1-L2 feature overlap, native-like performance may still not occur if that feature is realized differently in the L2.
Overall, there is evidence that certain domains may be more strongly impacted by transfer of features from the L1, while other processing strategies are commonly shared by L2 learners regardless of L1 background.

4.1.2.2 Proficiency

In addition to L2 effects, individual differences such as proficiency have also been examined in relation to L2 syntactic processing. Hopp (2007) investigated the processing of subject-object ambiguities in L2 German. He found that while near-native L2 speakers used verbal agreement and case to disambiguate garden paths in the same pattern as native speakers, advanced L2 speakers did not. Rossi, Gugler, Friederici, and Hahne (2006) also investigated the effect of proficiency on syntactic processing in L2 German and Italian. ERP responses were measured during auditory processing of sentences containing either a word category violation (e.g., missing noun in a PP) (52), an agreement violation (53), or a combination of the two (54).

(52) *Il signore nel _____ beve un caffè.  
    The man in-the _____ drinks a coffee.  
    (Italian)

(53) *Il signore nel bar bevo un caffè.  
    The man in-the bar drink a coffee.

(54) *Il signore nel _____ bevo un caffè.  
    The man in-the _____ drink a coffee.

The high-proficiency learners showed brain responses similar to those of native speakers: LAN and P600 effects for agreement violations and ELAN, LAN, and P600 effects for word category and combined violations. The low-proficiency learners, on the other hand, showed only a P600 effect for agreement violations, which was lower in amplitude and delayed in comparison to the other groups. For the word category and combined violations, their ELAN and P600 responses were also delayed and lower in amplitude.
These results indicate that L2 proficiency can determine alignment with native-like processing patterns, despite the lack of similar features or processing cues in the L1. Moreover, advanced levels of proficiency appear to reflect more efficient use of cognitive resources during processing.

4.1.2.3 Processing limitations

Interacting with proficiency are cognitive limitations and processing speed, which may also affect L2 performance. Kilborn (1992) suggests that task demand and the resulting limitation in resources is a critical factor in L2 sentence processing.

In a word-monitoring task, Kilborn measured the online integration of syntactic and semantic information by L1 and L2 English speakers in three conditions: (i) a ‘complete’ sentence condition, where both syntactic and semantic information were available, (ii) a ‘syntactic’ condition where syntactic structure was maintained but nouns and verbs were replaced and semantic cues were missing, and (iii) a ‘random’ word string condition where both syntactic and semantic cues were missing. The L2 group performed the task under normal conditions in both their L1 (German) and their L2 (English), while the native English group performed the task in normal and noise conditions.

When performing the task in their L1, the L2 group was faster at detecting the target word in the syntactic condition than in the random condition, and fastest in the complete condition. However, when performing the task in the L2, they were facilitated by the presence of syntactic information (‘syntactic’ condition), but were not further benefited by semantic information in the complete condition.
The native speakers mirrored these patterns in their normal and noise test conditions. In the normal condition, they responded to the target word most quickly in the complete condition, then the syntactic, then the random condition. In the noise condition, native speakers were facilitated by syntactic structure, but were not further facilitated by the addition of semantic information in the complete condition.

Results indicate that task demand affects processing speed and automaticity, and moreover, that task demand is directly responsible for the performance of advanced L2 speakers under normal conditions, and of native speakers in noise conditions.

4.1.2.4 Summary

The study of L2 syntactic processing, including agreement, has been approached from many angles, and although no consensus regarding the L2 processing profile has been reached, it is suggestive of some dominant characteristics. In comparison to L1 speakers, L2 speakers appear to be less sensitive to errors (Jiang, 2004, 2007; Keating, 2009; Sato & Felser, 2006), however this difference is mediated by L2 proficiency (Foote, 2011; Hopp, 2007) and does not necessarily indicate a qualitative difference in how L2 speakers process syntactic information (Tanner, 2011). Recent proposals such as the Shallow Structure Hypothesis (Clahsen & Felser, 2006) and the Declarative-Procedural Model (Ullman, 2004) have incorporated critical period effects into L2 processing models to account for differences from native speakers. While L2 processing may be slower and less automatic than native processing, its similarity to L1 processing under noise conditions suggests that increased burden incurred during L2 processing may be more related to common cognitive load effects. The ‘Good Enough’ Hypothesis (Ferreira et al., 2002; Ferreira & Patson, 2007; Ferreira, 2003) is applicable to both L1 and L2 processing, suggesting that the
mechanism is universal. In the following section, I discuss these models and how they may handle the findings in L2 syntactic processing.

4.2 **L2 processing models**

In this section, I summarize three prominent models of L2 processing, each of which attributes nonconvergence with native speakers to different sources: representation of the grammar (Shallow Structure Hypothesis, Clahsen & Felser, 2006), processing mechanisms (Declarative-Procedural Model, Ullman, 2004), and processing capacity (‘Good Enough’ Hypothesis, Ferreira et al., 2002; Ferreira, 2003).

4.2.1 **Shallow Structure Hypothesis**

Clahsen and Felser's (2006) Shallow Structure Hypothesis claims that L2 processing is qualitatively different from native processing in that L2 speakers rely on lexical and semantic information rather than complete computation of syntactic information. According to the model, native speakers have two processing routes at their disposal—a full parsing route and a shallow processing route. The full parsing route allows computation of complex syntactic phrase structure while the shallow route does not. According to Clahsen and Felser, basic sentence comprehension only requires segmentation of the input and semantic integration of the components. This can often be accomplished using semantic and pragmatic knowledge, without recourse to complex syntactic information. However, a more detailed phrase structure representation must be activated to process purely structural relations, or to make use of elements such as intermediate traces.

While in theory, L2 speakers may have access to both routes, this requires the development of a native-like grammatical representation, which is constrained by critical period
effects in acquisition. Thus, to the degree that L2 learners can attain a native-like grammatical representation, they can also make use of the full processing route. However, for most adult learners, critical period effects do not allow this, and so they are dependent on semantic or plausibility information to make up for their deficient syntactic representations.

The Shallow Structure Hypothesis is able to account for findings in L2 research that L2 speakers have difficulty integrating complex syntactic information, and do not transfer L1 processing strategies to the L2 (Keating, 2009; Marinis et al., 2005; Papadopoulou & Clahsen, 2003; Sato & Felser, 2006). However, the observed differences do not necessarily equate evidence for a qualitatively different representation. Citing the wh-extraction data from Marinis et al. (2005), Clahsen and Felser propose that while native speakers construct a complex representation of the incoming sentence, complete with intermediate traces (55), L2 speakers may process incoming material much more simply, relying on semantic/pragmatic information to assign thematic roles as soon as possible (56). The resulting structure does not project any intermediate traces, which may account for erroneous thematic assignment and other processing errors (Juffs, 2005; Williams et al., 2001).

(55) The nurse who the doctor argued e1 that the rude patient had angered e1 is refusing to work late.

(56) The nurse who the doctor argued that the rude patient had angered is refusing to work late.

Clahsen and Felser acknowledge these performance effects, despite evidence of L2 speakers’ knowledge of subjacency constraints. However, if L2 speakers do possess this grammatical knowledge or utilize an alternate yet equally complex grammatical construal (e.g., pro-binding,
as discussed in Dekydtspotter & Sprouse, 2003), it is difficult to maintain the view that they can never deploy this knowledge during parsing, or that their grammatical representation is necessarily shallow or syntactically simplified.

4.2.2 Declarative-Procedural Model

An alternative to variations in grammatical representation are models that point to differences in how that knowledge is accessed during processing. Models such as the Declarative-Procedural model (Ullman, 2004) draw upon memory systems which are well-established in human and animal models: the declarative memory system, which is responsible for the acquisition and processing of facts and events, and the procedural memory system, which is responsible for the acquisition and control of automatic motor and cognitive skills or habits. Applied to native linguistic processing, the two systems roughly align with the mental lexicon and the mental grammar. The declarative system handles lexical storage—form and meaning, irregular morphological forms, as well as other memorized forms. The procedural system handles sequential or hierarchical linguistic structures, drawing on syntax, regular morphological forms, and other implicit knowledge.

On this view, due to maturational factors, late L2 learners rely more heavily on declarative knowledge, using formulae and memorized forms to compensate for their procedural learning deficits. While this allows them to develop an extensive explicit knowledge of an L2’s vocabulary and grammar, it predicts difficulty in constructions that are not easily memorizable, such as long-distance dependencies. Despite this declarative-procedural imbalance, Ullman claims that with sufficient exposure and proficiency, L2 learners may eventually converge with native speakers, automatizing grammatical processes in procedural memory.
The Declarative-Procedural Model is notable in that it draws upon well-established neurophysiological memory models. However, it is unclear how learners transfer from declarative to procedural knowledge, and why implicit methods of learning can input directly into procedural memory for some learners but not others (Brill-Schuetz & Morgan-Short, 2014; Morgan-Short et al., 2014).

4.2.3 ‘Good-Enough’ Processing Hypothesis

A third type of model, such as the ‘Good-Enough’ Hypothesis, is based on differences in processing capacity and task goals (Ferreira et al., 2002; Ferreira, 2003). The ‘Good-Enough’ Hypothesis is premised on the assumption that most of the time, listeners are not required to perform a complete parse of the input, and so tend to generate representations based on both syntactic information and general heuristics. These heuristics prioritize thematic templates, pragmatics, animacy, and plausibility over morphological and syntactic processing. This strategy may be responsible for misinterpretations based on integrating elements in the input before all relevant structural information is available. As evidence, Christianson and colleagues (Christianson et al., 2001; Christianson, Zacks, & Ferreira, 2006) cite comprehension of garden-path sentences such as (57), where the baby is often first misinterpreted as the object of bathed.

(57) While Anna bathed the baby played in the crib.

They found that while comprehenders correctly responded ‘yes’ to the question Did the baby play in the crib?, they also responded ‘yes’ to Did Anna bath the baby?, indicating that the initial misparse persisted, despite correct reanalysis of baby as the subject rather than the object.

Beyond the effect in garden-path sentences, comprehenders also misinterpret noncanonical sentences if the interpretation clashes with plausibility factors and or an
expectation of agent-verb-patient structure. Ferreira (2003) presented sentences such as (58), asking listeners to indicate the agent of the verb.

(58) The dog was bitten by the man.

A significant proportion of participants responded incorrectly, selecting the dog as the agent. While in general, passives were comprehended less accurately overall than actives, the implausible ‘man bites dog’ passives such as (58) were the most difficult. Ferreira attributes this error pattern to erroneous reliance on an Agent-Verb-Patient thematic template, suggesting that despite the parser’s ability to use syntactic information, it often uses heuristic strategies to minimize processing effort (see also Townsend & Bever, 2001).

Christianson et al. (2006) find that this type of heuristic processing correlates with working memory, further supporting a resource-based motivation. Ferreira and colleagues (Ferreira & Patson, 2007; Ferreira, 2003) argue that the real-time constraints on processing—including time constraints, working memory limitations, or other processing load factors—result in ‘good-enough’ processing patterns. Comprehenders may sacrifice complete syntactic parsing, applying pragmatic or thematic template expectations, to economize these resources.

The ‘Good-Enough’ Hypothesis was not formulated to address differences in L1 and L2 processing. However, I suggest that for L2 processing, the model would predict greater and earlier reliance on heuristics over full parsing. L2 processing is typically slower and more error-prone than native processing, suggesting that it is more cognitively demanding. Whether it is due to working memory limitations, overall proficiency, or reduced automaticity in the L2, the greater processing load would manifest in heavier reliance on plausibility and templates.
Swets, Desmet, Clifton, and Ferreira (2008) investigated whether task goals affect parsing strategies. In a self-paced reading study, participants read ambiguous relative clause sentences, some globally ambiguous (59), and some containing a disambiguating reflexive (60)–(61).

(59) The maid of the princess who scratched herself in public was terribly humiliated.
(60) The son of the princess who scratched himself in public was terribly humiliated.
(61) The son of the princess who scratched herself in public was terribly humiliated.

When the comprehension probes were superficial (e.g., Was anyone humiliated?), they found an ambiguity advantage, where reading times were fastest for the globally ambiguous sentences. This follows previous studies which have suggested that comprehenders do not put effort to resolve ambiguities if the task does not require it, and there is not sufficient information to do so (Traxler, Pickering, & Clifton, 1998). However, reading times significantly increased when comprehension probes specifically targeted the relative clause interpretation (Did the maid/princess/son scratch in public?), and were slower than reading times for the disambiguated sentences. Swets et al. concluded that not only can parses remain incomplete (as for the globally ambiguous sentences), but that the nature of the comprehension task determines parsing strategy and depth.

Lim (2011) compared the predictions of the Shallow Structure Hypothesis and the ‘Good-Enough’ Hypothesis in L2 processing, looking at performance during comprehension and translations tasks. While L2 speakers were more susceptible than native speakers to errors based on plausibility, they were able to compute both syntactic and semantic information in a pattern similar to native speakers, suggesting perhaps quantitative, but not qualitative differences. Also, task goals affected performance: when processing for translation, L2 speakers were more
attentive to structure than when processing for comprehension alone. This argues directly against the predictions of the Shallow Structure Hypothesis, which claims that grammatical representations are either accessible to the L2 speaker or are not.

Crucially though, the ‘Good-Enough’ Hypothesis does not propose a different grammatical representation or architecture for native and L2 speakers. While the Shallow Structure Hypothesis and Declarative-Procedural Model indicate that there are fundamental differences in how language is represented and processed in the L2 (at least after the critical period), the ‘Good-Enough’ Hypothesis allows that, for both groups, the full syntactic algorithm is available along with semantic-pragmatic heuristics, but processing strategy or depth is dependent on the processing goals and task type.

While not directly incorporated into the ‘Good-Enough’ Hypothesis model, it is plausible that working memory and other processing constraints interact with the selection of parsing strategies. The additional demand of processing under noise conditions or within an L2 may promote greater reliance on heuristics in an attempt to conserve resources.

**4.2.4 Summary**

All three models are able to accommodate the evidence that L2 processing is often more effortful than native processing, and that L2 speakers tend to rely more heavily on semantic-pragmatic information during parsing, and do not always perform complete syntactic analysis of the input. However, they diverge in terms of the sources of these differences, and whether convergence with native performance is possible.

The Shallow Structure Hypothesis claims that L2 speakers differ in their grammatical representation of the language, and beyond the critical period, it is difficult for them to develop a
native-like representation that would allow access to the full processing route. The Declarative-Procedural Model postulates two mechanisms for both native and L2 speakers, which are accessible to both groups, but maturational and experiential factors lead L2 speakers’ processing to be more effortful, relying on rote knowledge. This can however improve with exposure and proficiency, and processing can become more automatized and native-like. The ‘Good-Enough’ Hypothesis does not call for attribution of L2 differences to different architectures or mechanisms, but to a greater processing burden which results in heavier reliance on heuristics. As with the Declarative-Procedural Model, as experience and proficiency increase, L2 speakers’ processing profile may converge with that of native speakers.

4.3 L2 prosody in processing

L2 prosody in processing has not been as extensively studied, nor is there a consensus as to how L2 learners may use prosodic information. While many unanswered questions remain in this field, several studies have set the groundwork for work in this area.

Harley, Howard, and Hart (1995) investigated age effects in the use of prosodic cues to sentence structure. Their experimental task was used to determine whether learners would favor prosodic or syntactic cues to structure when prosody and syntax were in conflict. Learners were presented with sentences as in (62) and (63) with either natural prosody (a), or conflicting prosody (b). The conflicting prosody items were created by splicing the first part of (a) sentences with the continuation of the (b) counterpart, and vice versa.

(62)  a. The new teacher’s watch | has stopped.
      b. The new teacher’s | watch has stopped.

(63)  a. The new teachers | watch baseball on TV.
      b. The new teachers watch | baseball on TV.
L1 and L2 participants ranging from 7–23 years were asked to identify the subject noun phrase of each sentence. The oldest native group had the highest accuracy, demonstrating the ability to use syntactic information to override conflicting prosodic cues. The younger native groups and all L2 groups however, were significantly influenced by prosodic cues. The authors attribute these results to the centrality of prosodic information, particularly at earlier stages of language development. They suggest that since prosodic cues are often salient and reliable in the input, overriding those cues requires an advanced skill level not available to younger or less experienced speakers.

Dekydtspotter, Donaldson, Edmonds, Fultz, and Petrush (2008) examined L2 relative clause attachment ambiguities using both aural and written stimuli. Beginning and intermediate L2 French learners were tested on their resolution of ambiguous constructions as in (64).

(64)  a. Nous adorons | le secrétaire | du psychologue qui se promène | (au centre ville).
    b. Nous adorons | le secrétaire du psychologue | qui se promène | (au centre ville).
    ‘We adore the secretary of the psychologist who takes a walk (downtown).’

Experiment 1 was a silent reading task manipulating RC length, where items contained either a long RC (the modifier in parentheses), or a short RC (no modifier). In Experiment 2, identical items were presented both aurally and in written form. Both RC length and intonation were manipulated such that the RC either formed a constituent with the NP2 (64a), or the NP1 and NP2 formed a constituent and the RC its own constituent (64b). Experiment 3 was a self-paced reading task where contextually disambiguating information was provided for each of the items.

This set of experiments was designed to test the assumptions of the Shallow Structure Hypothesis, where L2 learners are presumably limited to use of non-structural information during parsing. If this is true, then learners would be unable to revise an initial parse based on
prosodic cues (Experiment 2). Further, if it is the case that learners use contextual information rather than syntactic strategies such as minimal attachment, then the contexts provided in Experiment 3 should allow direct attachment of the RC to the appropriate noun, without evidence of a reanalysis effect.

The results for Experiment 1 indicated that beginner learners were not sensitive to RC length effects, although intermediate learners were. In Experiment 2, overall, the learners did not show an effect of intonation; however, a subset of intermediate learners did consistently make use of the prosodic cues. And finally in Experiment 3, all learner groups demonstrated a minimal attachment bias, where despite the disambiguating contexts, response times were shorter when the RC modified NP2 than when it modified NP1.

Overall, results demonstrated not only that learners can make use of prosodic cues, but that, contrary to the Shallow Structure Hypothesis, they are also capable of deploying and integrating syntactic and prosodic strategies during parsing.

4.4 L2 reading

Many L2 speakers read less rapidly and accurately in their second language than in their first, despite advanced oral fluency in the L2 (Favreau & Segalowitz, 1983). Even fluent L2 speakers often do not attain an equivalent level of reading fluency, despite success in L1 reading skill. This discrepancy has been attributed to factors such as limited experience with processing L2 text and underdeveloped L2 grammar or vocabulary (Koda, 1996). These are reasonable assumptions; however, other major differences in the experience and process must be taken into account.
First, L2 reading acquisition typically occurs simultaneously with L2 oral language acquisition, in contrast with L1, where many aspects of oral language have been mastered prior to reading instruction. Second, L2 readers are more likely to learn to read in a much wider range of settings, and for different purposes. Third, they often have already acquired reading proficiency in their L1 (Koda, 1996).

These major differences suggest that, despite an outward correspondence with L1 reading acquisition, L2 reading skill is often influenced by unique factors such as input type and quality, and transfer of processing skill from the L1. Nonetheless, it is assumed that the actual cognitive processes involved in reading remain essentially the same, regardless of L1 or L2 input text (Day & Bamford, 1998). Thus, attaining fluency in either setting requires the same coordination of word recognition, lexical access, and phrasing strategies.

As mentioned above, one critical difference in the portrait of L2 readers is that they have often already achieved reading fluency in their L1. Given that the cognitive processes leading to automaticity are not language-specific, it seems plausible that some skills already in place for L1 reading may be transferrable to the L2.

4.4.1 L1 transfer of reading skills

The Linguistic Interdependence Hypothesis (Cummins, 1979, 1991) claims that L2 reading skill is critically dependent on the transfer of established reading skills in L1. Thus, L2 reading skill, or fluency, is dependent on (i) whether skills have been sufficiently developed in the L1, and (ii) whether they have been successfully transferred to the L2.

The Linguistic Threshold Hypothesis (Clark, 1979) elaborates on the interdependence theory, claiming that a critical level of L2 linguistic knowledge must be attained before transfer
of L1 skills can take place. Hence, the main point of contention between the two theories is whether L1 reading skill or L2 proficiency is the more decisive factor in successful L2 reading and comprehension. Several L2 reading studies designed to investigate this relationship give evidence that, in fact, both factors do play a role in L2 reading ability (Bernhardt & Kamil, 1995; Carrell, 1991).

Carrell (1991) examined the reading comprehension of bilingual Spanish-English students, and native English students of Spanish in both languages in relation to measures of L1 oral reading ability, and proficiency level in the L2.

In the bilingual Spanish-English group, L2 reading ability was more strongly correlated to L1 reading ability than L2 proficiency. However, for the native English Spanish learners, L2 reading ability was more strongly correlated to L2 proficiency. The data suggest that L2 exposure may relate to L2 reading skill: those with high exposure to the L2 show a greater effect of L1 ability, and those with lower L2 exposure show a greater effect of L2 proficiency.

The results support the Linguistic Threshold Hypothesis, in that the high exposure group (bilinguals) may have reached a critical level of L2 linguistic knowledge, allowing for transfer of L1 reading skills. On the other hand, the low exposure group (learners) may not have reached a sufficient level of linguistic knowledge, making their L2 reading skill more reliant on their level of proficiency.

4.4.2 L2 automaticity in reading

When L2 reading is sufficiently developed, automaticity of certain processes is crucial to subsequent improvements in reading ability and comprehension (Favreau & Segalowitz, 1983). As in L1 reading, the basic subtasks of decoding and word recognition must be made efficient
enough to free up cognitive resources for higher-level processes linked to comprehension. This level of efficiency is not simply a more rapid implementation of the subtasks, but a reorganization of processing routines that does not require conscious effort (Fukkink et al., 2005).

The initial stages of L2 reading acquisition are similar to those of L1: slow, effortful decoding typically gives way to more rapid processing as experience and exposure to text increase. Additionally, despite differences in the skill sets available to L2 readers, cognitive adaptations similar to those in L1 reading must take place for reading fluency to emerge. However, do the observed patterns for L1 automaticity and comprehension also hold for L2 reading? Given that L2 readers have often already attained reading fluency in their L1, is automatization of lower-level processes in the L2 sufficient to improve reading comprehension? Some may assume that if higher-level, non-language-specific skills are transferrable from the L1, automatization of the lower-level language-specific processes may be sufficient to observe significant improvement in comprehension.

A study investigating this question examined the effect of word recognition training on comprehension in Dutch intermediate-level English learners (Fukkink et al., 2005). The researchers found that the training improved speed in list-word recognition, but had no significant effect on comprehension. Although all participants displayed fluency in L1 oral reading, transfer of the higher-order skills did not obtain, and training on lower-level skills did not improve comprehension.

From these data, it would seem that, just as in L1 reading, automatization of lower-level lexical processes does not directly influence comprehension. Although not definitive, the study provides additional evidence in support of the Verbal Efficiency Theory (Perfetti, 1985, 1988),
that it is the automatization of higher-level post-lexical processes rather than of word recognition and decoding that most directly influences comprehension.

Grabe and Stoller (2002), among others, emphasize the importance of post-lexical automatization—specifically of the ability to appropriately parse text into meaningful phrase units. This suggestion matches well with observations of L2 prosody and fluency in general.

It is well supported that L2 oral fluency is indicated by attainment of native-like prosodic patterns (Cucchiarini, Strik, & Boves, 2000a, 2000b; Missaglia, 2007). It would appear that a similar relationship holds for oral reading and prosodic phrasing; however, the path to achieving fluency in reading makes the process more complex.

4.4.3 L2 prosody and reading

In terms of major prosodic features, languages perhaps most notably differ in intonation and lexical stress patterns (see discussion in Cutler 2012). Languages may also differ in typical or acceptable prosodic phrase lengths, and may use prosody differently to indicate variations in information structure. For example, as noted by Holmes (1995), French speakers have been shown to use different syntactic strategies than English speakers to indicate topic and focus elements in spontaneous discourse. This use of distinct syntactic arrangements, such as clefting and dislocation, is, in turn, reflected in the prosodic pattern of the output in each language.

However, prosodic phrasing patterns appear to be more universal, at least when those patterns align with syntactic constituents. In several notable studies, even with no previous exposure to the test language, listeners were able to correctly identify pauses occurring at constituent boundaries (Endress & Hauser, 2010; Pilon, 1981; Wakefield, Doughtie, & Yom, 1974). In fact, L2 speakers may rely more heavily on prosodic rather than syntactic cues,
particularly when syntax and prosody are misaligned (Harley et al., 1995). Evidence would suggest, then, that where the performance of L1 and L2 speakers may diverge is not necessarily in prosodic phrasing itself, but in its relation to overall fluency and the availability of processing resources. A task manipulating the prosodic projection environment would thus be further complicated in reading, adding the processing burden of decoding and automatized word recognition in a second language.

Most L2 reading fluency and comprehension studies focus on processing speed in word recognition, leaving the role of prosody unexplored. This is perhaps partially due to the difficulty in accurate and consistent measures of prosodic patterns. However, since prosodic phrasing is typically indicated by pauses in oral production, measures of pause frequency and location can be used to measure development of native-like patterns in L2 speakers.

The prosodic aspects, or suprasegmentals, of a language are complex, and cover a broad range of function. In the literature, there is often a two-way distinction made in the features of its subcomponents. Specifically, many models differentiate between language-specific suprasegmentals typically acquired early, and less language-specific suprasegmentals, typically acquired later. There are five primary suprasegmentals constituting prosody: stress timing, peak alignment, speech rate, pause length, and pause frequency (Trofimovich & Baker, 2007).

The “prosody-based” features—stress timing and peak alignment—are acquired early, and are often language-specific (Anema, 2008; Botinis, Granström, & Möbius, 2001). The three “fluency-based” features—speech rate, pause length, and pause frequency—are less language-specific and in L1 acquisition, continue to develop much later into childhood (Anema, 2008; Holmes, 1995). The distinction of prosody- and fluency-based features may perhaps be linked to the distinction of lower-level lexical processes, and the higher-order post-lexical processes. Both
stress timing and peak alignment may apply on a more lexical level, while rate and pausing apply across a more extensive range, e.g., phrases, sentences, and paragraphs.

This distinction may also be helpful in elucidating the process of prosodic development in L2 learners. L2 learners, particularly those who have attained a level of automatization in lexical processes, such as word recognition, may presumably have also attained sufficient proficiency in the prosody-based suprasegmentals. However, given that the fluency-based suprasegmentals are late to arrive even in L1 development, many L2 learners may not yet have attained a native-like level of speech rate and pausing in spontaneous production or in oral reading. Thus, even advanced L2 speakers may lag in development of the fluency-based suprasegmentals, i.e., appropriate phrasing, and native-like pausing patterns.

An oral reading study investigated the relationship of pausing patterns and reading comprehension in advanced L2 speakers. Looking at Dutch L2 speakers of English, Anema (2008) found evidence supporting that more native-like prosody, as indicated by frequency and location of pauses, correlated with higher scores in reading comprehension. Two experimental participant groups were recruited for this study: Dutch-English bilinguals living for at least one year in the US, and Dutch-English bilinguals in the Netherlands.

The oral reading skills of all participants were assessed in both English and Dutch, and participants were rated on each passage based on four measures: (i) reading accuracy, (ii) reading rate, (iii) number of pauses, and (iv) pauses at nonnative locations. Reading comprehension measures included passage-based questions and independent sentence-question pair items.

Measures of L2 working memory span and list-generation ability were used to eliminate the potential effect of working memory capacity differences and evaluate lexical access speed (automaticity).
The two groups, immersed and non-immersed bilinguals, did not differ significantly in working memory span, and both performed at near-native like levels in the list-generation task. However, although both groups exhibited automaticity in lexical access, they performed significantly differently in both pausing patterns and reading comprehension. The participants who had spent at least one year immersed in the L2 environment performed better on reading comprehension measures than the non-immersed group. This group also displayed more native-like pausing patterns in oral reading, pausing less frequently at inappropriate positions, and pausing less frequently overall.

In addition to strengthening the evidence for a link between phrasing and comprehension, the study also suggests a strong role of immersion exposure to the development of both oral and reading fluency.

### 4.4.4 Text segmentation in L2 reading

Within the field of reading development, L2 reading fluency has not been given the same attention as L1 reading fluency. Much of the relevant research tends to focus on L2 reading comprehension, since this is often an area of difficulty for learners. However, there is a clear connection between reading fluency and comprehension (Fuchs et al., 2001, 1988; Jenkins et al., 2000; Nathan & Stanovich, 1991) which suggests that comprehension improves with the development of fluent reading patterns. One method of examining reading fluency is by manipulating how text is presented, and measuring the resultant effect on comprehension. Early work by Cromer (1970) examined this effect in L2 readers of varying reading proficiency levels. Some adaptations of this work to L2 readers are summarized below.
In a series of studies with Japanese learners of English, Kadota and colleagues examined the effect of text segmentation on reading comprehension. Kadota (1982) and Kadota and Tada (1992) found that text segmented into phrasal units improved comprehension and recall rates over sentence or word unit presentations (as cited in Yamashita & Ichikawa, 2010). In subsequent work, Kadota, Yoshida, and Yoshida (1999) presented text in three modes: word-by-word (65), phrase-by-phrase (66), and clause-by-clause (67).

(65) A glacier is a river of ice. It may be ten to thirty miles long and one or two miles wide.

(66) A glacier is a river of ice. It may be ten to thirty miles long and one or two miles wide.

(67) A glacier is a river of ice. It may be ten to thirty miles long and one or two miles wide.

Comprehension was higher and reading times faster in both the phrase and clause conditions than in the word condition.

Yamashita and Ichikawa (2010) further expanded on Cromer’s (1970) text segmentation design to L2 learners, presenting narrative texts to Japanese learners of English in four modes: Whole, Word-by-Word, Chunk, and Fragment. In the Chunk presentation mode, boundary positions roughly corresponded with phrasal boundaries (68), while in the Fragment presentation mode, boundary positions deliberately violated grammatical units (69).

(68) The origin of Australian Rules Football is unclear. Some people say it might have developed from an ancient game in which a ball made of kangaroo skin was kicked around.

(69) The origin of Australian Rules Football is unclear. Some people say it might have developed from an ancient game in which a ball made of kangaroo skin was kicked around.

Yamashita and Ichikawa predicted that lower proficiency learners’ comprehension would be facilitated by appropriately chunked text (Chunk mode), and disrupted by inappropriately
chunked text (Fragment mode), while advanced learners would not be affected by presentation mode. Due to ceiling effects in the comprehension measure, they were unable to confirm the facilitative effect of appropriate chunking in either test group; however, in the Fragment mode, comprehension was significantly lower for the lower proficiency group than in any other mode. There was no effect of mode on comprehension for the advanced learners, although reading times in the Word-by-Word mode were significantly longer.

The results suggest that the advanced learners’ typical phrasing patterns overcome the effect of text presentation, but lower proficiency learners’ underdeveloped phrasing patterns make their reading more susceptible to both disruptive or facilitative effects.

These studies suggest that phrasal chunking of text improves reading comprehension in both native and L2 readers. However, this effect varies with the skill level of the reader. Low proficiency readers, who may read at a word-by-word pace, are not facilitated or disrupted by phrasal manipulation, indicating that their reading skills are not sufficiently developed to project phrases onto the text, and so there is no effect. Intermediate-level readers who may have phrasal projection skills that are not yet robust, are affected by phrasing manipulation, suggesting that their projections are present, but fragile. Advanced readers are not only able to project their phrasing onto incoming text, but that projection is robust enough to resist influence based on presentation—regardless of whether that effect is facilitative or disruptive.

4.4.5 Summary

In this chapter, I reviewed the characteristic features of nonnative processing and areas of variation in ultimate attainment, and summarized the primary assumptions of several L2 processing models. I also introduced some research on the development of reading fluency,
focusing in particular on the role of phrasing ability and the effect of text segmentation on comprehension.

While there is ample evidence of processing-related differences between L1 and L2 speakers, it seems reasonable to suggest that many of these differences are not due to representational deficits, but to processing limitations that constrain performance. Particularly in the development of a complex skill such as reading, which requires application of linguistic knowledge to another modality, with its own accompanying set of proficiency criteria.

One goal of this research is to describe a processing model which is compatible with speakers of varying proficiencies, regardless of language profile, and which can accommodate factors that have received strong empirical validation. These include the effects of task demand and task goals on performance, evidence for a cue-based retrieval mechanism for long-distance dependencies, and the influence of prosody/implicit prosody manipulation on parsing. In the next chapter, I introduce the ‘Good-enough Cue’ model, which integrates these aspects, and present the experimental paradigm and data used to support it.
5. Empirical Evidence for a ‘Good-enough Cue’ Model

This dissertation outlines a comprehensive processing paradigm that joins syntactic and structural foundations, psycholinguistic evidence, and cognitive processing. The preceding chapters have described three critical facets of such a model: a cognitive model of memory retrieval, the role of factors such as prosody in parsing decisions, and a processing paradigm that outlines how resources may be deployed online. Pulling these aspects together, it is possible to develop a broader perspective on processing which is compatible with all of these considerations.

The general framework of this integrated Good-enough Cue (GC) model, sketched and supported empirically below, assumes the ‘Good-Enough’ Hypothesis and cue-based memory retrieval as central aspects. The ‘Good-Enough’ Hypothesis states that all speakers have access to two processing routes: a complete syntactic route, and a ‘good enough’ heuristic route. In the interest of conserving resources, speakers tend to rely more on heuristics and templates whenever the task allows, and may be required to rely on this fallback route when task demand is high. In the GC model, cue-based memory retrieval (CBMR) is the instantiation of the complete syntactic route for agreement and long-distance dependencies in particular. When retrieval fails using CBMR (due to cue overlap, memory trace decay, or some other factor), comprehenders may compensate by applying a ‘good-enough’ processing heuristic, which prioritizes general comprehension over detailed syntactic computation. Prosody (or implicit prosody) may reduce processing load by either facilitating syntactic processing or otherwise assisting memory retrieval, thus reducing reliance on the good-enough fallback route.
5.1 Proposal and review of predictions

To build empirical support for the GC model, I investigate the interactions of structural complexity and text presentation on grammatical and comprehension processing during reading in L1 and L2 speakers. Grammatical and ungrammatical relative clause constructions of two complexity levels were presented in one of three formats: whole sentence, word-by-word, or phrase-by-phrase. Data collected from both native English and L2 Spanish-English speakers included reading and response times, grammaticality ratings, and sentence comprehension responses for probes targeting either general message comprehension or relative clause interpretation.

The GC model predicts that while general message comprehension may remain high even for complex materials, the ability to detect grammatical errors may be reduced as complexity increases and contextual or plausibility cues become more informative to the parse. Performance in general message comprehension may remain stable despite complexity, while performance in relative clause interpretation (which requires syntactic processing to correctly determine agent-patient relations), may be reduced to a similar degree as other grammatical/syntactic processing.

Phrasal presentation may ease processing of grammatical features or aspects by aligning with syntactic structure. It is predicted to improve grammatical processing (e.g., processing of agreement) and possibly improve comprehension for probes targeting relative clause interpretation since a correct answer requires parsing of the structural relation between the two relevant nouns. Phrasal presentation is not predicted to significantly affect performance for probes targeting general comprehension, since overall comprehension is already prioritized by the parser.
Because the proposed processing paradigm is assumed to be a cognitive model operating independently of language background, L1 and L2 participants’ performance is predicted to be qualitatively similar across tasks. However, because L2 participants may be under higher task demand, they may show evidence of greater and earlier reliance on ‘good-enough’ strategies.

5.2 Development of current experimental design

A previously conducted study used a similar paradigm to examine the effect of structural complexity and prosodic disruption on processing and comprehension during reading in native English speakers (see Appendix A for complete description). Results demonstrated that when sentences were presented in such a way that allows normal prosody projection, agreement attraction occurred in subject relative clause constructions, but not in the more difficult object relative constructions. The lack of an attraction effect was attributed to a combination of processing complexity and retrieval cue overlap in the object relatives that made it difficult for readers to detect ungrammaticality.

A second experimental condition presented text in three-word chunks that did not align with syntactic constituents, and was intended to disrupt normal prosody projection. In this condition, no attraction effect was found for either relative clause type. Overall, grammaticality ratings for all sentences were higher when prosody was disrupted, suggesting that disruption interferes in the cue-matching procedure that would normally function in determining agreement.

Based on the findings from this study, I chose to expand the current investigation to examine performance in word-by-word and phrasal presentation formats. L2 participants were included to examine whether language background interacted with the effect of text segmentation. Object relatives were omitted from the current design since they were found to be
exceptionally difficult to parse, resulting in many participants performing at or close to chance in the comprehension measures.

5.3 Methods

5.3.1 Participants

*L1 participants:* 63 native English speakers were recruited (mean 21.6 yrs, SD 6.2 yrs; 37 female). All were Queens College students enrolled in an introductory psychology course and received course credit for their participation.

*L2 participants:* 24 Spanish-English bilinguals (L2) were recruited (mean 23.1 yrs, SD 5.8 yrs; mean AoA: 7.97 yrs; 15 female). 18 were Queens College students enrolled in an introductory psychology course and received course credit for their participation. 6 were recruited via flyer or word of mouth, and were compensated $15.00 for their time. The final pool of L2 participants were selected based on information provided in a background questionnaire—specifically, indication of Spanish as an L1, and age of arrival/age of first exposure to an English environment as 4 years old or older. Average reading proficiency (Woodcock-Johnson and Woodcock-Muñoz) and self-rated proficiency measures for L1 and L2 participants are given in Table 1 below.
Table 1. Average proficiency measures for L1 and L2 participants

<table>
<thead>
<tr>
<th>Measure</th>
<th>L1</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodcock-Johnson</td>
<td>39.3/47</td>
<td>38.6/47</td>
</tr>
<tr>
<td>Woodcock-Muñoz</td>
<td>--</td>
<td>38.1/47</td>
</tr>
<tr>
<td>EN Speaking</td>
<td>9.28/10</td>
<td>8.66/10</td>
</tr>
<tr>
<td>EN Comprehension</td>
<td>9.41/10</td>
<td>8.86/10</td>
</tr>
<tr>
<td>EN Reading</td>
<td>9.25/10</td>
<td>8.82/10</td>
</tr>
<tr>
<td>SP Speaking</td>
<td>--</td>
<td>8.51/10</td>
</tr>
<tr>
<td>SP Comprehension</td>
<td>--</td>
<td>8.95/10</td>
</tr>
<tr>
<td>SP Reading</td>
<td>--</td>
<td>8.04/10</td>
</tr>
</tbody>
</table>

Thirty-two (32) additional participants completed the study, but were excluded due to being native speakers of a language other than Spanish or English (n = 24), significant data loss due to software error (n = 6), or failure to follow instructions (n = 2).

5.3.2 Materials

Experimental materials consisted of subject relative clause sentence sets in which an intervening plural attractor noun either matched or mismatched the target verb. Given that agreement attraction occurs almost exclusively with singular heads and plural attractors, all items were of this configuration. The nouns and target verb were selected from a list of the 5000 most frequent words of each type in the Corpus of Contemporary American English (COCA), and all head and attractor nouns were animate, human, occupational NPs. The 2x2 design crossed the factors of structural complexity (Simple, Complex) and grammaticality (Grammatical, Ungrammatical). Sentences within each set were matched for length by either including an adjunct modifier for the simple items, or embedding an additional that- relative clause for the complex items (see Table 2). The complex items were constructed using proper nouns rather than an additional occupational NP, based on evidence that proper nouns do not interfere with agreement attraction...
Grammaticality was manipulated by varying whether the number feature of the main verb agreed with the singular subject. A sample set of experimental materials appears in Table 2, and the full set of experiment items in Appendix B.

**Table 2. Sample set of materials**

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Grammaticality</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Grammatical</td>
<td>a  The reporter who called the senators every so often writes awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>b  The reporter who called the senators every so often write awful stories for the newspaper.</td>
</tr>
<tr>
<td>Complex</td>
<td>Grammatical</td>
<td>c  The reporter who called the senators that Scott supported writes awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>d  The reporter who called the senators that Scott supported write awful stories for the newspaper.</td>
</tr>
</tbody>
</table>

Filler items contained either agreement violations within other configurations such as noun complements, *wh*-phrases, or pronouns, or violations relating to mass-count number or argument structure (see Appendix C for complete list of fillers by type).

Sixty-four (64) experimental items were distributed across four lists in a Latin Square design and combined with 16 practice items and 128 fillers. Each list was pseudorandomized into four blocks of 48 sentences each. Half of all experimental items were ungrammatical. Of the fillers, 47% were ungrammatical, resulting in 49% of all items seen by each participant being ungrammatical.

### 5.3.3 Procedure

Stimuli were presented on a PC using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Each participant was randomly assigned to a presentation paradigm and stimulus list upon recruitment. Each presentation paradigm was characterized by a unique presentation format and rate for the test items and fillers, as detailed in Table 3.
Table 3. Overview of presentation paradigms

<table>
<thead>
<tr>
<th>Presentation format</th>
<th>Presentation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SENTENCE</strong></td>
<td>whole sentence</td>
</tr>
<tr>
<td><strong>WORD</strong></td>
<td>word-by-word</td>
</tr>
<tr>
<td><strong>PHRASE</strong></td>
<td>phrase segments</td>
</tr>
<tr>
<td></td>
<td>self-paced</td>
</tr>
<tr>
<td></td>
<td>RSVP, 500 ms/word</td>
</tr>
</tbody>
</table>

In the **SENTENCE** presentation paradigm, sentences were presented individually on one line in their entirety, and reading was self-paced. This format most closely imitates natural reading, providing a baseline of comparison for the other presentation formats. Each sentence was preceded by a fixation cross appearing centrally on the screen for 1000 ms. After 7000 ms, a prompt to respond quickly appeared in the upper left corner of the display. A final timeout was set at 15000 ms from initial onset. Both the sentence and prompt remained on the screen until the timeout, or until the participant pressed the space bar to advance to the next screen.

In the **WORD** presentation paradigm, sentences were presented word-by-word, in rapid serial visual presentation (RSVP) format, at a fixed rate of 500 ms per word. While RSVP has often been utilized as a method of enhancing reading speed (Rahman & Muter, 1999) or testing processing ability at high presentation rates (Potter, 1984), presenting sentences at a rate of two words per second is strikingly slow, and has been shown to disrupt processing during reading (Fernández, 2007). Although this rate is often used in ERP reading studies, it is primarily used to avoid overlap in the measured potentials, not to simulate natural reading patterns (Swaab et al., 2011).

While RSVP has the potential to interfere with the projection of prosody onto the text (Castelhano & Muter, 2001; Fernández, 2007), the segmentation of sentences into appropriate phrasal units may significantly improve performance (O’Shea & Sindelar, 1983; Schreiber, 1987,
1991). The **Phrase** presentation paradigm was designed to validate this effect and facilitate processing by presenting each sentence with phrasal breaks after the head noun, and again after the relative clause, as shown in (70).

(70)  
   a. Simple: The reporter who called the senators every so often write(s)…  
   b. Complex: The reporter who called the senators that Scott supported write(s)…

In the **Phrase** presentation paradigm, sentences were presented in these three phrasal segments, and reading was self-paced. Each sentence was preceded by a fixation cross appearing centrally on the screen for 1000 ms. The first segment of the sentence then appeared centrally on the screen, and after 2000 ms, a prompt to respond quickly appeared in the upper left corner of the display. A final timeout was set at 5000 ms from initial onset. Both the segment and prompt remained on the screen until the timeout, or until the participant pressed the space bar to advance to the next segment.

For all presentation paradigms, following each sentence, participants were prompted to rate the sentence on a 6-point Likert scale (1 = “very bad”, 6 = “perfect”). To minimize low ratings due to general dislike for these types of complex structures, participants were instructed to rate the sentences based on whether a 300-level English professor would consider them grammatical. Participants were then prompted to respond to a true/false comprehension probe, and received speed and accuracy feedback on their responses. The comprehension probes for the experimental items targeted either the relative clause interpretation (71), or general comprehension (71).

(71) The reporter who called the senators every so often writes awful stories for the newspaper.
   a. RC target: The reporter called the senators.
   b. Other target: The reporter works in television.
Participants were instructed to respond as quickly and accurately as possible, and were allowed 5000 ms to rate each sentence, and an additional 5000 ms to respond to the comprehension probe.

Following the main experimental session, participants completed a Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007). All participants were administered the passage comprehension subtest of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001), where they silently read a short passage and verbally identified a missing word appropriate to the context.

L2 participants were also administered the equivalent passage comprehension subtest of the Batería III Woodcock-Muñoz Pruebas de Aprovechamiento (Woodcock et al., 2004).

5.4 Results

5.4.1 Analysis overview

Grammaticality ratings and comprehension accuracy data were analyzed, as well as response times for both measures, and reading times in the PHRASE and SENTENCE conditions.

Statistics are presented as the results of linear mixed-effects models with the maximal random-effects structures justified by the models, calculated over grammaticality ratings, comprehension question accuracy, and reaction times. Models were fit using R software (version 3.1.3, R Core Team, 2015) and the lme4 and lmerTest packages.

Correlations were also performed among the Woodcock-Johnson and Woodcock-Muñoz passage comprehension tests, and the self-rated proficiency measures from the LEAP-Q questionnaire.
For all response time (RT) measures, only items for which the comprehension probe was answered correctly were included in the analysis. A response timeout was set at 5000 ms for all measures, with the exception of Sentence RT, which was set at 15000 ms. Due to software error, some responses were recorded beyond these limits, and so were excluded from the analysis. Response times of less than 250 ms were also excluded, resulting in combined data loss of less than 5% per measure (see Table 4).

Table 4. Response time (RT) data exclusions

<table>
<thead>
<tr>
<th>Measure</th>
<th>% data excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammaticality RT</td>
<td>3.98</td>
</tr>
<tr>
<td>Comprehension RT</td>
<td>1.29</td>
</tr>
<tr>
<td>Segment 1 RT (PHRASE only)</td>
<td>2.85</td>
</tr>
<tr>
<td>Segment 2 RT (PHRASE only)</td>
<td>2.23</td>
</tr>
<tr>
<td>Segment 3 RT (PHRASE only)</td>
<td>2.77</td>
</tr>
<tr>
<td>Sentence RT (SENTENCE only)</td>
<td>4.03</td>
</tr>
</tbody>
</table>

All remaining response times that exceeded a threshold of ± 2 standard deviations were replaced by the cutoff value (equal to 2 standard deviations beyond the mean), and subsequently log-transformed prior to analysis.

5.4.2 Grammaticality ratings

A linear mixed-effects model was applied to the grammaticality ratings data, and included fixed effects of format (SENTENCE, WORD, PHRASE), group (L1, L2), grammaticality (Grammatical, Ungrammatical), and complexity (Simple, Complex), as well as all interactions. The results of this model and information on the random-effects structure appear in Table 5. As shown in Table 5, there was an overall effect of grammaticality, where grammatical sentences were rated higher than ungrammatical sentences, as well as an overall effect of complexity, where simple constructions were rated more highly than complex constructions. Complexity was also found to
interact with presentation format, indicating that participants rated complex items more highly when presented in the PHRASE format than in the SENTENCE format.\(^12\)

**Table 5.** Linear mixed-effects model of overall grammaticality ratings. Effects of format, group, grammaticality, and complexity.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.026</td>
<td>0.11</td>
<td>100</td>
<td>37.76</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.190</td>
<td>0.29</td>
<td>89</td>
<td>0.65</td>
<td>0.515</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>0.416</td>
<td>0.30</td>
<td>96</td>
<td>1.40</td>
<td>0.165</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>-0.119</td>
<td>0.21</td>
<td>92</td>
<td>-0.58</td>
<td>0.566</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.264</td>
<td>0.05</td>
<td>92</td>
<td>-5.41</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.147</td>
<td>0.03</td>
<td>80</td>
<td>-5.00</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Group</td>
<td>0.062</td>
<td>0.58</td>
<td>89</td>
<td>0.11</td>
<td>0.915</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group</td>
<td>-0.291</td>
<td>0.59</td>
<td>95</td>
<td>-0.49</td>
<td>0.624</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality</td>
<td>0.221</td>
<td>0.13</td>
<td>83</td>
<td>1.66</td>
<td>0.101</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality</td>
<td>-0.301</td>
<td>0.14</td>
<td>91</td>
<td>-2.18</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>0.039</td>
<td>0.07</td>
<td>80</td>
<td>0.57</td>
<td>0.570</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.360</td>
<td>0.07</td>
<td>81</td>
<td>5.20</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>-0.068</td>
<td>0.10</td>
<td>86</td>
<td>-0.71</td>
<td>0.476</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>0.031</td>
<td>0.05</td>
<td>82</td>
<td>0.64</td>
<td>0.525</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.038</td>
<td>0.02</td>
<td>4952</td>
<td>1.79</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Grammaticality</td>
<td>0.105</td>
<td>0.27</td>
<td>82</td>
<td>0.39</td>
<td>0.695</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Grammaticality</td>
<td>-0.413</td>
<td>0.27</td>
<td>84</td>
<td>-1.54</td>
<td>0.127</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity</td>
<td>-0.179</td>
<td>0.14</td>
<td>79</td>
<td>-1.30</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity</td>
<td>-0.328</td>
<td>0.14</td>
<td>80</td>
<td>-2.38</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × Complexity</td>
<td>0.029</td>
<td>0.06</td>
<td>4864</td>
<td>0.49</td>
<td>0.621</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality × Complexity</td>
<td>-0.033</td>
<td>0.06</td>
<td>4967</td>
<td>-0.56</td>
<td>0.574</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.030</td>
<td>0.05</td>
<td>4727</td>
<td>0.71</td>
<td>0.481</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>-0.033</td>
<td>0.12</td>
<td>4912</td>
<td>-0.29</td>
<td>0.774</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Grammaticality × Complexity</td>
<td>0.156</td>
<td>0.12</td>
<td>5003</td>
<td>1.34</td>
<td>0.181</td>
<td></td>
</tr>
</tbody>
</table>

12 The full mixed-effects model which included English reading proficiency (WJ proficiency) as a factor did not converge, so WJ proficiency was omitted here as a fixed effect, but included in the L1, L2, and presentation format by-group models.
The three-way interaction between format, group, and complexity was investigated further by conducting separate analyses on the L1 and L2 groups, and presentation formats. Both the L1 and L2 models included fixed effects of format, grammaticality, complexity, and English reading proficiency (WJ Proficiency), as well as all interactions. Results for L1 participants are reported in Table 6, and results for L2 participants are reported in Table 7.

While both groups rated complex items more highly in the PHRASE condition, only the L1 group reliably distinguished grammatical from ungrammatical sentences overall, and most robustly in the PHRASE condition (See Figure 2, as compared to L2 participants in Figure 3). There was a significant interaction between grammaticality and WJ proficiency, indicating that L1 participants with higher English reading proficiency were better able to correctly accept grammatical sentences. For the L2 participants, a format and WJ proficiency interaction revealed that higher WJ proficiency was associated with higher grammaticality ratings in the WORD condition only. No other results were significant.

**Figure 2.** L1 grammaticality ratings by presentation format. Errors bars indicate standard error.

A. L1 Grammaticality ratings, Simple items  
B. L1 Grammaticality ratings, Complex items
Table 6. Linear mixed-effects model of L1 grammaticality ratings. Effects of format, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.001</td>
<td>0.10</td>
<td>66</td>
<td>38.55</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.370</td>
<td>0.29</td>
<td>63</td>
<td>1.26</td>
<td>0.213</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>0.106</td>
<td>0.28</td>
<td>63</td>
<td>0.37</td>
<td>0.714</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.271</td>
<td>0.05</td>
<td>66</td>
<td>-5.52</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.132</td>
<td>0.03</td>
<td>62</td>
<td>-5.28</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>0.030</td>
<td>0.04</td>
<td>65</td>
<td>0.82</td>
<td>0.413</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality</td>
<td>0.192</td>
<td>0.14</td>
<td>63</td>
<td>1.39</td>
<td>0.170</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality</td>
<td>-0.446</td>
<td>0.14</td>
<td>63</td>
<td>-3.28</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>-0.057</td>
<td>0.07</td>
<td>63</td>
<td>-0.87</td>
<td>0.388</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.181</td>
<td>0.06</td>
<td>62</td>
<td>2.85</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × WJ Proficiency</td>
<td>0.106</td>
<td>0.09</td>
<td>63</td>
<td>1.14</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × WJ Proficiency</td>
<td>0.078</td>
<td>0.11</td>
<td>63</td>
<td>0.72</td>
<td>0.475</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.054</td>
<td>0.02</td>
<td>3583</td>
<td>2.90</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>-0.051</td>
<td>0.02</td>
<td>64</td>
<td>-2.97</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>-0.002</td>
<td>0.01</td>
<td>62</td>
<td>-0.31</td>
<td>0.757</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × Complexity</td>
<td>0.019</td>
<td>0.05</td>
<td>3258</td>
<td>0.36</td>
<td>0.719</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality × Complexity</td>
<td>0.037</td>
<td>0.05</td>
<td>3360</td>
<td>0.73</td>
<td>0.467</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × WJ Proficiency</td>
<td>0.075</td>
<td>0.04</td>
<td>63</td>
<td>1.70</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × WJ Proficiency</td>
<td>-0.058</td>
<td>0.05</td>
<td>63</td>
<td>-1.13</td>
<td>0.263</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WJ Proficiency</td>
<td>0.002</td>
<td>0.02</td>
<td>64</td>
<td>0.08</td>
<td>0.933</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity × WJ Proficiency</td>
<td>0.013</td>
<td>0.02</td>
<td>64</td>
<td>0.55</td>
<td>0.582</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.002</td>
<td>0.01</td>
<td>3479</td>
<td>0.32</td>
<td>0.748</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × Complexity × WJ Proficiency</td>
<td>-0.003</td>
<td>0.02</td>
<td>3334</td>
<td>-0.21</td>
<td>0.837</td>
<td></td>
</tr>
</tbody>
</table>

Sentence v. Word = Comparison of Baseline (SENTENCE) to WORD, Sentence v. Phrase = Comparison of Baseline (SENTENCE) to PHRASE; Random effects structure included intercept and effects of Grammaticality and Complexity for Subjects plus intercept and effects of Format, Grammaticality, Complexity and WJ Proficiency for Items. R-code used with lmerTest: Rating ~ Format * Grammaticality * Complexity * WJ Proficiency + (1 + Grammaticality * Complexity | Subject) + (1 + Format + Grammaticality + Complexity + WJ Proficiency | Item). Degrees of freedom (df) calculated using the Satterthwaite approximation.
Figure 3. L2 grammaticality ratings by presentation format. Error bars indicate standard error.

A. L2 grammaticality ratings, Simple items

B. L2 grammatical ratings, Complex items
Table 7. Linear mixed-effects model of L2 grammaticality ratings. Effects of format, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.272</td>
<td>0.18</td>
<td>25.60</td>
<td>23.50</td>
<td>&lt;0.001 *</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.048</td>
<td>0.50</td>
<td>25.59</td>
<td>0.096</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>0.115</td>
<td>0.50</td>
<td>25.67</td>
<td>0.228</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.200</td>
<td>0.09</td>
<td>24.67</td>
<td>-2.197</td>
<td>&lt;0.05 *</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.162</td>
<td>0.08</td>
<td>32.27</td>
<td>-2.151</td>
<td>&lt;0.05 *</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>-0.020</td>
<td>0.05</td>
<td>32.67</td>
<td>-0.423</td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality</td>
<td>0.106</td>
<td>0.23</td>
<td>19.04</td>
<td>0.461</td>
<td>0.650</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality</td>
<td>-0.100</td>
<td>0.23</td>
<td>19.01</td>
<td>-0.430</td>
<td>0.672</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>0.162</td>
<td>0.18</td>
<td>23.65</td>
<td>0.877</td>
<td>0.389</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.441</td>
<td>0.19</td>
<td>23.62</td>
<td>2.363</td>
<td>&lt;0.05 *</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × WJ Proficiency</td>
<td>0.384</td>
<td>0.12</td>
<td>24.14</td>
<td>3.171</td>
<td>&lt;0.01 *</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × WJ Proficiency</td>
<td>-0.017</td>
<td>0.12</td>
<td>24.26</td>
<td>-0.142</td>
<td>0.881</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.036</td>
<td>0.05</td>
<td>114.2</td>
<td>0.658</td>
<td>0.512</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>0.000</td>
<td>0.02</td>
<td>22.05</td>
<td>0.008</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>-0.003</td>
<td>0.02</td>
<td>25.37</td>
<td>-0.203</td>
<td>0.841</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × Complexity</td>
<td>0.069</td>
<td>0.14</td>
<td>101.7</td>
<td>0.484</td>
<td>0.670</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality × Complexity</td>
<td>-0.001</td>
<td>0.15</td>
<td>103.0</td>
<td>-0.067</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × WJ Proficiency</td>
<td>-0.014</td>
<td>0.06</td>
<td>19.24</td>
<td>-0.251</td>
<td>0.805</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality × WJ Proficiency</td>
<td>0.044</td>
<td>0.06</td>
<td>19.89</td>
<td>0.793</td>
<td>0.437</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WJ Proficiency</td>
<td>0.053</td>
<td>0.05</td>
<td>23.75</td>
<td>1.162</td>
<td>0.257</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity × WJ Proficiency</td>
<td>-0.073</td>
<td>0.05</td>
<td>26.39</td>
<td>-1.607</td>
<td>0.120</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.025</td>
<td>0.01</td>
<td>119.2</td>
<td>1.872</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × WJ Proficiency</td>
<td>-0.008</td>
<td>0.04</td>
<td>104.3</td>
<td>-0.224</td>
<td>0.823</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>0.024</td>
<td>0.04</td>
<td>108.5</td>
<td>0.694</td>
<td>0.489</td>
<td></td>
</tr>
</tbody>
</table>

Sentence v. Word = Comparison of Baseline (SENTENCE) to WORD, Sentence v. Phrase = Comparison of Baseline (SENTENCE) to PHRASE; Random effects structure included intercept and effects of Grammaticality and Complexity for Subjects plus intercept and effects of Format, Grammaticality, Complexity and WJ Proficiency for Items. R-code used with lmerTest: Rating ~ Format * Grammaticality * Complexity * WJ Proficiency + (1 + Grammaticality * Complexity | Subject) + (1 + Format + Grammaticality + Complexity + WJ Proficiency | Item). Degrees of freedom (df) calculated using the Satterthwaite approximation.

Grammaticality ratings for the L2 participants were also analyzed as above, including Spanish reading proficiency (WM proficiency) as a factor, see Table 8. There was a marginally significant interaction of grammaticality and WM proficiency, suggesting that higher Spanish reading proficiency may improve ability to correctly accept the grammatical sentences. There was a significant three-way interaction between format, grammaticality, and WM proficiency,
indicating that higher proficiency may be particularly beneficial for those in the PHRASE presentation condition.

**Table 8.** Linear mixed-effects model of L2 grammaticality ratings. Effects of format, grammaticality, complexity, and Spanish reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.059</td>
<td>0.25</td>
<td>24.7</td>
<td>16.033</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.051</td>
<td>0.66</td>
<td>24</td>
<td>0.077</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>0.221</td>
<td>0.83</td>
<td>24.4</td>
<td>0.267</td>
<td>0.792</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.061</td>
<td>0.10</td>
<td>24</td>
<td>-0.626</td>
<td>0.537</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.227</td>
<td>0.10</td>
<td>43.4</td>
<td>-2.383</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>-0.045</td>
<td>0.05</td>
<td>24.9</td>
<td>-0.856</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality</td>
<td>-0.167</td>
<td>0.24</td>
<td>18.5</td>
<td>-0.704</td>
<td>0.490</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality</td>
<td>0.449</td>
<td>0.30</td>
<td>18.7</td>
<td>1.511</td>
<td>0.148</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>0.253</td>
<td>0.22</td>
<td>28</td>
<td>1.175</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.222</td>
<td>0.28</td>
<td>30.3</td>
<td>0.807</td>
<td>0.426</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × WM Proficiency</td>
<td>0.153</td>
<td>0.13</td>
<td>24</td>
<td>1.137</td>
<td>0.267</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × WM Proficiency</td>
<td>-0.014</td>
<td>0.17</td>
<td>24.1</td>
<td>-0.079</td>
<td>0.937</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.083</td>
<td>0.07</td>
<td>1075.7</td>
<td>1.161</td>
<td>0.246</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WM Proficiency</td>
<td>0.036</td>
<td>0.02</td>
<td>20.2</td>
<td>1.854</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>-0.016</td>
<td>0.02</td>
<td>29.5</td>
<td>-0.951</td>
<td>0.350</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.074</td>
<td>0.17</td>
<td>1258.2</td>
<td>-0.435</td>
<td>0.664</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.164</td>
<td>0.22</td>
<td>1105.2</td>
<td>0.738</td>
<td>0.461</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality ×</td>
<td>-0.067</td>
<td>0.05</td>
<td>18.9</td>
<td>-1.373</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>0.160</td>
<td>0.06</td>
<td>20</td>
<td>2.489</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WM</td>
<td>0.039</td>
<td>0.05</td>
<td>32.8</td>
<td>0.852</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td>-0.080</td>
<td>0.06</td>
<td>38.2</td>
<td>-1.276</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WM</td>
<td>0.019</td>
<td>0.01</td>
<td>1044.5</td>
<td>1.358</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.042</td>
<td>0.04</td>
<td>987.6</td>
<td>-1.132</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>0.062</td>
<td>0.05</td>
<td>715.1</td>
<td>1.227</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality ×</td>
<td>0.114</td>
<td>0.02</td>
<td>1105.2</td>
<td>0.738</td>
<td>0.461</td>
<td></td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>-0.067</td>
<td>0.05</td>
<td>18.9</td>
<td>-1.373</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WM</td>
<td>0.039</td>
<td>0.05</td>
<td>32.8</td>
<td>0.852</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td>-0.080</td>
<td>0.06</td>
<td>38.2</td>
<td>-1.276</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WM</td>
<td>0.019</td>
<td>0.01</td>
<td>1044.5</td>
<td>1.358</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.042</td>
<td>0.04</td>
<td>987.6</td>
<td>-1.132</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>0.062</td>
<td>0.05</td>
<td>715.1</td>
<td>1.227</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality ×</td>
<td>0.114</td>
<td>0.02</td>
<td>1105.2</td>
<td>0.738</td>
<td>0.461</td>
<td></td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>-0.067</td>
<td>0.05</td>
<td>18.9</td>
<td>-1.373</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WM</td>
<td>0.039</td>
<td>0.05</td>
<td>32.8</td>
<td>0.852</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td>-0.080</td>
<td>0.06</td>
<td>38.2</td>
<td>-1.276</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WM</td>
<td>0.019</td>
<td>0.01</td>
<td>1044.5</td>
<td>1.358</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.042</td>
<td>0.04</td>
<td>987.6</td>
<td>-1.132</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>0.062</td>
<td>0.05</td>
<td>715.1</td>
<td>1.227</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality ×</td>
<td>0.114</td>
<td>0.02</td>
<td>1105.2</td>
<td>0.738</td>
<td>0.461</td>
<td></td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>-0.067</td>
<td>0.05</td>
<td>18.9</td>
<td>-1.373</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WM</td>
<td>0.039</td>
<td>0.05</td>
<td>32.8</td>
<td>0.852</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td>-0.080</td>
<td>0.06</td>
<td>38.2</td>
<td>-1.276</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WM</td>
<td>0.019</td>
<td>0.01</td>
<td>1044.5</td>
<td>1.358</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.042</td>
<td>0.04</td>
<td>987.6</td>
<td>-1.132</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>0.062</td>
<td>0.05</td>
<td>715.1</td>
<td>1.227</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality ×</td>
<td>0.114</td>
<td>0.02</td>
<td>1105.2</td>
<td>0.738</td>
<td>0.461</td>
<td></td>
</tr>
<tr>
<td>WM Proficiency</td>
<td>-0.067</td>
<td>0.05</td>
<td>18.9</td>
<td>-1.373</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × WM</td>
<td>0.039</td>
<td>0.05</td>
<td>32.8</td>
<td>0.852</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>Proficiency</td>
<td>-0.080</td>
<td>0.06</td>
<td>38.2</td>
<td>-1.276</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WM</td>
<td>0.019</td>
<td>0.01</td>
<td>1044.5</td>
<td>1.358</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality ×</td>
<td>-0.042</td>
<td>0.04</td>
<td>987.6</td>
<td>-1.132</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>Complexity × WM Proficiency</td>
<td>0.062</td>
<td>0.05</td>
<td>715.1</td>
<td>1.227</td>
<td>0.220</td>
<td></td>
</tr>
</tbody>
</table>

Sentence v. Word = Comparison of Baseline (SENTENCE) to WORD, Sentence v. Phrase = Comparison of Baseline (SENTENCE) to PHRASE; Random effects structure included intercept and effects of Grammaticality and Complexity for Subjects plus intercept and effects of Format, Grammaticality, Complexity and WM Proficiency for Items. R-code used with lmerTest: Rating ~ Format * Grammaticality * Complexity * WM Proficiency + (1 + Grammaticality + Complexity | Subject) + (1 + Format + Grammaticality + Complexity + WM Proficiency | Item). Degrees of freedom (df) calculated using the Satterthwaite approximation.

Grammaticality ratings for all participants were further examined by conducting separate analyses according to presentation format. All models included fixed effects of group,
grammaticality, complexity, and WJ proficiency, and all interactions. Results of the Sentence condition model and random effects structure are given in Table 9. In the Sentence presentation format, there were significant main effects of grammaticality, complexity, and WJ proficiency, such that ratings for grammatical items were higher than for ungrammatical, and higher for simple than complex items, and that ratings tended to be lower as WJ proficiency increased. There was also a significant interaction between group and complexity, indicating that the effect of complexity was particularly evident in the L2 group. See Figure 4.

**Figure 4.** Grammaticality ratings by group and complexity, Sentence condition only. Error bars indicate standard error.

In the Word presentation format, there were similarly significant main effects of grammaticality, complexity, and WJ proficiency, as well as a marginally significant interaction of grammaticality and complexity. See Table 10 and Figure 5. In the Phrase presentation format, there was a significant effect of grammaticality only; no other main effects or interactions were significant. See Table 11 and Figure 6.
**Figure 5.** Grammaticality ratings by group and complexity, *WORD* condition only. Error bars indicate standard error.

![Grammaticality ratings by group and complexity, *WORD* condition only](image1)

*Figure 6.** Grammaticality ratings by group and complexity, *PHRASE* condition only. Error bars indicate standard error.

![Grammaticality ratings by group and complexity, *PHRASE* condition only](image2)
Table 9. Linear mixed-effects model of grammaticality ratings in the SENTENCE condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.994</td>
<td>0.19</td>
<td>30.80</td>
<td>21.157</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>-0.465</td>
<td>0.37</td>
<td>29.60</td>
<td>-1.243</td>
<td>0.223</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.191</td>
<td>0.09</td>
<td>28.50</td>
<td>-2.200</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.334</td>
<td>0.06</td>
<td>28.60</td>
<td>-5.567</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>-0.139</td>
<td>0.05</td>
<td>29.50</td>
<td>-2.670</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.094</td>
<td>0.17</td>
<td>27.30</td>
<td>0.551</td>
<td>0.586</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>0.283</td>
<td>0.12</td>
<td>28.20</td>
<td>2.376</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>0.152</td>
<td>0.10</td>
<td>29.20</td>
<td>1.469</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.023</td>
<td>0.04</td>
<td>1427</td>
<td>0.660</td>
<td>0.509</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>-0.037</td>
<td>0.02</td>
<td>25.90</td>
<td>-1.557</td>
<td>0.132</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>-0.005</td>
<td>0.02</td>
<td>28.40</td>
<td>-0.271</td>
<td>0.788</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>-0.002</td>
<td>0.07</td>
<td>1325</td>
<td>-0.033</td>
<td>0.973</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>-0.043</td>
<td>0.05</td>
<td>25.30</td>
<td>-0.931</td>
<td>0.361</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>-0.014</td>
<td>0.03</td>
<td>28.50</td>
<td>-0.435</td>
<td>0.667</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.007</td>
<td>0.01</td>
<td>1318</td>
<td>0.750</td>
<td>0.454</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity × WJ Proficiency</td>
<td>-0.014</td>
<td>0.02</td>
<td>1197</td>
<td>-0.722</td>
<td>0.470</td>
<td></td>
</tr>
</tbody>
</table>


Table 10. Linear mixed-effects model of grammaticality ratings in the WORD condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.244</td>
<td>0.17</td>
<td>30.10</td>
<td>24.35</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>-0.123</td>
<td>0.35</td>
<td>29.20</td>
<td>-0.356</td>
<td>0.724</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.166</td>
<td>0.04</td>
<td>34.50</td>
<td>-3.949</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>-0.121</td>
<td>0.04</td>
<td>31.20</td>
<td>-2.965</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>0.127</td>
<td>0.05</td>
<td>29.10</td>
<td>2.811</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>-0.019</td>
<td>0.08</td>
<td>30.40</td>
<td>-0.245</td>
<td>0.808</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>-0.074</td>
<td>0.08</td>
<td>28.20</td>
<td>-0.958</td>
<td>0.346</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>0.087</td>
<td>0.09</td>
<td>29.00</td>
<td>-0.963</td>
<td>0.343</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.054</td>
<td>0.03</td>
<td>1621</td>
<td>1.885</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>-0.009</td>
<td>0.01</td>
<td>28.70</td>
<td>-0.867</td>
<td>0.393</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>0.011</td>
<td>0.01</td>
<td>28.40</td>
<td>1.102</td>
<td>0.280</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>0.034</td>
<td>0.06</td>
<td>1521</td>
<td>0.605</td>
<td>0.545</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>-0.013</td>
<td>0.02</td>
<td>29.70</td>
<td>-0.655</td>
<td>0.517</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>-0.024</td>
<td>0.02</td>
<td>28.30</td>
<td>-1.181</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.005</td>
<td>0.01</td>
<td>1557</td>
<td>0.720</td>
<td>0.471</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Linear mixed-effects model of grammaticality ratings in the PHRASE condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.194</td>
<td>0.17</td>
<td>33.10</td>
<td>24.177</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>-0.276</td>
<td>0.338</td>
<td>30.00</td>
<td>-0.817</td>
<td>0.421</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.387</td>
<td>0.11</td>
<td>30.80</td>
<td>-3.568</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>0.003</td>
<td>0.05</td>
<td>84.20</td>
<td>0.056</td>
<td>0.956</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>0.022</td>
<td>0.05</td>
<td>43.70</td>
<td>0.415</td>
<td>0.680</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>-0.265</td>
<td>0.21</td>
<td>26.70</td>
<td>-1.270</td>
<td>0.215</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>-0.091</td>
<td>0.10</td>
<td>88.40</td>
<td>-0.932</td>
<td>0.354</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>0.094</td>
<td>0.10</td>
<td>29.00</td>
<td>0.989</td>
<td>0.331</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.041</td>
<td>0.05</td>
<td>1713</td>
<td>0.901</td>
<td>0.368</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>-0.031</td>
<td>0.03</td>
<td>27.20</td>
<td>-1.023</td>
<td>0.315</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>-0.016</td>
<td>0.01</td>
<td>89.90</td>
<td>-1.142</td>
<td>0.257</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>0.033</td>
<td>0.09</td>
<td>1711</td>
<td>0.364</td>
<td>0.716</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>-0.059</td>
<td>0.06</td>
<td>33.10</td>
<td>-0.912</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>0.036</td>
<td>0.03</td>
<td>91.70</td>
<td>1.273</td>
<td>0.206</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.017</td>
<td>0.01</td>
<td>1692</td>
<td>1.262</td>
<td>0.207</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity × WJ Proficiency</td>
<td>0.029</td>
<td>0.03</td>
<td>1737</td>
<td>1.126</td>
<td>0.260</td>
<td></td>
</tr>
</tbody>
</table>


5.4.3 Comprehension accuracy

The comprehension accuracy data model included fixed effects of format, group, complexity, and comprehension target (RC, Other), as well as all interactions. Model results and random-effects structure are given in Table 12.

Overall comprehension accuracy was higher for L1 participants (mean 73.4%) than L2 participants (mean 66.6%) and accuracy was higher for simple constructions than for complex constructions. There was a significant interaction between format and group, where accuracy for L1 participants was similar in the SENTENCE and PHRASE conditions, but lower in the WORD condition (Figure 7), while accuracy for L2 participants was similar in the SENTENCE and WORD conditions, but lower in the PHRASE condition (Figure 8). The effect of target was also
significant: accuracy was lower when the probe targeted the relative clause interpretation versus general comprehension (Figure 9 and Figure 10).

**Table 12.** Generalized linear mixed-effects model of overall comprehension accuracy. Effects of format, group, complexity, and target.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>z</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.117</td>
<td>0.14</td>
<td>7.855</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>−0.178</td>
<td>0.17</td>
<td>−1.071</td>
<td>0.284</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>−0.104</td>
<td>0.17</td>
<td>−0.630</td>
<td>0.529</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.422</td>
<td>0.12</td>
<td>3.628</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Complexity</td>
<td>−0.202</td>
<td>0.06</td>
<td>−3.206</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Target</td>
<td>−0.347</td>
<td>0.14</td>
<td>−2.532</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Group</td>
<td>−0.798</td>
<td>0.33</td>
<td>−2.444</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Phrase × Group</td>
<td>0.863</td>
<td>0.34</td>
<td>2.550</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>0.147</td>
<td>0.14</td>
<td>1.025</td>
<td>0.305</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>−0.157</td>
<td>0.14</td>
<td>−1.090</td>
<td>0.276</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Target</td>
<td>0.052</td>
<td>0.12</td>
<td>−0.380</td>
<td>0.704</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Target</td>
<td>−0.190</td>
<td>0.13</td>
<td>0.413</td>
<td>0.680</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>−0.038</td>
<td>0.10</td>
<td>−1.507</td>
<td>0.132</td>
<td></td>
</tr>
<tr>
<td>Group × Target</td>
<td>0.055</td>
<td>0.09</td>
<td>0.616</td>
<td>0.538</td>
<td></td>
</tr>
<tr>
<td>Complexity × Target</td>
<td>0.097</td>
<td>0.06</td>
<td>1.689</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity</td>
<td>0.031</td>
<td>0.30</td>
<td>0.103</td>
<td>0.918</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity</td>
<td>−0.014</td>
<td>0.29</td>
<td>−0.047</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Target</td>
<td>−0.221</td>
<td>0.25</td>
<td>−0.895</td>
<td>0.371</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Target</td>
<td>0.113</td>
<td>0.26</td>
<td>0.429</td>
<td>0.668</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × Target</td>
<td>0.033</td>
<td>0.12</td>
<td>0.268</td>
<td>0.789</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity × Target</td>
<td>0.011</td>
<td>0.13</td>
<td>0.089</td>
<td>0.929</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × Target</td>
<td>−0.007</td>
<td>0.09</td>
<td>−0.073</td>
<td>0.942</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity × Target</td>
<td>0.119</td>
<td>0.27</td>
<td>0.436</td>
<td>0.663</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity × Target</td>
<td>−0.141</td>
<td>0.25</td>
<td>−0.564</td>
<td>0.573</td>
<td></td>
</tr>
</tbody>
</table>

*Sentence v. Word = Comparison of Baseline (SENTENCE) to WORD, Sentence v. Phrase = Comparison of Baseline (SENTENCE) to PHRASE; Random effects structure included intercept and effects of Complexity and Target for Subjects plus intercept and effects of Format, Group, and Complexity for Items. R-code used with glmer: Comprehension Accuracy ~ Format * Group * Complexity * Target + (1 + Complexity * Target | Subject) + (1 + Format * Group * Complexity | Item).*
Figure 7. L1 comprehension accuracy by presentation format and complexity. Error bars indicate standard error.

Figure 8. L2 comprehension accuracy by presentation format and complexity. Error bars indicate standard error.
Figure 9. L1 comprehension accuracy by presentation format and comprehension target. Error bars indicate standard error.

A. L1 comprehension accuracy, Simple items

B. L1 comprehension accuracy, Complex items

Figure 10. L2 comprehension accuracy by presentation format and comprehension target. Error bars indicate standard error.

A. L2 comprehension accuracy, Simple items

B. L2 comprehension accuracy, Complex items

5.4.4 Response times (RTs)

5.4.4.1 Grammaticality rating RTs

The grammaticality rating RT model included fixed effects of format, group, grammaticality, and complexity, as well as all interactions. The results of this model and information on the random-effects structure appear in Table 13. The only significant finding was a main effect of format, as well as an interaction of format and complexity, such that response times were slower in the WORD condition, and particularly so for complex constructions.
Table 13. Linear mixed-effects model of overall grammaticality rating response times. Effects of format, group, grammaticality, and complexity.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.123</td>
<td>0.04</td>
<td>87</td>
<td>170.008</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.372</td>
<td>0.12</td>
<td>86</td>
<td>3.149</td>
<td>&lt;0.01</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>0.034</td>
<td>0.12</td>
<td>87</td>
<td>0.284</td>
<td>0.777</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.033</td>
<td>0.08</td>
<td>87</td>
<td>0.399</td>
<td>0.691</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.004</td>
<td>0.01</td>
<td>76</td>
<td>-0.484</td>
<td>0.630</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.002</td>
<td>0.01</td>
<td>86</td>
<td>0.202</td>
<td>0.840</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group</td>
<td>-0.031</td>
<td>0.24</td>
<td>86</td>
<td>-0.130</td>
<td>0.897</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group</td>
<td>-0.191</td>
<td>0.24</td>
<td>86</td>
<td>-0.811</td>
<td>0.202</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality</td>
<td>0.006</td>
<td>0.02</td>
<td>133</td>
<td>0.258</td>
<td>0.797</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Grammaticality</td>
<td>0.008</td>
<td>0.02</td>
<td>139</td>
<td>0.365</td>
<td>0.716</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>0.038</td>
<td>0.02</td>
<td>138</td>
<td>2.385</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.003</td>
<td>0.02</td>
<td>221</td>
<td>0.123</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.011</td>
<td>0.02</td>
<td>232</td>
<td>0.503</td>
<td>0.615</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>-0.020</td>
<td>0.02</td>
<td>229</td>
<td>-1.179</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>0.009</td>
<td>0.01</td>
<td>3447</td>
<td>1.232</td>
<td>0.218</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Grammaticality</td>
<td>0.016</td>
<td>0.04</td>
<td>134</td>
<td>0.369</td>
<td>0.712</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Grammaticality</td>
<td>-0.002</td>
<td>0.04</td>
<td>140</td>
<td>-0.054</td>
<td>0.957</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity</td>
<td>0.021</td>
<td>0.04</td>
<td>220</td>
<td>0.474</td>
<td>0.636</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity</td>
<td>0.021</td>
<td>0.04</td>
<td>231</td>
<td>0.480</td>
<td>0.631</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Grammaticality × Complexity</td>
<td>-0.004</td>
<td>0.02</td>
<td>3538</td>
<td>-0.201</td>
<td>0.841</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>Sentence v. Phrase × Grammaticality × Complexity</td>
<td>-0.002</td>
<td>0.02</td>
<td>3529</td>
<td>-0.096</td>
<td>0.923</td>
</tr>
<tr>
<td>Complexity</td>
<td>Group × Grammaticality × Complexity</td>
<td>0.006</td>
<td>0.02</td>
<td>2968</td>
<td>0.366</td>
<td>0.714</td>
</tr>
<tr>
<td>Complexity</td>
<td>Sentence v. Word × Group × Grammaticality × Complexity</td>
<td>0.013</td>
<td>0.04</td>
<td>3545</td>
<td>0.302</td>
<td>0.762</td>
</tr>
<tr>
<td>Complexity</td>
<td>Sentence v. Phrase × Group × Grammaticality × Complexity</td>
<td>0.005</td>
<td>0.05</td>
<td>3545</td>
<td>0.116</td>
<td>0.908</td>
</tr>
</tbody>
</table>

Sentence v. Word = Comparison of Baseline (Sentence) to Word, Sentence v. Phrase = Comparison of Baseline (Sentence) to Phrase; Random effects structure included intercept and effects of Grammaticality and Complexity for Subjects plus intercept and effects of Format, Group, Grammaticality, and Complexity for Items. R-code used with lmerTest: Log.RatingRT ~ Format * Group * Grammaticality * Complexity + (1 + Grammaticality + Complexity | Subject) + (1 + Format + Group + Grammaticality + Complexity | Item). Degrees of freedom (df) calculated using the Satterthwaite approximation.

5.4.4.2 Comprehension probe RTs

The comprehension RT data model included fixed effects of format, group, complexity, and comprehension target, as well as all interactions—see Table 14. Overall, L2 participants were faster than L1 participants in responding to the comprehension probe. Response times were also faster for simple constructions than for complex constructions. While overall response times were only marginally faster in the Phrase condition, there was a significant interaction between
format and group, such that L2 participants responded more quickly to the comprehension probe in the PHRASE condition only.

**Table 14.** Linear mixed-effects model of overall comprehension probe response times. Effects of format, group, complexity, and target.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.620</td>
<td>0.03</td>
<td>130</td>
<td>281.526</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word</td>
<td>0.115</td>
<td>0.07</td>
<td>87</td>
<td>1.761</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase</td>
<td>-0.117</td>
<td>0.07</td>
<td>88</td>
<td>-1.786</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.127</td>
<td>0.05</td>
<td>87</td>
<td>2.748</td>
<td>&lt;0.01*</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.017</td>
<td>0.01</td>
<td>77</td>
<td>2.534</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>-0.005</td>
<td>0.02</td>
<td>62</td>
<td>-0.339</td>
<td>0.735</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group</td>
<td>0.077</td>
<td>0.01</td>
<td>86</td>
<td>0.590</td>
<td>0.557</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group</td>
<td>-0.030</td>
<td>0.01</td>
<td>87</td>
<td>-2.304</td>
<td>&lt;0.05*</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity</td>
<td>-0.021</td>
<td>0.02</td>
<td>93</td>
<td>-1.163</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity</td>
<td>0.008</td>
<td>0.02</td>
<td>97</td>
<td>0.428</td>
<td>0.670</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>-0.004</td>
<td>0.01</td>
<td>95</td>
<td>-0.273</td>
<td>0.785</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Target</td>
<td>0.007</td>
<td>0.02</td>
<td>98</td>
<td>0.429</td>
<td>0.669</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Target</td>
<td>0.015</td>
<td>0.02</td>
<td>74</td>
<td>0.875</td>
<td>0.384</td>
<td></td>
</tr>
<tr>
<td>Group × Target</td>
<td>-0.017</td>
<td>0.01</td>
<td>81</td>
<td>-1.470</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td>Complexity × Target</td>
<td>-0.006</td>
<td>0.01</td>
<td>119</td>
<td>-0.954</td>
<td>0.342</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity</td>
<td>0.015</td>
<td>0.04</td>
<td>93</td>
<td>0.407</td>
<td>0.685</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity</td>
<td>0.012</td>
<td>0.04</td>
<td>96</td>
<td>0.316</td>
<td>0.752</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Target</td>
<td>0.018</td>
<td>0.03</td>
<td>101</td>
<td>0.593</td>
<td>0.555</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Target</td>
<td>-0.012</td>
<td>0.03</td>
<td>108</td>
<td>-0.371</td>
<td>0.711</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Complexity × Target</td>
<td>-0.006</td>
<td>0.02</td>
<td>3626</td>
<td>-0.400</td>
<td>0.689</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Complexity × Target</td>
<td>0.003</td>
<td>0.02</td>
<td>3641</td>
<td>0.208</td>
<td>0.835</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × Target</td>
<td>-0.003</td>
<td>0.01</td>
<td>3636</td>
<td>-0.255</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Word × Group × Complexity × Target</td>
<td>0.009</td>
<td>0.03</td>
<td>3621</td>
<td>0.299</td>
<td>0.765</td>
<td></td>
</tr>
<tr>
<td>Sentence v. Phrase × Group × Complexity × Target</td>
<td>-0.029</td>
<td>0.03</td>
<td>3657</td>
<td>-0.934</td>
<td>0.351</td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sentence v. Word = Comparison of Baseline (SENTENCE) to WORD, Sentence v. Phrase = Comparison of Baseline (SENTENCE) to PHRASE; Random effects structure included intercept and effects of Complexity and Target for Subjects plus intercept and effects of Format, Group, and Complexity for Items. R-code used with lmer: Log.Comprehension RT ~ Format * Group * Complexity * Target + (1 + Complexity + Target | Subject) + (1 + Format + Group + Complexity | Item). Degrees of freedom (df) calculated using the Satterthwaite approximation.

**5.4.4.3 PHRASE format: Segment 1 RTs**

**PHRASE presentation: Segment 1**

<table>
<thead>
<tr>
<th></th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>The reporter who called the senators every so often</td>
<td>write/s awful stories for the newspaper.</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>The reporter who called the senators that Scott supported</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Segment 1 RT model included fixed effects of group, grammaticality, complexity, and English reading proficiency, as well as all interactions. See Table 15. There were no significant main effects of grammaticality or complexity. The only significant result was an interaction between group and WJ proficiency, indicating that reading times were faster for L2 participants with higher English reading proficiency. No reading time differences were found for the L1 participants.

Table 15. Linear mixed-effects model of Segment 1 reading times in the PHRASE condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>6.834</td>
<td>0.11</td>
<td>29.2</td>
<td>63.503</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>0.082</td>
<td>0.22</td>
<td>29.1</td>
<td>0.383</td>
<td>0.704</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>-0.002</td>
<td>0.02</td>
<td>147.5</td>
<td>-0.145</td>
<td>0.885</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.013</td>
<td>0.02</td>
<td>200.8</td>
<td>0.849</td>
<td>0.397</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>-0.060</td>
<td>0.03</td>
<td>28.9</td>
<td>-1.975</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.006</td>
<td>0.03</td>
<td>355.4</td>
<td>0.202</td>
<td>0.840</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>0.023</td>
<td>0.03</td>
<td>407.5</td>
<td>0.701</td>
<td>0.484</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>-0.019</td>
<td>0.02</td>
<td>1044</td>
<td>-1.206</td>
<td>0.228</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>-0.131</td>
<td>0.06</td>
<td>28.9</td>
<td>-2.144</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>0.005</td>
<td>0.01</td>
<td>296.6</td>
<td>1.086</td>
<td>0.278</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>-0.007</td>
<td>0.01</td>
<td>361.7</td>
<td>-1.695</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>-0.031</td>
<td>0.03</td>
<td>1080</td>
<td>-0.968</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>0.003</td>
<td>0.01</td>
<td>243.3</td>
<td>0.347</td>
<td>0.729</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>0.001</td>
<td>0.01</td>
<td>366.7</td>
<td>0.151</td>
<td>0.880</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>-0.002</td>
<td>0.01</td>
<td>1220</td>
<td>-0.495</td>
<td>0.621</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity × WJ Proficiency</td>
<td>-0.001</td>
<td>0.01</td>
<td>930.7</td>
<td>-0.093</td>
<td>0.926</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.4.4.4  **Phrase format**: Segment 2 RTs

**Phrase presentation**: Segment 2

<table>
<thead>
<tr>
<th></th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simple</strong></td>
<td>The reporter</td>
<td>who called the senators every so often</td>
<td>write/s awful stories for the newspaper.</td>
</tr>
<tr>
<td><strong>Complex</strong></td>
<td>The reporter</td>
<td>who called the senators that Scott supported</td>
<td>write/s awful stories for the newspaper.</td>
</tr>
</tbody>
</table>

The Segment 2 RT data model included fixed effects of group, grammaticality, complexity, and English reading proficiency, as well as all interactions. See Table 16 for summary. Overall, reading times were faster for the L2 participants than the L1 participants. A group by WJ proficiency interaction revealed that, similarly to the Segment 1 RT data, reading times were faster for L2 participants with higher English reading proficiency. No other results were significant.

**Table 16.** Linear mixed-effects model of Segment 2 reading times in the Phrase condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.662</td>
<td>0.07</td>
<td>29.4</td>
<td>109.153</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>-0.314</td>
<td>0.14</td>
<td>29.3</td>
<td>-2.236</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Grammaticality</td>
<td>0.003</td>
<td>0.01</td>
<td>94.3</td>
<td>0.267</td>
<td>0.790</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.013</td>
<td>0.01</td>
<td>43</td>
<td>1.163</td>
<td>0.251</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>-0.037</td>
<td>0.02</td>
<td>28.9</td>
<td>-1.864</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.018</td>
<td>0.02</td>
<td>171.4</td>
<td>0.859</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>-0.009</td>
<td>0.02</td>
<td>42.3</td>
<td>-0.424</td>
<td>0.674</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>-0.019</td>
<td>0.01</td>
<td>1217</td>
<td>-1.908</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>-0.088</td>
<td>0.04</td>
<td>28.8</td>
<td>-2.206</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>0.003</td>
<td>0.01</td>
<td>135.9</td>
<td>1.154</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>0.002</td>
<td>0.01</td>
<td>34.6</td>
<td>0.524</td>
<td>0.604</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>-0.018</td>
<td>0.10</td>
<td>991.7</td>
<td>-0.891</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>-0.008</td>
<td>0.01</td>
<td>133.6</td>
<td>-1.325</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>-0.002</td>
<td>0.01</td>
<td>35.3</td>
<td>-0.367</td>
<td>0.716</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.00</td>
<td>0.01</td>
<td>1035</td>
<td>0.074</td>
<td>0.941</td>
<td></td>
</tr>
</tbody>
</table>

5.4.4.5 Phrase format: Segment 3 RTs

Phrase presentation: Segment 3

<table>
<thead>
<tr>
<th></th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>The reporter</td>
<td>who called the senators every so often</td>
<td>write/s awful stories for the newspaper.</td>
</tr>
<tr>
<td>Complex</td>
<td>The reporter who called the senators that Scott supported</td>
<td></td>
<td>write/s awful stories for the newspaper.</td>
</tr>
</tbody>
</table>

The Segment 3 RT data model included fixed effects of group, grammaticality, complexity, and English reading proficiency, as well as all interactions. However, no effects were found to be significant. See Table 17.

Table 17. Linear mixed-effects model of Segment 3 reading times in the Phrase condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.596</td>
<td>0.08</td>
<td>29.9</td>
<td>98.579</td>
<td>&lt;0.001  *</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>−0.182</td>
<td>0.15</td>
<td>29.5</td>
<td>−1.186</td>
<td>0.245</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>0.001</td>
<td>0.01</td>
<td>96.2</td>
<td>0.045</td>
<td>0.964</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.006</td>
<td>0.01</td>
<td>481.8</td>
<td>0.579</td>
<td>0.563</td>
<td></td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>−0.034</td>
<td>0.02</td>
<td>28.9</td>
<td>−1.582</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.015</td>
<td>0.02</td>
<td>182.9</td>
<td>0.654</td>
<td>0.514</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>0.003</td>
<td>0.02</td>
<td>912.6</td>
<td>0.135</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>−0.004</td>
<td>0.01</td>
<td>1196</td>
<td>−0.346</td>
<td>0.730</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>−0.059</td>
<td>0.04</td>
<td>28.9</td>
<td>−1.358</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>0.002</td>
<td>0.01</td>
<td>143.9</td>
<td>0.829</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>−0.001</td>
<td>0.01</td>
<td>992.0</td>
<td>−0.196</td>
<td>0.845</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>−0.011</td>
<td>0.02</td>
<td>972.1</td>
<td>−0.517</td>
<td>0.605</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>0.006</td>
<td>0.01</td>
<td>143.8</td>
<td>1.052</td>
<td>0.294</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>−0.001</td>
<td>0.01</td>
<td>847.9</td>
<td>−0.026</td>
<td>0.979</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>−0.004</td>
<td>0.01</td>
<td>954.5</td>
<td>−1.505</td>
<td>0.133</td>
<td></td>
</tr>
</tbody>
</table>


5.4.4.6 Sentence format: Sentence RTs

Finally, the Sentence RT data model included fixed effects of group, grammaticality, complexity, and English reading proficiency, as well as all interactions. See Table 18. As anticipated, simple
constructions were read more quickly than complex constructions. There was also a significant interaction between group and WJ proficiency, indicating that for L1 participants, higher English proficiency was associated with slower reading times, but for L2 participants was associated with faster reading times.

Table 18. Linear mixed-effects model of reading times in the Sentence condition. Effects of group, grammaticality, complexity, and English reading proficiency.

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Estimate</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>8.945</td>
<td>0.09</td>
<td>29</td>
<td>102.04</td>
<td>&lt;0.001</td>
<td>*</td>
</tr>
<tr>
<td>Group</td>
<td>0.249</td>
<td>0.18</td>
<td>29</td>
<td>1.420</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td>Grammaticality</td>
<td>0.010</td>
<td>0.01</td>
<td>80</td>
<td>0.807</td>
<td>0.422</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>0.003</td>
<td>0.01</td>
<td>43.6</td>
<td>2.346</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>WJ Proficiency</td>
<td>0.001</td>
<td>0.02</td>
<td>29</td>
<td>0.020</td>
<td>0.984</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality</td>
<td>0.012</td>
<td>0.02</td>
<td>93.3</td>
<td>0.570</td>
<td>0.570</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity</td>
<td>−0.031</td>
<td>0.02</td>
<td>41.6</td>
<td>−1.261</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity</td>
<td>−0.008</td>
<td>0.01</td>
<td>1059</td>
<td>−0.760</td>
<td>0.447</td>
<td></td>
</tr>
<tr>
<td>Group × WJ Proficiency</td>
<td>−0.126</td>
<td>0.05</td>
<td>28.9</td>
<td>−2.591</td>
<td>&lt;0.05</td>
<td>*</td>
</tr>
<tr>
<td>Grammaticality × WJ Proficiency</td>
<td>0.001</td>
<td>0.01</td>
<td>93.2</td>
<td>0.403</td>
<td>0.688</td>
<td></td>
</tr>
<tr>
<td>Complexity × WJ Proficiency</td>
<td>−0.002</td>
<td>0.01</td>
<td>39.8</td>
<td>−0.547</td>
<td>0.587</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity</td>
<td>−0.002</td>
<td>0.02</td>
<td>993.5</td>
<td>−0.095</td>
<td>0.925</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × WJ Proficiency</td>
<td>−0.001</td>
<td>0.01</td>
<td>92.6</td>
<td>−0.091</td>
<td>0.928</td>
<td></td>
</tr>
<tr>
<td>Group × Complexity × WJ Proficiency</td>
<td>0.002</td>
<td>0.01</td>
<td>40.5</td>
<td>0.241</td>
<td>0.811</td>
<td></td>
</tr>
<tr>
<td>Grammaticality × Complexity × WJ Proficiency</td>
<td>0.002</td>
<td>0.01</td>
<td>1032</td>
<td>0.517</td>
<td>0.605</td>
<td></td>
</tr>
<tr>
<td>Group × Grammaticality × Complexity × WJ Proficiency</td>
<td>−0.005</td>
<td>0.01</td>
<td>1051</td>
<td>−0.768</td>
<td>0.443</td>
<td></td>
</tr>
</tbody>
</table>


5.4.5 Proficiency correlations

A correlation series was performed to determine the relationship between the standardized proficiency measures (Woodcock-Johnson and Woodcock-Muñoz), and the self-reported proficiency ratings. Results of the correlations are reported in Table 19.
Table 19. Correlation table for Woodcock-Johnson/Woodcock-Muñoz passage comprehension and self-rated proficiency (English and Spanish)

<table>
<thead>
<tr>
<th></th>
<th>Woodcock-Johnson</th>
<th>Woodcock-Muñoz</th>
<th>EN Speaking</th>
<th>EN Comp</th>
<th>EN Reading</th>
<th>SP Speaking</th>
<th>SP Comp</th>
<th>SP Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Corr</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Woodcock-Johnson</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>87</td>
<td>24</td>
<td>87</td>
<td>87</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>Woodcock-Muñoz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>EN Speaking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>EN Comp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>EN Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>SP Speaking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>SP Comp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>SP Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Corr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

For the L2 participants, Woodcock-Johnson scores were significantly correlated with Woodcock-Muñoz scores. Interestingly however, Woodcock-Johnson scores were not correlated with any of the self-reported English proficiency measures. On the other hand, Woodcock-Muñoz scores were found to correlate with Spanish speaking and Spanish reading proficiency ratings. No other significant correlations were of interest.

5.4.6 Tradeoff effects

Finally, to examine the possibility of tradeoff effects in processing for comprehension versus grammaticality, grammaticality ratings in the SENTENCE condition were recoded as binary grammatical accuracy responses (>3 = ‘grammatical’, <4 = ‘ungrammatical’), and Pearson
correlations were performed by item type to assess the relationship between grammatical rating accuracy and comprehension accuracy across both L1 and L2 participants. For the simple items, there was no significant correlation between the two measures ($r = 0.170, n = 29, p = 378$). A scatterplot of the results is given in Figure 11, and accuracy measures are plotted together by individual in Figure 12. For the complex items however, there was a negative correlation between grammatical rating and comprehension accuracy ($r = −0.412, n = 29, p = 0.01$), such that higher comprehension accuracy was associated with lower grammatical rating accuracy. A scatterplot of the results is given in Figure 13, and accuracy measures are plotted together by individual in Figure 14. For comparison purposes, a correlation analysis was performed for the filler items as well. A positive correlation was found in the filler items, such that higher comprehension accuracy was associated with higher accuracy in grammaticality ratings ($r = 0.337, n = 29, p = 0.037$). A scatterplot of the results is given in Figure 15 and accuracy measures are plotted together by individual in Figure 16.
**Figure 11.** Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy. Simple experimental items only, SENTENCE condition.

**Figure 12.** Average comprehension accuracy and grammatical rating accuracy by individual. Simple experimental items only, SENTENCE condition.
**Figure 13.** Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy, Complex experimental items only, SENTENCE condition.

**Figure 14.** Average comprehension accuracy and grammatical rating accuracy by individual. Complex experimental items only, SENTENCE condition.
**Figure 15.** Scatterplot illustrating relationship between comprehension accuracy and grammatical rating accuracy, Filler items only, SENTENCE condition.

**Figure 16.** Average comprehension accuracy and grammatical rating accuracy by individual. Filler items only, SENTENCE condition.
5.4.7 Summary of results

5.4.7.1 Grammaticality

The taskload incurred by processing the experimental items made it difficult for all participants to appropriately distinguish grammatical from ungrammatical constructions. While this was true for both L1 and L2 participants in the baseline SENTENCE condition, L1 participants were best able to reliably distinguish grammatical and ungrammatical items presented in the PHRASE format. The PHRASE format also resulted in greater acceptability for the typically disliked complex constructions in both groups, resulting in higher ratings overall.

5.4.7.2 Comprehension

Comprehension accuracy was higher for simple than complex items, and higher when the probes targeted general comprehension, rather than the relative clause interpretation. Comprehension accuracy was higher overall for L1 participants than L2 participants; however, the groups were differently affected by presentation format. L1 participants were significantly disrupted by the WORD format presentation, particularly when the comprehension probe targeted the relative clause interpretation. On the other hand, L2 participants were significantly disrupted by the PHRASE format presentation, again, most strongly when the comprehension probe targeted the relative clause.

5.4.7.3 Response times (RTs)

Response times for grammaticality ratings were fastest for the WORD format overall, however, since the rating prompt immediately followed the items in all conditions, this effect could be due to the additional processing time allowed by the slower presentation rate.
The L2 participants had faster reading times in the PHRASE condition for Segment 2, which contained the critical relative clause structure. And interestingly, L2 participants’ reading times decreased at higher proficiency levels for Segments 1 and 2 of the PHRASE condition. This effect was found in the SENTENCE condition reading times as well. However, this proficiency effect did not translate into similar improvements in grammatical processing, since the L2 participants did not rate grammatical sentences differently from ungrammatical sentences in either format.

5.4.7.4 Tradeoff effects

Finally, there was evidence of tradeoff effects in comprehension versus grammatical processing, such that for both L1 and L2 participants in the SENTENCE condition, higher comprehension accuracy was correlated with lower grammatical rating accuracy in the complex items, but not in the simple items.
6. Discussion

This study investigated how text presentation format may affect readers’ comprehension and ability to detect subject-verb agreement errors in both simple and complex relative clause constructions. The overarching goal has been to integrate both implicit prosody and cue-based memory retrieval into a framework such as the ‘Good-Enough’ Hypothesis, and investigate how text presentation format interacts with algorithmic versus heuristic processing strategies in comprehension and grammatical measures. To substantiate this proposal, I have provided support for the following points:

i. a cue-based retrieval mechanism at work in agreement

ii. processing strategies that align with the ‘Good-Enough’ Hypothesis model

iii. an effective role of prosody in parsing, specifically in reading

In the sections below, I introduce the basic framework of a ‘Good-enough Cue’ model, which incorporates cue-based retrieval into the ‘good-enough’ processing schema, then discuss and interpret the results of the study with a view to these issues. Some implications of the results in relation to agreement, implicit prosody, and L1/L2 processing will also be discussed.

6.1 Elaboration of the Good-enough Cue (GC) integrated model

6.1.1 Overview

The ‘Good-Enough’ Hypothesis identifies two processing routes: a full syntactic route, and a ‘good-enough’ heuristic route. Elaborating on this framework, I propose the ‘Good-enough Cue’ (GC) model, an integrated processing paradigm described in more detail below.
The full syntactic processing route involves computation of all syntactic aspects including agreement and interpretation of structurally based relations. However, as long as the goal of a task does not require this level of detailed processing, the parser prefers to conserve resources by using heuristics such as semantic and pragmatics, as well as thematic templates and plausibility, to maintain a generally sufficient level of comprehension. However, if it becomes necessary to calculate structure or featural information more completely, the full computational parser may be deployed. In the case of agreement in particular, I propose this takes the form of a cue-based retrieval mechanism. High taskload—as created via item/syntactic complexity or processing under noise or other duress, can reduce the ability to fully compute structure. Under these circumstances, the parser would then be forced to fall back on the good-enough strategy in order to preserve general message comprehension.

I propose that one of the functions of prosody (or implicit prosody during reading) is to facilitate processing and reduce task demand. This may be partially due to the relation between syntactic and prosodic structure, and partially as a memory aid—by way of prosodic contour, and by creating phrasal ‘edges’, which may strengthen the ability to retrieve elements associated with these edges. The fallout of this would be that under circumstances that normally would derail the full computational route, the application of prosodic phrasing (or other relevant prosodic features) would reduce the necessity of resorting to a ‘good-enough’ strategy.

A visualization of a comprehensive processing model is given in Figure 17, which is adapted from the auditory language processing model outlined in Friederici (2011), and will serve as the framework for the GC model. Friederici’s original three-phase model aligned functional processes with data from ERP and fMRI studies to outline the neural basis and time course of major linguistic processing features. The first phase involves the parallel processing of
local syntactic structure and prosodic phrasing. At this stage, syntactic word categories are identified, and the initial local phrase structure is built. These effects correspond to the early left anterior negativity (ELAN) ERP component. Prosodic phrasing features are also processed early, based on ERP evidence of a closure positive shift (CPS) associated with intonational phrase boundaries. During this initial phase, prosody and syntax may interact, since prosodic cues can be used to mark syntactic phrase boundaries, and identify syntactic constituents.

The second phase corresponds to the computation of syntactic and semantic relations, as reflected in LAN and N400 effects, respectively. Semantic features such as animacy, or selectional restrictions on a verb are incorporated here, as well as grammatical processes such as subject-verb agreement and Case marking. It is likely during this phase that structural complexity effects would be observed, since the early phrase structure building processes appear to be less affected by task demand (Hahne & Friederici, 2002). Input from declarative knowledge stores may also apply at this stage, which is associated with not only semantic information, but thematic relations as well. The third phase is the point of integration of syntactic and semantic information (corresponding to P600 effects), as well as the processing of prosodic features related to information structure, such as pitch accents and focus intonation. The multiple streams of information are then pulled together to form the final interpretation.
Using this model as a framework, in the following sections, I discuss the main features of the GC model in greater detail: cue-based retrieval, the ‘good-enough’ processing route selection, and the role of prosody in performance.

### 6.1.2 Cue-based retrieval mechanism

While agreement has been approached from many linguistic standpoints, recent arguments have contended that the memory architecture for its computation is similar to that of general memory—specifically a cue-based, content-addressable system (Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003; Wagers et al., 2009; Wagers, 2008). This cue-based memory retrieval (CBMR) approach works within syntactic phrase structure configurations, but does not order its feature-match search by hierarchical relations. Instead, it uses a probe-goal search to specifically isolate constituents with a particular set of features. This allows for the rapid resolution of dependencies needed for online processing (McElree et al., 2003), but also accommodates error patterns found in production and comprehension—where, based on feature overlap, multiple candidate constituents may be retrieved, even though not all are grammatically licensed.
Evidence from speed-accuracy tradeoff (SAT) measures suggests that adjacent subjects and verbs can occupy the focus of attention simultaneously (McElree et al., 2003), and retrieval may not typically even be necessary. In fact, it is more likely that the CBMR mechanism functions primarily as a reanalysis tool—i.e., deployed only in the case of a mismatch of the verb with the predicted form. This is supported in part by findings that attraction errors seem to occur only in ungrammatical sentences (Wagers et al., 2009; Wagers, 2008), and is particularly practical considering the syncretic agreement patterns of English lexical verbs, which would be too often uninformative in identifying the appropriate controller.

Solomon and Pearlmutter (2004) consider semantic integration, or relatedness of the head and attractor nouns, to be the primary determiner of agreement attraction. In their model, this translates into both nouns being simultaneously active at the time of verb selection or checking, which increases the chances of the wrong noun being identified as the controller. There is strong evidence of the role of semantic relations in agreement; however, the semantic relatedness effect can be accommodated within the CBMR mechanism in that both positional and featural similarity between two elements can increase the likelihood of interference.

While several instantiations of cue-based retrieval models have been proposed, there is no general consensus on which may best be applied to most configurations, and I will not attempt to make that determination here. However, I develop a basic framework that incorporates formulae from Wagers (2008) and Lewis and Vasishth (2005), who have approached this issue for similar constructions. For the sake of illustration, I use the sample items given in (72) and (73) to demonstrate how such a mechanism may be applied.
(72) Simple relative clause construction:

\[
\begin{array}{c}
| \text{500 ms} | [\text{IP} [\text{NP } \text{The reporter}] [\text{CP } \text{who} [\text{IP } \text{t} [\text{VP } \text{t} \text{ called} [\text{NP} \text{ the senators}] [\text{AP} \text{ every so often}]]]]] [\text{VP } \text{t} \text{ writes...}]]] \\
| \text{3000 ms} |
\end{array}
\]

(73) Complex relative clause construction:

\[
\begin{array}{c}
| \text{500 ms} | [\text{IP} [\text{NP } \text{The reporter}] [\text{CP } \text{who} [\text{IP } \text{t} [\text{VP } \text{t} \text{ called} [\text{NP} \text{ the senators}] [\text{CP} \text{ that} [\text{IP} \text{ Scott} [\text{VP} \text{ supported } \text{t}]]]]]]] [\text{VP } \text{t} \text{ writes...}]]] \\
| \text{3000 ms} |
\end{array}
\]

As a first step, I adopt a cue-based retrieval frame derived from Wagers (2008), where the goal (in this case, the verb) is marked with a set of features that are used to identify an appropriate controller. These features may include relevant cues such as Case specification, structural position or clause identifier, morphological number, or perhaps semantic restrictions such as animacy. In the basic model given here, each of the relevant cues is weighted equally, and NPs are categorized as either a match (1 – 0.01 noise factor = 0.99) or a mismatch (0 + 0.01 noise factor = 0.01) to each of the cues (See Table 17).

**Table 17.** Feature cue weights and associative strengths for grammatical and ungrammatical items

<table>
<thead>
<tr>
<th>Goal cue</th>
<th>Grammatical – ‘writes’</th>
<th>Ungrammatical – ‘write’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cue Weight NP1 NP2</td>
<td>Cue Weight NP1 NP2</td>
</tr>
<tr>
<td>Case: NOM</td>
<td>0.25 0.99 0.01</td>
<td>0.20 0.99 0.01</td>
</tr>
<tr>
<td>Spec, TP</td>
<td>0.25 0.99 0.01</td>
<td>0.20 0.99 0.01</td>
</tr>
<tr>
<td>Clause Subject</td>
<td>0.25 0.99 0.01</td>
<td>0.20 0.99 0.01</td>
</tr>
<tr>
<td>Num: PL</td>
<td>-- -- --</td>
<td>0.20 0.01 0.99</td>
</tr>
<tr>
<td>Animate</td>
<td>0.25 0.99 0.99</td>
<td>0.20 0.99 0.99</td>
</tr>
</tbody>
</table>

Multiplying match with cue weight generates a cue match level for each of the features and controller candidates, which are then summed to find the total cue match (See Tables 18 and 19). Factoring in the effect of decay over time generates a measure of controller strength. The controller strength equation used here is adapted from the log odds of retrieval formula in Lewis...
and Vasishth (2005), where $d$ is the decay rate (default value of 0.5) and $t$ is the time elapsed since the last retrieval (in ms), summed for the total number of retrievals (see (74)). While not specified in either of the models I draw from, I will assume that retrievals occur at trace positions (with retrievals being triggered at or following the word adjacent to the trace), but that the effects of retrieving immediately adjacent traces are not additive. Also, for simplicity in this example, I will assume a 500 ms per word input rate to be used in the calculation. Following this schema, although overall time elapsed since initial presentation for NP1 and NP2 is the same in both simple and complex items (3500 ms for NP1, 2000 ms for NP2), the calculated strength level still varies based on number of retrievals.

(74) Calculation of controller strength

\[ \sum_{j=1}^{n} t_j^{-d} \]

a. Simple, NP1 ‘the reporter’: $S_i = (500^{-0.5}) + (3000^{-0.5}) = 0.063$

b. Simple, NP2 ‘the senators’: $S_j = (2000^{-0.5}) = 0.022$

c. Complex, NP1 ‘the reporter’: $S_i = (500^{-0.5}) + (3000^{-0.5}) = 0.063$

d. Complex, NP2 ‘the senators’: $S_j = (1500^{-0.5}) + (500^{-0.5}) = 0.044$

Table 18. Simple RCs: Probability of NP1 vs. NP2 retrieval in grammatical and ungrammatical items

<table>
<thead>
<tr>
<th>Goal cue</th>
<th>Grammatical – ‘writes’</th>
<th>Ungrammatical – ‘write’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP1</td>
<td>NP2</td>
</tr>
<tr>
<td>Case: NOM</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Spec, TP</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Clause Subject</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Num: --</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Animate</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td><strong>Total cue match:</strong></td>
<td><strong>0.992</strong></td>
<td><strong>0.256</strong></td>
</tr>
<tr>
<td>Strength</td>
<td>0.063</td>
<td>0.022</td>
</tr>
<tr>
<td>Activation</td>
<td>0.062</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Probability</strong></td>
<td><strong>0.912</strong></td>
<td><strong>0.088</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goal cue</th>
<th>Grammatical – ‘writes’</th>
<th>Ungrammatical – ‘write’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP1</td>
<td>NP2</td>
</tr>
<tr>
<td>Case: NOM</td>
<td>0.198</td>
<td>0.002</td>
</tr>
<tr>
<td>Spec, TP</td>
<td>0.198</td>
<td>0.002</td>
</tr>
<tr>
<td>Clause Subject</td>
<td>0.198</td>
<td>0.002</td>
</tr>
<tr>
<td>Num: PL</td>
<td>0.002</td>
<td>0.198</td>
</tr>
<tr>
<td>Animate</td>
<td>0.198</td>
<td>0.198</td>
</tr>
<tr>
<td><strong>Total cue match:</strong></td>
<td><strong>0.794</strong></td>
<td><strong>0.402</strong></td>
</tr>
<tr>
<td>Strength</td>
<td>0.063</td>
<td>0.022</td>
</tr>
<tr>
<td>Activation</td>
<td>0.050</td>
<td>0.009</td>
</tr>
<tr>
<td><strong>Probability</strong></td>
<td><strong>0.847</strong></td>
<td><strong>0.153</strong></td>
</tr>
</tbody>
</table>
Table 19. Complex RCs: Probability of NP1 vs. NP2 retrieval in grammatical and ungrammatical items

<table>
<thead>
<tr>
<th>Goal cue</th>
<th>Grammatical – ‘writes’</th>
<th>Ungrammatical – ‘write’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP1</td>
<td>NP2</td>
</tr>
<tr>
<td>Case: NOM</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Spec, TP</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Clause Subject</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td>Num: --</td>
<td>--</td>
<td>0.0025</td>
</tr>
<tr>
<td>Animate</td>
<td>0.248</td>
<td>0.0025</td>
</tr>
<tr>
<td><strong>Total cue match:</strong></td>
<td><strong>0.992</strong></td>
<td><strong>0.256</strong></td>
</tr>
<tr>
<td>Strength</td>
<td>0.063</td>
<td>0.044</td>
</tr>
<tr>
<td>Activation</td>
<td>0.062</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Probability</strong></td>
<td><strong>0.849</strong></td>
<td><strong>0.151</strong></td>
</tr>
</tbody>
</table>

Activation is calculated by multiplying item strength level by total cue match, which can then be used to determine the probability of retrieval. In this example, while erroneous retrieval of NP2 ‘the senators’ as a controller is unlikely in the grammatical condition of the simple construction (probability of 0.088), it is much more likely in the ungrammatical condition, where the number feature match increases the probability of retrieval (probability of 0.153). The probability of erroneous retrieval increases in the complex construction, with a probability of 0.151 in the grammatical condition, and 0.254 in the ungrammatical condition.

This model is relatively underspecified—other feature cues or factors may be added to the schema, and the relative cue weights may be adjusted as appropriate. However, it is clearly able to capture the general error patterns discussed within the agreement attraction literature, and can be easily modified to accommodate more specific feature sets and configurations.

6.1.3 Good-enough processing strategy

The ‘Good-Enough’ Hypothesis is fundamentally an efficiency-based model which claims that although a complete parse of the input is possible, it is usually unnecessary. Comprehenders tend to favor heuristic strategies when constructing a representation, unless the processing task
requires more complete parsing. There are specific consequences to applying this strategy in certain parsing conditions; however, the general outcome is realized as a prioritization of overall message comprehension over strict grammatical or syntactic processing.

The model presents this strategy as one that is initiated and implemented by the parser in order to conserve processing resources (or perhaps by an executive component that manages and coordinates the parser with regard to memory and taskload). However, it stands to reason that external influences such as a difficult parse may also force the parser to take the heuristic route, since the maintenance of general comprehension requires more resources than otherwise expected.

6.1.4 The role of prosody

Prosody has been shown to contribute to multiple components of auditory processing: beginning with segmentation of the speech stream, and on to disambiguation of syntactic structures and facilitation of working memory and memory for recall. The implicit prosody projected during reading has also been demonstrated to affect syntactic processes such as agreement, where the segmentation of text into non-constituent chunks has been shown to reduce readers’ ability to detect subject-verb agreement errors (Kreiner, 2005). The assumption is that readers typically project prosody onto text (Implicit Prosody Hypothesis), which may allow them to better track dependency relations such as agreement, particularly across long distances. This effect may be due to alignment with syntactic constituents, which eases structurally based computations, and/or to facilitation of processing by scaffolding memory.

In addition to syntactic processes of agreement, even earlier work has linked facilitation and disruption of implicit prosody with general comprehension. Cromer (1970) found that segmentation of text influences readers’ comprehension, albeit differently based on their level of
reading proficiency. While ‘good’ readers (high comprehension and vocabulary scores) were generally resistant to the influence of text manipulation, ‘poor’ readers were affected by variations in presentation format. ‘Difference’ readers, who scored high on vocabulary measures but not comprehension, performed best when text was presented in phrases, but were disrupted when text was presented word-by-word, or in non-phrasal fragments. ‘Deficit’ readers, who scored low on both vocabulary and comprehension measures, performed best when text was presented word-by-word, but were not affected by other presentation formats.

The integration of prosody, and prosodic phrasing in particular, into a model utilizing cue-based retrieval has not yet been addressed, although it is well-motivated. In speech production, clauses typically function as planning units, and pauses tend to occur at clause boundaries (Butterworth, 1980; Garrett, 1982; McDaniel, McKee, & Garrett, 2010). This is reasonable, given that each clause typically carries its own verb or tense features, along with other relevant information structure, and the parser must be able to track relationships within the context of the clause. In the case of a cue-based retrieval model, this may translate into a clause-identifying cue—where the feature set on the verb specifies that it expects a controller from the same clause.

In addition, items at clausal edges may be reinforced in memory, so that a PHRASE presentation format not only pre-segments the relative clause in a way that facilitates parsing by inserting a break after the head noun, but by doing so creates an ‘edge’. This may help to offset the cue decay, increasing the chance of correct cue-matching with the target verb.

6.2 Discussion of results

Expanding on this overview, the next sections apply aspects of the GC model to the current study and discuss the results in the light of that paradigm.
6.2.1 Cue-based memory retrieval in agreement

Following previous studies of this mechanism, this study used subject-verb agreement licensing as a measure of online grammatical processing during reading. While relative clauses are typically considered less susceptible to agreement attraction effects, items were configured to maximize the potential for agreement attraction (singular subject-plural attractor), and vary structural complexity while minimizing sentence length differences that would affect the time course of cue decay (see (75) and (76)).

(75) Simple relative clause construction:
The reporter who called the senators every so often write(s) awful stories for the newspaper.
(76) Complex relative clause construction:
The reporter who called the senators that Scott supported write(s) awful stories for the newspaper.

As discussed in Chapter 2.4, Wagers and colleagues support CBMR over agreement attraction models based on erroneous number representation, due to its ability to account for non-intervening attractor scenarios, as well as the lack of attraction found in grammatical constructions (i.e., grammatical constructions appearing ungrammatical due to the presence of an attractor). While the base outline of the model does not address structural distance or clause-bounding effects, they suggest that structural features may be incorporated as feature-cues—a proposal that is explored further in Franck and Wagers (2015).
This study was not designed to exhaustively test the validity of CBMR, and so the overall difficulty of the experimental items (both simple and complex relative clause constructions) somewhat obscures potential evidence in support of it. However, many others have more thoroughly investigated its application to agreement phenomena, and we can demonstrate that the current results are compatible with a cue-based retrieval mechanism.

For the current data, we might predict that attraction effects would be stronger (i) in complex items than in simple items, and (ii) in the L2 participants than in the L1 participants—if retrieval is influenced by reactivation at (intermediate) trace positions, and if L2 speakers favor a pro-binding strategy instead of movement (Dekydtspotter & Sprouse, 2003).

Regarding the first prediction, looking solely at activation and cue decay across structure, we may anticipate higher probability of erroneous retrieval (i.e., fewer ungrammatical ratings) in complex items than in simple items. Overall, neither L1 nor L2 participants were able to reliably differentiate between grammatical and ungrammatical constructions in the Sentence condition. However, results were similar for both the simple and complex items, so it is not possible to confirm or invalidate the prediction for the baseline responses. However, in the Phrase condition, in which L1 participants were able to discriminate between grammatical and ungrammatical constructions, an effect of complexity was observed, such that ratings for grammatical and ungrammatical were less distinct for the complex items.

The inability to distinguish between grammatical and ungrammatical constructions in the Sentence condition could reflect an attraction effect, and/or uncertainty about which NP is the correct controller. If this is the case, then the Phrase condition may have allowed the L1 participants to respond with more certainty to the grammatical items—particularly for the simple items—because one of the following is true: (i) alignment of segments with syntactic
constituents directly affects retrieval, either by strengthening the cue(s) for the correct controller, enhancing trace reactivation, or slowing cue decay; (ii) phrasal segmentation otherwise assists memory, which mitigates the drain on cognitive processing resources, allowing retrieval to proceed as usual; (iii) some combination of these two effects.

The data also do not contradict the second prediction that L2 participants will be more susceptible to attraction effects. Not only are L2 participants unable to differentiate grammatical from ungrammatical items in the SENTENCE condition, they are unable to differentiate between the two in any other condition as well. If the L2 participants are encumbered with a greater processing load at the outset, the rate of cue decay prior to onset of the target verb may make retrieval of the appropriate controller extremely difficult, particularly as complexity and subject-verb distance increase.

6.2.2 ‘Good-enough’ processing strategy

The ‘Good-Enough’ Hypothesis is not proposed specifically as an L1 or L2 processing model; however, given its status as a general cognitive strategy, I will assume that it functions similarly in both L1 and L2 populations, although how the application of heuristics and algorithms are balanced may vary due to processing load differences.

Following this assumption, we would expect that under normal conditions, while grammatical computations may take place, comprehension is prioritized over grammatical processing, particularly if that particular computation is not critical to the representation of the parse\textsuperscript{13}. This prioritization would be seen to even greater extent when processing demand is high—whether due to item complexity, or task goal difficulty.

\textsuperscript{13} This is not to say that the parser is not capable of processing the grammatical information, or that it does not actually compute relations such as clause structure or agreement unless necessary. The issue is whether—considering task goal and task complexity—there are enough resources available for
Based on the results, this does seem to be the case. In the Sentence condition, comprehension is relatively high considering the complexity of the materials, but differentiation between grammatical and ungrammatical sentences is low, indicating fewer resources dedicated to grammatical processing. However, text presentation clearly influences performance in both L1 and L2 participants, with Phrase presentation leading to increased differentiation between grammatical and ungrammatical items, and Word presentation leading to decreased comprehension accuracy in the L1 participants, and Phrase presentation leading to decreased comprehension accuracy in the L2 participants. If a process is already prioritized by the parser, it should be the case that very little can be done to improve performance in that area, although it may be possible to disrupt performance in some way. For processes that are not prioritized, manipulations that reduce taskload in some way may result in improvements for those non-prioritized processes. This reasoning fits with the results, providing support for a ‘good-enough’ strategy: because comprehension is prioritized by the parser, none of the presentation formats is able to improve it, but it is disruptable, as seen in both L1 and L2 participants. On the other hand, grammatical processing is not prioritized, and is more susceptible when the parse is complex, or otherwise made difficult. However, factors such as presentation format may allow more resources to be dedicated to this deprioritized process, improving performance. The specific factors that would allow this, and the degree to which they improve performance, may additionally depend on individual differences in processing burden and reading fluency.

There is also evidence of tradeoff effects at the individual level, which supports the claim that grammatical processing may be sacrificed in exchange for maintenance of comprehension. Following the correlational analyses of grammatical rating and comprehension accuracy in the grammatical information to be processed on a level such that behavioral measures can capture the distinction.
SENTENCE condition, performance for comprehension and grammatical rating accuracy measures tracked together for the structurally simpler filler items (Figure 15 and Figure 16). Compare this to performance in the experimental items, particularly the complex relative clause items, where higher comprehension accuracy was correlated with lower accuracy in grammatical ratings (Figure 13 and Figure 14). Performance in the simple relative clause items fell in between these two patterns (Figure 11 and Figure 12) with no correlation between the accuracy measures, suggesting that tradeoff effects become more evident as processing difficulty increases. This pattern further supports a strategy-selection tradeoff effect based on individual processing differences.

6.2.3 The role of prosody during reading

The results of this current study substantiate the findings of previous studies regarding the interaction of text segmentation, prosody, and processing. By examining both comprehension and agreement under the same conditions, we can more confidently claim an effect of prosody (or implicit prosody in particular) on parsing. From the data, we see conclusively that text segmentation affects processing—enhancing grammatical processing in the case of the L1 participants, and either enhancing or disrupting comprehension processing in both L1 and L2 participants. If text segmentation modulates implicit prosody, and if implicit prosody is reflective of explicit prosodic patterns, what is the role of prosody/implicit prosody during reading? Is it beneficial to grammatical processing because it aligns with the syntax, reduces memory load, or eases processing in some other way? When it disrupts processing, why does it do so? Is it because it conflicts with the prosody projected onto the text by the reader?

As discussed above, in the SENTENCE condition, neither the L1 nor L2 participants were able to consistently distinguish grammatical from ungrammatical items, which may be partially
due to an attraction effect, i.e., the illusion of grammaticality based on the presence of a plural intervener. However, even grammatical ratings were low (mean grammatical rating = 3.85/6) in contrast with ratings for fillers (mean grammatical rating = 4.70/6), suggesting that these constructions are generally disliked due to complexity, independent of grammaticality. This dispreference may relate to proficiency, as seen within the L1 participant group, where higher English reading proficiency (as indicated by Woodcock-Johnson passage comprehension scores) was associated with higher ratings. At higher proficiency levels, readers are better able to parse complex materials and thus are less likely to reject them based on difficulty alone.

Results were similar in the WORD condition: although overall ratings were numerically higher than in the SENTENCE condition, this difference was not significant, nor were the ratings for grammatical and ungrammatical items significantly different. However, the performance of the L1 participants in the PHRASE condition suggests that pre-segmentation of the text eases the processing load caused by item complexity. This increases the ability to distinguish grammatical from ungrammatical items, which is demonstrated by correctly accepting the grammatical sentences—an effect found for the L1 participants in both the simple and complex constructions. This effect is also seen with the complex items: despite generally higher ratings for simple items over complex items, the PHRASE format mitigates this dispreference, and acceptability is higher for both L1 and L2 participants.

In the comprehension measure, accuracy was higher for the L1 participants than the L2 participants. Accuracy was also higher for simple items over complex items, and higher for general comprehension probes over relative clause interpretation probes. However, a significant interaction of format and group indicated that while the WORD presentation format was disruptive for L1 participants, the PHRASE presentation format was disruptive for L2 participants.
While it follows prediction that the WORD format may disrupt processing, as it did for the L1 participants, it is not immediately obvious why the PHRASE format would result in lower comprehension accuracy for the L2 participants. Since it is assumed that the alignment of prosodic and syntactic phrases is relatively stable crosslinguistically, we would not necessarily expect group differences to come from prosodic phrasing patterns.

Although performance differences in the L2 participants may be unexpected on the view that prosodic phrasing is similar across languages, the results suggest that the problem is based on processing and taskload, rather than differences in representation. Differences in processing load may be related to the participants’ stage of reading fluency development. While both the L1 and L2 participants were equivalent in reading proficiency (as measured by the Woodcock-Johnson task), this is not necessarily an indicator of equivalence in reading fluency. Following the Verbal Efficiency Theory (Perfetti, 1985, 1988), attainment of reading fluency requires not only automatization of low-level processes such as lexical access and decoding, but of higher-level processes such as phrasing as well. If this is true, it follows that variations in text presentation may then affect readers of varying fluency levels in different ways.

In the results of the current study, the WORD presentation format was disruptive to comprehension for the L1 participants—an effect similar to that found for Cromer’s ‘difference’ readers. For the L2 participants however, comprehension was disrupted in the PHRASE condition. This finding does not correspond with effects for any of the reading fluency levels delineated by Cromer; however, because the L2 participants were not facilitated in the WORD condition (as would be expected for low-level readers who might benefit from assistance during the decoding and lexical access stages of reading), but were not impervious to presentation format altogether (as Cromer’s ‘good’ readers), it seems that these L2 participants may also be classified as
‘difference’ readers. However, due to other factors (e.g., processing load), their reading fluency manifests differently than the L1 participants.

Because RT analyses contained reading times for correct trials only, these results are compatible with the idea that higher proficiency/fluency allows faster, accurate processing, at least with regard to comprehension. In further support of this, for the L1 participants, reading times in the SENTENCE condition increased with proficiency, suggesting that the additional processing time reflects instantiation of the more complete syntactic processing route, which would allow determination of grammaticality, as well as higher accuracy for relative clause comprehension probes, both of which are supported by the data. A systematic measure of reading fluency would be necessary in order to clarify the connection between reading fluency and susceptibility to text manipulation. While the currently available measures of reading fluency are fairly subjective, some quantifiable measures may still shed light on group differences such as these.

6.2.3.1 L2 proficiency and use of prosody

For native speakers of a language, the clause typically operates as a planning unit in speech, and pauses tend to occur at clause boundaries (Butterworth 1980; Garrett 1982; McDaniel, McKee, & Garrett 2010). However, for low-fluency L2 learners, there is no evidence of the clause as a planning unit (Temple 2005). Pauses are distributed across the utterance, and there is less hesitation at clause boundaries than in native speech. However, as the development of fluent L2 speech progresses, pauses at clausal boundaries increase, and may begin to converge with native speech patterns. Thus, despite crosslinguistic similarities in prosodic and syntactic phrase alignment, something specific in the development of L2 fluency is needed before native-like
pausing patterns can emerge. If implicit prosody is the projection of prosodic contours onto text, it would follow that similar effects of fluency would be found for reading as well.

Liljestrand Fultz (2009) examined the effect of prosodic phrasing on L2 ambiguity resolution, and found differences based on syntactic structure and complexity. While the L2 learners consistently used prosodic information to disambiguate conjunct modification and PP-attachment constructions, no clear preference was found for relative clause ambiguities. She suggests three possible explanations for this result: (i) learners cannot perceive the prosodic cues for certain structures, and so the parser cannot incorporate this information during processing, (ii) learners can perceive the prosodic cues, but the parser cannot incorporate the information for independent reasons, or (iii) learners perceive and incorporate the cues, but for certain structures are not able to use that information in conjunction with other information from the parse. The first two possibilities do not fit her results: the L2 learners were able to both perceive and incorporate prosodic cues for conjuncts and PP constructions. This leaves the third possibility: that learners do perceive and use prosodic cues, but the parser may not be able to integrate that information in complex computations required by relative clauses. At earlier stages of proficiency, learners may be able to effectively integrate all information when parsing a simpler structure, but not a complex one. As proficiency increases and processing routines develop, the parser is able to more efficiently integrate information from multiple cues, even when the computation is complex.

Interpreting the results with this view, it may be that the L2 participants are generally able to perceive and utilize prosodic cues in certain circumstances. However, when complex computations are required, such as for relative clauses (and particularly for the complex items in
the current study), those cues, along with others, may not be effectively integrated at the point of interpretation.

6.3 Conclusions

In this dissertation, I have introduced the Good-enough Cue (GC) model—a comprehensive processing paradigm that joins syntactic and structural foundations, psycholinguistic evidence, and cognitive processing. In the GC model, cue-based memory retrieval is the instantiation of the complete syntactic route for agreement and long-distance dependencies in particular. However, high task load—as created via item/syntactic complexity or processing under noise or other duress, can reduce the ability to fully compute structure. Under these circumstances, the parser would then be forced to fall back on a ‘good-enough’ strategy which uses heuristics such as plausibility and thematic templates in order to preserve general message comprehension. Prosody (or implicit prosody) may reduce processing load by either facilitating syntactic processing or otherwise assisting memory retrieval, thus reducing reliance on the good-enough fallback route.

The data presented here are compatible with the proposal of a cue-based memory retrieval system for agreement, and provide support for the claims that comprehension is prioritized over grammatical processing, and that implicit prosody contributes to parsing, with varying effects based on reading fluency.

Developing a comprehensive processing model that incorporates general cognitive and linguistic considerations, and that can be applied to language users of all profiles, and across modalities, is a formidable task. However, by proposing the Good-enough Cue paradigm, I hope to set the groundwork for future elaboration on what form an ideal comprehensive model might take. This introduction opens several avenues of future research, including further refinement of
the model, particularly with regard to relevant feature cues and weights, the specific role of prosody, and other factors that may contribute to individual processing differences.

A crucial consideration is clarifying what types of features inform the cue-based memory retrieval process. Are some features weighted more heavily than others? It is likely that certain features may be more informative in one language than another, thus may be weighted differently. If so, how does this affect processing in an L2? Despite the universality of the GC model as a whole, language-specific differences in cue weighting may contribute to performance variation in L2 speakers of a language.

Another goal is to clarify the role of prosody/implicit prosody in performance. Does it act mainly as a cue to structure, or as a memory aid, or both? How can these effects be teased apart? A related concern is linking text presentation effects with oral prosody. If text presentation is tapping into (implicit) prosodic effects, we may be able to elicit similar results using aural stimuli. This would not only provide support in favor of the GC model, but would reinforce the link between the application of explicit and implicit prosody as well.

Finally, in this study, text presentation format was clearly shown to affect L1 and L2 participants differently, even though self-rated proficiency levels were equivalent, and the groups did not differ significantly in passage comprehension measures. It is crucial to determine where those differences originate. Potential points of divergence could include working memory, reading speed, or reading fluency, or a general aspect of prosodic phrasing patterns.

Along with more detailed development of the features of the model, future work may pursue its application to other areas of sentence processing—most straightforwardly to other long-distance relations such as wh-agreement, Case licensing, NPI licensing, and pronoun agreement. Even further, given the compatibility of the GC model with the processing time
course summarized in Friederici’s (2011) model, future work may adapt the study design for
direct testing with ERP measures. The validity of the model would be greatly strengthened by
confirmation of the neural effects and timing of comprehension and grammatical processing in
the paradigm used here.

Finally, the results of the study suggest that taking advantage of the relationship between
prosody and processing may provide innovative approaches to improving comprehension and
grammatical processing via prosodic training. Given the link between explicit and implicit
prosody, a training intervention for developing readers and L2 learners would likely improve
both oral and reading fluency, with subsequent improvements for reading comprehension as well.
A sample training paradigm for the development and productive use of prosodic features and
phrasing is outlined below.

- **Stage 1: Awareness training**

  Awareness training would consist of listening first to prosodically differentiated
  constructions, and identifying the prosodic cues used (e.g., pause placement, pitch
  accents). Prosodic contours (contours presented with no words) could be presented
  visually, and matched to auditory stimulus.

- **Stage 2: Meaning matching**

  After sufficient awareness training to identify pattern differences, the next stage
  would involve matching prosodic patterns to meaning differences, and identifying
  what contours or patterns may indicate a particular interpretation.
- **Stage 3: Production practice**

  The next stage would involve practice in producing sentences with appropriate prosodic features for a particular meaning or interpretation.

- **Stage 4: Transfer to reading and integration**

  The final step would involve transfer of these skills to the reading modality. First, by using cues to prosodic features, such as segmentation for phrasing, and markings for pitch accents or sentence-level prosodic contours, followed by practice in applying prosodic features to standard (non-marked) text.

  Training could advance progressively or cyclically through increasingly complex constructions that are often differentiated based on prosodic patterns, such as PP-attachments, relative clause attachments, and center embedded constructions.

  Thus, beyond the immediate relevancy of this work to sentence processing research, it has clear application to pedagogical concerns, including the development and testing of interventions for readers with lower fluency and reading comprehension, as well as techniques for presenting text to learners in a facilitative way.

  The results of this study have provided insight into the online processing of complex sentences in both L1 and L2 speakers, and demonstrated the importance of prosodic considerations, taskload, and reading fluency in both comprehension and grammatical computations. However, much work still remains in clarifying what types of information contribute most to the parse, how the general course of the model interacts with individual differences in processing capacity and taskload, and how this information may be applied productively to developing readers and L2 learners. The development of a comprehensive
processing paradigm such as the GC model is a critical first step that sets the stage for many avenues of future research.
Appendix A – Pilot Study

Introduction

The purpose of this study was to determine the effect of disrupting prosodic projection during reading. Embedded subject relative and object relative constructions were used to examine how this effect may differ in relation to structural complexity and processing load.

Participants

For this experiment, 46 native English speakers were recruited. All participants were enrolled in a Queens College introductory psychology course, and received course credit for their participation.

Materials

Thirty-two experimental items were distributed across 8 lists in a Latin Square design and combined with 8 practice items and 64 fillers. Approximately half of all items were ungrammatical. Each list was pseudorandomized into two blocks of 48 sentences each. Experimental items consisted of relative clause sets as in Table 1, in a 2x2x2 design crossing the factors of relative clause type (Subject-Object), grammaticality (Grammatical-Ungrammatical), and local NP number (Singular-Plural).
Table 1. Sample Set of Materials

<table>
<thead>
<tr>
<th>RC Type</th>
<th>Grammaticality</th>
<th>Attractor Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Grammatical</td>
<td>The reporter who called the senator that Scott supported writes awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>The reporter who called the senator that Scott supported write awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Singular</td>
<td>The reporter who called the senator that Scott supported write awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>The reporter who called the senators that Scott supported write awful stories for the newspaper.</td>
</tr>
<tr>
<td>Object</td>
<td>Grammatical</td>
<td>The reporter who the senator that Scott supported called writes awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical</td>
<td>The reporter who the senators that Scott supported called write awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Singular</td>
<td>The reporter who the senator that Scott supported called write awful stories for the newspaper.</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>The reporter who the senators that Scott supported called write awful stories for the newspaper.</td>
</tr>
</tbody>
</table>

All nouns and verbs were selected from the 5000 most frequent words of each type in the Corpus of Contemporary American English (COCA), and all head and attractor nouns were animate.

Sentences within each set were matched for length and syllable count.

Procedure

Sentences were presented on a PC using E-Prime 2.0 software (Psychology Software Tools, 2012). Each participant was randomly assigned to a stimulus list upon recruitment.

Part 1 – Normal prosody

In Part 1, sentences were presented individually on one line in their entirety. Each sentence was preceded by a fixation cross which appeared centrally on the screen for 1000 ms. The sentence then appeared and remained on the screen until the participant pressed the space bar.

Following each sentence, participants were prompted to rate the sentence on a 6-point Likert scale (1 = “very bad”, 6 = “perfect”). Following this, participants were then prompted to
respond to a comprehension question targeting the agent-patient relation within the relative clause. An example is given in (1).

(1)  a. The reporter who the senator that Scott supported called writes awful stories for the newspaper.
    b. Did the reporter call the senator?

Reading times for each sentence were recorded, as well as response times for rating and comprehension responses.

Part 2 – Disrupted prosody

In Part 2, each sentence was presented in five segments, each consisting of three words and 5–6 syllables, as in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Disrupted prosody sample materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC Type</td>
</tr>
<tr>
<td>Sub</td>
</tr>
<tr>
<td>Obj</td>
</tr>
</tbody>
</table>

Each sentence was preceded by a fixation cross which appeared centrally on the screen for 1000 ms. The first segment of the sentence then appeared and remained on the screen until the participant pressed the space bar to advance to the next segment. As in Part 1, following each sentence, participants were prompted to rate each sentence for grammaticality, and then respond to a comprehension question. Following the experimental session, participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007).
Results

Part 1 – Normal prosody

In the normal prosody condition, comprehension accuracy was higher for the subject relatives than the object relatives (F(1,45) = 25.54, p<.001), and interestingly, within the subject relatives, comprehension accuracy was highest for the ungrammatical sentences (p<.038). Possibly because readers tend to generally disprefer object relatives, ratings for subject relatives were higher than for object relatives, regardless of grammaticality (F(1,45) = 59.03, p<.001). Ratings for grammatical sentences were significantly higher than for ungrammatical sentences in the subject relatives only (p<.001), and, demonstrating the agreement attraction effect, ungrammatical items with a plural local NP were rated significantly higher than those with a singular local NP (p<.007).

Part 2 – Disrupted prosody

As in the normal prosody condition, comprehension accuracy was higher in the subject relatives than in the object relatives (F(1,45) = 12.11, p<.001). There was a significant interaction between grammaticality and attractor noun number (F(1,45) = 4.71, p<.035), where comprehension accuracy was higher in grammatical versus ungrammatical sentences when the local noun was plural (p<.039), but not when it was singular (p<1).

In grammaticality judgments, overall, subject relatives were rated higher than object relatives, regardless of grammaticality (F(1,45) = 14.39, p<.001). While there was a significant difference between grammatical and ungrammatical items in the subject relatives (p<.001), there was no difference found in the object relatives (p<1). Unlike in the normal prosody condition, no
attraction effect was found for the subject relatives: ungrammatical items with a plural local NP were not rated significantly higher than those with a singular local NP (p<1).

Summing reaction times across all segments, subject relatives were read significantly faster than object relatives (F(1,45) = 4.17, p<.05), and sentences containing a plural local NP were read more slowly than those with a singular local NP.

Breaking down the results by segments, RTs were slower for object relatives in Segment 2 (F(1,45) = 9.16, p<.004) and Segment 3 (F(1,45) = 7.76, p<.018). RTs in Segment 2 were also slower for items with plural local nouns than singular (F(1,45) = 12.01, p<.001). In Segment 4, which contained the target verb, RTs were significantly slower for the ungrammatical items than the grammatical; however, this effect was found only in the subject relatives (F(1,45) = 6.66, p<.013), not the object relatives, see Figures 1 and 2.

![Figure 1: Subject relative reading time results](image-url)
Comprehension accuracy

In the normal prosody condition, comprehension accuracy was higher for ungrammatical than grammatical sentences—a result which was found to hold only within the subject relatives, but not the more complex object relatives. Also within the subject relatives, accuracy was higher for sentences containing a plural attractor than those with a singular attractor. However, when prosody was disrupted, there was no significant difference based on grammaticality or attractor number. This may be partially because participants could not go back to reread a previously presented segment. Additionally, the effect may not be seen in the object relatives because they are more difficult to process, and the ungrammaticality is not as readily noticed. In support of this processing hypothesis, ungrammatical subject relatives with a plural attractor in Part 1 have the highest comprehension accuracy of all, suggesting that plural induces additional processing, and ungrammaticality (if noticed) increases processing as well, thereby enhancing comprehension (See Figure 3).
Ratings

Across the normal and disrupted conditions, ratings for subject relatives were consistently higher than object relatives, regardless of grammaticality. Ratings increased significantly from the normal to disrupted prosody condition, indicating that, overall, prosodic disruption decreases the ability to detect agreement errors, and masks the effect of processing difficulty, regardless of grammaticality.

Most importantly, the attraction effect found with normal reading prosody, specifically in the subject relatives, is not present in the disrupted prosody condition. While the attraction difference between subject and object relatives in Part 1 can be attributed both to complexity and cue-matching differences, the loss of attraction effect in the subject relatives of Part 2 is the result of prosodic disruption alone.

Overall, sentences are rated higher than when read with normal prosody; ungrammaticality is detected, but a plural attractor does not improve the rating, i.e., there is no attraction effect.
Conclusions

These results demonstrated that when sentences are read with normal prosody, agreement attraction occurs in subject relatives, but not in the more difficult object relatives. The lack of an attraction effect was attributed to a combination of processing complexity and retrieval cue overlap in the object relatives that makes it difficult for readers to detect ungrammaticality—regardless of the number feature of an attractor noun.

This conclusion was confirmed in the disrupted prosody experiment, where no attraction effect was found in either relative clause type. Since overall, ratings for all sentences were higher when prosody was disrupted, it was concluded that the disruption interferes in the cue-matching procedure that would normally function in determining agreement, as well as allowing any attraction errors.

Finally, while disrupting normal reading prosody as done here seems to increase reading speed, it is at the cost of both comprehension, and grammatical error detection. This effect is likely due to prosody’s influence in memory retrieval.
Appendix B – Experimental Items

a. Simple, Grammatical
b. Simple, Ungrammatical
c. Complex, Grammatical
d. Complex, Ungrammatical

1. a. The advisor who quickly picked the defendants from a list keeps legal documents for many clients.
   b. The advisor who quickly picked the defendants from a list keep legal documents for many clients.
   c. The advisor who picked the defendants that Gary accused keeps legal documents for many clients.
   d. The advisor who picked the defendants that Gary accused keep legal documents for many clients.

2. a. The ambassador who happily presented the diplomats to the queen organizes events for American visitors.
   b. The ambassador who happily presented the diplomats to the queen organize events for American visitors.
   c. The ambassador who presented the diplomats that Lisa interviewed organizes events for American visitors.
   d. The ambassador who presented the diplomats that Lisa interviewed organize events for American visitors.

3. a. The artist who interviewed the dealers from the gallery wins many awards for her shows.
   b. The artist who interviewed the dealers from the gallery win many awards for her shows.
   c. The artist who interviewed the dealers that Todd suggested wins many awards for her shows.
   d. The artist who interviewed the dealers that Todd suggested win many awards for her shows.

4. a. The attorney who interviewed the analysts before the trial prepares every case with special experts.
   b. The attorney who interviewed the analysts before the trial prepare every case with special experts.
   c. The attorney who interviewed the analysts that Lynn requested prepares every case with special experts.
   d. The attorney who interviewed the analysts that Lynn requested prepare every case with special experts.

5. a. The author who confronted the publishers about the contract refuses to revise any of her manuscripts.
   b. The author who confronted the publishers about the contract refuse to revise any of her manuscripts.
   c. The author who confronted the publishers that Jack admired refuses to revise any of her manuscripts.
   d. The author who confronted the publishers that Jack admired refuse to revise any of her manuscripts.

6. a. The banker who brought the clerks to the boardroom meets with the shareholders every month.
   b. The banker who brought the clerks to the boardroom meet with the shareholders every month.
   c. The banker who brought the clerks that Carrie hired meets with the shareholders every month.
   d. The banker who brought the clerks that Carrie hired meet with the shareholders every month.

7. a. The candidate who briefly met the senators from Indiana receives secret funding from many banks.
   b. The candidate who briefly met the senators from Indiana receive secret funding from many banks.
   c. The candidate who met the senators that Lee introduced receives secret funding from many banks.
   d. The candidate who met the senators that Lee introduced receive secret funding from many banks.

8. a. The chairman who kindly thanked the directors in a letter donates money to the local shelter.
   b. The chairman who kindly thanked the directors in a letter donate money to the local shelter.
   c. The chairman who thanked the directors that Clark assisted donate money to the local shelter.
   d. The chairman who thanked the directors that Clark assisted donates money to the local shelter.

9. a. The chef who disliked the critics from the magazine has many years of experience in restaurants.
   b. The chef who disliked the critics from the magazine have many years of experience in restaurants.
   c. The chef who disliked the critics that Jeff sent has many years of experience in restaurants.
   d. The chef who disliked the critics that Jeff sent have many years of experience in restaurants.

10. a. The coach who loved the players on the soccer team encourages everyone before each big game.
    b. The coach who loved the players on the soccer team encourage everyone before each big game.
    c. The coach who loved the players that Billy trained encourages everyone before each big game.
    d. The coach who loved the players that Billy trained encourage everyone before each big game.
11. a. The columnist who suddenly attacked the fans in the crowd blames entertainers for corrupting the city.
   b. The columnist who suddenly attacked the fans in the crowd blame entertainers for corrupting the city.
   c. The columnist who attacked the fans that Diane invited blames entertainers for corrupting the city.
   d. The columnist who attacked the fans that Diane invited blame entertainers for corrupting the city.

12. a. The consultant who eagerly helped the architects from the firm likes colorful and unique designs.
   b. The consultant who eagerly helped the architects from the firm like colorful and unique designs.
   c. The consultant who helped the architects that Mario hired likes colorful and unique designs.
   d. The consultant who helped the architects that Mario hired like colorful and unique designs.

13. a. The contractor who regularly paid the inspectors from the office approves the expenses for every project.
   b. The contractor who regularly paid the inspectors from the office approve the expenses for every project.
   c. The contractor who paid the inspectors that Alison hired approves the expenses for every project.
   d. The contractor who paid the inspectors that Alison hired approve the expenses for every project.

14. a. The counselor who briefly met the survivors of the flood prefers to organize group therapy sessions.
   b. The counselor who briefly met the survivors of the flood prefer to organize group therapy sessions.
   c. The counselor who met the survivors that Kelly rescued prefers to organize group therapy sessions.
   d. The counselor who met the survivors that Kelly rescued prefer to organize group therapy sessions.

15. a. The criminal who greatly admired the detectives on the case feels guilty after committing a crime.
   b. The criminal who greatly admired the detectives on the case feel guilty after committing a crime.
   c. The criminal who admired the detectives that Kate picked feels guilty after committing a crime.
   d. The criminal who admired the detectives that Kate picked feel guilty after committing a crime.

16. a. The customer who completely ignored the managers of the store pays for everything using only coins.
   b. The customer who completely ignored the managers of the store pay for everything using only coins.
   c. The customer who ignored the managers that Jonathan called pays for everything using only coins.
   d. The customer who ignored the managers that Jonathan called pay for everything using only coins.

17. a. The designer who called the retailers several times a week buys expensive silks for her customers.
   b. The designer who called the retailers several times a week buy expensive silks for her customers.
   c. The designer who called the retailers that Max sponsored buys expensive silks for her customers.
   d. The designer who called the retailers that Max sponsored buy expensive silks for her customers.

18. a. The driver who openly insulted the tourists from the city enjoys traveling through the countryside.
   b. The driver who openly insulted the tourists from the city enjoy traveling through the countryside.
   c. The driver who insulted the tourists that Karen accompanied enjoys traveling through the countryside.
   d. The driver who insulted the tourists that Karen accompanied enjoy traveling through the countryside.

19. a. The editor who hated the reporters from the TV station submits fake reports to the agency.
   b. The editor who hated the reporters from the TV station submit fake reports to the agency.
   c. The editor who hated the reporters that Claire suspected submits fake reports to the agency.
   d. The editor who hated the reporters that Claire suspected submit fake reports to the agency.

20. a. The employee who reported the vendors to the bureau has many children to support alone.
   b. The employee who reported the vendors to the bureau have many children to support alone.
   c. The employee who reported the vendors that Fay interviewed has many children to support alone.
   d. The employee who reported the vendors that Fay interviewed have many children to support alone.

21. a. The engineer who sent the suppliers to the warehouse assembles all the models of the equipment.
   b. The engineer who sent the suppliers to the warehouse assemble all the models of the equipment.
   c. The engineer who sent the suppliers that Anita met assembles all the models of the equipment.
   d. The engineer who sent the suppliers that Anita met assemble all the models of the equipment.

22. a. The executive who intimidated the competitors from China hosts elegant lunches twice a year.
   b. The executive who intimidated the competitors from China host elegant lunches twice a year.
   c. The executive who intimidated the competitors that Carl funded hosts elegant lunches twice a year.
   d. The executive who intimidated the competitors that Carl funded host elegant lunches twice a year.
23. a. The fisherman who questioned the merchants from the market handles the finances of the group.
b. The fisherman who questioned the merchants from the market handle the finances of the group.
c. The fisherman who questioned the merchants that Maria used handles the finances of the group.
d. The fisherman who questioned the merchants that Maria used handle the finances of the group.

24. a. The guard who attacked the inmates in the prison yard punishes anyone who steps out of line.
b. The guard who attacked the inmates in the prison yard punish anyone who steps out of line.
c. The guard who attacked the inmates that Mitch escorted punishes anyone who steps out of line.
d. The guard who attacked the inmates that Mitch escorted punish anyone who steps out of line.

25. a. The guide who approached the travelers at the station warns his clients about dangerous areas.
b. The guide who approached the travelers at the station warn his clients about dangerous areas.
c. The guide who approached the travelers that Bill left warns his clients about dangerous areas.
d. The guide who approached the travelers that Bill left warn his clients about dangerous areas.

26. a. The instructor who disliked the observers from Chicago believes that all teenagers are spoiled.
b. The instructor who disliked the observers from Chicago believe that all teenagers are spoiled.
c. The instructor who disliked the observers that Isabel sent believes that all teenagers are spoiled.
d. The instructor who disliked the observers that Isabel sent believe that all teenagers are spoiled.

27. a. The judge who recognized the attorneys before the trial records every meeting in the courthouse.
b. The judge who recognized the attorneys before the trial record every meeting in the courthouse.
c. The judge who recognized the attorneys that Sam paid records every meeting in the courthouse.
d. The judge who recognized the attorneys that Sam paid record every meeting in the courthouse.

28. a. The manager who approved the associates for a promotion closes the office late every Friday night.
b. The manager who approved the associates for a promotion close the office late every Friday night.
c. The manager who approved the associates that Jane hired closes the office late every Friday night.
d. The manager who approved the associates that Jane hired close the office late every Friday night.

29. a. The manufacturer who once cheated the sellers on a shipment evaluates every business deal carefully.
b. The manufacturer who once cheated the sellers on a shipment evaluate every business deal carefully.
c. The manufacturer who cheated the sellers that Gordon referred evaluates every business deal carefully.
d. The manufacturer who cheated the sellers that Gordon referred evalua every business deal carefully.

30. a. The mechanic who sent the agents to the auto show restores vintage cars for several private buyers.
b. The mechanic who sent the agents to the auto show restore vintage cars for several private buyers.
c. The mechanic who sent the agents that Erica knew restores vintage cars for several private buyers.
d. The mechanic who sent the agents that Erica knew restore vintage cars for several private buyers.

31. a. The minister who helped the farmers at the town fair is always eager to meet new people.
b. The minister who helped the farmers at the town fair are always eager to meet new people.
c. The minister who helped the farmers that Mary visited is always eager to meet new people.
d. The minister who helped the farmers that Mary visited are always eager to meet new people.

32. a. The musician who angrily fired the producers before the show plays several concerts every week.
b. The musician who angrily fired the producers before the show play several concerts every week.
c. The musician who fired the producers that Roger recommended plays several concerts every week.
d. The musician who fired the producers that Roger recommended play several concerts every week.

33. a. The nominee who admired the investors at the fundraiser loves people with a lot of money.
b. The nominee who admired the investors at the fundraiser love people with a lot of money.
c. The nominee who admired the investors that Mark recruited loves people with a lot of money.
d. The nominee who admired the investors that Mark recruited love people with a lot of money.

34. a. The officer who calmly saved the soldiers during the battle is quite popular with the civilians.
b. The officer who calmly saved the soldiers during the battle are quite popular with the civilians.
c. The officer who saved the soldiers that Steven accompanied is quite popular with the civilians.
d. The officer who saved the soldiers that Steven accompanied are quite popular with the civilians.
35. a. The owner who discussed the developers at the meeting appreciates honesty in negotiations.
b. The owner who discussed the developers at the meeting appreciate honesty in negotiations.
c. The owner who discussed the developers that Pat advised appreciates honesty in negotiations.
d. The owner who discussed the developers that Pat advised appreciate honesty in negotiations.

36. a. The painter who knew the buyers from the museum arranges every purchase through an accountant.
b. The painter who knew the buyers from the museum arrange every purchase through an accountant.
c. The painter who knew the buyers that Jane contacted arranges every purchase through an accountant.
d. The painter who knew the buyers that Jane contacted arrange every purchase through an accountant.

37. a. The person who replaced the employees in the business office needs two jobs to support the family.
b. The person who replaced the employees in the business office need two jobs to support the family.
c. The person who replaced the employees that Paul recommended needs two jobs to support the family.
d. The person who replaced the employees that Paul recommended need two jobs to support the family.

38. a. The photographer who impressed the journalists at the conference requests new assignments once a month.
b. The photographer who impressed the journalists at the conference request new assignments once a month.
c. The photographer who impressed the journalists that Matthew sent requests new assignments once a month.
d. The photographer who impressed the journalists that Matthew sent request new assignments once a month.

39. a. The pilot who rudely approached the passengers in first class is always nervous before a flight.
b. The pilot who rudely approached the passengers in first class are always nervous before a flight.
c. The pilot who approached the passengers that Christine brought is always nervous before a flight.
d. The pilot who approached the passengers that Christine brought are always nervous before a flight.

40. a. The politician who financed the activists from France admits to forging several hundred documents.
b. The politician who financed the activists from France admit to forging several hundred documents.
c. The politician who financed the activists that Bill caught admits to forging several hundred documents.
d. The politician who financed the activists that Bill caught admit to forging several hundred documents.

41. a. The prisoner who seriously injured the policemen in the fight has a history of being violent.
b. The prisoner who seriously injured the policemen in the fight have a history of being violent.
c. The prisoner who injured the policemen that Frank called has a history of being violent.
d. The prisoner who injured the policemen that Frank called have a history of being violent.

42. a. The producer who found the sponsors for the new season promises to make every show a success.
b. The producer who found the sponsors for the new season promise to make every show a success.
c. The producer who found the sponsors that Julie wanted promises to make every show a success.
d. The producer who found the sponsors that Julie wanted promise to make every show a success.

43. a. The reporter who called the senators every so often writes awful stories for the newspaper.
b. The reporter who called the senators every so often write awful stories for the newspaper.
c. The reporter who called the senators that Scott supported writes awful stories for the newspaper.
d. The reporter who called the senators that Scott supported write awful stories for the newspaper.

44. a. The researcher who rarely sent the doctors to the clinic creates reports for the governor.
b. The researcher who rarely sent the doctors to the clinic create reports for the governor.
c. The researcher who sent the doctors that Megan trusted creates reports for the governor.
d. The researcher who sent the doctors that Megan trusted create reports for the governor.

45. a. The resident who respected the nurses in the retirement home uses charm to get favors from the staff.
b. The resident who respected the nurses in the retirement home use charm to get favors from the staff.
c. The resident who respected the nurses that Kim supervised uses charm to get favors from the staff.
d. The resident who respected the nurses that Kim supervised use charm to get favors from the staff.

46. a. The scientist who hired the technicians to the project develops the plans for deadly weapons.
b. The scientist who hired the technicians to the project develop the plans for deadly weapons.
c. The scientist who hired the technicians that Rick called develops the plans for deadly weapons.
d. The scientist who hired the technicians that Rick called develop the plans for deadly weapons.
47. a. The secretary who clearly ignored the clients in the lobby threatens to quit her job twice a week.
   b. The secretary who clearly ignored the clients in the lobby threaten to quit her job twice a week.
   c. The secretary who ignored the clients that Thomas sent threatens to quit her job twice a week.
   d. The secretary who ignored the clients that Thomas sent threaten to quit her job twice a week.

48. a. The senior who recruited the freshmen at the career fair encourages student involvement in politics.
   b. The senior who recruited the freshmen at the career fair encourage student involvement in politics.
   c. The senior who recruited the freshmen that Tom advised encourages student involvement in politics.
   d. The senior who recruited the freshmen that Tom advised encourage student involvement in politics.

49. a. The speaker who warmly greeted the guests at the seminar insists on meeting everyone in the audience.
   b. The speaker who warmly greets the guests at the seminar insist on meeting everyone in the audience.
   c. The speaker who greeted the guests that Carol invited insists on meeting everyone in the audience.
   d. The speaker who greeted the guests that Carol invited insist on meeting everyone in the audience.

50. a. The specialist who appointed the physicians from the clinic recruits new employees every year.
   b. The specialist who appointed the physicians from the clinic recruit new employees every year.
   c. The specialist who appointed the physicians that John met recruits new employees every year.
   d. The specialist who appointed the physicians that John met recruit new employees every year.

51. a. The spokesman who introduced the experts to the sales team gets nervous in front of an audience.
   b. The spokesman who introduced the experts to the sales team get nervous in front of an audience.
   c. The spokesman who introduced the experts that Judy called gets nervous in front of an audience.
   d. The spokesman who introduced the experts that Judy called get nervous in front of an audience.

52. a. The spy who suspected the officers in the Pentagon hides his gear in a hidden closet.
   b. The spy who suspected the officers in the Pentagon hide his gear in a hidden closet.
   c. The spy who suspected the officers that Sean consulted hides his gear in a hidden closet.
   d. The spy who suspected the officers that Sean consulted hide his gear in a hidden closet.

53. a. The student who truly liked the professors in the program studies for many hours every night.
   b. The student who truly liked the professors in the program study for many hours every night.
   c. The student who liked the professors that Tony fired studies for many hours every night.
   d. The student who liked the professors that Tony fired study for many hours every night.

54. a. The supervisor who welcomed the aides from the field office motivates others to work diligently.
   b. The supervisor who welcomed the aides from the field office motivate others to work diligently.
   c. The supervisor who welcomed the aides that Barbara transferred motivates others to work diligently.
   d. The supervisor who welcomed the aides that Barbara transferred motivate others to work diligently.

55. a. The surgeon who visited the patients after their treatments dislikes traditional medicine and procedures.
   b. The surgeon who visited the patients after their treatments dislike traditional medicine and procedures.
   c. The surgeon who visited the patients that Lois treated dislikes traditional medicine and procedures.
   d. The surgeon who visited the patients that Lois treated dislike traditional medicine and procedures.

56. a. The teacher who advised the principals at the conference has no interest in teaching music.
   b. The teacher who advised the principals at the conference have no interest in teaching music.
   c. The teacher who advised the principals that Anne respected has no interest in teaching music.
   d. The teacher who advised the principals that Anne respected have no interest in teaching music.

57. a. The teacher who noticed the parents in the principal's office endorses discipline in the classroom.
   b. The teacher who noticed the parents in the principal's office endorse discipline in the classroom.
   c. The teacher who noticed the parents that Marcie greeted endorses discipline in the classroom.
   d. The teacher who noticed the parents that Marcie greeted endorse discipline in the classroom.

58. a. The technician who chose the assistants for the renovations arrives every day after five o'clock.
   b. The technician who chose the assistants for the renovations arrive every day after five o'clock.
   c. The technician who chose the assistants that Caroline hired arrives every day after five o'clock.
   d. The technician who chose the assistants that Caroline hired arrive every day after five o'clock.
59. a. The therapist who never consulted the psychologists on staff refuses help from other professionals.
b. The therapist who never consulted the psychologists on staff refuse help from other professionals.
c. The therapist who consulted the psychologists that Tara saw refuses help from other professionals.
d. The therapist who consulted the psychologists that Tara saw refuse help from other professionals.

60. a. The trainer who inspired the athletes at the high school supports community service projects.
b. The trainer who inspired the athletes at the high school support community service projects.
c. The trainer who inspired the athletes that Julie coached supports community service projects.
d. The trainer who inspired the athletes that Julie coached support community service projects.

61. a. The visitor who slowly approached the generals from Washington distrusts anyone in a uniform.
b. The visitor who slowly approached the generals from Washington distrust anyone in a uniform.
c. The visitor who approached the generals that Nick guarded distrusts anyone in a uniform.
d. The visitor who approached the generals that Nick guarded distrust anyone in a uniform.

62. a. The volunteer who trained the assistants in only two weeks works every Monday at the shelter.
b. The volunteer who trained the assistants in only two weeks work every Monday at the shelter.
c. The volunteer who trained the assistants that Diego brought works every Monday at the shelter.
d. The volunteer who trained the assistants that Diego brought work every Monday at the shelter.

63. a. The widow who bravely rescued the teenagers from the fire attracts attention everywhere in town.
b. The widow who bravely rescued the teenagers from the fire attract attention everywhere in town.
c. The widow who rescued the teenagers that Adam brought attracts attention everywhere in town.
d. The widow who rescued the teenagers that Adam brought attract attention everywhere in town.

64. a. The witness who easily trusted the jurors in the room acknowledges that she didn't see the thief.
b. The witness who easily trusted the jurors in the room acknowledge that she didn't see the thief.
c. The witness who trusted the jurors that James selected acknowledges that she didn't see the thief.
d. The witness who trusted the jurors that James selected acknowledge that she didn't see the thief.
Appendix C – Filler Items

**Argument Structure:**
During the show the famous magician appeared suddenly on the stage of the theater.
This past month the head scientist proved the theory after many years of active research.
During every class the top student smiled to himself when he realized the correct answer.
Every day Marcie sailed the boat across the lake to her house on the island.
The soccer player kicked the ball and immediately fell down holding his knee.
Before the test the college student borrowed lecture notes from several people in the class.
The friend that Joe offended fought him angrily about it for several weeks.
In hot weather the young teenager swims every day in the pool behind the house.
When Ashley heard the news she came quickly to help save the school from closing.
The coach who managed the team has had twenty years of experience in the sport.
During the summer the tired nanny sleeps for hours in the garden every afternoon.
The maniac who hit the dumptruck ran two red lights before the police caught him.
On Monday morning the sneaky employee stole some money and met his friends at the corner.
The man sent by the studio photographed the president and sent the prints to the tabloids.
Before the exhibit the excited artist wandered for hours in the gallery and waited nervously.
The pilot who flew the plane across the Atlantic Ocean just earned his license last month.

*During the party the annoying guest spilled with everyone in the living room and told bad jokes.
*Before the party the excited children carried together and played games in the backyard.
*After the race the tired athlete grabbed on the ground and talked with the other runners.
*When it rained the anxious father brought every morning to the museum with the car.
*After hearing the speech from the boss Gabby lied the story about the missing supplies.
*During the class the distracted boy opened suddenly when he heard the teacher call.
*Bobby jumped over the table and broke shamelessly only five minutes after arriving at school.
*Whenever Nate worked late he sat the tree in the park and read a book.
*Cassie shopped the new dress only three days before the big event.
*Two hours ago the heroic guard waited every room in the building and caught the thief.
*After the dinner the embarrassed guest sneezed the hostess before falling on the ice sculpture.
*At five o'clock Abby arrived the cake and set up the drinks for the reception.
*After the meeting the unhappy boss remembered the clerk that his birthday was last week.
*Without any warning Frank fell the baby and ran inside the house.
*The little boy held playfully and ran around the playground for hours.
*The old carpenter made quietly and never bothered his neighbors with loud noises.

**Complex Subject:**
The stereo for the luxury cars was stolen before the technician had finished the repairs.
The copy of the valuable document was lost after the terrible fire in the town hall.
The engraving of the famous captain is currently kept in the basement of the museum.
The generous gift from the guests were used to pay for the family's medical bills and other expenses.
The corridor to the conference chambers was expanded to display the statue of the president.
The impressive plan from the architect was submitted to the senior trustees before the meeting.
The urgent request from the institute was presented to the head accountant of the company.
The gift bags for the attendees are filled with expensive samples from many companies.

*The secret memo for the senators were quickly handed across the aisle before the debate.
*The reinforced door to the armory prevent thieves from stealing the weapons.
*The ceremony for the grand opening are scheduled during the busiest month for retail shopping.
*The elephant from the nearby zoo have often escaped despite extra locks on the main gate.
*The corroded key to the cabinets were barely able to turn properly without getting stuck.
*The passageway to the lower level were partially blocked by fallen debris after a rockslide.
*The large can of peas are being served with steak at the family dinner tonight.
*The portrait of the youngest sisters were in the gallery for many years before being sold.

**Mass/Count Nouns:**
Mary's father always makes a lot of spaghetti for the family whenever he cooks dinner.
The professor of the college freshmen planned to assign a lot of homework during the semester.
The young teacher at the school remembered to bring a lot of chalk for the classroom.
The airline worker was too busy to notice that a lot of suitcases were still unloaded.
The old gardener showed us that all plants need a lot of nutrients to grow properly.
The manager of the boutique counts on selling a lot of scarves during the holidays.
The scientist saw that the experiment was ruined by a lot of clouds filling the sky.
The zoo was happy to report that the endangered turtle laid a lot of eggs this summer.
When the landlord inspected the building he had found a lot of boxes in the hallway.
The sailors bailed out a lot of water from the bottom of the boat before it could sink.
The new owner of the mansion needed to buy a lot of furniture before the gala.
The construction workers removed a lot of dirt from the empty lot after the project was finished.

*The thoughtful server at the diner always brought a lot of butters to the customers.
*The family discovered that a lot of chair had been broken during their move to California.
*The practical hostess made sure that there would be a lot of chip at the party.
*When the chef prepared his soup he always added a lot of pickle to the recipe.
*The top athletes at the university prefer to eat a lot of rices before every game.
*The young children living on the farm love to eat a lot of corns in the summer.
*George's sister likes to give a lot of advices to her coworkers at the office.
*When the contractor visited the worksite he discovered a lot of shovel had been left behind.
*On Sunday mornings the oldest children take turns preparing a lot of toasts for the family.
*The car dealer ordered a lot of truck from the manufacturer before the end of the season.
*The busy father of the toddlers always tried to pack a lot of milks for their picnics.
*After the storm the residents saw that a lot of tree had been damaged by the wind.

**Noun Complement Clause:**
The evidence that the crime took place at night proves that the defendant is innocent.
The lie that only wealthy people can afford a nice vacation needs to be corrected.
The suggestion that Brian should have apologized for the mistake is unforgivable.
The prediction that Lauren's brother would win the election was shown to be false.
The assumption that Tom's mother will do his laundry for him is foolish.
The hope that Steven will be a successful doctor is comforting to his mother.
The conclusion that potassium is good for you has led to a rise in banana imports.
The idea that George might become a police officer makes his family proud.

*The claim that Shakespeare didn't write all of the plays attributed to him are disturbing.
*The fact that the vaccine prevented many illnesses are well known in the community.
*The thought that the stolen gems could still be in the country are encouraging.
*The promise that Bill would love Jane forever still bring joy to her heart.
*The intention that the book would be finished before the holidays were unrealistic.
*The proposal that the city build several new parks by the river are going to be expensive.
*The theory that the sun will explode within a few years are unlikely to be true.
*The decision that all students must attend the seminars are wildly unpopular.

**Pronoun agreement:**
The father of the recent widow prepared his own speech before the memorial last Tuesday.
The nephew of the stingy bride is bringing his own food to the reception.
The grandmother of the daring cowboy grooms her own horse for hours after every show.
The grandson of the count inherited his own wealth from his mother.
The mother of the wealthy landlord bought her own property in the Bahamas several years ago.
The aunt of the nervous groom polished her own shoes two hours before the wedding.
The daughter of the snobby butler washed her own wine glasses after the celebration.
The nephew of the pious nun manages his own charity for young orphans and widows.
The brother of the sophisticated duchess writes his own letters to the people in the countryside.
The son of the professional waiter makes his own money betting on horse races.
The daughter of the stewardess flies her own planes across the country every week.
The niece of the empress runs her own zoo for exotic animals from around the world.
The father of the duke revealed his own secrets in the interview with the reporter.
The nephew of the ancient wizard learned his own spells from a secret book.
The sister of the countess hired her own household staff more than five years ago.
The grandmother of the landlady invested her own money in stocks and government bonds.

*The sister of the popular king organizes his own dinners for the princes and other nobles.
*The brother of the longtime actress produced her own show for network TV two years ago.
*The uncle of the noble lady calls her own doctor whenever the children are sick.
*The son of the loyal maid mended her own uniform before the party last night.
*The granddaughter of the emperor oversees his own territory in the east.
*The niece of the well-known actor watched his own movies for the first time last week.
*The grandson of the busy waitress brings her own soda to the restaurant.
*The mother of the lonely bachelor invited his own friends to the beach house this year.

**Wh Relative Clauses:**

The director fired the actors who the producer agrees are not at all talented.
The detective questioned the suspects who the client claims have stolen the records.
Sally wrote to the reporters who the editor trusts are honest and thorough.
Kathy interviewed the applicants who the mayor predicts will do well in politics.
The president appointed the judges who the senator believes are well qualified.
The manager scolded the employees who the clerk knows are very lazy.
John admired the teachers who the principal thinks are too demanding.
Mark introduced the visitors who the boss expects will tour the factory all day.

*Bill met with the painters who the contractor believe are the best in the city.
*Donna thanked the delegates who the ambassador hope will return soon.
*The doctor examined the patients who the nurse think need additional treatment.
*Max called the engineers who the inspector predict will be fired soon.
*The officer arrested the women who the witness claim were in the stolen car.
*The judge pardoned the criminals who the lawyer know were guilty.
*Laura worked with the assistants who the director trust are familiar with the projects.
*The waiter laughed at the customers who the chef expect will never visit again.
References


