THE REANALYSIS AND INTERPRETATION OF GARDEN-PATH SENTENCES BY NATIVE SPEAKERS AND SECOND LANGUAGE LEARNERS

BY

ZHIYING QIAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in East Asian Languages and Cultures
with a concentration in Second Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Jerry Packard, Chair, Co-Director of Research
Associate Professor Susan Garnsey, Director of Research
Associate Professor Kiel Christianson
Associate Professor Annie Tremblay
Associate Professor Chilin Shih

ABSTRACT

This dissertation examines factors (verb bias and plausibility) that influence reanalysis processes in native and non-native processing of English and Mandarin garden-path sentences (Chapters 2 and 3) and the relationship between the amount of reanalysis and final interpretation of such sentences (Chapter 4).

Verb bias refers to the likelihood of a particular verb taking a particular argument structure, such as a direct object (DO) or a sentential complement (SC). Previous research has demonstrated that native speakers of English are able to use verb bias information fast enough to generate predictions about the upcoming syntactic structure and that verb bias plays a larger role than plausibility in this predictive process (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997). However, little is known about the relative importance of verb bias and plausibility in second language sentence processing. A prevailing view in the L2 psycholinguistic literature claims that L2 learners underuse structural cues during real time processing, and that to compensate, they rely predominantly on lexical-semantic cues (Clashen & Felser, 2006). What has not been considered on this view is the use of lexically-associated structural cues, such as verb bias. Since such information is both lexical and structural, it is unclear whether L2 learners would be able to use these cues in real-time processing. In two self-paced reading experiments, Chapter 2 compared L1-Mandarin speakers of L2 English and L1-Korean speakers of L2 English with native English speakers on the resolution of temporary DO/SC ambiguity in sentences. Results showed that similar to native speakers, both L2 groups were able to use verb bias cue to predict the likely type of following structure, but were unable to use the plausibility cue predictively

when the verb bias cue was present, challenging the view that L2 learners rely more on plausibility than syntax during parsing.

While substantial research has been conducted on verb bias effect in English, few studies have examined such effects in other languages, especially in languages that have been found to rely more on plausibility than structural information, such as Mandarin (Su, 2001a, 2001b, 2004). In one self-paced reading experiment, Chapter 3 compared the relative contributions of verb bias and plausibility in processing Mandarin sentences that bore the surface level resemblance to English sentences with temporary DO/SC ambiguity. Since Mandarin allows null subjects, such a structure is temporarily ambiguous between an embedded clause and a blended structure, in which the object of the first clause is also the subject of the second clause. Results showed that verb bias trumped plausibility in Mandarin, such that readers made use of verb bias cues to anticipate the following structure and were only sensitive to plausibility information when verb bias allowed it, contrary to the claim that Mandarin relies heavily on plausibility in sentence comprehension.

In Chapters 2 and 3, reading time (RT) at the disambiguating region in sentences was used as the diagnostic in determining the effects of verb bias and plausibility, based on the assumption that RT at the disambiguation reflects the amount of reanalysis work. In two self-paced reading and two event-related brain potential (ERP) experiments, Chapter 4 demonstrated that RT and ERP on-line measures at the disambiguation might not reflect primarily reanalysis, since both RTs and the amplitudes of the P600 and N400 ERP components were found to be unrelated to the accuracy of the final interpretation of garden-path sentences, as measured by responses to post-sentence questions, thus calling into question traditional assumptions about the meaning of traditional measures. The original prediction was that more time/effort spent

reanalyzing at the disambiguation would lead to more success in question responses. Instead, whenever there was any trend toward a relationship between the online measures and question responses, it was opposite the predicted direction, i.e., when more time/effort was spent on the disambiguation, questions tended to be answered less accurately. Chapter 4 thus proposed that the RTs and ERP component amplitudes at the disambiguation may reflect the amount of confusion about and/or competition between different possible interpretations, rather than or in addition to any reanalysis triggered there. Overall, this dissertation examined the reanalysis processes at the disambiguation in garden-path sentences in both native and non-native sentence processing and the link between the reanalysis processes and the final interpretation in native sentence processing. It paved way for conducting similar research on the final interpretation of garden-path sentences by L2 learners.

ACKNOWLEDGEMENTS

I started my journey of Ph.D. study in the Department of East Asian Languages and Cultures at the University of Illinois at Urbana-Champaign with curiosity, enthusiasm, and yet also with fear and uncertainty. Coming from a background of language pedagogy, I had never imagined that I would end this journey with a research focus on psycholinguistic and neuroscientific approaches to languages. It was the extraordinary faculty here at Illinois who shaped my research interest, and who provided me with the best training I could ever ask for.

First and foremost, I would like to express my most sincere thanks to my advisor, Dr. Jerry Packard, for his valuable guidance, continuous encouragement and unwavering belief in me. Several times along the journey when I was about to deviate, he directed me back onto the right track. I thank him for the opportunity to collaborate with him on my very first eye-tracking experiment, and for giving careful consideration to my ideas even when they were naïve. I have learned a lot from his editing of my grant proposals and conference abstracts. He is always very supportive of my participation in activities that will lead to better job opportunities. When I had visa problems and almost lost the chance of working for the best language teaching university in the US, Middlebury College, he worked patiently with the International Students Office and eventually persuaded them to allow me to take the position. I am also indebted to his thoughtful and detailed comments on my job application materials. I also thank him for having introduced many good papers to me, among which, the Ueno and Garnsey (2008) paper led to my four-year close collaboration with Dr. Susan Garnsey.

No words can adequately express my gratitude to my co-advisor, Dr. Susan Garnsey, whose wisdom, sincerity, generosity and integrity have and will continue to have tremendous influence on me, both in academia and in personal life. Four years ago when I could not even tell

apart EEG and ERPs, Susan welcomed me to her lab whole-heartedly. From then on, there were numerous afternoons when I was so obsessed in learning from her until the beautiful rays from the sunset passing through the windows reminded me that I should let my professor go home. I have learned from her meticulousness and integrity about research. I will always remember her saying that she would prefer not to find an effect that is there rather than find one that is not. This dissertation would not have been possible without her wisdom and endless support. Before my talk at the CUNY 2015 conference on Chapter 4 of my dissertation, Susan spent several hours a day for a week on making slides together with me. She paid attention to font size of every text, color of every bar and design of every slide. She shared with me her expertise about how to deliver a good talk. She accompanied me to the conference, even though she had already retired and even though that meant another trip shortly before her long flight to Hawaii. At other conferences, she put up posters together with me and brought me coffee and cookies. In the field of neuroscience of language where the genealogy tree and the old apprenticeship model still exists, I am very happy to be Susan's scientific daughter. I consider Susan more than my co-advisor, she is also my lifelong friend, a friend whom I can resort to at any time of difficulty.

I am deeply indebted to Dr. Annie Tremblay, for teaching me so much about psycholinguistics. In the second language processing course that I took with her, I learned to design and program experiments, collect and analyze data, and make a poster to present research, all in one semester. Her class was one of the hardest I have ever taken, and yet we all loved her and secretly said that she must have two brains. Annie is a role model to me for being so smart, passionate, kind, productive and willing to share and help. I only hope that I can live up to her high standard of excellence.

Similarly, I am very fortunate to have had the chance to take a course with Dr. Kiel Christianson, who has broadened my horizons. Kiel's wisdom has inspired several of my research ideas. His kindness and support have also led me to turn some of those ideas into experiments. I would like to thank Kiel for providing me with the facilities to do an eye-tracking study on auditory perceptual simulation, and for his bringing back from conferences prints of posters related to my research. Although extremely busy, Kiel always responded to my emails very quickly, even at late night, and was always willing to spend time checking my experimental procedures before data collection.

I thank Dr. Chilin Shih for her help with constructing the Mandarin stimuli used in my dissertation, for thinking hard to find a Mandarin structure that does not allow pro-drop when I needed it, and for critically looking for alternatives to my interpretation of the data. It was she who told me that I was interested in the brain and language even before I noticed myself. Dr. Shih always responds to my emails quickly and offers to meet with me at the earliest possible time. I am grateful to her for bringing insights to my dissertation as a native speaker and a linguist of Chinese.

Other faculty I would like to thank are Dr. Kara Federmeier, for her insights and brilliance in interpreting data; Dr. Darren Tanner, for his invaluable suggestions for analyzing data; Dr. Sarah Brown-Schmidt, for teaching me mixed-effects models; Dr. Cindy Fisher and Dr. Duane Watson, for their valuable advice at the psycholinguistic joint lab meetings; Dr. Silvina Montrul, for her interesting introductory course on psycholinguistics; the late Dr. David Goodman, for his extremely hard course on literature theory; and Dr. Gary Dell, for some very interesting conversations. Faculty at University of Illinois will surely be what I will miss most after I finish this Ph.D. journey.

I will also miss my cohorts at University of Illinois, who have helped me in nearly every aspect of research. I would like to thank Erika Hussey, Joe Toscano, Zhenghan Qi, Eunah Kim, Mallory Stites, Cassie Palmer, Danielle Dickson, Alex Fine, Cybelle Smith, Jui Namjoshi, Si On Yoon and Sun-A Kim for their help and friendship.

I am fortunate to be a member of the Language and Brain Lab, where the cozy and relaxing atmosphere has sparked my research ideas. I am thankful to our lab undergraduate students whose senior research I guided: I thank Raeann Shelly and Ayaka Misui for constructing experimental stimuli, and Gabrielle Elizalde-Ocasio for help with data collection. Through supervising these students, I learned how to guide undergraduate research. I cherish this experience and sincerely thank Dr. Garnsey for having given me these opportunities.

This dissertation would not have been possible without the help from Dr. Lin Hui, Xie Yinghua and Gu Ying from Shanghai Fudan University, and Mike Hudgins, Young Jae Lee, Keqi Wei, Gabrielle Elizalde-Ocasio and Ayaka Misui from the Language and Brain Lab at University of Illinois. I thank them for their help recruiting participants and collecting clean brainwave data.

My deep appreciation also goes to several institutions that offered me fellowships and research grants. I am grateful to the Department of East Asian Languages and Cultures for a first-year graduate student fellowship, the Graduate College Intersect Neuro-Cultures Program for a two-year fellowship, and the Graduate College for a one-year Dissertation Completion Fellowship. I also thank the Inter-University Center for a one-year fellowship to study in Japan, where I enjoyed one of the happiest years in my life. I also thank the Department of East Asian Languages and Cultures, the Beckman Institute and the Graduate College for several research grants, without which data collection in China would not have been possible.

Last but not least, I am deeply indebted to my friends and family, who have accompanied me through every moment of happiness and sadness along this journey. I thank my best friend, Yanhua Sun, for many years of friendship. I truly cherish the time when we could share our joys and tears, celebrate at good times and comfort each other at the most difficult ones. I thank my advisor and friend in Japan, Kiyomi Kushida, for our wonderful time together. I deeply thank my parents for taking care of my very young daughter, for unceasing faith and confidence in me, and for teaching me to be perseverant and hard-working. I am deeply indebted to my husband for his unconditional support, love, encouragement, guidance and caring. He earned a Ph.D. degree in Industrial Engineering and is currently a SAS programmer specializing in big data. For a long period of time before I learned computational programming using R, he came back home after a long day of work and started to work on another full-time job helping me analyze my big data. I am also grateful for his many valuable insights on my research projects, conference presentations, dissertation, and my career path in general.

I will miss you all, my professors and friends, after I finish my Ph.D. journey. I will embark on the next step in my journey with all the knowledge and the qualities of being hard-working and perseverant that I have learned from University of Illinois.

TABLE OF CONTENTS

CHAPTER 1

Introduction

Sentence processing proceeds incrementally, with each incoming word being syntactically analyzed and integrated into the structure as soon as it is encountered. The big question that has dominated psycholinguistic research for decades concerns the timing at which non-syntactic information is used by the parser to influence the construction of syntactic structure. The two most influential classes of theories, i.e., the serial parsing models and the parallel parsing models, differ on whether non-syntactic information can influence parsing as soon as it is available.

According to the serial, two-stage models, which are best represented by the Garden-Path Model, first-stage parsing is restricted to the use of syntactic information, with non-syntactic information only affecting the later, reanalysis stage. Contrary to this view, parallel, constraint-based models argue that all sources of information start to influence parsing from the beginning. Parsing occurs in one stage, with multiple possible structures remaining active at the same time. Potential structures are ranked according to the amount of support they receive from various constraints (Garnsey et al., 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1994; Trueswell et al., 1993). One way to test between these two classes of theories is to examine whether lexical frequency information can influence first-pass parsing, as lexical frequency information is not the type of information that the first-stage parser considers according to the original version of the Garden-Path model.

One type of lexical frequency information is the frequency with which a particular verb appears with a particular type of argument structure (termed verb bias). Verbs differ in their structural biases. Some verbs are more frequently followed by direct objects (DO-bias) while others are more frequently followed by sentential complements (SC-bias).

Substantial research has been conducted on the effect of verb bias on the processing of English sentences. Although researchers differ on how fast such information becomes available to the parser, i.e., during the first-stage or the reanalysis stage, they converge on the view that verb bias affects the processing of the subsequent words in the sentences (Ferreira & Henderson, 1990; Garnsey et al., 1997; Kennison, 2001; Osterhout, Holcomb, & Swinney, 1994; Pickering & Traxler, 1998; Pickering & Traxler, 2003; Pickering, Traxler, & Crocker, 2000; Traxler, 2005; Trueswell & Kim, 1998; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993; Sturt, Pickering, & Crocker, 1999). For instance,

(1) a. *The club members understood the bylaws would be applied to everyone.* (DO-bias)
b. *The club members understood that the bylaws would be applied to everyone.*

(2) a. *The ticket agent admitted the mistake might be hard to correct.* (SC-bias)
b. *The ticket agent admitted that the mistake might be hard to correct.*

In (1), readers initially interpret *the bylaws* as the direct object of *understood*, and subsequently experience processing difficulty at *would*, because *would* signals that the sentential complement *would be applied to everyone* lacks a subject and therefore *the bylaws* cannot be the direct object of *understood*, but instead must serve as the subject of *would be applied to everyone*. At *the bylaws*, this sentence is temporarily ambiguous between a direct object or a sentential complement structure (termed DO/SC ambiguity), and such ambiguity is eliminated at the sentential complement verb *would*. In (1b), the initially incorrect direct object interpretation is eliminated by the complementizer *that*. Processing difficulty in (1a) is reflected in the slower

reading time at *would* compared to the reading time of *would* in (1b). Sentence (2a) has exactly the same structure as (1a), but differs from (1a) in that the main clause verb *admitted* has SC-bias. Temporarily ambiguous sentences like (1a) are also termed "garden-path" sentences, because the parser is misled into one structural analysis and has to subsequently revise that incorrect analysis.

Previous studies on English sentences have found that when the main clause verb has DO-bias, as in (1a), readers experience more processing difficulty at *would* than when the main clause verb has SC-bias, as in (2a). This is because when the main clause verb biases towards taking direct objects, the parser analyzes the following noun as the direct object and thus experience processing difficulty when such interpretation turns out to be incorrect. On the contrary, when the main clause verb biases towards taking sentential complements, the parser analyzes the following noun as part of a sentential complement and thus does not experience processing difficulty when later information turns out to be consistent with such interpretation.

Not only have researchers found that verb bias affects the reading of the subsequent words, several studies have provided evidence that the effect of verb bias on parsing occurs rapidly (e.g., Garnsey et al., 1997; Trueswell & Kim, 1998; Wilson & Garnsey, 2009). For instance, in an eye-tracking experiment using sentences like (1) and (2), Garnsey et al. (1997) found evidence in the first-pass reading times that the disambiguating verb *would* in (1a) was read slower than in (1b), but the disambiguating verb *might* in (2a) was read as fast as in (2b), indicating that SC-bias verbs were sufficient in guiding the parser away from considering the direct object analysis. The same rapid effect of verb's biases was observed in sentences that turned out to have direct object endings, as in (3) and (4) (Wilson & Garnsey, 2009).

(3) The club members understood the bylaws because they had read them. (DO-bias)
(4) The ticket agent admitted the mistake because she had been caught. (SC-bias)

First-pass reading times at the disambiguating word *because* in (4) was read slower than in (3). In (4), the parser integrated the ambiguous noun *the mistake* as part of the sentential complement after encountering the SC-bias verb *admitted*, which turned out to be the incorrect analysis at *because*. In (3), in contrast, the initial direct object interpretation, i.e., *understood the bylaws*, turned out to be the correct interpretation. Since first-pass reading time is considered to be an early measure that most likely reflects the underlying parsing processes that occur at the first-stage of parsing, the rapid effect of verb bias described above has been taken to support the constraint-based models.

Because of the clear evidence that verb bias has an early and strong effect on the processing of English sentences, this dissertation does not aim to examine verb bias effects on English sentences again. Rather, I explored the effect of verb bias in both second language (L2) sentence processing in English and in first-language (L1) sentence processing in Mandarin as a way to test important theories and assumptions about language processing. Throughout this dissertation, I use the term *reanalysis* to refer to the reanalysis processes in the Garden-Path Model and the re-ranking processes in the constraint-based models.

Few studies have examined how verb bias affects sentence processing by L2 learners, although the answer to this question is informative to the on-going debate about the differential use of syntactic and semantic information in the real-time parsing of L2 sentences. The debate in the L2 psycholinguistic literature concerns whether the L2 parser is qualitatively different from the L1 parser in terms of the way syntactic information is used. Some researchers claim that syntactic information is not accessible to the L2 parser during online processing. Rather, the L2 parser is restricted to the use of lexical-semantic information only (the Shallow Structure

Hypothesis, Clahsen & Felser, 2006a, 2006b). The claim is that L2 learners underuse syntactic information so the syntactic structures they build during parsing are shallower and less detailed than those built by native speakers (see Chapter 2 for a review of the evidence for and against this view). What has not been considered on this view is L2 learners' use of lexically-associated structural cues, such as verb bias.

Verb bias is the frequency with which a verb is used in sentences with particular types of structure. It seems to lie between the type of structural information that L2 learners have been argued to underuse and the type of lexical-semantic information that they have been argued to rely heavily on. Only a few studies have examined verb bias effects in L2 processing, and so far they have converged to show that L2 learners are able to learn verb bias that is specific to the L2 and are capable of using such cues to guide online processing of L2 sentences (Dussias & Cramer Scaltz, 2008; Frenck-Mestre & Pynte, 1997; Lee, Lu, & Garnsey, 2013). However, the evidence so far has come from a limited number of studies on a limited number of languages (French, Spanish, and Korean). In Chapter 2, I seek to add another piece of evidence to this line of research by testing L1-Mandarin speakers of L2-English. Unlike the other languages tested so far, an important feature of Mandarin is that there is no complementizer that marks sentential complements. Therefore, it is unclear whether L2 learners would be able to learn and use the cue provided by the complementizer *that* in English, and whether they are able to use verb bias and the complementizer interactively in the way that native speakers do.

In Chapter 2, I also explored the relative importance of verb bias and plausibility in L2 sentence processing, which is a question that has not been investigated before, and yet may shed some light on the ongoing debate. Research from English showed that native English speakers do not use the plausibility cue when the verb bias cue is available for them to rely on in the

processing of DO/SC ambiguity (Garnsey et al., 1997). In contrast, since L2 learners have been argued to rely heavily on semantic information to guide on-line parsing, they may show a different pattern from native speakers.

As mentioned earlier, research on verb bias effects is important to the understanding of how sentences are processed, because it provides a good test case to distinguish between the two major classes of parsing theories. However, studies on verb bias effects have been conducted predominantly on English sentences. Little is known about whether verb bias is used in the same way in other languages that are typologically different from English. One reason this might not be true is that some languages have been found to rely more on plausibility than on syntax, such as Mandarin (Su, 2001a, 2001b, 2004). Therefore, more work is needed to examine the effect of verb bias in other languages so as to know whether the verb bias effect observed in English is a universal phenomenon, rather than a feature specific to English. To this end, Chapter 3 of this dissertation examined how Mandarin speakers use verb bias and plausibility cues to disambiguate DO/SC sentences in Mandarin.

When investigating verb bias effects in the L2 and in Mandarin sentence processing, I rely on reading times at the disambiguating word as a way to illustrate how much reanalysis effort is needed for the parser to recover from garden-pathing. For instance, in (1a), the parser initially interprets *the bylaws* as the direct object of *understood*, since direct object analysis is the simpler of the possible structures and *understood* is a verb that most frequently takes direct objects. Reading time at the disambiguating verb *would* is taken to indicate how much the parser has committed to such misinterpretation, based on the assumption that readers slow down at the disambiguation because they spend extra effort on reanalyzing the syntactic structure. It is a

commonly held assumption in psycholinguistic research that reading time (and other on-line measures) at the disambiguation indexes the amount of syntactic reanalysis.

However, if this is true, there should be a relationship between the reading time at the disambiguation and successful recovery from the initial misinterpretation, such that the more time readers spend on syntactic reanalysis of the sentence, the more likely they are to successfully recover from the initial misanalysis. To the best of my knowledge, no empirical evidence has been provided to specifically support this assumption, and yet it is important for this assumption to be tested, because the majority of psycholinguistic research, including Chapters 2 and 3 of this dissertation, is based on it. In Chapter 4, I explored the link between on-line measures at the disambiguation and off-line interpretation of garden-path sentences to test this assumption.

To summarize, this dissertation seeks to provide evidence relevant to the following unanswered questions in the sentence processing literature: 1) whether L2 learners of English are capable of using verb bias and plausibility cues to predict the upcoming syntactic structure; 2) how verb bias and plausibility are used in the processing of Mandarin sentences, given that Mandarin has been found to rely more on plausibility than on syntax; and 3) whether on-line measure at the disambiguation in garden-path sentences is a good indicator of the amount of syntactic reanalysis. In what follows, I outline the design and major findings of each chapter.

In response to the first question, Chapter 2 compared L2 learners of English to native speakers on the resolution of the DO/SC ambiguity in English sentences. In self-paced reading Experiment 1, verb bias and ambiguity were manipulated and L1-Mandarin speakers of L2-English were tested, as well as native speakers. Results for native speakers replicated previous findings, showing that the verb bias and complementizer cues were each sufficient for

disambiguation. For L1-Mandairn speakers, both cues were helpful for the recovery from garden-pathing, but the optimally efficient native-like pattern was not yet achieved. Self-paced reading Experiment 2 additionally manipulated the plausibility of the ambiguous noun as the direct object of the main clause verb (*the club members understood the bylaws*… vs *the club members understood the pool*…) and tested native speakers, L1-Mandarin and L1-Korean speakers of L2-English. Results for the native speakers replicated previous studies showing no effect of plausibility, and also showed the same pattern for both L2 groups, thus challenging the claim that L2 learners rely more on plausibility than syntax during on-line sentence processing.

In response to the second question, Chapter 3 conducted one self-paced reading experiment to examine how Mandarin speakers use verb bias and plausibility cues to process Mandarin sentences that are similar to English sentences with DO/SC ambiguity, such as *The proud mother announced the wedding would be a big event.* Whereas in English, *the wedding* can serve as either the direct object of the main clause verb *announced* or the subject of the embedded clause *would be a big event*, in Mandarin, it is temporarily ambiguous between being the direct object of *announced* or both the direct object of *announced* and the subject of *would be a big event*. Mandarin allows such a structure whereby a noun serves as both the object of the first clause and the subject of the second clause, when the noun is plausible as the direct object of the first clause. In cases when the noun is implausible as the direct object of the first clause, a sentential complement reading results, where the noun is analyzed as the subject of the second clause, as in English. Verb bias and plausibility were manipulated. Results showed that verb bias trumped plausibility in processing Mandarin sentences, just as it does in English. Readers constructed syntactic structures that were consistent with verbs' biases, but benefited from the plausibility cue only when verb bias allowed it, thus challenging the view that Mandarin relies

heavily on plausibility for sentence comprehension.

In response to the third question, Chapter 4 conducted two self-paced reading and two ERP experiments to explore the link between on-line measures at the disambiguating region of sentences and the final interpretation of garden-path sentences with early/late closure ambiguity, such as *While the man hunted the deer that was brown and graceful ran into the woods*. In all experiments, participants read the sentences word by word and answered a question after each sentence that probed whether they discarded the initial misanalysis (i.e., *Did the man hunt the deer? or, Did the sentence explicitly say that the man hunted the deer?*). Results from the four experiments converged to show that slower reading time and larger P600/N400 amplitudes at the disambiguating verb *ran* did not lead to better comprehension of these sentences, indicating that on-line measures at the disambiguation were unrelated to the correct interpretation of these sentences. However, if on-line measures at the disambiguation are good indicators of the amount of reanalysis work, slower reading time and larger ERP components should have led to better comprehension. Therefore, the results of the studies in Chapter 4 challenge the traditional view that the time or effort spent on the disambiguation is caused primarily by the effort of reanalysis, and suggest that on-line measures at the disambiguation may instead indicate a combination of the amount of reanalysis and other factors such as the confusion resulting from having competing structural possibilities. These results in Chapter 4 qualify the interpretation of the results in Chapters 2 and 3 by adding the possibility that readers may have slowed down at the point where sentences were disambiguated towards the sentential complement structure not because they were successfully revising their initial interpretation, but rather because they remained confused about which of the competing possible analyses to adopt. That is, it is possible that slowing down at the disambiguation in garden-path sentences in those studies, and by extension many other

studies in the field, does not index successful reanalysis, as has typically been assumed.

CHAPTER 2

Verb Bias and Plausibility in Non-native Sentence Processing

Verbs differ in the type of complements that they can take. Consider (5),

(5) The scientist read the article…
(a) ……..at lunch time.
(b) ……..had been published two months ago.

The syntactic role of *the article* is temporarily ambiguous. The sentence proceeds with *the article* having the direct object role in (5a) but having the role of subject of an embedded clause in (5b). In (5a), the scientist did read the article, while in (5b) the scientist read something about the article, but not necessarily the article itself. Such temporary structural ambiguity at *the article* arises because English allows the complementizer *that* to be dropped before an embedded sentential complement clause. In what follows, this type of structural ambiguity will be called the direct object/sentential complement (DO/SC) ambiguity because *the article* is temporarily ambiguous between being the direct object of the main clause or the subject of the embedded clause.

Readers typically slow down at reading the first verb in the embedded clause (*had* in 5b), because they have initially interpreted *the article* as the direct object of the main clause verb *read* under the guidance of the universal parsing heuristic *the minimal attachment principle* (Frazier & Fodor, 1978), which posits that the parser favors the syntactically simpler structure. When *had* is encountered, the initial direct object analysis must be revised to accommodate for the fact that *had* lacks a subject and therefore the preceding noun *the article* must be removed from the direct object role of *read* and be attached as the subject of *had*. Such slowing down in reading, which

has been taken in the psycholinguistic literature to reflect reanalysis processes is termed a garden-path effect.

Sentences like (5b) can be disambiguated by including the complementizer *that* after the main clause verb *read*, as shown in (6).

(6) The scientist read that the article had been published two months ago.

The frequency with which the main clause verb appears with a particular type of complement (termed verb bias) has been found to influence the garden-path effect (Ferreira & Henderson, 1990; Garnsey et al., 1997; Kennison, 2001; Osterhout et al., 1994; Pickering & Traxler, 1998; Pickering & Traxler, 2003; Pickering et al., 2000; Traxler, 2005; Trueswell & Kim, 1998; Trueswell et al., 1994; Trueswell et al., 1993; Sturt et al., 1999). Consider (7), in which the verb *understand* biases towards taking a direct object (DO-bias verb) and (8), in which the verb *admit* biases towards a sentential complement (SC-bias verb),

(7) The club members understood the bylaws would be applied to everyone. (DO-bias)
(8) The ticket agent admitted the mistake might be hard to correct. (SC-bias)

In (7), the parser anticipates a direct object after encountering *understood*, and thus experiences garden-path effect at *would*, which is the earliest point in the sentence that signals that the analysis of *the bylaws* as the direct object of *understood* is incorrect. In contrast, the parser expects an embedded clause when encountering *admitted* in (8) and thus is less committed to the analysis of *the mistake* as the direct object of *admitted*. As a result, there is less difficulty at the subordinate clause verb *might*.

Previous studies have shown that verb bias has a rapid effect on the processing of the subsequent words (e.g., Garnsey et al., 1997; Trueswell & Kim, 1998; Wilson & Garnsey, 2009).

For instance, in a self-paced reading with fast priming experiment, Trueswell and Kim (1998) showed that the structural biases of verbs could be retrieved rapidly and be used to influence the interpretation of the following structure. In this study, readers read sentences in which the main clause verb had DO-bias and the sentences were ended with sentential complements, as in (7). Readers pressed a button to read each word of the sentence. Before the main clause verb (e.g., *understood*) was displayed, a prime verb was displayed for 39 milliseconds, which was too brief for participants to recognize. The structural biases of the prime verbs were manipulated. Trueswell and Kim found that processing difficulty at the disambiguating verb *would* was alleviated when the prime verb had SC-bias and exacerbated when it had DO-bias.

In contrast to the large number of studies on verb bias effects, only a few have compared the effects of verb bias and plausibility on the resolution of DO/SC ambiguity (Garnsey et al., 1997; Trueswell, 1996). For instance, Garnsey et al. (1997) manipulated the structural biases of the main clause verbs (DO-bias, Equi-bias, and SC-bias) and the plausibility of the ambiguous noun as the direct object of the preceding verb (*The club members understood the bylaws…* vs. *The club members understood the pool…*) in DO/SC sentences. Equi-bias verbs were those that were used equally often with DO and SC structures (e.g., *declare*). They found that verb bias trumped plausibility in guiding sentence interpretation. When verbs did not bias toward either type of continuation, reading time at the disambiguating verb was affected by plausibility. Reading times were faster when the ambiguous noun was implausible as the direct object than when it was plausible, suggesting that the parser committed less to the direct object analysis when such analysis was implausible, leading to the relative ease of recovery at the disambiguating verb. However, when the verb biased towards either direct object or embedded

clause, the plausibility manipulation did not show any effect, indicating that plausibility did not have a chance to influence parsing in the presence of verb bias.

In the type of sentences that Garnsey et al. investigated, verbs appeared earlier in the sentence than the temporarily ambiguous noun, raising the possibility that verb bias trumped noun plausibility because the verb came first. Trueswell (1996) eliminated such concern by using sentences with main clause/reduced relative clause ambiguities, as in *The room searched by the police contained the missing weapon*, in which the first verb *searched* is temporarily ambiguous between being the main clause verb or the verb in the reduced relative clause. The temporary ambiguity arises because the past participle form of the verb, which is what is required in the reduced relative clause, is identical to the past tense form, which is what is required when it is the main verb. Note that not all verbs are ambiguous in this way. Some have different past and past participle forms, such as *saw* and *seen*. Since the main clause analysis is simpler than the reduced relative clause analysis, readers typically experience processing difficulty at *by the lawyer*, which is the earliest signal in the sentence that the main clause interpretation is incorrect. Trueswell manipulated the plausibility of the noun preceding the verb as the agent of the verb (*The room searched…* vs *The thief searched…*) together with the frequency of the verb's usage as a past tense or a past participle verb. They found that plausibility had an effect only when the verb was biased towards past participle form but not when it was biased towards past tense form. When the verb was more often used as a past participle, reading times at the disambiguating words were faster when the noun was implausible as the agent of the verb (*The room searched…*) than when it was plausible (*The thief searched…*). However, when the verb was more often used as a past tense verb, implausible-as-subject noun and verb combinations did not alleviate the processing difficulty at the disambiguating word relative to plausible-as-subject noun and verb

combinations. Trueswell (1996) and Garnsey et al (1997) converged to show that plausibility had more restricted effects than verb bias on the initial interpretation. One explanation is that verb bias information is retrieved as soon as the verb is recognized, but plausibility must be computed and evaluated online for particular word combinations. Several researchers have argued that plausibility is most likely to have an effect when other constraints have narrowed the number of structural possibilities down to a limited number and plausibility can play a role in choosing one over the others (MacDonald et al. 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Trueswell, 1996; Trueswell & Tanenhaus, 1994). This is consistent with Garnsey and colleagues' finding that plausibility played a determining role when there were just two possibilities and no other cues made one of the possible structures more likely.

Studies on the effects of verb bias and plausibility on English sentences showed that verb bias plays a larger role than plausibility in guiding sentence interpretation. What about in L2 sentence processing? Since verb subcategorization information is implicit knowledge that probably cannot be taught in classroom, but instead can only be acquired through substantial exposure to the target language, whether second language learners are able to learn verb bias information and use it fast enough to generate predictions about the upcoming structure in the way that the native speakers do has attracted some attention in the L2 psycholinguistic literature. In terms of how second language learners parse sentences in their L2, there is a hypothesis that they use syntactic information qualitatively differently from native speakers. The claim is that L2 learners underuse syntactic information and consequently that the syntactic structure they build is shallower and less detailed than those built by native speakers. To compensate, they rely on lexical-semantic cues such as plausibility more than native speakers do (the Shallow Structure Hypothesis, Clahsen & Felser, 2006a, 2006b). However, how learners use the frequency

information about verbs' subcategorization preferences has not been addressed in formulations of the Shallow Structure Hypothesis. On the one hand, verb bias is lexically-associated information that is stored in the lexicon and retrieved when words are recognized. Such information might be considered to be part of the lexical information the Shallow Structure Hypothesis claims that L2 learners rely on. On the other hand, verb bias is about structure, so L2 learners may not use it to the extent that native speakers do. Several studies on L2 learners' use of verb bias information thus far have revealed that L2 learners are able to learn verb bias information that is specific to their L2 and use it fast enough to guide on-line parsing in the L2 (Dussias & Cramer Scaltz, 2008; Dussias, Marful, Gerfen, & Bajo Molina, 2010; Frenck-Mestre & Pynte, 1997), even if such information cannot be used in the same way in their L1 because the L1 and L2 use different word orders (Lee, Lu, & Garnsey, 2013).

Dussias and Cramer Scaltz (2008) tested native speakers of Spanish who were and were not learning English as L2. Native speakers were asked to complete forty sentence fragments in Spanish that started with a proper noun and a verb. The forty Spanish verbs were translations of twenty DO-bias and twenty SC-bias English verbs normed in Garnsey et al., (1997). For those native speakers, fewer than half of the translated Spanish verbs had the same bias as the English versions had been found to have for native English speakers. Thus, meaning appears not to be the primary determinant of verb bias. If it were, then verbs with approximately equivalent meanings in the two languages should also have the same structural biases. When advanced L1-Spanish learners of L2-English completed the norming in English, they showed similar verb biases as native English speakers, showing that they were able to learn the biases for English verbs when they conflicted with the biases of Spanish verbs with similar meanings. In addition, L2 learners

showed native-like pattern in a self-paced reading experiment in English, experiencing processing difficult when the sentence continuation was incongruent with the verb's bias.

Similarly, Frenck-Mestre and Pynte (1997) found that verb bias could be retrieved by L2 learners fast enough to generate predictions about upcoming structure. Experiment 1 used French sentences with prepositional phrase attachment ambiguity, as in *They accused the ambassador of espionage/Indonesia but nothing came of it*, and manipulated the lexical property of the main clause verb (ditransitive vs monotransitive). L1-English learners of L2-French behaved like native French speakers in preferring VP attachment following ditransitive verbs and NP attachment following monotransitive verbs, indicating that they anticipated two arguments after ditransitive verbs but only one argument after monotransitive verbs, just like native speakers. Experiment 2 showed in addition that L2 learners were able to use properties of verb argument structure that were specific to their L2 even when such properties conflicted in their L1 and L2. For instance, in English the verb *obey* is optionally transitive and *bark* is intransitive, while both verbs are intransitive in French. When reading sentences in French, L1-English speakers of L2-French did not slow down at the disambiguating word *showed* when processing French sentences like *Every time the dog obeyed/barked the pretty girl showed her approval* with either type of verb, indicating that they used verb argument structure information specific to the L2 in processing their L2 French, even when it conflicted with information from their L1.

Spanish and French both have the same default SVO word order as English, so L1-speakers of both languages have L1 experience with verbs preceding everything except the subject and thus becoming available early enough to generate predictions about what might follow. Lee et al., (2013) investigated whether L2 learners were able to learn L2-specific verb bias information if such information was not useful in the same way in their L1. Korean is an

SOV language that places verbs at the ends of clauses, so verb-based information cannot be used to generate predictions about likely upcoming structure in Korean. In addition, unlike in English, where the complementizer *that* is optional, a clause-final complementizer particle *ko* is obligatory in Korean. Thus, L1-Korean speakers do not have L1 experience with predicting upcoming structure based on either verb bias or the complementizer, but they do have experience with an end-of-clause complementizer that is a perfect cue to an embedded clause. Since the complementizer is a perfect cue on its own, it is possible that Korean speakers would never learn to associate structural biases with verbs, which is a much less reliable cue. In Lee et al.'s study, higher proficiency L1-Korean speakers of L2-English were compared with lower proficiency learners in their use of verb bias and complementizer cues. Results showed that lower proficiency L2-learners reading English sentences needed both cues to be present to derive any benefit, while higher proficiency learners were able to derive benefit from each of the cues separately. In addition, the higher proficiency learners were able to combine the cues to some extent, since they gained more benefit from the presence of the complementizer when the verb was DO-biased than when it was SC-biased, but they did not achieve the optimally efficient interactive pattern seen in the native English speakers, for whom either cue alone was entirely sufficient. So, L1-Korean learners of L2-English proved to be able to learn to associate structural biases with verbs in English and use them predictively in spite of the fact that they had no L1 experience with predicting upcoming structure based on verb bias, and in spite of the fact that in Korean the complementizer is completely reliable cue while verb bias is less reliable.

The fact that the higher proficiency L1-Korean learners did not achieve the optimal native pattern might well be true for any L2 learners, regardless of the properties of their L1, simply because they don't have as much experience as native English speakers. Alternatively,

however, it is possible that the fact that verb bias is not available early enough in the sentence to base predictions on in Korean that is responsible for the failure to achieve the native pattern, making it important to test L2-English learners whose L1 has a word order placing verbs earlier in the sentence. Mandarin differs from Korean and is similar to English in two important respects. First, Mandarin places verbs early in the sentence, with the same SVO order as English, thus allowing verb bias to be used to develop expectations about the upcoming structure. Second, Mandarin has no complementizer in sentences like the ones used here, although DO/SC ambiguity can be disambiguated by adding an optional comma after the main clause verb (e.g., 那个售票员承认，错误还没被发现。*That ticket agent admitted "that" the mistake had not been caught*.). However, the optional comma is used less often in Mandarin than the optional *that* is in English. Thus, DO/SC sentences in Mandarin are rather similar in Mandarin and English in both the early availability of verb bias and the optionality of a disambiguating cue, though the kind of disambiguating cue is different and it is less often available in Mandarin. There is one important way, however, that these types of sentences differ between the two languages. In English, the verb in the embedded clause completely disambiguates the sentence. It simply must have a subject so the noun preceding it has to be its subject and cannot be the object of the preceding verb. In Mandarin, in contrast, because subjects can be dropped if they are recoverable from context, the embedded verb might not have an expressed subject, so the noun preceding it can remain the object of the preceding verb. Another option is that the critical noun is both the object of the verb preceding it and the subject of the verb following it. This is an issue that will become important in Chapter 3.

These differences between Korean and Mandarin suggest that L1-Mandarin learners of L2-English might use verb bias and complementizer cues differently from L1-Korean learners.

L1-Mandarin learners may find it easier to learn and use the biases of English verbs earlier. As for their ability to use the complementizer *that*, it is not clear what to predict. Given that they have experience with similar but much less frequent usage of commas, it may be easy for them to learn to rely on the complementizer when it is present. Alternatively, though, the fact that the English cue is a separate word, and furthermore one that has many other usages besides a complementizer (pronoun, demonstrative, relative pronoun, …) may make it a difficult cue for Mandarin-L1 learners to learn to rely on. Experiment 1 in this chapter aims to test these predictions.

As described earlier, native English speakers are more influenced by verb bias than by the plausibility of the temporarily ambiguous noun as the direct object of the preceding verb (Garnsey et al. 1997; Trueswell, 1996). It is possible that the same would not be true for L2-learners because L2 learners have been argued to rely heavily on lexical-semantic information rather than structure. The evidence supporting that argument has come primarily from studies of relative clause attachment ambiguity (Felser, Roberts, Marinis, & Gross, 2003; Papadopoulou, 2005; Papadopoulou & Clahsen, 2003) and filler-gap dependencies (Dussias & Pinar, 2010; Felser & Roberts, 2007; Marinis, Roberts, Felser, & Clahsen, 2005; Williams, Möbius, & Kim, 2001). For example, Papadopoulou and Clahsen (2003) examined relative clause attachment preference in L1-Spanish, L1-German, and L1-Russsian learners of L2-Greek when reading Greek sentences equivalent to *Someone shot the servant of the actress who was on the balcony*. The relative clause *who was on the balcony* could be attached to either the high noun, *the servant*, or the low noun, *the actress*. Native Greek speakers showed an on-line preference for attachment to the high noun in this sentence, while they showed a preference for attachment to the low noun if the sentence had the preposition *with* instead of the genitive *of (Someone shot the servant with*

the actress who was on the balcony.). The relative clause is attached to the low noun when the low noun is preceded by *with*, because the low noun receives the theta role from the preposition *with* and the relative clause is processed within that thematic domain (see Frazier & Clifton, 1996, for the processing of non-primary structures such as relative clauses). None of the L2 learner groups showed attachment preference in on-line measures in *of* sentences, although they all showed high-attachment preference in the off-line interpretation. In contrast, they preferred low-attachment in both on-line processing and off-line interpretation in *with* sentences (see also Felser et al., 2003, for similar results). These results suggested that during on-line parsing, native speakers were able to use both the lexical-semantic cue, such as *with*, and the syntactic parsing strategy, which in this case was the default high-attachment preference parsing strategy in Greek, to guide their on-line building of syntactic structures. On the contrary, non-native speakers relied on lexical-semantic cues to guide on-line parsing, as in the *with* sentences. When such lexical-semantic cues were missing, as in the *of* sentences, they were not capable of using the syntactic parsing strategy that was specific to their L2. Furthermore, Marinis et al. (2005) showed that the syntactic structures built by second language learners were less detailed than those built by native speakers. In this study, L2 learners were compared with native English speakers on reading sentences like *The manager who the secretary claimed that the new salesman had pleased will raise company salaries* and sentences like *The manager who the consultant's claim about the new proposal had pleased will hire five workers tomorrow*. Native speakers were slower at reading *claimed* in the first sentence than *claim* in the second sentence, but were faster at reading *pleased* in the first sentence than *pleased* in the second sentence. L2 learners, however, did not show this pattern. This result indicated that native speakers posited an intermediate gap after *claimed* in the first sentence and subsequently were facilitated in integrating the filler to its

subcategorizer (*pleased*). L2 learners, in contrast, were insensitive to the intermediate gap site and their gap-filling processes were guided by the association of lexical items instead (see also Felser & Roberts, 2007).

With respect to the use of semantic information, Williams, Möbius and Kim (2001) showed that non-native speakers used plausibility information to recover from garden-pathing just like native speakers. They asked native speakers and L2 learners of English with Korean, Chinese and German native language background to read sentences like *Which girl/river did the man push the bike into late last night?*, in which the plausibility of the displaced *wh*-phrase was either plausible or implausible for the first potential gap site (after *push*). Results from a self-paced reading experiment showed that when the *wh*-phrase was implausible (*which river*), both L2 learners and native speakers were slower at *pushed* and faster at *the bike*, compared to when the *wh*-phrase was plausible (*which girl*). Thus, L2 learners were similar to native speakers in their ability to use plausibility to assist recovery from garden-pathing. Felser and Cunnings (2012) reported that L2 learners relied on discourse, but native speakers relied on syntactic information to interpret reflexive pronouns. Online reflexive interpretation is constrained by the binding principle for native speakers, but L2 learners initially associated reflexives with the most salient antecedent in the discourse, even if that violated the binding principle. Thus discourse plausibility seemed to override syntactic principles for L2 learners. Discourse effects have also been found in L2 processing of PPs that could be attached to either a VP or an NP, as in *The policeman watched the spy with binoculars*. In Pan and Felser (2011), L1-Chinese learners of L2-English, but not native English speakers, exhibited attachment preferences congruent to biases from the discourse, such that they preferred VP or NP attachment when the context biased

towards such attachment, suggesting that L2 learners relied more on discourse semantic information in their construction of syntactic structures.

One problem with much of the previous work showing that L2 learners cannot use syntactic information to the fullest degree has come mostly from null results, and has not gone unchallenged. Counter-evidence has been found in a number of studies (Cunnings, Batterham, Felser, & Clahsen, 2009; Omaki & Schulz, 2011; Pliatsikas & Marinis, 2013; Witzel, Witzel, & Nicol, 2012). There is also evidence showing that observed L1-L2 differences in the apparent use of syntactic information during sentence processing may actually be due to factors such as differences in the availability of the required cognitive resources, proficiency, task demands, and properties of the L1, rather than fundamental differences in L2 parsing mechanisms (Hopp, 2006; Jackson, 2008; Jackson & Bobb, 2009; Jackson & Dussias, 2009; Juffs, 1998; Kim, Baek, & Tremblay, 2015; Lim & Christianson, 2013a, 2013b; van Hell & Tocowicz, 2010; Sabourin & Stowe, 2008). For instance, using the same sentence structures as in Marinis et al. (2005), Pliatsikas and Marinis (2013) demonstrated that L2 learners' sensitivity to intermediate gaps depended on how the L2 was acquired. Learners who acquired the second language in naturalistic settings showed evidence of on-line use of intermediate gaps, while learners with primary classroom exposures did not. Cunnings et al. (2009) and Omaki and Schulz (2011) found that L2 learners made use of relative clause island constraints to guide their filler-gap formation in the same way as native speakers. They compared sentences like *The city/book that the author wrote regularly about was named for an explorer* (no island) vs *The city/book that the author who wrote regularly saw was named for an explorer* (island), in which plausibility of the filler (*the city* vs *the book*) and presence or absence of the relative clause island were manipulated. Native speakers and L2 learners both slowed down at the earliest potential gap

position (after *wrote*) if the filler was an implausible direct object (*the city*) than if it was plausible (*the book*) as the direct object for the verb (*wrote*) only in non-island conditions, indicating that both native speakers and L2 learners of English built detailed syntactic representation of the sentences and respected the island constraint on forming filler-gap dependencies. Kim et al. (2015) showed that properties in the L1 affected L2 learners' sensitivity to island constraints. While L1-Spanish learners of L2-English were able to use their knowledge of island constraints to avoid the formation of ungrammatical *wh*-dependencies during on-line parsing, L1-Korean learners of L2-English were unable to do so.

The aforementioned studies illustrate that the evidence is quite mixed about the extent to which L2 learners are able to make online use of syntactic constraints during sentence processing, and about whether the structures they build are shallower than those built by native speakers. However, all of the previous studies have found that L2 learners use lexical-semantic information during on-line processing, and some have suggested that they rely on it more heavily than native speakers do. With respect to L2-English learners' processing of DO/SC ambiguous sentences, it is not obvious what to predict about the relative contributions of verb bias and plausibility. Verb bias is lexically specific information about what structures a verb is most likely to appear in. Should that be considered syntactic information that L2-learners might have trouble with, or should it be considered lexical-semantic information that would be easy to use? A verb's meaning certainly contributes a great deal to its structural biases, but the differences in biases found by Dussias et al. (2010) for English and Spanish verbs with very similar meanings shows that meaning cannot be the only determinant. Lee et al. (2013) has already shown that L1-Korean learners of L2-English did learn to make use of verb bias in such sentences. However, they did not also consider the role that plausibility might play. In Experiment 2 in this chapter, verb bias

and plausibility are both manipulated and pitted against each other to examine the relative importance of plausibility and verb bias in L2-English sentence processing. Native English speakers rely more heavily on verb bias than plausibility, but the opposite might be true for L2-English learners.

Experiment 1

Method

Participants

A control group of 32 native English speakers (22 males, mean age 20) and an L2 group of 78 L1-Mandarin learners of L2-English (26 males, mean age 24) participated in the experiment. All were undergraduate or graduate students at the University of Illinois at Urbana-Champaign, had normal or corrected-to-normal vision, gave written informed consent, and received course credit or payment for taking part. The native speaker control group was the group reported in Lee et al. (2013).

All L2 learners completed at least their middle school education in Mainland China or Taiwan and lived in English-speaking countries for less than five years. None of them lived in the U.S. before the age of 15. Additional language background information is summarized in Table 1. L2 learners' proficiency was assessed using a cloze test (i.e., fill-in-the-blanks test) that contained forty blanks (adopted from P. Dussias at Pennsylvania State University, personal communication; see Tremblay, 2011, for validity and reliability of using the cloze test to assess L2 proficiency). L2 learners were divided into two proficiency groups based on a median split of their cloze test scores (lower proficiency group <32; higher proficiency group: ≥32) to examine

whether their ability to use verb bias and the complementizer *that* cue improved as their proficiency increased.

Table 1. Language background information of the L1-Mandarin L2-English group in Experiment 1. Ranges are shown in parentheses.

	All Learners	Lower Proficiency Group	Higher Proficiency Group
Number of Participants	78	40	38
Age	24 (18-37)	23 (18-37)	24 (18-35)
Proficiency score	31 (21-37)	28 (21-31)	34 (32-37)
Age at start of English classroom instruction	10 (4-16)	10 (5-16)	10 (4-16)
Age at first residence in English-speaking countries	21 (15-33)	21 (15-33)	22 (15-30)
Duration of residence in English-speaking countries	30 months (6-60)	28 months (6-60)	32 months (6-60)
Daily use of English	50% (5%-95%)	47% (10%-85%)	54% (5%-95%)

Materials and Design

Ten DO-bias and ten SC-bias verbs were each used four times to construct 80 sets of sentences, with each set containing ambiguous and unambiguous versions of the same sentence, as shown in (9) (see Lee et al., 2013, for a full list of experimental sentences). Unambiguous sentences were disambiguated by adding the complementizer *that* after the main clause verb. In the ambiguous version, the ambiguous noun (e.g., *the bylaws*) was temporarily ambiguous between being the direct object of the preceding verb (e.g., *understood*) or the subject of an upcoming embedded clause, whereas in the unambiguous version, such temporary ambiguity

was blocked by the presence of *that* between the verb and the ambiguous noun. All sentences started with a subject noun phrase that contained three words (e.g., *the club manager*), followed by a main clause verb that was either biased towards taking direct objects or embedded clause complements. The ambiguous noun following the main clause verb contained two words (e.g., *the bylaws*), which were then followed by the disambiguating region that contained the subordinate clause verb and the word immediately following it. Care was taken when selecting the two words for the disambiguating region in each sentence. All disambiguating verbs and the words immediately following them were auxiliary verbs such as *were*, *could*, *would* and *had*, so that the properties of the disambiguating words did not differ between items with DO bias and SC bias verbs. All critical sentences turned out to have the embedded clause structures.

(9) Example stimuli in Experiment 1:

DO-bias verb
Ambiguous: *The club members understood the bylaws would be applied to everyone.*
Unambiguous: *The club members understood that the bylaws would be applied to everyone.*

SC-bias verb
Ambiguous: *The ticket agent admitted the mistake might be hard to correct.*
Unambiguous: *The ticket agent admitted that the mistake might be hard to correct.*

Verbs used in the experiment all met the following criteria: DO-bias verbs were followed at least twice as often by direct object completions as by sentential complement completions in the sentence completion norming task reported in Garnsey et al. (1997), which asked 108 native English speakers to complete one hundred sentence fragments that began with a proper name and a verb that could take both direct objects and embedded clauses (e.g., *Bill believed…*). The reverse was true for SC-bias verbs: there were at least twice as many sentential complement completions as direct object completions generated by participants in the norming task. The ten

DO-bias verbs and ten SC-bias verbs used in the present study were matched on the number of letters, $F<1$, and frequency (Francis & Kucera, 1982), $F<1$. Verb properties are summarized in Table 2.

Table 2. Properties of the verbs used in Experiments 1 and 2.

	DO bias strength (%)	SC bias strength (%)	Mean length	Mean log frequency
DO-verbs	76	13	8.1	1.9
SC-verbs	17	59	7.9	1.7

To ensure that any effect found at the disambiguating region was caused only by the biases of the verbs, two plausibility norming tasks were conducted to examine whether the ambiguous nouns were equally plausible as the direct object of the preceding verb and as the subject of the embedded clause between DO-bias and SC-bias items. The plausibility of the ambiguous noun as the direct object was rated by asking a separate group of 56 native speakers of English to judge the plausibility of the subject, verb and ambiguous noun combinations on a 1 (very implausible) to 7 (very plausible) scale, as shown below in (10). Results showed that the ambiguous nouns following DO-bias verbs were rated as slightly more plausible than those following SC-bias verbs (6.5 vs. 6.2, $F(1,78)=5.4$, $p<.05$). This replicated previous findings that plausibility ratings of the ambiguous noun as the direct object were affected by biases of the verbs. Participants tended to rate the nouns as more plausible when they followed verbs that preferred direct object completions (Garnsey et al. 1997). This raised a possibility that the effects found at the disambiguating region would reflect the combination of the effects from verb bias and plausibility. According to constraint-based parsing models (e.g., MacDonald, 1994; Trueswell, Tanenhaus, & Kello, 1993), the activation of the direct object structure should be

ranked higher than the embedded clause structure after the parser encounters a DO-bias verb. This ranking would be further strengthened by the higher plausibility of the ambiguous noun as the direct object following DO-bias verbs. The reverse was true for SC-bias items, in which both verb bias and plausibility worked in the same direction to rank the embedded clause to be the more likely structure that the sentence would develop into. However, it is unlikely that the small difference in plausibility would have a detectable effect in sentences with strongly biased verbs, given previous findings from studies specifically manipulating plausibility (Garnsey et al. 1997).

A separate norming study that assessed the plausibility of the ambiguous noun as the subject of the embedded clause was conducted with twelve native English speakers, who rated on a 1 (very implausible) to 7 (very plausible) scale the plausibility of sentence fragments such as (11) as the beginning of a sentence. The ambiguous nouns for both DO- and SC-items were both rated as highly plausible (mean DO: 6.1; mean SC: 6.1) and did not differ between verb types, $F<1$. The properties of the ambiguous nouns in Experiment 1 are summarized in Table 3.

(10) *The club members understood the bylaws.*
The ticket agent admitted the mistake.

(11) *The club members understood that the bylaws…*
The ticket agent admitted that the mistake…

Table 3. Properties of the ambiguous nouns used in Experiment 1.

	Mean length	Log frequency	Plausibility as the direct object	Plausibility as the clause subject
DO-items	7.4	1.3	6.5	6.1
SC-items	7.1	1.4	6.2	6.1

Critical sentences were distributed over two lists according to a Latin Square design, so that all participants saw an equal number of trials of each condition and no one saw two versions from the same sentence set. A total of eighty distractors was added to each list for a total of 160 trials per list. In twenty distractors, the noun phrase immediately following the main clause verb turned out to be the direct object, so that across the experiment, it was not always the case that the noun following the main clause verb must be reanalyzed as the subject of the embedded clause. The rest of the sixty distractors had various syntactic structures. All sentences, including critical items and distractors, were followed a comprehension question that targeted the content of various parts of the sentences (e.g., *Were the bylaws applied fairly?*). The comprehension questions to critical sentences did not probe the initial misanalysis (e.g., *Did the club members understand the bylaws?*). All sentences were pseudo-randomized once so that no two critical sentences from the same condition appeared consecutively and were presented to all participants in the same order in all lists.

Procedure

Participants sat in a dimly lit sound-attenuated booth in front of a 23-inch LCD monitor. A total of 160 sentences was presented word by word in white 26-point Arial font on a black background in a non-cumulative moving window self-paced reading paradigm, controlled by the Presentation software package. Each trial began with a trial number that stayed at the left side of the screen for one second. The participant then pressed a button on a Cedrus-830 response box to read each word successively at their own speed. Each time they pressed the button, the next word appeared and the previous word reverted to the mask character. All sentences were presented on a single line. Following each sentence, a comprehension question was presented and participants

pressed *yes* or *no* buttons to indicate their answers. Feedback about question response accuracy was given after each trial. A "Too Slow" message was presented when participants did not make a response within four seconds. Sentences were divided into four blocks with forty sentences each, and participants took a break after each block. A practice block of five trials was added at the beginning. The entire experiment took 30-45 minutes to complete.

Results

Comprehension Accuracy

Trials on which participants pressed the *yes* or *no* button before the comprehension questions were displayed, and trials on which participants did not make a response within the four second limit were excluded from the analysis of comprehension accuracy. On average, the accuracy rate for native English speakers was 92% (range 87%-96%) and for L1-Mandarin learners of L2-English was 86% (74%-96%).

Comprehension accuracy for the experimental sentences was analyzed using a logit mixed-effect model (Jaeger, 2008) that included ambiguity, verb bias, language group and their interactions as fixed effects, and random intercepts and slopes for subjects and items. Language groups were coded so that the model compared native speakers with non-native speakers and high proficiency non-native speakers with low proficiency non-native speakers. Results showed main effects of language group, with native speakers answering the comprehension questions more accurately than non-native speakers (90% vs. 87%, $p<.001$) and high proficiency L2 learners answering more accurately than low proficiency L2 learners (88% vs. 85%, $p<.01$). The main effect of ambiguity was marginal ($p=.1$), with unambiguous sentences being answered more accurately than ambiguous sentences (88% vs. 86%).

Reading Times

Prior to data analysis, word-by-word reading times that were faster than 100 milliseconds (ms) or slower than 2000 ms were excluded, leading to a loss of 0.4% of the native speakers' data and 1.7% of the L1-Mandarin speakers' data. Reading times were also excluded from further analysis for sentences that participants pressed the *yes* or *no* button before the comprehension questions were displayed or failed to respond to the questions within the four second limit, affecting 2.3% of the native speakers' data and 6.1% of the L2 learners' data. To remove individual differences in reading speed, statistical results reported below were based on length-corrected residual reading times computed separately for each participant by entering their reading times for each word in all sentences (including distractors) into a regression equation that took reading times as the dependent variable and word length as the independent variable, and then subtracted the predicted reading times from the actual reading times (Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994). In Table 4 below both raw reading times and residual reading times at the disambiguating region are reported.

Statistical analyses were performed on the disambiguating region and the ambiguous noun region, since these two regions reflected the parser's commitment to the initial misanalysis and the ease of recovery from such misanalysis. Reading times at the disambiguating region were obtained by averaging across the reading times of the disambiguating verb (e.g., *would*) and the word immediately following it (e.g., *be*) to capture spill-over effects that often occur in self-paced reading experiments. Similarly, reading times at the ambiguous noun region were calculated by averaging across the reading times of the determiner and the head noun (e.g., *the bylaws*). Reading times in each region were then analyzed using linear mixed-effect models in R (R Development Core Team, 2008). For all analyses, the initial model included a maximal

random effects structure that included all fixed effects and interactions among them, random intercepts and random slopes for all fixed effects and their interactions for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013). If the maximal model failed to converge, the random slopes of fixed effects were removed, one at a time, based on the values in the Hessian matrix. All categorical fixed effects were coded using contrast coding and continuous fixed effects were centered to avoid colinearity. The final models reported here were the most complex model that converged. Estimates, standard errors, and *t*-values were reported, with $t>2$ being interpreted as significant.

The disambiguating region. The multi-level model at this region that included ambiguity, verb bias, language group (native vs. non-native, high proficiency L2 group vs. low proficiency L2 group) and their interactions as fixed effects, and random intercepts and random slopes of the ambiguity factor for subjects and items revealed a main effect of ambiguity, with ambiguous sentences being read slower than unambiguous sentences (425 vs. 414 ms), an interaction between verb bias and ambiguity, with the ambiguity effect of DO-items (15 ms) being larger than that of SC-items (5 ms; DO: ambiguous 436 ms, unambiguous 421 ms, SC: ambiguous 413 ms, unambiguous 408 ms), and an interaction between verb bias and language group (native vs. non-native), with the difference between DO- and SC-items being bigger in the non-native speakers group (21 ms) than the native speakers group (10 ms; native speakers: DO 352 ms, SC 342 ms; non-native speakers: DO 461 ms, SC 440 ms). Statistical results are summarized in Table 5. Residual reading times at the disambiguating region for all conditions are plotted separately for native and non-native groups in Figure 1 and summarized in Table 4.

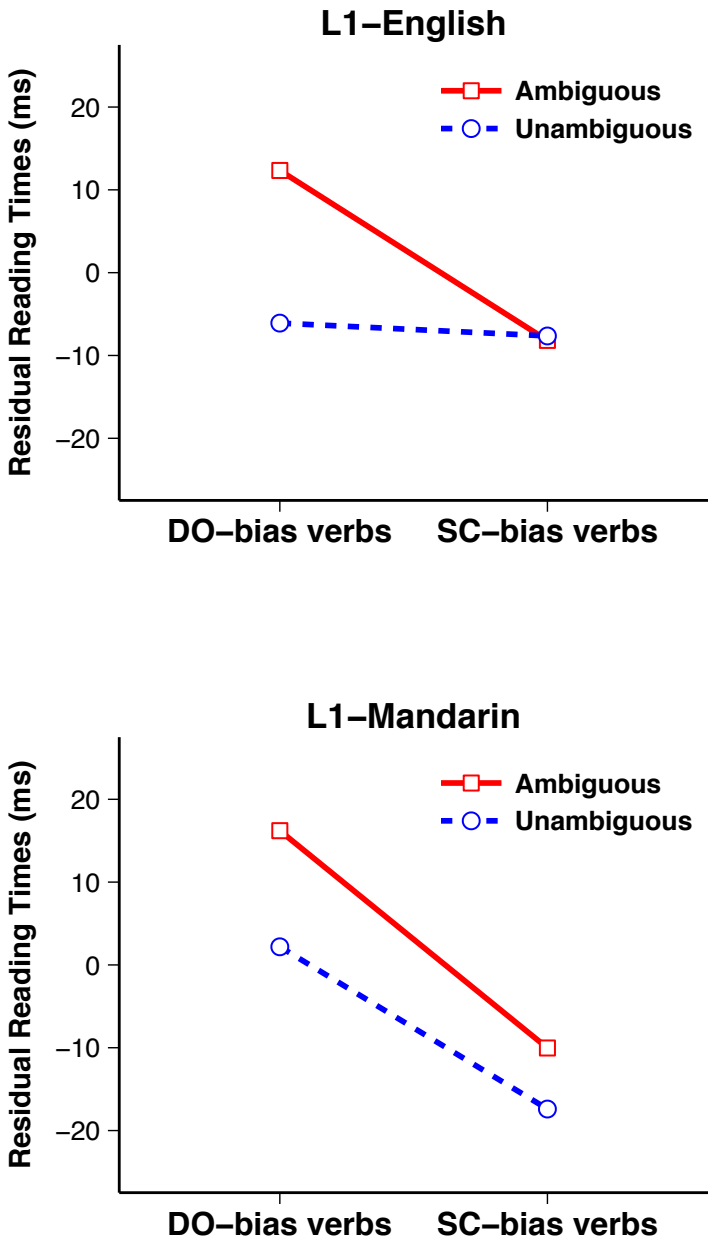

Figure 1. Residual reading times at the disambiguating region for native English speakers and L1-Mandarin learners of L2-English.

Table 4. Raw reading times and residual reading times at the disambiguating region in Experiment 1.

	DO-bias verbs			SC-bias verbs		
	Ambig	Unambig	Ambiguity Effect	Ambig	Unambig	Ambiguity Effect
Raw RT (ms)						
L1-English	362	343	19	341	342	-1
L1-Mandarin	469	454	15	444	436	8
Residual RT (ms)						
L1-English	12	-6	18	-8	-8	0
L1-Mandarin	16	2	14	-10	-17	7

Table 5. Fixed effects of the mixed-effect model on the residual reading times at the disambiguating region. The model compared native speakers to non-native speakers (Group 1) and high proficiency non-native speakers with low proficiency non-native speakers (Group 2).

Fixed Effects	**Coefficient**	***SE***	***t*-value**
(Intercept)	0.03	5.61	0.01
Verb bias	19.06	10.78	1.77
Ambiguity	10.18	2.86	3.56*
Group1	0.01	5.94	0.00
Group2	-6.89	4.87	-1.41
Verb bias x Ambiguity	10.89	5.35	2.04*
Verb bias x Group 1	-15.16	7.57	-2.00*
Verb bias x Group 2	0.04	6.27	0.01
Ambiguity x Group 1	-2.57	8.13	-0.32
Ambiguity x Group 2	3.13	6.74	0.47
Verb bias x Ambig x Group 1	15.89	15.14	1.05
Verb bias x Ambig x Group 2	13.91	12.57	1.11

Since there was no effect associated with the comparison between the high proficiency and low proficiency L1-Mandarin L2-English groups, which suggested that proficiency did not have any effect on either the ambiguity and verb bias manipulations, further analyses involving the L2 learners did not break down into high and low proficiency groups. Analyses within each language group were conducted to examine the ambiguity, verb bias and the interaction between them for native speakers and Mandarin learners of English separately. Within the native speakers control group, there was a main effect of ambiguity, with ambiguous sentences being read slower

than unambiguous sentences (352 vs. 342 ms), and an interaction between verb bias and ambiguity, which resulted because the disambiguating region was read slower in ambiguous than in unambiguous sentences only when the main clause verb had DO bias (ambiguity effect 19 ms; β=18.53, *SE*=5.61, t=3.30), but not when it had SC bias (ambiguity effect -1 ms; β=.43, *SE*=4.39, t<1), as shown in Table 6. This finding replicated results from earlier studies on verb bias effects in English (Garnsey et al., 1997; Kim & Trueswell, 1998; Trueswell et al. 1994; Wilson & Garnsey, 2009), which suggests that native English speakers can actively anticipate the upcoming structure based on the verb's subcategorization bias and are only garden-pathed when the sentence develops into a structure that is incongruent with the verb's preference.

Analyses on the L1-Mandarin L2-English group revealed a main effect of verb bias, with DO items read slower than SC items (461 vs. 440 ms) and a main effect of ambiguity (Ambiguous: 455 ms; Unambiguous 445 ms), as shown in Table 7. Although the disambiguating region of ambiguous sentences were read slower than unambiguous sentences only after DO bias verbs (t=2.74) but not after SC bias verbs (t=1.55), this difference was not big enough to produce an interaction between verb bias and ambiguity in the L1-Mandarin group. These results indicated that L1-Mandarin learners of L2-English were able to use both verb bias and complementizer cues, but that neither cue alone was sufficient in the way that it is for native speakers.

Table 6. Fixed effects of the mixed-effect model on the residual reading times at the disambiguating region for the native English speakers control group.

Fixed Effects	Coefficient	*SE*	*t*-value
(Intercept)	0.04	6.50	0.01
Verb bias	11.42	11.96	0.96
Ambiguity	8.98	3.47	2.59*
Verb bias x Ambiguity	19.06	6.57	2.90*

Table 7. Fixed effects of the mixed-effect model on the residual reading times at the disambiguating region for the L1-Mandarin L2-English group.

Fixed Effects	Coefficient	*SE*	*t*-value
(Intercept)	0.04	6.50	0.01
Verb bias	22.65	11.10	2.04*
Ambiguity	10.67	3.72	2.87*
Verb bias x Ambiguity	6.06	6.92	0.88

The ambiguous noun region. The linear mixed-effect model on the residual reading times at the ambiguous noun region including verb bias, ambiguity, language groups (native vs. non-native speakers; high proficiency level vs. low proficiency level non-native speakers) and their interactions revealed a main effect of ambiguity (β=19.84, *SE*=3.56, t>5), with the ambiguous condition being read slower than the unambiguous condition (448 vs. 428 ms), a main effect of group 1 (native vs. non-native, t>2), a main effect of group 2 (high proficiency group vs. low proficiency group, t>3), because English native speakers read the ambiguous noun faster than L2 learners and learners with high proficiency read this region faster than those with low proficiency, and a significant interaction between verb bias and ambiguity (β=14.63, *SE*=5.95, t>2), which was caused by a bigger ambiguity effect at the ambiguous noun region following SC-bias verbs than DO-bias verbs (28 vs 11 ms). The ambiguity effect was significant in both DO and SC items

(DO items: t>2; SC items: t>6). Residual reading times for both native and non-native groups are plotted in Figure 2.

Separate analyses on native speakers and L2 learners revealed that the high level interaction between ambiguity and verb bias came almost entirely from the L1-Mandarin group. Native speakers of English showed a main effect of ambiguity (β=23.84, SE=3.96, t>6), but no interaction between verb bias and ambiguity (t<1). The L1-Mandarin group showed a main effect of ambiguity (β=18.47, SE=4.62, t>4), which was also modulated by an interaction between ambiguity and verb bias (β=22.07, SE=7.50, t=3). The interaction was caused by the ambiguity effect being significant only in sentences with SC-bias verbs (t>5), but not in sentences with DO-bias verbs (t<2), suggesting that verb bias information carried by the main clause verb influenced the integration of the subsequent noun phrase into the preceding structure when L1-Mandarin learners of L2-English processed their L2. When the main clause verb biased towards taking a direct object, L2 learners were more willing to attach the following noun phrase as its direct object than when the main clause verb biased towards a sentential complement structure. However, this interaction between verb bias and ambiguity in L1-Mandarin L2-English learners was not strong enough to result in a higher level interaction between verb bias, ambiguity and language group (native vs. non-native).

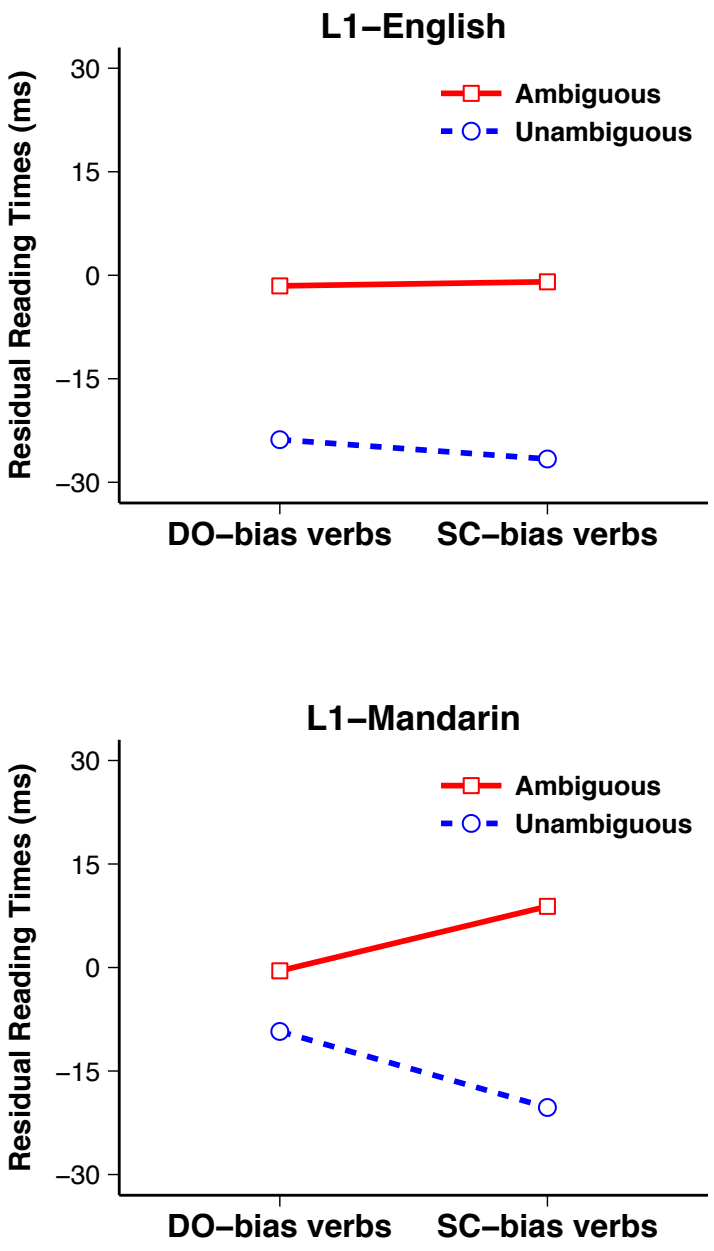

Figure 2. Residual reading times at the ambiguous noun for native English speakers and L1-Mandarin learners of L2-English.

Discussion

Experiment 1 manipulated verb bias and the presence of the complementizer *that* to compare the use of these two cues by L1-Mandarin speakers of L2-English and native English speakers. Native speakers showed the usual interaction between verb bias and ambiguity, with the ambiguity effect appearing at the disambiguating region only after DO-bias but not after SC-bias verbs, replicating previous studies with English native speakers (e.g., Garnsey et al. 1997; Trueswell et al., 1994). Furthermore, the interaction suggested an optimal, efficient and interactive use of the two cues, such that either one of them is sufficient to avoid garden-pathing, with no additional benefit when both cues are available.

With respect to L1-Mandarin learners of L2-English, the prediction was that they would be able to make use of the verb bias cue in English because verbs appear early in Mandarin sentences, too, and is therefore useful in generating predictions about the structure that a sentence is likely to develop into. This prediction was borne out, as evidenced by slower reading times on the disambiguation after DO-bias verbs than after SC-bias verbs. In addition, L2 learners' use of verb bias information was not modulated by their proficiency, suggesting that even lower proficiency L1-Mandarin learners of L2-English made use of verb bias information, perhaps because they already do the same in their L1. Learning the subcategorization preferences for English verbs and using them predictively may be easy for them because they have already done so in their first language.

Given that Mandarin does not have a complementizer that functions similarly to the complementizer *that* in English to signal an upcoming embedded clause, the prediction was that L1-Mandarin learners of L2-English might not be sensitive to such a cue. Contrary to that prediction, L1-Mandarin learners did use the complementizer *that* cue in disambiguating English

sentences, as shown by a main effect of ambiguity at the disambiguating region that was caused by faster reading times when the complementizer *that* was present. Furthermore, usage of the complementizer *that* cue was not modulated by proficiency. Lower proficiency learners of English patterned with higher proficiency learners and native English speakers in using the complementizer to disambiguate sentences. Presumably, this is because the complementizer *that* cue is a salient cue and therefore is easy to learn, leading L2 learners to be fast at acquiring such a cue even when there is no equivalent in their native language.

The findings also revealed some tendency for L1-Mandarin learners of L2-English to combine the verb bias and complementizer cues interactively, with a numerically larger ambiguity effect for DO-bias than for SC-bias verbs, but the interaction did not reach significance. Thus, the L2 learners tested in Experiment 1 showed that they were able to use both verb bias and complementizer information, but had not yet learned to put the two cues together in the optimally efficient way that native speakers do.

At the ambiguous noun region, L1-Mandarin learners of L2-English showed some evidence of using verb bias information to modulate their commitment to the initial direct object analysis, whereas native speakers did not. One possible explanation for the pattern at the ambiguous noun region in the L1-Mandarin group is that they had some expectation of a *that* after a SC-bias verb and thus were slower when *the*+Noun appeared instead, while they had the opposite expectation after DO-bias verb. That is, they expected *the*+Noun after a DO-bias verb and so when *that* appeared instead, they were slower. Regardless of the explanation, the results at the ambiguous NP region showed that verb bias was already affecting reading patterns at that point, and not just later at the disambiguation. Since no revision was required yet at the

ambiguous noun, the difference between DO-items and SC-items could only be accounted for by first-stage parsing rather than reanalysis.

Experiment 2

Experiment 1 showed that L1-Mandarin learners of L2-English were able to use both verb bias and complementizer cues to anticipate upcoming syntactic structure. The next question is how their usage of those cues compares with the usage of the kinds of lexical-semantic cues that have been proposed to be especially important for second language sentence processing (e.g., Clahsen and Felser, 2006). To evaluate that, Experiment 2 manipulated the plausibility of the ambiguous noun as the direct object of the main clause verb to examine the relative importance of verb bias and complementizer cues on the one hand and plausibility cues on the other. Native English speakers have been previously found to rely more heavily on verb bias and complementizer cues than on plausibility in DO/SC sentences, but it could be the opposite for L2-learners, if they do generally rely on semantic more than syntactic cues. In addition to a native English speakers control group and a L1-Mandarin L2-English group, Experiment 2 also tested a group of L1-Korean learners of L2-English. In Experiment 2, both L2-English groups might be expected to show strong effects of semantic cues.

Method

Participants

65 native English speakers, 70 L1-Mandarin speakers of L2-English, and 69 L1-Korean speakers of L2-English participated in Experiment 2. All were undergraduate or graduate

students at the University of Illinois at Urbana-Champaign, had normal or corrected-to-normal vision, gave written informed consent, and received course credit or payment for taking part. All L2 learners of English completed the cloze test after the self-paced reading experiment to assess their proficiency. L2 groups were divided into higher and lower proficiency groups based on median splits in each group (L1-Korean: lower proficiency group<33, higher proficiency group≥33; L1-Mandarin: lower proficiency group<35; higher proficiency group≥35). Notice that the L1-Mandarin group had somewhat higher proficiency overall than the L1-Korean group in this study.

Materials and Design

The same ten DO-bias and ten SC-bias verbs were each used four times to construct eighty sets of sentences, forty with DO-bias and forty with SC-bias verbs. Within each verb type, the plausibility and ambiguity factors were fully crossed, resulting in four conditions: Ambiguous Plausible, Ambiguous Implausible, Unambiguous Plausible, Unambiguous Implausible, as shown below in example (12). Unambiguous sentences were disambiguated using the complementizer *that*. The ambiguous noun immediately following the main clause verb was either quite plausible or quite implausible as the direct object. All experimental sentences turned out to have the sentential complement structure. Sentences in the plausible condition were identical to the sentences used in Experiment 1, for the most part.

(12) Example stimuli for Experiment 2:

DO-bias verb
Plausible: *The club members understood (that) the bylaws would be applied to everyone.*
Implausible: *The club members understood (that) the pool would be closed on Mondays.*

SC-bias verb
Plausible: *The ticket agent admitted (that) the mistake might be hard to correct.*
Implausible: *The ticket agent admitted (that) the kiosk might be difficult to find.*

Plausible and implausible ambiguous nouns were selected based on the results of a plausibility norming task, which asked 56 native English speakers to rate the plausibility of a sentence in which the noun was the direct object of the verb on a 1 (very implausible) to 7 (very plausible) scale, as shown in (13). The ambiguous nouns used in Experiment 2 all met the criterion that within any sentence set, the plausible noun was rated at least 2.5 points more plausible than the implausible noun. The properties of the ambiguous nouns are summarized in Table 8. The plausible nouns were rated as significantly more plausible than the implausible ones (6.4 vs 2.1, $F(1,156)=1481$, $p<.001$). Consistent with previous studies reporting that plausibility ratings reflected verbs' biases, such that nouns tend to be rated as more plausible after a DO-bias verb than after a SC-bias verb (Garnsey et a. 1997), the plausible nouns in sentences with DO-bias verbs were rated slightly more plausible than those in sentences with SC-bias verbs (6.5 vs 6.2, $F(1,78)=5.4$, $p<.05$). In addition, the implausible nouns in sentences with DO-bias verbs were also rated slightly more plausible than those in sentences with SC-bias verbs (2.3 vs 1.9, $F(1,76)=5.4$, $p<.05$), also consistent with previous findings. By mistake, two implausible nouns used in sentences with DO-bias verbs were not rated in the norming study (*The construction worker observed the morning*; *The navy veterans protested the ocean*), so the mean plausibility rating values shown in Table 8 do not include those items. Plausible nouns in sentences with DO-bias verbs did not differ from those in SC-bias sentences in the number of letters (7.2 vs 7.1,

F<1) and log frequency (2.9 vs 3.2, F<1; from SUBTL-$_{EN}$ corpus, Brysbaert & New, 2009). The same was true for implausible nouns (length: 6.1 vs 5.8, F<1; log frequency: 3.2 vs 3.0, F<1). On average, plausible nouns were about 1 letter shorter than implausible nouns (7.1 vs 6.0, $F(1,79)=11.61$, $p<.01$). If there is any effect of this small difference in length, it could contribute to faster reading times for plausible nouns, but length effects will be removed in the length-corrected residual reading time measure that is submitted to statistical analysis.

A separate group of twelve native English speakers rated the plausibility of the ambiguous noun as the subject of an embedded clause on a 1 (very implausible) to 7 (very plausible) scale, as shown in (14). The plausible nouns for DO sentences were rated as equally plausible as the subject of an embedded clause as the plausible nouns for SC sentences (6.1 vs 6.1, F<1), and the same was true for the implausible nouns (DO 5.2 vs SC 5.3, F<1). Notice that nouns that were implausible as direct objects were also slightly more implausible as embedded clause subjects than were nouns that were plausible as direct objects (5.25 vs 6.1), but this was equally true for both verb types. The properties of the ambiguous nouns in Experiment 2 are summarized in Table 8.

(13) *The club members understood the bylaws.*
The club members understood the pool.

(14) *The club members understood that the bylaws…*
The club members understood that the pool…

Table 8. Properties of the ambiguous nouns used in Experiment 2.

	Mean length	Log frequency	Plausibility as the direct object	Plausibility as the clause subject
DO-items				
Plausible Noun	7.2	2.9	6.5	6.1
Implausible Noun	6.1	3.2	2.3	5.2
SC-items				
Plausible Noun	7.1	3.2	6.2	6.1
Implausible Noun	5.8	3.0	1.9	5.3

Critical sentences were distributed over four lists according to a Latin Square design, so that all participants saw an equal number of trials in each condition and never saw two sentence versions from the same sentence set. Eighty distractors were added to each list for a total of 160 trials per list. The distractors were identical to those used in Experiment 1. A comprehension question was asked following each sentence and the question did not probe the comprehension of the initial misinterpretation (e.g., *Did the club members understand the bylaws?*). Sentences were pseudo-randomized once so that no two critical sentences from the same condition appeared consecutively. Participants were randomly assigned to one of the four lists and saw the same order of all sentences in each list.

Procedure

The procedure for Experiment 2 was exactly the same as Experiment 1.

Results

Comprehension Accuracy

Trials on which participants pressed the *yes* or *no* button before the comprehension questions were displayed, and trials on which participants did not make a response within the four second limit were excluded from the analysis of comprehension accuracy. Comprehension accuracy was analyzed using a logit mixed-effect model that included ambiguity, verb bias, language groups and their interactions, and random intercepts for subjects and items. Language groups were coded so that the model compared native vs non-native speakers, L1-Mandarin vs L1-Korean speakers of L2-English, high proficiency L1-Mandarin vs low proficiency L1-Mandarin speakers of L2-English, and high proficiency L1-Korean vs low proficiency L1-Korean speakers of L2-English. Results revealed main effects of all the four comparisons of language groups, because native speakers answered the questions more accurately than non-native speakers (93% vs 86%, $p<.001$), L1-Mandarin learners of L2-English were more accurate than L1-Korean learners of L2-English (87% vs 85%, $p<.05$), high proficiency L1-Korean were more accurate than low proficiency L1-Korean learners of English (87% vs 83%, $p<.01$), and high proficiency L1-Mandarin were more accurate than low proficiency L1-Mandarin learners of English (89% vs 85%, $p<.01$). There was also a significant interaction between verb bias, plausibility and Group 1 (native vs. non-native, $p<.05$), because native speakers were slightly more accurate in answering sentences that contained plausible than implausible ambiguous noun following DO-bias verbs (94% vs. 93%) and this pattern was reversed in SC-bias sentences (92% vs. 93%), whereas there was no such interaction with non-native speakers (DO plausible: 85%; DO implausible 86%; SC plausible 86%, SC plausible 86%), which indicated that native

speakers found the sentences slightly easier to understand when the plausibility of the ambiguous noun was congruent with a verb's preferred continuation, but non-native speakers did not manifest their sensitivity to verb bias in their answers to comprehension questions.

Reading Times

Disambiguating region. Data were analyzed using a linear mixed-effect model with maximal random effects structure, which included ambiguity, verb bias, plausibility, language group and the interactions among them, random intercepts and random slopes of fixed effects for subjects and items. Language groups were coded so that the model compared native with non-native speakers, native speakers with L1-Mandarin speakers of L2-English, native speakers with L1-Korean speakers of L2-English, L1-Mandarin with L1-Korean speakers of English. Results revealed a main effect of ambiguity, with ambiguous sentences being read slower than unambiguous sentences (429 vs 412 ms, $t>9$), and an interaction between verb bias and ambiguity ($t=3$), which was caused by the ambiguity effect being bigger in sentences with DO-bias verbs (22 ms, $t>7$) than those with SC-bias verbs (11 ms, $t>4$). There were no effects of plausibility in the high level analysis, but when the results were broken down by verb type, the analysis with DO-items showed an interaction between ambiguity and plausibility ($t>2$), which resulted because the ambiguity effect was bigger after implausible nouns (30 ms, $t>5$) than after plausible nouns (15 ms, $t>3$), as shown in Figure 4. There was no such interaction after SC-bias verbs ($t<1$). However, this difference between sentences with DO-bias and SC-bias verbs was not strong enough to produce a significant interaction between verb bias, ambiguity and plausibility. The observed plausibility effect was in the opposite direction as expected. After DO-bias verbs, readers should be slower at reading an implausible noun than a plausible noun, but

faster at recovering from garden-pathing when the noun was implausible because there should be less commitment to it being the direct object. Thus the plausibility effect observed here seems likely to be a spill-over effect from reading the implausible noun itself. There was no effect involving the language group factor (ts<2). Overall, the analysis including all language groups suggested that L2 learners did not reliably differ from native speakers in their use of verb bias, complementizer *that*, and plausibility cues. There was no main effects or interactions involving the plausibility of the noun as a direct object, but all language groups were sensitive to both verb bias and complementizer cues, and there was a reliable interaction between them in the high-level analysis. However, examination of the reading times at the disambiguation collapsed over plausibility plotted separately for the three language groups in Figure 3 shows that the native English speakers showed the same optimally efficient interactive pattern between verb bias and ambiguity that was found in Experiment 1, while the two L2-English groups showed additive effects of each cue. Thus, further analyses were performed on the language groups separately.

Analysis on the residual reading times at the disambiguating region for native speakers revealed a main effect of ambiguity, with ambiguous sentences read slower than unambiguous sentences (356 vs 349 ms; t>2), and an interaction between verb bias and ambiguity (t>2). The interaction resulted because ambiguous sentences were read slower than unambiguous sentences only after DO-bias verbs (364 vs 351 ms, t>2) but not after SC-bias verbs (348 vs 347 ms, t<1), as shown in Figure 3. There was no effect involving the plausibility factor, indicating that the interaction between ambiguity and plausibility within DO-items in the higher level analysis that included all language groups was not reliable in the native English speakers group. This is consistent with a previous finding that plausibility of the noun as a direct object had no effect on

disambiguation region reading times in native English speakers when verbs were strongly biased (Garnsey et al, 1997).

The linear mixed-effect model with maximal random effect structures performed on L1-Mandarin L2-English speakers that included proficiency as a categorical predictor variable (higher proficiency group vs lower proficiency group) revealed the same pattern as native English speakers. There was a main effect of ambiguity, with ambiguous sentences being read slower than unambiguous sentences (447 vs 425 ms; $t>5$), a main effect of proficiency, with the higher proficiency group reading faster than the lower proficiency group (421 vs 451 ms; $t=3$), and an interaction between verb bias and ambiguity ($t>2$), which was caused by the ambiguous sentences being read slower than unambiguous sentences after DO-bias verbs (462 vs 432 ms, $t>4$), but not after SC-bias verbs (433 vs 419 ms, $t<2$), as shown in Figure 3. There was no effect involving the plausibility factor (ts<2), suggesting that the ambiguity by plausibility interaction in DO-items that emerged in the higher level analysis with all language groups was not reliable in the L1-Mandarin learners of L2-English, either.

Analysis of L1-Korean speakers of L2-English that included proficiency as a fixed effect in addition to other fixed effects revealed a main effect of ambiguity, with ambiguous sentences read slower than unambiguous sentences (482 vs 460 ms, $t>5$), and a main effect of verb bias, with sentences with DO-bias verbs being read slower than those with SC-bias verbs (479 vs 463 ms, $t>2$). For this group, there was no interaction between verb bias and ambiguity, nor were there were any effects involving the proficiency and plausibility factors (ts<2).

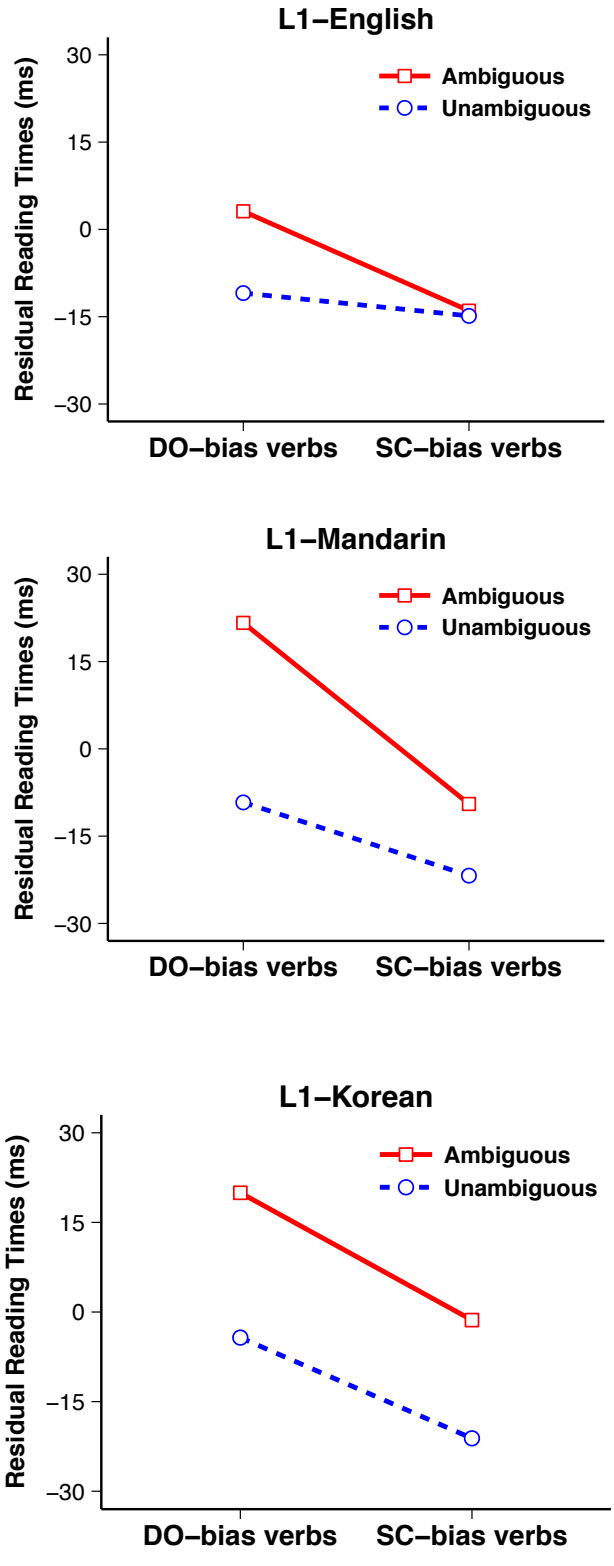

Figure 3. Residual reading times at the disambiguating region for native English speakers, L1-Mandarin learners of L2-English and L1-Korean learners of L2-English, collapsing over plausibility.

In summary, all groups read the disambiguating region more slowly in ambiguous sentences than in unambiguous ones. Native English speakers and L1-Mandarin speakers of L2-English both showed a significant interaction between ambiguity and verb bias while L1-Korean speakers of L2-English did not, but instead showed a main effect of verb bias. However, the difference between L1-Korean group and the other two groups was not strong enough to produce a higher level interaction between ambiguity, verb bias and language group. Neither of the L2 learner groups showed the maximally efficient interactive pattern that the native speakers showed, in which either cue alone was sufficient, but the L1-Mandarin group was closer to the native pattern than the L1-Korean group was.

There were no effects involving the plausibility manipulation in the highest level analysis, but when the verb types were analyzed separately, there was a tendency for slower reading times on the disambiguation following a DO-bias verb and an implausible noun. The same numeric pattern was present in all three language groups, though it did not reach significance in any of the groups tested separately. This effect was hypothesized to be spillover from reading the implausible noun itself.

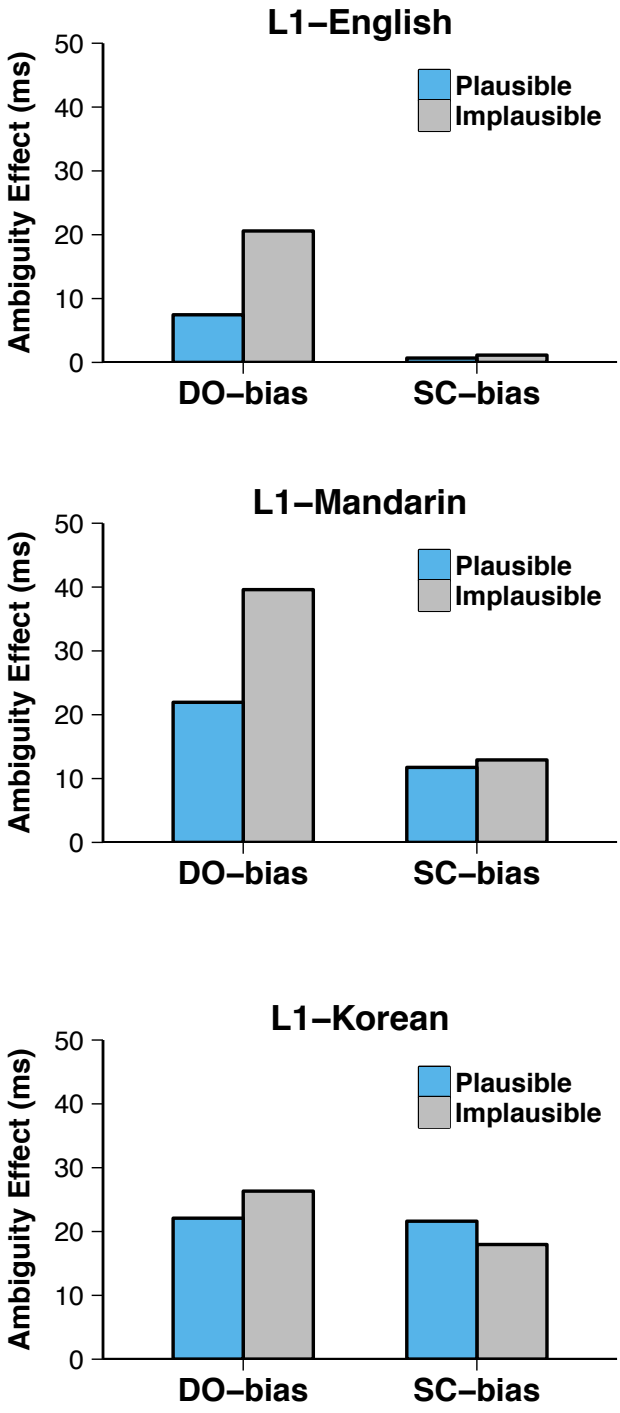

Figure 4. Ambiguity effect at the disambiguating region. Ambiguity effect was computed by subtracting the residual reading times of unambiguous sentences from those of ambiguous sentences. Plausibility effect was not significant in any of the three groups.

Ambiguous noun. Data analysis that included all language groups revealed a main effect of ambiguity, with ambiguous sentences being read slower than unambiguous sentences (439 vs 415 ms, t=5), and a main effect of language group (native vs. L1-Korean), with L1-Korean speakers being slower than native speakers (444 vs 347 ms, t>3). There were no reliable effects involving the language group or plausibility factors (ts<1.3). Nonetheless, separate analyses were conducted on native speakers, L1-Mandarin and L1-Korean groups to examine whether the observed ambiguity effect was present in all groups.

Native English speakers showed a main effect of ambiguity, with unambiguous sentences being read faster than ambiguous sentences (ambiguity effect 20 ms, t>6). There were no effects involving plausibility (ts<1.2). L1-Mandarin speakers of L2-English showed a main effect of ambiguity (ambiguity effect 26 ms, t>5) and an interaction between ambiguity and verb bias (t>2), which resulted because there was an ambiguity effect after SC-bias verbs (ambiguity effect 34 ms, t>4) but not after DO-bias verbs (t<2). There were no effects involving the proficiency or plausibility factors in this group (ts<2). Similarly, L1-Korean speakers of L2-English showed a main effect of ambiguity, with ambiguous sentences being read slower than unambiguous sentences (ambiguity effect 27 ms, t>5), and an interaction between ambiguity and verb bias (t>2), because the ambiguity effect was bigger after SC-bias verbs (ambiguity effect 36 ms, t=4) than after DO-bias verbs (ambiguity effect 17 ms; t=3), just as it was for the L1-Mandarin group. Also just as for the L1-Mandarin group, there were no effects involving the plausibility or proficiency factors (ts<1.5). The ambiguity effects in all three language groups are plotted in Figure 5.

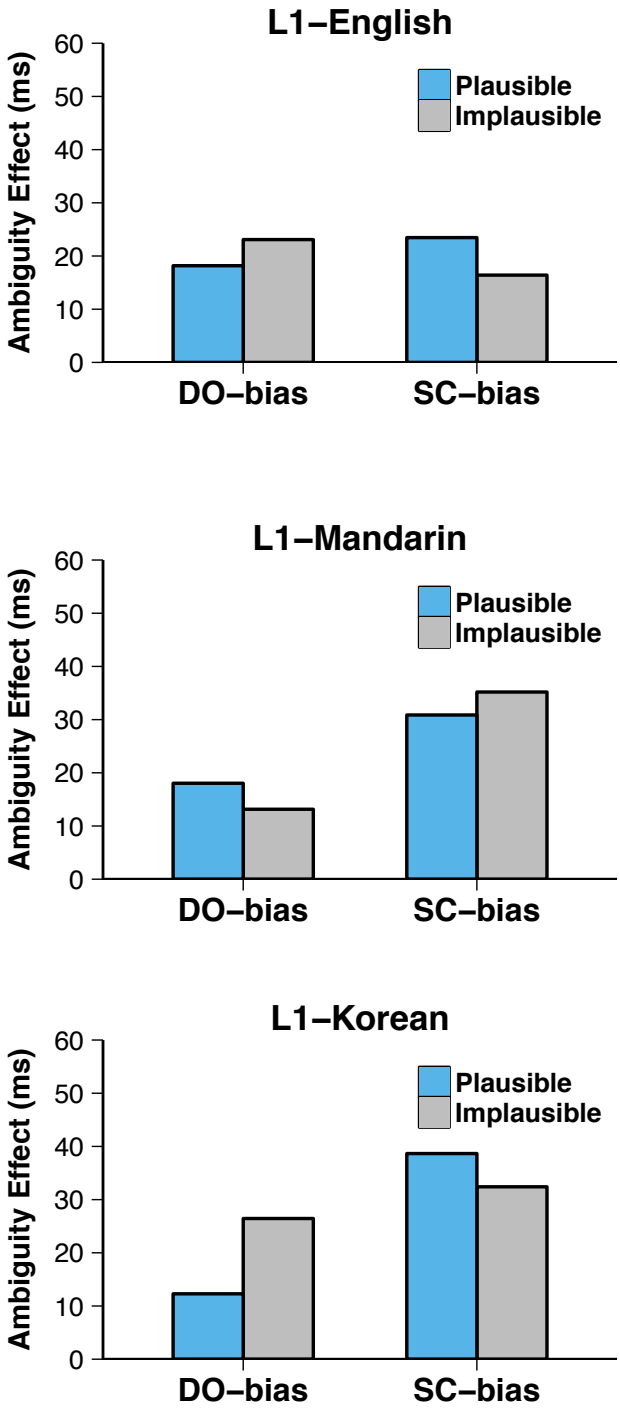

Figure 5. Ambiguity effect at the ambiguous noun region. Ambiguity effect was computed by subtracting the residual reading times of unambiguous sentences from those of ambiguous sentences.

Discussion

Experiment 2 manipulated the plausibility of the ambiguous noun as the direct object of the main clause verb together with the structural biases of the verbs and complementizer presence to examine the relative importance of plausibility, verb bias, complementizer cues in the processing of English DO/SC sentences by native speakers and L2 learners of English. Native speakers' reading times at the disambiguating verb and the ambiguous noun were compared with those of non-native speakers whose native languages were Mandarin and Korean. The comparison of plausibility effects between native and non-native speakers was of particular interest, given proposals that L2 learners tend to rely more heavily on plausibility than on other more syntactic cues in sentence processing in the second language (e.g., Clahsen and Felser, 2006). A previous study of native English speakers had found that plausibility had no effect when verbs were strongly biased (Garnsey et al., 1997), but it seemed possible that non-native speakers might. The absence of plausibility effects in native English speakers was replicated here, and somewhat surprisingly the same pattern was found for both non-native groups. Contrary to expectation, non-native speakers were just as unaffected by plausibility as the native speakers. Thus, the results provide no evidence supporting the claim that non-native speakers rely on plausibility more than other kinds of cues.

Consistent with Lee et al. (2013) and Experiment 1 here, both native and non-native speakers of English were affected by the presence of the complementizer *that* and the structural biases of the verbs. Also consistent with both of those studies, the native English speakers showed the optimally efficient use of verb bias and complementizer cues, while the non-native groups showed additive effects of both cues, with the L1-Mandarin group getting closer to the pattern for the native speakers than the L1-Korean group.

Garnsey et al. (1997) argued that a possible reason that native speakers rely more on verb bias than on plausibility is that verb bias is retrievable information that comes with recognizing the verb and is thus available very rapidly, while plausibility must be computed online for particular verb-noun combinations and so may not be available quickly enough to influence reading times on the disambiguating region. The same is apparently true for those learning English as a second language. At least at the proficiency levels tested here, both L1-Mandarin and L1-Korean learners of English have accumulated information about the frequency with which different verbs are used in sentences with different kinds of structure and furthermore have rapid enough access to that information for it to influence their processing of sentences containing those verbs.

L1-Mandarin learners of L2 English in Experiment 2 showed a significant interaction between verb bias and ambiguity, indicating that they were able to use both the verb bias and the complementizer cues interactively, just like the native speakers. A comparison between native speakers and L1-Mandarin speakers' pattern at the disambiguating region, however, revealed that whereas native speakers did not benefit from having both the verb bias and the complementizer cues at the same time, L1-Mandarin learners of L2-English did benefit from having two cues, as shown in their fastest reading time at the disambiguating verb in the unambiguous sentences with SC-bias verbs. This finding suggested that although L1-Mandarin learners of L2-English in Experiment 2 have learned to combine the two cues interactively, they have not yet learned to use either cue as efficiently as native speakers. In addition, L1-Mandarin speakers in Experiment 1 differed from those in Experiment 2 in showing the main effects of verb bias and ambiguity, but no reliable interaction between them. This was probably because the Mandarin group in Experiment 1 was less proficient than in Experiment 2 (mean proficiency

score 31 vs 34), and thus they have not yet learned to combine the two cues interactively. At the ambiguous noun, the ambiguity effect was only significant when the verb had SC biases but not DO biases, indicating that L1-Mandarin speakers used the verb bias information at the main clause verb fast enough to influence the integration of the subsequent ambiguous noun. They were slower at integrating the ambiguous noun with the verb if the verb biased towards taking sentential complement than when it biased towards taking direct objects. The same as native speakers and consistent with Experiment 1, the plausibility of the initial misanalysis of the ambiguous noun as the direct object of the main clause verb, did not have an impact on the reading time of either the ambiguous noun or the disambiguating region. This result indicated that L1-Mandarin learners behaved just like the native speakers in that they did not use the plausibility information to guide their interpretation of the sentence structure at the ambiguous noun and consequently did not show plausibility effects on the reanalysis processes.

L1-Korean learners of L2-English showed main effects of verb bias and ambiguity but no interaction between them at the disambiguating verb in both high proficiency and low proficiency groups, suggesting that they made use of both the verb bias and the complementizer *that* cues but have not yet been able to combine them interactively. At the ambiguous noun, L1-Korean speakers showed the same interaction between verb bias and ambiguity as L1-Mandarin speakers, indicating that they used verb bias information to constrain the interpretation of the word immediately following it. They were more reluctant to analyze the noun as the direct object of the main clause verb if the verb more often takes sentential complements than direct objects. The sensitivity to verb bias information and its effect on the noun immediately following the verb was seen in L2 learners of English but not in native speakers, presumably because L2-learners read the sentences more slowly and thus verb bias effect had the opportunity to affect

the reading of the ambiguous noun only in L2-learners. At the ambiguous noun, L1-Korean speakers showed the same pattern as native speakers and L1-Mandarin speakers in that they did not make use of the plausibility information to guide their interpretation of the structure of the sentences.

Overall, neither L2 group was sensitive to the plausibility of the initial direct object analysis of the sentences when the main clause verb had strong structural biases, which is not consistent with claims that L2 learners rely more on semantics than syntax in their real-time analysis of sentences.

Conclusion

In two self-paced reading experiments, the present study investigated whether L1-Mandarin speakers of L2-English were able to use verb bias, complementizer *that,* and plausibility to predict upcoming sentence structure, and whether L2 learners relied more on semantic cues than lexically-associated syntactic information in guiding the on-line construction of syntactic structures. Mandarin follows SVO word order to place verbs early in the sentence and so verbs are useful in constraining expectations about upcoming syntactic structure. The complementizer cue, on the other hand, is not an available cue in Mandarin to signal an upcoming embedded clause. Thus, it was possible that L1-Mandarin speakers of L2-English would be able to use the verb bias cue but not the complementizer *that* cue when processing sentences in English. The results of both experiments showed that Mandarin learners of English were able to interactively combine the verb bias and complementizer cues, though the interaction did not reach the optimally efficient pattern seen in native speakers. They kept track of the frequency of the structures following particular English verbs and were able to use such

information rapidly to constrain their on-line interpretation of sentence structures. They were also able to learn the use of the complementizer *that* to generate predictions about the upcoming embedded clause in spite of the unavailability of such a cue in their native language.

Experiment 2 did not support the claim in the L2 sentence processing literature that L2 learners rely more on semantics to compensate for their underuse of syntax in real-time sentence processing. Neither L1-Mandarin nor L1-Korean speakers of L2-English used plausibility information to modulate their adoption of the initial direct object interpretation in sentences with temporary DO/SC ambiguity, which replicates the pattern found for native English speakers, both here and in Garnsey et al. (1997). A noun that was implausible as a direct object did not promote the embedded clause analysis sufficiently to alleviate processing difficulty at the disambiguating region for either the native speakers or the L2 learners. It was suggested that it may simply take too long, even for native speakers, to put the verb and noun meanings together in the way that is required for it to influence parsing decisions.

CHAPTER 3

Verb Bias and Plausibility in the Processing of Mandarin Sentences

While it has been well-established that English speakers rely more on verb bias than plausibility in interpreting temporarily ambiguous sentences (Garnsey et al. 1997; Trueswell, 1996), little is known about the relative weight of verb bias and plausibility in parsing other languages. Given that some languages, such as Mandarin, have been claimed to rely more on plausibility than on syntax, it is possible that Mandarin speakers rely more on plausibility when interpreting sentences in Mandarin.

Very few studies have examined the role of verb bias in languages other than English (see, however, Dussias & Cramer Scaltz, 2008 for a study on Spanish). In a study examining verb bias effects on the processing of Mandarin relative clauses, Lin and Garnsey (2011b) found that readers tended to expect shorter sentence continuations following DO-bias verbs and longer continuations following SC-bias verbs. When reading sentences like *The teacher disliked* $_{RC}$*[that parent scolded de-REL student]*, where *dislike* has DO-bias, and *The teacher believed* $_{RC}$*[that parent scolded de-REL student] made progress*, where *believe* has SC-bias, participants were slower at the relative clause verb *scolded* and the relative clause head noun *the student* after DO-bias than after SC-bias verbs. This was taken to indicate that readers anticipated that the sentence would end at *that parent* after DO-bias verbs and thus slowed down when it did not end. After SC-bias verbs, they anticipated that the sentence would take a sentential complement and thus experienced less processing difficulty when the sentence did not end at that point. However, sentences with DO-bias verbs in this study ended with direct object continuations and those with SC-bias verbs ended with sentential complement continuations, which did not allow direct comparison between the processing of sentential complements after DO-bias and SC-bias verbs.

Verb bias effects were also complicated by the content of the embedded relative clause structure to some extent, leading to the possibility that the plausibility of the combinations of the subject, the main clause verb, first noun in the relative clause and the relative clause verb (*The teacher disliked that parent scolded…* vs. *The teacher believed that parent scolded…*) might all have affected the reading time of the critical verb *scolded*.

Mandarin has been found to rely more on plausibility than syntax during sentence comprehension (Li, Bates, Liu, & MacWhinney, 1993; Su, 2001a, 2001b, 2004). Direct comparisons between Mandarin and English came from Su (2001a, 2001b, 2004), who used the "identifying agent" paradigm that was commonly used in studies within the framework of the Competition Model (e.g., MacWhinney, Bates, & Kliegl, 1984). Su (2001b) asked participants to listen to three-word sentences that manipulated word orders (NVN, NNV, VNN) and the animacy of the nouns (Animate Animate; Animate Inanimate; Inanimate Animate), such as *rabbit bites tiger* (AVA) and identify the agent of the action. Results showed that in the IVA word order, such as *the door bumps the rabbit*, English speakers almost always chose *the door* as the agent while Mandarin speakers only did that 30% of the time. In AIV (*the mouse the kite follows*) and VAI word orders (*pulls the pig the balloon*), Mandarin speakers chose the animate noun as the agent more than 90% of the time while English speakers did so only about 10% of the time. In the NVN order in general, English speakers always chose the first noun as the agent regardless of animacy, while Mandarin speakers did so much less often, 60% of the time. These findings were interpreted with respect to cue reliability in the two languages. Since English usually does not omit the subject and word order is relatively rigid, English speakers relied more on the word order cue than speakers of other languages such as Spanish (Kail, 1989) and Mandarin. In contrast to English, Mandarin is a null subject language and also allows any part of

a sentence to be moved to the beginning to serve as the topic (Li and Thompson, 1976). In addition, word order is relatively flexible compared to English, and thus is not as reliable a cue as in English. As a result, Mandarin speakers rely on plausibility instead to figure out thematic roles in the three-word sequences.

In a comparison of the use of contextual information between English and Mandarin speakers, Su (2001a, 2004) found that Mandarin speakers relied more on contextual information than English speakers. In both studies, prior to the three-word sentences manipulated in the same way as in Su (2001b), a short (Su, 2001a) or long (Su, 2004) context that biased towards either the first noun or the second noun as the agent was added (e.g., biasing the first noun: *The rabbit is angry*. *The rabbit bites the tiger*; biasing the second noun: *The tiger is hungry. The rabbit bites the tiger*.). Results showed that, when the context was short (one sentence), it had a bigger effect on Mandarin speakers than English speakers, although animacy still had the biggest effect for the Mandarin speakers and word order for the English speakers. When the context was long (three sentences), however, the contextual cue became the biggest cue that Mandarin speakers relied on, whereas word order was still the most important cue for English speakers.

The offline choose-the-agent task is somewhat unnatural, but there is also some evidence from online measures showing that plausibility guides the real-time interpretation of Mandarin sentences (Lin & Garnsey, 2011a; Wu, Kaiser & Anderson, 2012). For instance, Lin and Garnsey (2011a) created ambiguity about early vs late closure by dropping the head noun of a topicalized relative clause, and then manipulated the plausibility of the subject of the main clause as the head noun of the relative clause, as in $_{RC}$*[Interrogate councilman de-REL ___] reporter started to report* (plausible because *reporter can interrogate councilman*). vs $_{RC}$*[Interrogate councilman de-REL ___] newspaper started to report* (implausible because in Mandarin the verb

translated here as *interrogate* cannot take a non-human agent, not even one like *newspaper*). They found that readers followed the late closure principle to initially interpret the subject of the main clause as the relative clause head noun if it was plausible (*the reporter*) and subsequently experienced difficulty at the main clause. On the contrary, when the subject was an implausible head noun for the relative clause, readers did not attach it to the relative clause and therefore did not have trouble reading the main clause. This study indicated that in Mandarin, plausibility plays a role in ambiguity resolution when verb bias information is not available, just as in English (see Pickering & Traxler, 1998 for similar results in English).

Existing research on the effects of verb bias and plausibility in Mandarin leaves open two questions: 1) whether and how Mandarin speakers use verb bias information to guide their on-line interpretation of sentences; and 2) whether plausibility trumps verb bias in Mandarin, opposite to English, given that Mandarin has been shown to rely more on plausibility compared to other languages such as English.

(15) *The angry reporter revealed the truth…*
(a) *…… in his article.*
(b) *…… would not be discovered.*

(16) *愤怒的记者揭露真相...*
The angry reporter revealed the truth…
(a) *……以后很高兴。*
……..then he was happy.
(b) *……..已经被封锁了。*
…….had already been hidden.

(17) *愤怒的记者揭露公园已经被拆除了。*
The angry reporter revealed the park had already been demolished.

(18) *愤怒的记者揭露，真相已经被封锁了。*
The angry reporter revealed that the truth had already been hidden.

The present study uses sentences that bear surface resemblance to English sentences with DO/SC ambiguity, as shown in (15). Sentence (15) can continue into (15a), in which *the truth* is the direct object of the preceding verb *revealed*, or (15b), in which *the truth* is the subject of the sentential complement *would not be discovered*. Thus when the noun following the main clause verb *revealed* is encountered, it is temporarily ambiguous between being the direct object and being the subject of the embedded clause. In English, the ambiguity is resolved at the embedded clause verb *would* in (15b), because *would* lacks a subject and therefore a reanalysis process is triggered to remove *the truth* from the direct object role and attach it to the subject role of the embedded clause. Reanalysis is triggered because English does not allow the dropping of subjects. The embedded clause verb *would* must be preceded by a subject or the sentence is ungrammatical. However, this is not the case in Mandarin.

Consider the same sentence in Mandarin, as shown in (16) (small changes were made to the lexical items to make the sentence natural in Mandarin). *The truth* turns out to be the direct object of *revealed* in (16a), but in (16b), it turns out to be both the direct object of *revealed* and the subject of *had already been hidden*. Such a structure is allowed in Mandarin, because Mandarin is a null-subject language and thus the verb in the embedded clause does not require an overt subject. Reanalysis is not triggered at the embedded clause verb, resulting in *the truth* remaining both the direct object of the main clause verb and the subject of the embedded clause. As reanalysis is not triggered, the increased reading time in English at the embedded clause verb would not be expected. Indeed, a study investigating the processing of the same type of sentences in Spanish revealed that Spanish speakers did not slow down at the embedded clause verb, because Spanish is also a null subject language that does not require the embedded clause verb to be preceded by an overt subject (Jegerski, 2012). In what follows, this type of structure

will be called the “blended structure”. The blended structure is a combination of a direct object structure and a sentential complement structure, with both structures sharing the same noun. The blended structure is an allowable and frequently used structure in Mandarin. Therefore, in Mandarin, the ambiguous noun *the truth* is temporarily ambiguous between the direct object analysis and the blended structure analysis in (16).

With respect to the ultimate interpretation of sentences with blended structure, the meaning of the rest of the sentence determines whether the temporary ambiguity can be interpreted as a blended structure or instead only with an embedded clause analysis. For instance, in *愤怒的记者揭露真相是市长已经辞职了(The angry reporter revealed the truth is that the major had already resigned.),* the ultimate interpretation is consistent with the blended structure, because the angry reporter revealed the truth and also revealed what the truth was. In (16b), however, the ultimate interpretation is only consistent with an embedded clause structure, because the angry reporter did not reveal the truth; rather, he only revealed something about the truth. As the meaning of the second half of the blended structure (*the truth had been hidden*) contradicts the first half (*the angry reporter revealed the truth*), the ultimate interpretation is consistent with the embedded clause structure and not the blended structure. The same syntactic structure can lead to different interpretations in Mandarin depending on the meanings of later words in the sentence. Plausibility thus plays an important role in deciding the final interpretation.

One requirement for the blended structure is that the noun must be plausible as the direct object of the main clause verb; otherwise the first half of the blended structure (i.e., the direct object analysis) is broken, resulting in only the second half of the blended structure (i.e., an embedded clause structure). For instance, (17) is disambiguated towards an embedded clause

structure by the noun *the park*, because *the park* cannot be the direct object of *revealed*. Consequently, it can only be the subject of the subsequent sentential complement clause. In short, a sentence beginning like (16) is temporarily ambiguous between a direct object structure and a blended structure when the noun is plausible as the direct object. It is most likely to continue into a sentential complement if the noun is implausible as the direct object, because an implausible noun breaks the direct object part of the blended structure.

Both sentence (16b) and (17) can be disambiguated towards the sentential complement structure by adding a comma after the main clause verb *revealed*, as shown in (18). There is no ambiguity regarding the syntactic role of *the truth*. It can only serve as the subject of the sentential complement *the truth had already been hidden*, because both the direct object analysis and the blended analysis are eliminated by the comma. In the present study, the comma-disambiguated version was used as the baseline, to which temporarily ambiguous structures were compared.

Although no revision process is triggered at the embedded clause verb because there is no syntactic incompatibility at that point, this region is still informative in terms of revealing the parser's preference among multiple possible syntactic structures. If the parser generally prefers the blended structure to the sentential complement structure, the blended structure should be read faster than the sentential complement version. The reverse is true if the parser prefers sentential complements to blended structures. The embedded clause verb is referred to as the critical region in this study rather than the disambiguating region because, unlike English, it is not the appearance of a verb in this position that disambiguates the sentence structure. Instead, it is combination of the plausibility of the verb plus noun preceding the critical word together with the fact that the critical word is a verb that provides the disambiguation. In the experimental

sentences, the critical region consisted of two words, including an auxiliary verb that does not carry much meaning (e.g., *bei-PASSIVE* in 16b) and an adverb that preceded it (e.g., *already* in 16b).

Sentences beginning like (16) can also develop into a different kind of sentential subject structure, as in *愤怒的记者揭露真相（这件事）感动了很多人 (The angry reporter revealed the truth [this event]moved many people.)*, in which *the angry reporter revealed the truth* is the topic (or subject, as in *That the angry reporter revealed the truth moved many people.*) of the sentence. However, this type of structure is usually used with verbs that are followed by aspectual markers, such as *–zhe*, *-le*, *-guo* (roughly translated to *–ing*, *-ed*, *-ed*; aspectual markers make Mandarin verbs similar to tensed verbs in English), and the sentential subject is often followed by the word *the event* to avoid ambiguity. Since none of the main clause verbs in our experimental sentences are followed by aspectual markers, this structure is not very likely to be considered by the participants in the experiment.

The present study manipulated the verb bias of the main clause verb and the plausibility of the ambiguous noun as its direct object, in order to compare the relative importance of the two cues in Mandarin. If Mandarin speakers use verb bias information to expect the upcoming structure, they would expect either the direct object continuation or the blended structure continuation following DO-bias verbs, but the sentential complement continuation after SC-bias verbs. At the critical region, where the direct object continuation has already been ruled out (because the sentence does not end at the direct object), the blended structure is the only possible structure when the ambiguous noun is plausible, but the sentential complement structure is the only possible one when the ambiguous noun is implausible. Comparing the processing of the blended structure with the comma-disambiguated sentential complement structure, and

comparing the processing of the implausible-noun-disambiguated sentential complement structure with the comma-disambiguated sentential complement structure should reveal the parser's preferences among the three syntactic structures. Following SC-bias verbs, however, a sentential complement should be anticipated regardless of the plausibility of the noun as a direct object. Therefore, there should be no difference between the reading times for these structures following SC-bias verbs. If Mandarin speakers rely on plausibility, however, after encountering the implausible ambiguous noun, they should anticipate a sentential complement structure and should therefore not experience processing difficulty when the sentence unfolds into a sentential complement, which is the case in all experimental sentences.

Method

Participants

48 native speakers of Mandarin (14 males, mean age 23, range 19-28) who had completed at least a high school education in Mainland China participated in Experiment 3. They were undergraduate or graduate students at the University of Illinois at Urbana-Champaign, all had normal or corrected-to-normal vision, gave written informed consent and received payment for taking part.

Materials and Design

11 DO-bias verbs and 11 SC-bias verbs were each used three times (except 1 DO verb and 1 SC verb, which were each used twice) to create sixty-four sets of sentences, with each set fully crossing the plausibility and ambiguity factors, resulting in four conditions in each set:

Ambiguous Plausible, Ambiguous Implausible, Unambiguous Plausible, Unambiguous Implausible, as shown below in (19) (with the critical region underlined). Unambiguous sentences were disambiguated by placing a comma after the main clause verb, which is somewhat similar to using the complementizer *that* to disambiguate this type of sentences in English, except that such comma usage is much less frequent than *that*-inclusion in English . The ambiguous noun immediately following the main clause verb was plausible as its direct object in the plausible conditions and implausible as the direct object in the implausible conditions (*The angry reporter revealed the truth* vs *The angry reporter revealed the park*). All experimental sentences turned out to have the sentential complement structure. The two words in the critical region were identical between plausible and implausible conditions, and were words that did not carry much meaning, such as adverbs (e.g., *already*) and auxiliary verbs (e.g., *could*).

(19) Example sentences:

DO-bias, Plausible, Ambiguous and Unambiguous
愤怒的记者揭露(,)真相已经被封锁了。
The angry de-MOD reporter revealed(,) the truth already bei-PASSIVE hidden.
"The angry reporter revealed (that) the truth had already been hidden."

DO-bias, Implausible, Ambiguous and Unambiguous
愤怒的记者揭露(,)公园已经被拆除了。
The angry de-MOD reporter revealed(,) the park already bei-PASSIVE demolished.
"The angry reporter revealed (that) the park had already been demolished."

SC-bias, Plausible, Ambiguous and Unambiguous
狡猾的罪犯否认(,)事实已经被警察知道了。
The tricky de-MOD criminal denied(,) the fact already bei-PASSIVE the police knew.
"The tricky criminal denied (that) the fact had already been found out by the police."

SC-bias, Implausible, Ambiguous and Unambiguous
狡猾的罪犯否认(,)汽车已经被他卖掉了。
The tricky de-MOD criminal denied(,) the car already bei-PASSIVE him sold.
"The tricky criminal denied (that) the car had already been sold by him."

The verbs used in the experiment were chosen based on a norming study (modeled after Garnsey et al., 1997), which asked a separate group of 102 native speakers of Mandarin to complete one hundred sentence fragments that started with a proper noun and a verb that could take either direct object or sentential complement endings, such as *张红发现…*(*Zhanghong discovered…*). Completions were then categorized into 1) direct object; 2) sentential complement; and 3) other types of completions. Verbs selected to be used in the experiment all met the criteria that they were completed at least twice as often in one structure than the other. DO-verbs had at least twice as many direct object completions as SC-verbs and the reverse was true for SC-verbs. All verbs were composed of two characters. DO-verbs had more strokes per word than SC-verbs (19 vs 16; $F(1,62)=4.5$, $p<.05$) and lower log frequency (based on SUBTLEX-$_{CH}$ Corpus, Cai & Brysbaert, 2010; 3.0 vs. 3.5, $F(1,62)=9.8$, $p<.01$). DO-verbs had higher DO-bias than SC-verbs

(80% vs 13%, $F(1,62)=1113$, $p<.001$) and SC-verbs had higher SC-bias than DO-verbs (80% vs 12%, $F(1,62)=2406$, $p<.001$). Verb properties are summarized below in Table 9.

Table 9. Properties of the verbs used in Experiment 3.

	DO bias strength (%)	SC bias strength (%)	Mean number of characters	Mean number of strokes	Mean log frequency
DO-verbs	80	13	2	19	3.0
SC-verbs	12	80	2	16	3.5

A plausibility norming study was conducted to select plausible and implausible nouns for the main clause verbs for each item by asking 48 participants to rate on a 1 (very implausible) to 7 (very plausible) scale the combinations of subject, verb and ambiguous noun, as shown in (20). The selected plausible nouns were rated at least 2.5 points more plausible than the implausible nouns for each item. Across all items, plausible nouns were more plausible than implausible nouns (6.7 vs 2.2, $F(1,126)=2158$, $p<.001$). The degree of plausibility or implausibility of ambiguous nouns were matched between DO-items and SC-items (DO plausible 6.7, SC plausible 6.7, $F<2$; DO implausible 2.2, SC implausible 2.3, $F<1$). Log frequency of the noun used after DO-verbs were higher than SC-verbs (3.1 vs 2.8, $F(1,62)=4.2$, $p<.05$). Plausible nouns were more frequent than implausible nouns after DO-verbs (3.3 vs 2.9, $F(1,31)=21$, $p<.001$), but this difference should not affect the results because plausible and implausible ambiguous conditions were always compared with their unambiguous counterparts, which only differed in the addition of the comma, with all lexical items being identical. There was no difference between the frequency of plausible and implausible nouns after SC-verbs (2.8 vs 2.8, $F<1$). Noun properties are summarized in Table 10.

(20) 愤怒的记者揭露真相。 (The angry reporter revealed the truth.)
愤怒的记者揭露公园。 (The angry reporter revealed the park.)

Table 10. Properties of the ambiguous nouns used in Experiment 3.

	Plausible nouns		**Implausible nouns**	
	Plausibility as direct object	Plausibility as subject of embedded clause	Plausibility as direct object	Plausibility as subject of embedded clause
DO-items	6.7	6.5	2.2	5.2
SC-items	6.7	6.5	2.3	5.3

Another plausibility norming study was conducted to ensure that the ambiguous nouns were equally plausible as the subject of the embedded clauses between sentences with DO verbs and those with SC verbs. A separate group of 13 native speakers rated the plausibility of sentence beginnings, as shown in (21). Ambiguous nouns that were more plausible as the direct object were also more plausible as the subject of the embedded clause than the implausible-as-DO nouns (6.5 vs 5.3, $F(1,126)=85$, $P<.001$). There were no differences between DO and SC items (DO plausible 6.5, SC plausible 6.5, $F(1,62)<1$; DO implausible 5.2, SC implausible 5.3, $F(1,62)<1$).

(21) 愤怒的记者揭露，真相…… (The angry reporter revealed that the truth…)
愤怒的记者揭露，公园…… (The angry reporter revealed that the park…)

Critical sentences were distributed over four lists according to a Latin Square design, so that each participant saw only one version of each item and an equal number of items in each condition across the experiment. Sixty-four distractors were added to each list for a total 128 trials/list. Verbs used in the experiment were each used once to create one distractor that ended

with direct object continuation (e.g., *The assistant to the boss revealed the fact.*). The rest of the distractors had a variety of structures. All sentences were followed by a comprehension question. Questions after the critical sentences did not ask about the initial misinterpretation (e.g., *Did the reporter reveal the truth?*). Answers to comprehension questions were half *yes* and half *no* for both critical sentences and distractors. Sentences were pseudo-randomized once so that no two critical sentences from the same condition appeared consecutively and were presented to all participants in the same order in each list.

Procedures

Participants read a total of 128 sentences from a 23-inch LCD monitor in a dimly lit and sound-attenuated booth. Sentences were presented word-by-word in white SimSung font on a black screen in a non-cumulative moving window paradigm, controlled by the Presentation Software. All sentences were presented on a single line and each word was masked using dots before they were revealed. Each trial began with a "+"sign on the left side of the screen that remained on the screen for one second. Every time participants pressed a button on a Cedrus-830 response box, the next word was revealed and the previous word reverted back to dots. A comprehension question was displayed all at once after each trial, and participants pressed the *yes* or *no* button to answer the questions. Feedback regarding response accuracy was not given. However, a "Too Slow" message appeared if no answer was made within four seconds. The lists were divided into two blocks and participants took a break between blocks. A practice block of ten sentences was added at the beginning. The entire experiment took about 15 minutes.

Results

Reading Times

Residual reading times at the ambiguous noun and the critical region (averaged across two words) were analyzed using linear mixed-effect models with maximal random effects structure for both subjects and items (Barr et al., 2013), using the lme4 package in R. Prior to data analysis, reading times that were faster than 100 ms or slower than 2000 ms were discarded. Reading times above or below 2.5 sd from the mean were replaced by the 2.5 sd cut-off value for each participant. To remove individual differences in reading speed, statistical results reported below were based on length-corrected residual reading times computed separately for each participant by entering their reading times for every word in all sentences (including distractors) into a regression equation that took reading times as the dependent variable and word length as the independent variable, and then subtracted the predicted reading times from the actual reading times (Ferreira & Clifton, 1986; Trueswell et al., 1994). The graphs, however, show reading times without this correction procedure.

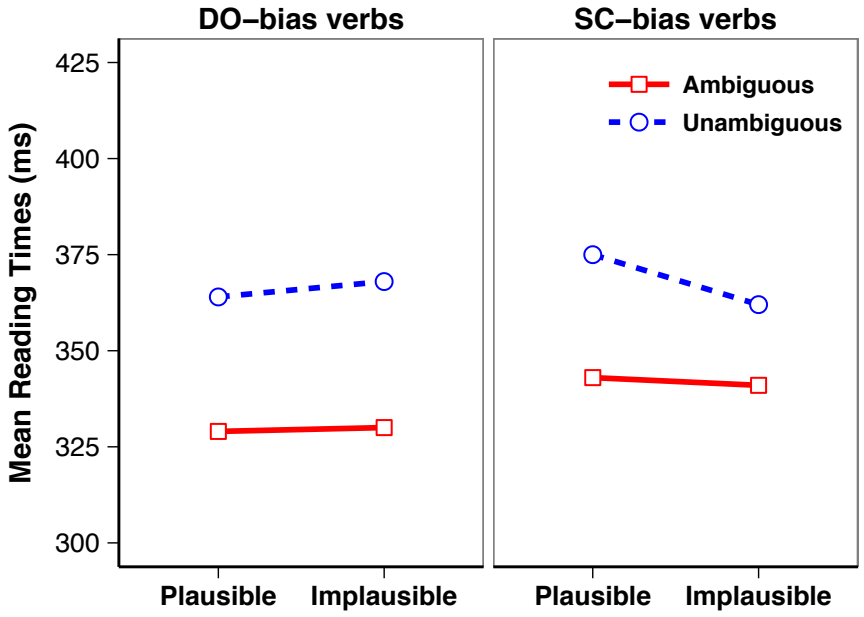

Figure 6. Mean reading time at the ambiguous noun.

At the ambiguous noun, there was a main effect of ambiguity, with the ambiguous condition read 28 ms faster than the unambiguous condition (336 vs 368 ms, β=30.78, *SE*=5.58, t>5), as shown above in Figure 6. This ambiguity effect was significant in all lower level analyses (DO plausible, DO implausible, SC plausible, SC implausible, all *t*s>2.6). The ambiguity effect could be explained in two ways. First, it might reflect spillover effect from the previous main clause verb, which was followed by a comma in the unambiguous condition but not in the ambiguous condition. So the slower reading time at unambiguous condition could be ascribed to the reading time of the additional comma. Another possibility was that after the unambiguous condition, the parser was ready to build a sentential complement structure starting with the ambiguous noun, and getting ready to build this complex structure needed more processing effort than not building an upcoming embedded clause structure, which should not yet

be happening in the ambiguous conditions. The slower reading times for the unambiguous condition probably reflected some combination of these two factors.

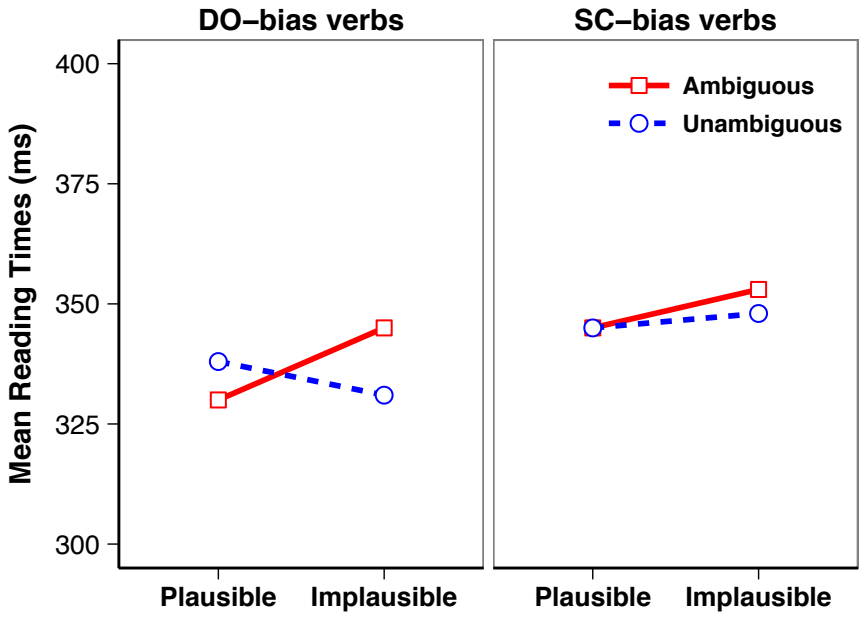

Figure 7. Mean reading time at the critical region (averaged across two words).

At the critical region, there was an interaction between ambiguity and plausibility (β=12.69, *SE*=5.63, t>2). Analyses breaking reading times down by verb type revealed that this interaction came entirely from DO-sentences. Within DO-sentences, there was an interaction between ambiguity and plausibility (β=21.14, *SE*=7.90, t>2), which resulted because the DO plausible ambiguous condition did not differ from the DO plausible unambiguous condition (t<2), but the DO implausible ambiguous condition was read slower than the DO implausible unambiguous condition (β=13.00, *SE*=6.29, t>2). There were no differences between SC plausible ambiguous and unambiguous conditions (t<1) and between SC implausible ambiguous and unambiguous conditions (t<1).

Following DO-bias verbs, there was a numeric trend that the blended structure was read more quickly than the comma-disambiguated sentential complement structure when the

ambiguous noun was plausible, but this difference did not reach reliability. However, this is consistent with the author's intuition that blended structures are easier to process than sentential complements because they are used very frequently in Mandarin. The implausible-noun-disambiguated sentential complement was read significantly slower than the comma-disambiguated sentential complement, suggesting that plausibility is less effective than the comma in disambiguating this type of sentences towards sentential complement endings. No plausibility by ambiguity interaction was seen in sentences with SC-bias verbs. The fact that the ambiguous and unambiguous conditions after SC-bias verbs did not differ in any of the comparisons suggested that at the ambiguous noun, neither the direct object analysis nor the blended structure analysis were ever seriously considered by the parser. The parser expected the words following the SC-bias verbs to continue into sentential complements, and therefore did not experience difficulty in processing ambiguous structures relative to comma-disambiguated sentential complement structures.

Discussion

The present study investigated the relative importance of verb bias and plausibility cues in processing Mandarin sentences. Results showed that Mandarin speakers relied more on the verb bias cue to predict the upcoming structure, since there were no ambiguity effects following SC-bias verbs. It was only after DO-bias verbs that there were any effects of ambiguity or plausibility. After DO-bias verbs, reading times were slowest for the implausible-noun-disambiguated condition, suggesting a preference for the blended structure when the noun's plausibility allowed it. Since there were no differences between conditions after SC-bias verbs,

verb bias seems to have dominated. Thus, just like English speakers, Mandarin speakers rely on verb bias information when processing ambiguous sentences.

Plausibility of the ambiguous noun as the direct object of the main clause verb appeared to influence the processing of the following words, such that the following words were processed as a sentential complement when the ambiguous noun was implausible and as part of a blended structure when the ambiguous noun was plausible, but only after DO-bias verbs. This result indicated that when verb bias allowed the following structure to be ambiguous between a blended structure and a sentential complement, then plausibility played a role in selecting between the two. Thus verb bias trumps plausibility in processing Mandarin sentences. Plausibility only influences parsing when verb bias allows it to.

These findings contrast with those from English verb bias studies. In English, plausibility has been found to have no influence on parsing when a strong verb bias cue is present, as evidenced by the absence of a plausibility effect at the disambiguating region in both Garnsey et al., (1997) and in Chapter 2 of this dissertation. It is likely that the crucial difference between the English and Mandarin studies lies in the nature of the ambiguity and what plausibility can contribute. In the type of sentence investigated in Mandarin here, unlike in English, plausibility actually is the deciding factor between two possible structures. When a DO-bias verb predicts that the noun following it should be its direct object, then the blended structure is allowed, with the noun being both the DO-bias verb's object and the subject of the embedded clause. However, when a DO-bias verb predicts that the noun following it should be its direct object but then the noun is not plausible in that role, then the blended structure is ruled out. Thus, plausibility is the cue that determines the structure. In contrast, in English the cue that absolutely determines the structure is the presence of an embedded verb, rather than anything about plausibility. The

difference between the two languages in the informativeness of the plausibility cues explains the different effects of plausibility. No similar reading time differences appear after SC-bias verbs, however, showing that verb bias is still the dominant cue, just as it is in English.

CHAPTER 4

Reanalysis and Lingering Misinterpretation

It has been well established that when reading sentences like (22), readers slow down at the main clause verb *ran*, presumably because they have initially interpreted the noun phrase *the deer that was brown and graceful* as the object of the subordinate clause verb *hunted*. At the main clause verb *ran*, the parser realizes that *ran* lacks a subject and triggers reanalysis processes. This is termed the garden-path effect. Successful reanalysis would lead to the noun phrase being deleted from the object role of the subordinate clause verb *hunted* and attached to the main clause verb *ran* to be its subject. Garden-path sentences like (22) have been studied extensively in psycholinguistic research (Ferreira & Henderson, 1990; Garnsey et al., 1997; Pickering & Traxler, 2003; Pickering et al., 2000; Trueswell et al., 1993) as a way to distinguish among theories of sentence processing.

(22) While the man hunted the deer that was brown and graceful ran into the woods.
(23) Did the man hunt the deer?
(24) Did the deer run into the woods?

Traditional sentence processing models differ on the timing of the parser's use of non-syntactic information to constrain the building of the syntactic structure, but they all assume that the parser always reaches the correct interpretation that is faithful to the linguistic input when parsing is completed. According to serial two-stage models, which are best represented by the Garden-path Model (Ferreira & Clifton, 1986; Frazier, 1979, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983), initial decisions about syntactic structure are based solely on syntactic information. When there are multiple possible syntactic structures, the parser prefers the simplest structure that includes the fewest number of nodes in

the syntactic tree (the Minimal Attachment Principle) and prefers to attach the incoming word to the current constitutent, rather than to initiate building a new constitutent (the Late Closure Principle). Once a single syntactic structure is selected, other sources of information, including non-syntactic ones, are used in the second stage to evaluate that selection (Frazier & Rayner, 1982). Interactive one-stage models, such as constraint-based models, on the other hand, state that multiple structures are activated simultaneously, and non-syntactic information comes into play from the beginning to determine the relative activation levels of these structures (Garnsey et al., 1997; MacDonald et al., 1994; McRae et al., 1998; Trueswell et al., 1994; Trueswell et al., 1993). The two types of models differ with respect to what happens at the subordinate verb *ran*. In the serial models, the main clause verb *ran* triggers the reanalysis processes that eliminates the direct object reading of *the deer* and reanalyzes it as the subject of the main clause, while in the interactive models, both the direct object and the subject analyses of *the deer* are activated when *the deer* is encountered, with the direct object analysis being ranked higher than the subject analysis. The verb *ran* then triggers re-ranking of the two analyses. In the present study, we use the term *reanalysis* to mean both the reanalysis process in the serial models and the re-ranking process in the interactive models. Despite differences in the two major processing theories, both of them assume that the parser eventually builds a complete, detailed and correct structure and attains the correct interpretation by the end of a sentence.

However, since the seminal work of Christianson, Hollingworth, Halliwell and Ferreira, (2001) and Ferreira, Christianson and Hollingworth (2001), there is increasing evidence that readers do not always reach the correct interpretation of garden-path sentences. After reading sentences like (22), readers often answer *yes* incorrectly to questions like (23), indicating that they interpret the sentence as meaning that *the man hunted the deer and the deer ran into the*

woods, although this interpretation is not licensed by the syntax (Christianson et al. 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira et al., 2001). The error rate was over 50% in Christianson et al. (2001) and this was replicated in follow-up studies (Christianson et al., 2006; Ferreira & Patson, 2007). In contrast to the high error rate to questions like (23), readers are highly accurate in answering *yes* to questions like (24), leading Christianson and colleagues to conclude that reanalysis processes started and were performed to such an extent that *the deer* was successfully attached to the main clause, but reanalysis was not carried out to the fullest degree to erase it from the direct object role in the subordinate clause. Such incomplete reanalysis results in the lingering misinterpretation derived from the initial misparse (i.e., *the man hunted the deer*).

One criticism of interpreting the high error rate to comprehension questions like (23) as reflecting lingering initial misinterpretation is that readers might have answered the questions based on inferences that they drew after reading this type of sentences. This possibility is supported by the high error rate to questions following unambiguous versions of the sentences. In Christianson et al. (2001), the error rate to comprehension questions was about 70% after reading (22) and about 50% after reading the comma-disambiguated version, *While the man hunted, the deer that was brown and graceful ran into the woods*, and the reverse-order disambiguated version, *The deer that was brown and graceful ran into the woods while the man hunted*. Although the difference in question accuracy between ambiguous and unambiguous versions could be ascribed to garden-pathing, the still-high error rate for unambiguous sentences suggested that at least a portion of the misinterpretation was not caused by garden-pathing. Readers might have answered the questions based on inferences, because after all, when a man is hunting and a deer is running into the woods, what else would the man be hunting?

To reduce the likelihood of answering questions based on inferences, Christianson et al. (2001) used sentences and questions like (25) and (26).

(25) While Anna dressed the baby who was cute and small spit up on the bed.
(26) Did Anna dress the baby?

Unlike in (22), in which the subordinate clause verb is an Optionally Transitive (OPT) Verb that can be either transitive or intransitive, the subordinate clause verb *dressed* in (25) is a Reflexive Absolute Transitive (RAT) verb, which takes its subject as the object if an object is not explicitly mentioned. Successful reanalysis of (22) results in an unspecified direct object of the subordinate verb *hunted*, leading to the interpretation that *the man hunted something unknown*. In contrast, complete reanalysis of (25) would result in the interpretation that *Anna dressed herself* rather than somebody unknown, thus leaving little room for inference. Indeed, Christianson et al. found that readers make 20% fewer errors after reading ambiguous sentences with RAT verbs like (25) than ambiguous sentences with OPT verbs like (22). However, readers still make more errors to ambiguous than unambiguous sentences with both RAT and OPT verbs. These extra errors are probably due to garden-pathing. However, the high error rates for unambiguous sentences show that garden-pathing is not the only reason for responding incorrectly.

A potential criticism of Christianson and colleagues' interpretation of their results is the possibility that the initial misinterpretation may have been discarded after complete reanalysis, but then was reactivated by comprehension questions that directly probed the misinterpretation (*Did the man hunt the deer?*), which was similar at surface-level to the ambiguous condition (*While the man hunted the deer…*) but less so to the unambiguous condition because of the presence of the comma in the unambiguous condition (*While the man hunted, the deer…*). This surface-level resemblance might have caused readers to answer questions more incorrectly

following ambiguous than unambiguous conditions (Sturt, 2007; van Gompel, Pickering, Pearson, & Jacob, 2006)

To address this issue, several studies employed more indirect and implicit measures to examine the existence of lingering misinterpretation, such as syntactic priming (van Gompel et al., 2006), a grammaticality judgment task after speech repair (Lau & Ferreira, 2005), processing newly-learned structures (Kaschak & Glenberg, 2004), paraphrasing (Patson, Darowski, Moon, & Ferreira, 2009) and processing subsequent sentences (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013). For example, Van Gompel et al. (2006) asked participants to read sentences that were either ambiguous (*While the man was visiting the children played outside*) or unambiguous (disambiguated with a comma) and then complete a sentence fragment (*While the doctor was visiti…*). Participants produced more transitive structures following ambiguous than unambiguous sentences. This result was interpreted as showing that the initial misparse remained active even after reanalysis was conducted and primed the structure produced in the sentence completion task. Similarly, when asked to paraphrase the sentences they have just read, participants produced more paraphrases that retained the meaning of the initial misanalysis (e.g., *The man hunted the deer and it ran into the woods*) after reading ambiguous than unambiguous garden-path sentences (Patson et al., 2009).

In two eye-tracking experiments, Slattery et al. (2013) examined whether reanalysis was completed and whether semantics from the initial misanalysis persisted after full reanalysis. In one experiment, readers read ambiguous and unambiguous sentences, in which the gender of the reflexives in the main clause either matched or did not match the ambiguous noun, as in *After the bank manager telephoned(,) David's father/mother grew worried and gave himself approximately five days to reply.* They found that readers slowed down at the reflexive *himself*

when the gender did not match the ambiguous noun (*David's mother*) in both ambiguous and unambiguous conditions, indicating that in the ambiguous condition, the ambiguous noun had been successfully reanalyzed as the subject of the main clause by the time the reflexive was reached, thus allowing the parser to analyze it as the antecedent of the reflexive. In another experiment, Slattery et al. used two-sentence paragraphs to examine whether the initial misinterpretation affected the processing of the second sentence. Experimental sentences contained RAT verbs and crossed ambiguity and plausibility, as in *While Frank dried off(,) the truck/grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.* A plausibility manipulation concerned whether the ambiguous noun was plausible as the direct object of the subordinate clause verb (*dry off the truck* vs. *dry off the grass*). The rationale was that if the initial misinterpretation, *Frank dried off the truck/grass*, was successfully abandoned after reanalysis and the reflexive reading, *Frank dried off himself*, was attained, readers would not slow down at the reflexive *himself* in the second sentence. On the other hand, if the initial misinterpretation was retained even after reanalysis was performed, readers would slow down at *himself*, because *drying himself* in the second sentence would be incompatible with the semantics of the misanalysis *dried off the truck/the grass*. Slattery et al. found that readers slowed down at *himself* only in the ambiguous plausible condition (*While Frank dried off the truck…*), which showed that semantics from the initial parse persisted and conflicted with the semantics of the second sentence, and that the degree of persisting misinterpretation was affected by plausibility. In addition, this study also showed that reanalysis could occur very fast, within several words following the disambiguating verb, and that even after reanalysis was completed, the parser still failed to erase the semantics

derived from the initial misinterpretation (see Sturt, 2007, for similar conclusion using a different structure).

Slattery et al.'s results complemented Christianson et al. (2001) and Ferreira et al. (2001)'s question response findings by providing evidence using more implicit measures. Christianson et al. and Ferreira et al. concluded that the ambiguous noun is successfully attached to the main clause after reanalysis is performed because readers answer highly accurately to the question *Did the deer run into the woods?*. But the ambiguous noun remains as the direct object of the subordinate clause verb, because readers make many errors answering the question *Did the man hunt the deer?*. This conclusion is based on comprehension accuracy to questions that directly probe the initial misinterpretation, an approach that has been questioned by some researchers (Nakamura & Arai, in press; Sturt, 2007; van Gompel et al., 2006). However, Slattery et al.'s findings, along with the other studies mentioned above, provided evidence that lingering misinterpretation is unlikely to be an artifact of the type of questions asked by Christianson et al. (2001) and Ferreira et al. (2001).

The effect of lingering misinterpretation has also been found in sentences with other types of syntactic ambiguity (Kaschak & Glenberg, 2004; Lau & Ferreira, 2005; Sturt, 2007, see Nakamura & Arai, in press, for lingering misinterpretation in Japanese). Lingering misinterpretation appears to be a universal phenomenon rather than occurring only with the direct object/main clause ambiguity such as (22). For instance, Sturt (2007) constructed direct object/sentential complement type of garden-path sentences in which reanalysis was relatively straightforward (Grodner, Gibson, Argaman, & Babyonyshev, 2003; Sturt et al., 1999). The final segment of experimental sentences were either consistent or inconsistent with the initial misinterpretation, as in *The explorers found the South Pole was actually right at their feet. vs The*

explorers found the South Pole was impossible to reach. In this study, an ambiguity effect was found to be localized to the disambiguating verb, which Sturt (2007) interpreted as indicating that reanalysis occurred and was completed quickly. Crucially, despite reanalysis efforts, readers still read final segments that conflicted in meaning with the initial misanalysis (*impossible to reach*) slower than those that did not (*right at their feet*), suggesting that the initial misinterpretation lingered although reanalysis was completed.

Even misinterpretation that is activated very briefly before being abandoned can persist and affect the processing of the following text (Kaschak and Glenberg, 2004; Lau & Ferreira, 2005). In Kaschak and Glenberg (2004), a group of participants read sentences that contained a novel structure like (27) and a control group read sentences like (28) in the training session. *Cleaned* in (27) could be analyzed temporarily as a modifier as in *The wood floor needs cleaned corners* while *cleaned* in (28) could not. At the testing session when both groups read sentences that contained a modifier (*cooked*), the group that had been exposed to the novel construction read *cooked* faster than the group that did not. This result indicated that when *cleaned* was misanalysed as the modifier in at least some trials in the training session, the misanalysed structure remained activated and facilitated the reading of *cooked* in (29).

(27) The wood floor needs cleaned before our parents get here.
(28) The wood floor needs to be cleaned before our parents get here.
(29) The meal needs cooked vegetables so the guests will be happy.

Similar results were reported by Lau and Ferreira (2005) in a disfluency study, in which listeners rated sentences like *The girl chosen, uh, selected for the role celebrated with her parents and friends* as more acceptable than sentences like *The girl picked, uh, selected for the role celebrated with her parents and friends*, because *chosen* activated a reduced relative clause structure, which remained activated after the error correction signal *uh* and primed the reduced

relative clause reading of *selected*, whereas *picked* was structurally ambiguous between main clause and reduced relative clause readings, just like *selected*. Listeners also rated *The little girl picked, uh, selected the right answer, so her teacher gave her a prize* as more acceptable than *The little girl chosen-uh selected the right answer, so her teacher gave her a prize*, because the reduced relative clause structure activated by *chosen* lingered after it was corrected and interfered with the main clause structure activated by *selected*.

The studies above showed that misinterpretation from an initially built syntactic structure tends to linger after reanalysis. While it has reached consensus that interpretation from the first-pass parse sometimes linger, researchers differ on what causes this to happen. Lingering misinterpretation has been ascribed to shallow syntactic processing (Clahsen & Felser, 2006; Frisson, 2009), underspecified syntactic structure built by the parser (Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002; Swets, Desmet, Clifton, & Ferreira, 2008), memory traces left from the process of computing the initial parse (Kaschak & Glenberg, 2004), shallow and underspecified semantic processing (Barton & Sanford, 1993), and fast-decaying syntactic structure (Sachs, 1967; Sturt, 2007).

Several processing accounts, however, have ascribed the lingering misinterpretation to incomplete reanalysis, including the Attach Anyway and Adjust Principle (Fodor & Inoue, 1998), lexically guided tree-adjoining grammar (Ferreira, Lau, & Bailey, 2004; Lau and Ferreira, 2005) and the Good-Enough Processing Account (Christianson et al. 2001; Ferreira et al. 2001). According to the Attach Anyway and Adjust Principle, the parser attaches every incoming word into the existing structure even if such integration results in syntactic incompatibility. When syntactically illicit structure results, the parser starts to revise the structure step by step in a backward manner. In sentences like (22), *ran* is initially analyzed as the matrix verb although it

lacks a subject. The parser then revises the already-built structure by stealing *the deer* from the subordinate clause and attaching it to the main clause. It then proceeds to reinterpret *hunted* as an intransitive verb. However, reanalysis may cease before it is completed, resulting in what Fodor and Inoue (1998) call the Thematic Overlay Effect, which has *the deer* remain both as the patient of *hunted* and the agent of *ran*. Similarly, Ferreira and colleagues' lexically guided tree-adjoining grammar (LTAG) account proposes that the correct structure built after reanalysis is overlain onto the initial incorrect structure because the initial incorrect structure has not decayed in memory. The not-yet-decayed incorrect structure competes with the correct structure to influence the processing of subsequent sentences until such decay is completed. This process results in a "tree-splicing" structure that has the correct structure spliced onto the initial incorrect structure (Christianson et al. 2001).

Most relevant to the present study is the Good-Enough Processing Account (Christianson et al., 2001; Ferreira et al., 2002; Ferreira et al., 2001; Ferreira & Patson, 2007), which states that when the interpretation derived from the initial misanalysis is sensible, the parser does not bother to fully reanalyze the structure even though later information is syntactically incompatible with the existing structure. The Good-Enough Processing Account assumes the dual-pathways processing model, in which the semantic processing route and the morphosyntactic processing route operate independently. Each of the two routes outputs its own interpretation. When the interpretations delivered by the two routes fail to converge, the parser reconciles them, resulting in a final interpretation that is not completely faithful to the linguistic input. In the case of garden-path sentences like (22), the sensible meaning derived from the initial misanalysis cancels out the need of computing detailed structure via the morphosyntactic processing route,

leading to incomplete syntactic reanalysis and the resultant lingering misinterpretation from the initial misparse.

The idea that world-knowledge heuristics may terminate parsing before a detailed representation is reached is supported by the Moses Illusion (Erickson & Mattson, 1981; Kamas, Reder, & Ayers, 1996). When asked "How many of each kind of animal did Moses take on the ark?", listeners typically answer "two" without pointing out that it is Noah rather than Moses who put animals on the ark. Similarly, after reading "The authorities were trying to decide where to bury the survivors.", readers usually do not realize that "survivors should not be buried" (Barton & Sanford, 1993).

In a series of experiments, Ferreira and colleagues also demonstrated that the parser sometimes opts for the interpretation derived from the semantic heuristics, especially when the syntactic algorithm is demanding and the syntactically licensed interpretation is implausible (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003). In Ferreira (2003), participants listened to sentences like *The dog bit the man*; *The man bit the dog*; *The man was bitten by the dog* and *The dog was bitten by the man*, and then answered questions about the agent and patient roles of these sentences. They made errors to implausible passives (*The dog was bitten by the man*.), but not plausible and implausible actives and plausible passives. Most of the errors involved flipping the thematic roles. In English, the NVN word order usually maps onto Agent-Verb-Patient thematic roles. In the case of implausible passives, the word-order heuristics delivers an analysis with *the dog* being the agent and *the man* being the patient, which is in conflict with the output from the syntactic processing route. Because NVN word-order is a very powerful heuristic and the nouns fit well with its usual thematic role assignments, it overrides the interpretation from the syntactic route, resulting in misinterpretation. Ferreira et al.'s (2003) findings were replicated by

Christianson et al., (2010) in a structural priming task, in which participants produced passive structures after reading implausible active sentences and produced active structures after reading implausible passive sentences. This is because outputs from the syntactic route that are not consistent with world-knowledge are "normalized" by the plausibility heuristics to make the sentence sensible (Bever, 1970; Ferreira, 2003; Townsend & Bever, 2001).

Evidence from electrophysiological studies also indicated that the semantic processing route sometimes cancels out or wins over the syntactic processing route (Kuperberg, 2007). In Kim and Osterhout (2005), participants read sentences like *The hearty meal was devouring the kids.* Since a meal cannot devour something, *devouring* should elicit an N400 effect, which is an ERP component that indexes semantic incongruence (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). However, the response to *devouring* showed an effect on the P600 component, which is usually elicited by syntactic violations (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), despite the fact that there was no syntactic violation or ambiguity in the sentence. Such "semantic P600" effects have been found typically with role-reversal sentences (Kolk, Chwilla, van Herten & Oor, 2003; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; van Hertan, Chwilla, & Kolk, 2006; van Herten, Kolk, & Chwilla, 2005; see Kuperberg, 2007 for a review). The semantic P600 effect provides evidence for the existence of the syntactic and semantic dual processing routes. Most importantly, the absence of the N400 effect suggests that information derived from the semantic route can be strong enough to even cause the parser to "normalize" the syntax to make it consistent with the semantics (Kim & Osterhout, 2005; Kuperberg, 2007).

Existing literature on Good-Enough Processing has proposed two mechanisms that could account for lingering misinterpretation. The first mechanism is that the semantics of the initial

misinterpretation cancels out the need to fully reanalyze the syntactic structure, resulting in lingering misinterpretation (Christianson et al. 2001; Ferreira et al. 2001). The second mechanism is that reanalysis is completed, but the interpretations from both the initial analysis and reanalysis co-exist (Ferreira, 2003; Slattery et al. 2013). In other words, according to the Good-Enough Processing Account, if either reanalysis of the syntactic structure is unsuccessful, or syntactic reanalysis succeeds, but both analyses linger, then the initial misinterpretation might linger. For the rest of the chapter, we will refer to the first mechanism as the "Incomplete Reanalysis" version and the second mechanism as the "Lingering Interpretations" version of the Good-Enough Processing Account.

The Good-Enough Processing Account is an important sentence processing model that addresses the issue of persistent misinterpretation that is not accounted for by the traditional sentence processing theories. However, to the best of our knowledge, no empirical study has specifically tested the Good-Enough Processing Account by examining the relation between on-line measures of the disambiguating verb and the off-line comprehension accuracy. The present study aims to do so. In two self-paced reading and two ERP experiments, we specifically test the "Incomplete Reanalysis" version of the Good-Enough Processing Account.

Assuming that responses to comprehension questions reflect whether or not the initial misinterpretation persists to the end of the sentence, then the incomplete reanalysis explanation of question responses predicts more evidence of reanalysis in trials with correctly answered questions than in those with incorrectly answered questions. If longer reading time at the disambiguating verb is evidence of more complete reanalysis and complete reanalysis leads to correct question responses, then longer reading time should predict better question accuracy. The P600 ERP component has been found to be associated with syntactic reanalysis (Osterhout et al.,

1994) or other types of syntactic processing (detection of syntactic anomaly, Frisch, Schlesewsky, Saddy, & Alpermann, 2002; syntactic integration, Kaan, Harris, Gibson, & Holcomb, 2000), and its amplitude has been found to be larger when reanalysis is more difficult (Osterhout et al. 1994) . If bigger P600 amplitude at the disambiguating verb in ambiguous sentences indexes amount of reanalysis work and complete reanalysis leads to better comprehension, then bigger P600 amplitude should predict better comprehension accuracy. Following this logic, the present study compares reading times (two experiments) and P600 amplitudes (two experiments) at the disambiguating verb between trials that are answered correctly and those that are answered incorrectly to test the incomplete-reanalysis explanation for lingering misinterpretation. Slower reading times and bigger P600 amplitudes for ambiguous sentences that are answered correctly than for those that are answered incorrectly would support such an explanation. However, there is also a possibility that slower reading times and/or bigger P600 amplitude may indicate amount of confusion rather than success of reanalysis, and we will return to this point in the Discussion.

The severity of garden-pathing and the likelihood of recovery from garden-pathing have been found to be affected by the distance between the ambiguous noun and the error signal, because the parser assigns thematic roles to syntactic structures when encountering the head of a noun phrase (Ferreira & Henderson, 1991; 1998, Frazier & Clifton, 1998; Tabor & Hutchins, 2004; Van Dyke & Lewis, 2003; Warner & Glass, 1987). Previous studies showed that readers judged the sentence *While the man hunted the deer that was brown and graceful ran into the woods* as less acceptable than *While the man hunted the brown and graceful deer ran into the woods*, because in the former, the parser has committed to the incorrect direct object analysis for a longer time compared to the latter by the time the disambiguating verb is reached, and therefore it is harder to abandon it (Ferreira & Henderson, 1991; 1998). Readers also make more

errors to the comprehension questions that target the misinterpretation following ambiguous sentences with post-noun-modification (*the deer that was brown and graceful*) than those with pre-noun-modification (*the brown and graceful deer*) (Christianson et al. 2001).

In the present study we need enough trials with correct and incorrect question responses to be able to compare reading times and P600 amplitude for correctly-answered and incorrectly-answered trials, so we will use garden-path sentences with post-noun modification, which has been found to elicit more incorrect responses than those with pre-noun modification. As the garden-pathing effect in both the on-line measures and the off-line comprehension accuracy is bigger in sentences with post-noun modification than those with pre-noun modification, it is more likely that we will find a difference in reading times and P600 amplitude between trials with correct answers and incorrect answers when using garden-path sentences with post-noun modification.

Experiment 4

Method

Participants

Thirty-two undergraduate students (12 males; mean age 18.5; range 18-21) at the University of Illinois at Urbana-Champaign participated in Experiment 4. All were native speakers of English, had normal or corrected-to-normal vision and gave written informed consent.

Materials and Design

Experimental sentences consisted of forty sets of sentences with OPT verbs and twenty-four sets of sentences with RAT verbs, with each set containing an ambiguous and a comma-disambiguated unambiguous version, as illustrated below in (30) and (31). In all sentences, the ambiguous noun was followed by a relative clause that comprised two adjectives (e.g., *that was brown and graceful*). Across the experiment, each OPT verb was used in just one item set and each RAT verb was used in two item sets, because there are fewer RAT verbs than OPT verbs. All sentences with OPT verbs and half of the sentences with RAT verbs were taken from Christianson et al. (2001).

(30) Critical sentence with OPT verb:

a. Ambiguous:
While the man hunted the deer that was brown and graceful ran into the woods.
b. Unambiguous:
While the man hunted, the deer that was brown and graceful ran into the woods.
Comprehension question:
Did the man hunt the deer?

(31) Critical sentence with RAT verb:

a. Ambiguous:
While Anna dressed the baby who was cute and small spit up on the bed.
b. Unambiguous:
While Anna dressed, the baby who was cute and small spit up on the bed.
Comprehension question:
Did Anna dress the baby?

Critical sentences were distributed over two lists using a Latin Square design, so that each participant saw only one version from each item set and an equal number of sentences in

each condition. Each sentence was followed by a comprehension question that directly probed the misinterpretation.

Ninety-two distractors were added to each list for a total of 156 trials/list. There were three types of distractors: (1) unambiguous sentences with subordinate-matrix clause order (e.g., *While Jenifer held the cigar that was aged and expensive she told bad jokes*; 40 sentences); (2) unambiguous sentences with matrix-subordinate clause order (e.g., *The mother comforted the toddler who was chubby and scared while the clown handed him a balloon*; 40 sentences); and (3) ambiguous and unambiguous versions of sentences using reciprocal verbs such *met*, which are similar to RAT verbs in that their subject is also their object when no other object is specified [e.g., *As Jane and Mary met(,) the men from Florida drove past them*; 12 items]. Comprehension questions to the first two types of distractors asked about the content of various parts of the sentences, and questions to the third type of distractors asked about misinterpretation. Answers to the first two types of distractors were half *yes* half *no* across the experiment. All sentences were pseudo-randomized once and presented to all participants in the same order across all lists. No two experimental items appeared consecutively.

Procedures

Participants sat in a dimly lit sound-attenuated booth in front of a 23-inch LCD monitor. To make presentation mode comparable for the self-paced reading and ERP experiments, sentences were presented one word at a time in white 26-point Arial font on a black background in the center of the screen. Each trial began with a "Ready" prompt that stayed on the screen for one second. Each time participants pressed a button on a Cendrus-830 response box, the next word appeared to replace the previous word in the center of the screen. Following each sentence,

a comprehension question was presented and participants pressed one of two buttons to indicate their answers. Feedback about question accuracy was not given. However, a "Too Slow" message prompted when participants did not make a response within four seconds. A total of 156 sentences was divided into four blocks of thirty-nine sentences each, and participants took a short break after each block. A practice block of seven trials was added at the beginning. The entire experiment took approximately forty minutes to complete.

Results

Reading times were analyzed at two sentence regions: 1) the disambiguating region, consisting of the disambiguating verb (e.g., *ran*) and the word following it, and 2) the post-disambiguating region, consisting of the 1-3 words following the disambiguating region through the end of the sentence. The post-disambiguating region was analyzed to address the possibility that reanalysis effects might spill over onto subsequent words, as often happens with self-paced reading times. Linear mixed effect models were used to analyze the reading times, with ambiguity as a fixed effect and subjects and items as random effects.

Comprehension accuracy was analyzed using logit mixed-effect models with binomial function (Jaeger, 2008) in R (R Development Core Team, 2008), including ambiguity and reading time at the disambiguating region as well as their interaction as fixed effects and subjects and items as random effects, with random slopes and intercepts for subjects and items.

For all analyses, the initial model included a maximal random effects structure that included all fixed effects, random intercepts and random slopes for all fixed effects for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013). If the maximal model failed to converge, the random slopes of fixed effects were removed, one at a time, based on the values in

the Hessian matrix. All fixed effects were centered to avoid colinearity. The final models reported here were the most complex models that converged. For logit mixed-effect models used to analyze question response accuracy, estimates, standard errors, and *z*- and *p*-values for fixed effects are reported. For linear mixed-effect models used to analyze reading times, estimates, standard errors and *t*-values are reported, with $t>2$ in linear mixed-effect models being interpreted as significant. Items with OPT verbs and RAT verbs were analyzed separately and the results are reported separately for the two verb types.

Prior to data analysis, word-by-word reading times that were faster than 100 milliseconds (ms) or slower than 2000 ms were excluded, leading to a loss of 0.5% of the data. Reading times were also excluded from further analysis for sentences after which participants failed to respond to the comprehension question within four seconds, affecting 2% of the data. Reading times above or below 2.5 standard deviations (sd) from the mean were replaced by the 2.5 sd cut-off value for each participant, affecting 3% of the data. To remove individual differences in reading speed, statistical results reported below were based on length-corrected residual reading times computed separately for each participant by entering their reading times for every word in all sentences (including distractors) into a regression equation that took reading time as the dependent variable and word length as the independent variable, and then subtracted the predicted reading times from the actual reading times (Ferreira & Clifton, 1986; Trueswell et al., 1994). The graphs, however, show reading times without this correction procedure.

Comprehension accuracy to distractors was used to examine whether participants were paying attention to the sentences. All participants were above 80% (range 80%-97%, mean 90%), indicating that they were attending to these sentences. Thus all participants' data were included in the analyses.

OPT verbs. The disambiguating region was read 30 ms slower in ambiguous (449 ms) than in unambiguous (419 ms) sentences (β=29.99, *SE*=8.53, *t*=3.52, $p<.01$), as shown below in Figure 8. (Standard errors have been adjusted for the within-subjects design in all figures [Morey, 2008; see also Cousineau, 2005; Loftus & Masson, 1994]). Question response accuracy was also affected by ambiguity, with 16% more erroneous "yes" responses to ambiguous (67%) than unambiguous (51%) sentences (β=1.1, *SE*=0.23, *z*=4.73, $p<.001$), as shown in Figure 9. Reading times at the post-disambiguating region was not affected by ambiguity (β=8.25, *SE*=7.99, *t*=1.03, $p>.1$), suggesting that there were no spillover effects of ambiguity on the post-disambiguating region in this study.

When reading times on the disambiguating region were broken down by the accuracy of the responses to the questions following the sentences, there was a numeric trend such that for ambiguous sentences, longer reading times were associated with correct responses (ambiguous correct: 455 ms; ambiguous incorrect 445 ms), while for unambiguous sentences, longer reading times were instead associated with incorrect responses (unambiguous correct: 409 ms; unambiguous incorrect: 430 ms). However, the analysis of question response accuracy revealed that there was neither a main effect of sentence reading time (β=0.05, *SE*=0.09, *z*=0.57, $p>.05$) nor any interaction between ambiguity and reading time (β=0.19, *SE*=0.16, *z*=1.18, $p>.05$) affecting the comprehension question responses. The lack of a significant effect of reading time or interaction between reading time and accuracy indicates that the amount of time readers spent on the disambiguating region was unrelated to their question response accuracy.

RAT verbs. The results for items with RAT verbs were the same as for items with OPT verbs. At the disambiguating region, reading times were 28 ms longer for ambiguous (432 ms) than unambiguous (404 ms) sentences (β=26.67, *SE*=10.11, *t*=2.64, $p<.05$), as shown in Figure 8.

By the post-disambiguating region, the effect of ambiguity on reading time was over (β=3.05, SE=11.07, $t<1$).

Analyses of question response accuracy for items with RAT verbs also showed a main effect of ambiguity, with 25% more errors for ambiguous (54%) than unambiguous (29%) sentences (β=1.87, SE=0.38, z=4.93, $p<.001$). Just as for items with OPT verbs, there was neither a main effect of reading time on the disambiguating region (p>.1) nor any interaction between reading time and ambiguity (ambiguous correct: 426 ms; ambiguous incorrect: 437 ms; unambiguous correct: 401 ms; unambiguous incorrect: 411 ms; p>.1) affecting question response accuracy, again indicating that reading time was unrelated to question response accuracy.

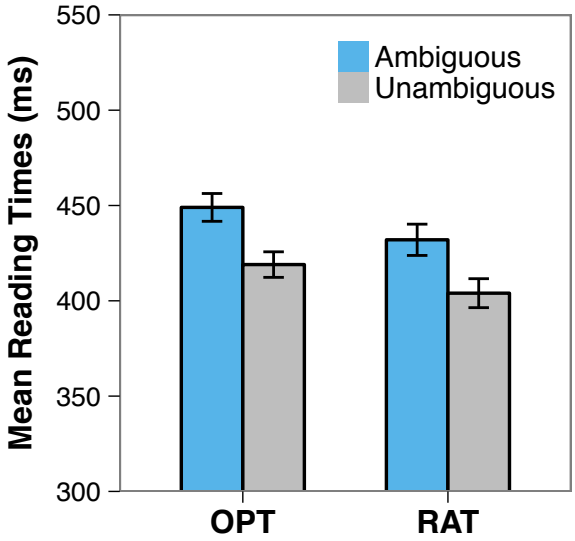

Figure 8. Reading time at the disambiguating region in Experiment 4, collapsing over question response accuracy. Error bars in all figures indicate standard errors.

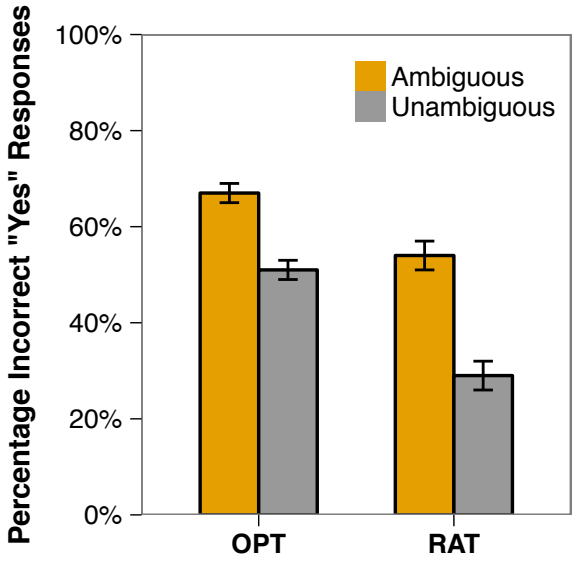

Figure 9. Error rates for question responses in Experiment 4.

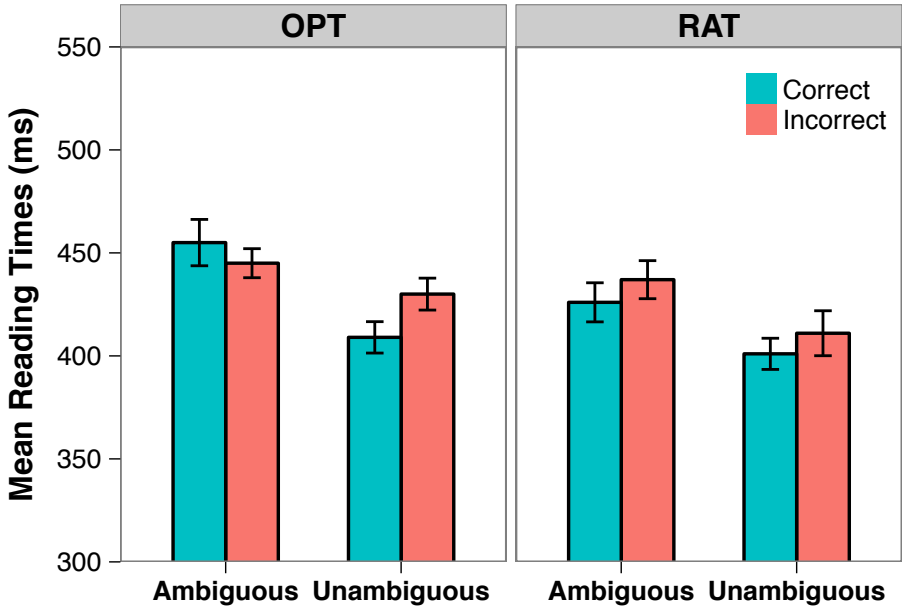

Figure 10. Reading time at the disambiguating region in Experiment 4 separately by question accuracy.

Discussion

The results of Experiment 4 showed that contrary to the prediction of the "Incomplete Reanalysis" version of the Good-Enough Processing Account, there was no evidence that more

time spent on the disambiguating region led to more correct question responses at the end of the sentence. In fact, in three out of the four comparisons (OPT unambiguous correctly vs. incorrectly answered trials; RAT ambiguous and unambiguous correctly vs. incorrectly answered trials), there was a numeric trend that longer reading time on the disambiguation was associated with more incorrect responses, as shown in Figure 10, which was in the opposite direction from the prediction of the Good-Enough Processing Account.

The fact that ambiguity only had an effect on the reading time at the disambiguating region but not at the region following it suggested that reanalysis was completed quickly, which is consistent with findings using eye-movement measures for similar sentences (Slattery et al. 2013; Sturt, 2007).

Why did reading times at the disambiguating region not predict question response accuracy? One possibility mentioned earlier is that people might at least sometimes answer the comprehension questions incorrectly based on inferences they draw from the content of these sentences. It is possible that the reason there was no relationship between reading time at the disambiguation and comprehension accuracy was that some incorrect responses were due to inferences rather than incomplete recovery from garden-pathing. Participants might have taken the time to fully reanalyze the sentence but then still respond incorrectly to the question because they also drew an inference. In *While the man hunted the deer ran into the woods*., they may have successfully reanalyzed *the deer* as the subject of *ran* rather than the object of *hunted* but still have inferred that the deer was what the man was hunting and answered the question based on that inference. To try to reduce the impact of inference, Experiment 5 asked questions like *Did the sentence explicitly say that the man hunted the deer?* to cue readers not to draw inferences when reading these sentences. It is possible that there would be a cleaner relationship

between reading times at the disambiguating region and question response accuracy when explicit questions discourage answering based on inferences that can easily be drawn from the sentence.

Experiment 5

Experiment 5 differed from Experiment 4 only in the type of questions asked after each sentence. In Experiment 4, non-explicit questions like *Did the man hunt the deer?* were asked, while in Experiment 5, explicit questions like *Did the sentence explicitly say that the man hunted the deer?* were asked to try to reduce effects of inference.

Method

Participants

Forty undergraduate students (16 males; mean age 20; range 18-25) at the University of Illinois at Urbana-Champaign participated in Experiment 5. All were native speakers of English, had normal or corrected-to-normal vision, gave written informed consent and received course credit for taking part.

Materials and Design

Critical sentences in Experiment 5 were exactly the same as Experiment 4, and were distributed over two lists according to a Latin Square design. 120 distractors were added so that there was more variety in sentence types. There were four types of new distractor sentences: (1)

ambiguous sentences in which the noun immediately following a verb turns out to be the subject of an embedded sentential complement rather than the direct object of the main clause, along with their unambiguous versions [e.g., *The naïve girl believed (that) the urban myth could teach her the real history*; 40 sentences]; (2) sentences matrix-subordinate clause order in which the noun immediately following the main clause verb is its direct object (e.g., *The union leader implied the raise when he met with strikers*; 50 sentences); (3) sentences with subordinate-matrix clause order like the experimental items, but containing both a direct object and a main clause subject (e.g., *While Janis watched the fish she cleaned the tank*; 20 sentences); and (4) unambiguous sentences with matrix-subordinate clause order (e.g., *The mother served the broccoli while the kids banged the table*; 10 sentences). Distractor types 2-4 were added so that the overall proportion of trials on which the noun immediately following a verb turned out to be its direct object, rather than needing to be reanalyzed as the subject of a subsequent clause, was higher. (Sentences of distractor type 1 were actually items for another experiment, not reported here.) For distractor types 2-4, the explicit question targeted various parts of the sentences. Correct answers to those distractors were half *yes* half *no*. All sentences were randomized once and then adjusted so that no two critical sentences appeared consecutively. Participants saw the same order of all sentences in all lists. A total of 184 trials was divided into four blocks with forty-six sentences each.

Procedure

Procedures in Experiment 5 were exactly the same as in Experiment 4.

Results

Average comprehension accuracy to distractors was 84% (range: 97%-70%), which was slightly lower than in Experiment 4. Presumably answering the explicit questions correctly required participants to suppress meanings derived from inference, which was harder than answering the non-explicit questions in Experiment 4. There were three participants who made over 25% errors to distractor items, but the results reported below include them since analyses with and without them yielded the same pattern of results. (All effects were slightly bigger when they were excluded.)

Data trimming and analyses were the same as for Experiment 4. Removing word-by-word reading times faster than 100 ms or slower than 2000 ms led to loss of 1% of the data. Removing reading times for trials on which participants failed to respond to the comprehension question affected 0.2% of the data. Replacing reading times that were above or below 2.5 sd away from the mean with the cut-off values for each participant affected 3% of the data.

For critical items, the most striking difference between the results of Experiments 4 and 5 was a drop in the overall error rate in question responses in Experiment 5 (Experiment 4: 50%; Experiment 5: 30%). Using explicit questions apparently succeeded, at least to some extent, in pushing participants to respond based on what they understood the sentence to have actually said had happened, rather than on inferences they could easily draw from the sentences.

OPT verbs. For sentences with OPT verbs, reading times on the disambiguating region showed a bigger effect of ambiguity in Experiment 5, with 50 ms longer reading times for ambiguous (449 ms) than for unambiguous (399 ms) sentences (β=53.99, *SE*=9.75, *t*=5.54, $p<.001$), as shown in Figure 11, compared to a 30 ms ambiguity effect in Experiment 4.

Question response error rates decreased for both ambiguous (48%) and unambiguous (19%) items compared to Experiment 4 (ambiguous: 67%, unambiguous: 51%), but did so especially for unambiguous sentences, as shown in Figure 12. As a result, the effect of ambiguity on question response accuracy was also bigger in Experiment 5 (29%) than in Experiment 4 (16%). A logit mixed-effect model with maximal random effect structure revealed a main effect of ambiguity on question response accuracy (β=2.24, *SE*=0.18, z=12.25, p<.001). One way that the results of the two experiments differed is this analysis also showed a main effect of disambiguation region reading time on question response accuracy (β=0.29, *SE*=0.08, z=3.40, p<.001) in Experiment 5, with longer reading times associated with incorrect question responses in both ambiguous and unambiguous conditions, as shown in Figure 13. (In Experiment 4, the trend was in the same direction.) Crucially, there was still no interaction between ambiguity and disambiguating region reading time (β=0.21, *SE*=0.16, z=1.27, p>.05) on question response accuracy.

Another way that the results of the two experiments differed was that the ambiguity effect in reading times persisted into the post-disambiguating region (ambiguous: 438 ms; unambiguous: 419 ms; β=22.25, *SE*=10.61, t=2.10, p<.05). Since there was no ambiguity effect on reading time at this region in Experiment 4, the explicit questions seem to have led to a longer lasting effect of ambiguity on reading times. However, the reading times at the post-disambiguating region did not affect question response accuracy the way the reading times at the disambiguating region itself did (ps>.05).

RAT verbs. Analyses of items with RAT verbs yielded similar results, with one exception noted below for question response accuracy. Reading times at the disambiguating region were 55 ms longer for ambiguous (455 ms) than unambiguous (400 ms) sentences

(β=58.62, *SE*=10.58, *t*=5.54, *p*<.001), which was a larger ambiguity effect than was found for the same sentences in Experiment 4 (28 ms). Just as for items with OPT verbs, the effect of ambiguity persisted into the post-disambiguating region in this experiment, with the ambiguous (458 ms) sentences read 30 ms slower than the unambiguous (428 ms) sentences.

The error rate for question responses decreased compared to Experiment 4, again especially for unambiguous sentences (ambiguous: 40%; unambiguous: 12%), as shown in Figure 12. The effect of ambiguity on response accuracy was significant (β=2.18, *SE*=0.60, *z*=3.66, *p*<.001), just as it was for items with OPT verbs. Different from items with OPT verbs, however, reading time at the disambiguating region did not affect question response accuracy (β=0.06, *SE*=0.13, *z*=0.45, *p*>.05). Like items with OPT verbs, there was no interaction between disambiguating region reading time and ambiguity (β=0.01, *SE*=0.25, *z*=0.03, *p*>.05) in the analysis of question response accuracy, as shown in Figure 13. There were also no effects of post-disambiguating region reading on question response accuracy.

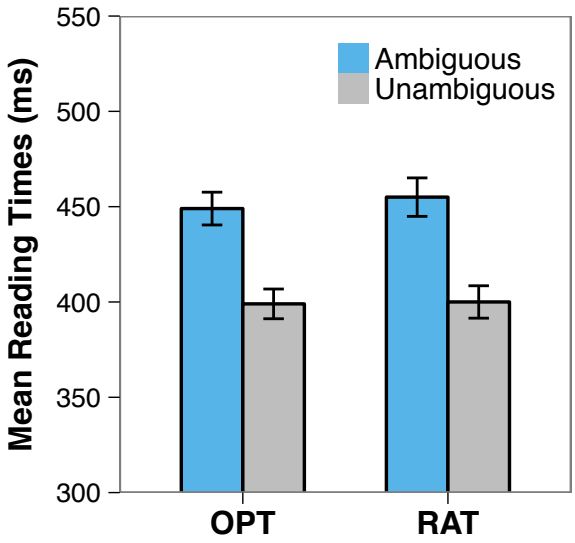

Figure 11. Reading time at the disambiguating region in Experiment 5, collapsing over question accuracy.

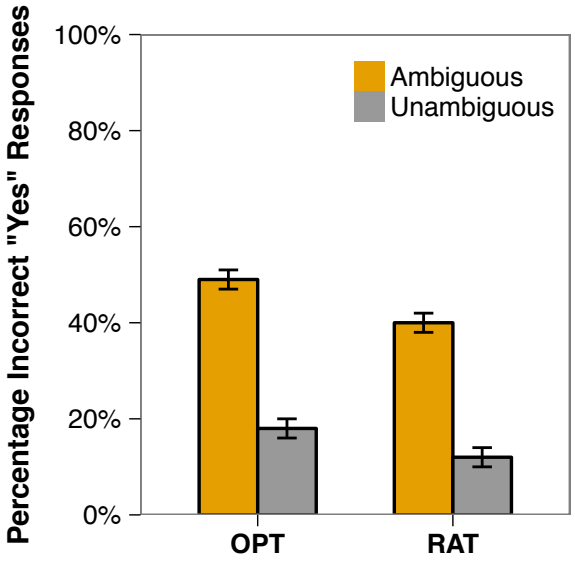

Figure 12. Error rates to question responses in Experiment 5.

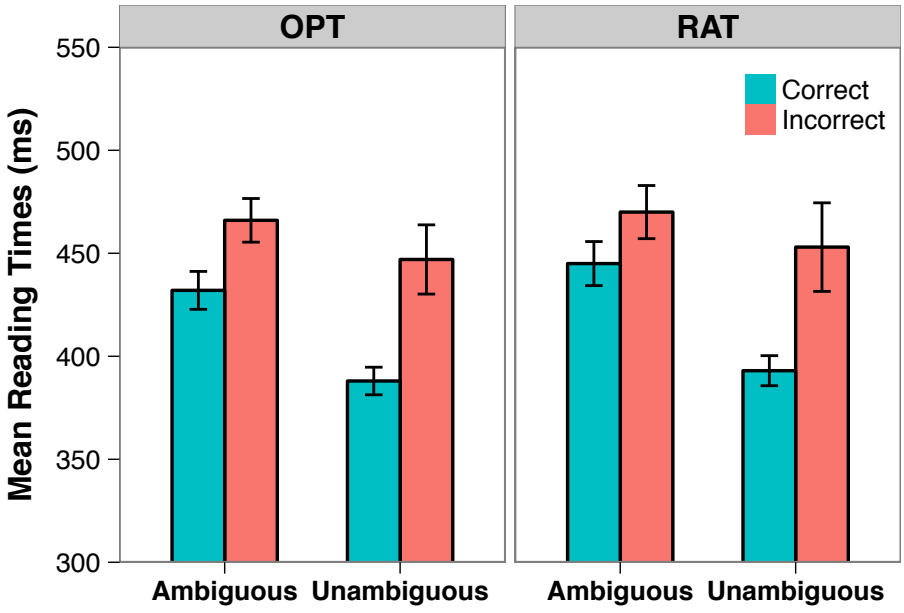

Figure 13. Reading time at the disambiguating region in Experiment 5 separately by question accuracy.

Discussion

There were several important differences between the results of Experiments 4 and 5. First, the overall error rate in question responses decreased substantially, from 50% in Experiment 4 to 30% in Experiment 5, suggesting that the explicit questions had the desired effect of reducing responses based on easily-drawn inferences. The decrease was bigger for unambiguous sentences, leading to a bigger effect of ambiguity on question response accuracy in Experiment 5. There was also a bigger effect of ambiguity on reading times at both the disambiguating and post-disambiguating regions in Experiment 5 than in Experiment 4. The explicit questions clearly led people to both read the sentences more carefully and rely more on what the sentence actually said had happened in responding to the questions. However, in spite of this, there was still very little relationship between reading times at the disambiguating region and question responses. It's not that there was no relationship at all between reading time and question response accuracy in Experiment 5, as was the case in Experiment 4. In Experiment 5, reading time on the disambiguating region did reliably predict question response accuracy, but the direction of the effect was opposite that predicted by the "Incomplete Reanalysis" version of the Good-Enough Processing account. Instead of being more likely to answer the question correctly when they spent longer reading the disambiguating region, which might index more work done to reanalyze the garden path, they were less likely to respond correctly on trials where they spent longer reading the disambiguation, suggesting that they were just more confused all around on those trials. Furthermore, in none of the analyses of question response accuracy in either study has there been any interaction between ambiguity and reading time at the disambiguation, which is what should happen according to the "Incomplete Reanalysis" version of the Good-Enough Processing account. Time spent reading the disambiguating region

specifically in the ambiguous sentences is what should reflect amount of reanalysis work, which should lead to an interaction between ambiguity and reading time in predicting question response accuracy, but there was not even the hint of any such interaction in either study. Thus, there is no evidence from the reading times studies to support a claim that people should be more likely to respond to the questions correctly if they spend more time reanalyzing garden path sentences, even in Experiment 5 where responding based on easily-drawn inferences was successfully reduced.

This line of reasoning assumes that time spent reading the disambiguating region indexes amount of reanalysis of garden paths. It is clear, though, that reading times are influenced by many factors in addition to garden path reanalysis. The fact that longer reading times at the disambiguation were associated with more errors in the question responses in Experiment 5 suggests that one thing influencing reading times is overall confusion. Thus, it is worth testing the Good-Enough Processing account using a measure that is believed to be more specific to structural processing of sentences.

Experiment 6

Event-related brain potentials (ERPs) may provide a more specific tool for examining the predictions of the Good-Enough Processing account. In particular, the P600 component could be useful because it is believed to specifically index structure processing. In sentences like the ones used in Experiments 4 and 5, P600 should be elicited by the disambiguation verb, and its amplitude may be related to the amount of work required to reanalyze the garden path.

There is currently some controversy regarding what the P600 component indexes, but all of the accounts involve structure processing. P600 has been interpreted as reflecting syntactic reanalysis of garden-path sentences (Osterhout & Holcomb, 1992, 1993, see also Osterhout et al., 1994), repair of syntactic violations in sentences (Friederici, 1998), and syntactic integration in structurally complex sentences (Kaan et al., 2000). Controversy has arisen recently because P600 has also been found when the N400, a meaning-related component, was expected. The studies finding "semantic P600" effects have all used sentences in which the subject and object nouns would be plausible arguments for the verb but those nouns appear in the wrong position or with the wrong morphosyntactic markers for the role that fits their meaning (Kolk et al., 2003; Kuperberg et al., 2006; Van Herten et al., 2006; Van Herten et al., 2005). For example, Kim and Osterhout (2005) found P600 in response to sentences beginning like *The hearty meal devoured …* , where *meal* is a good theme of *devour*ing but not a good agent and the syntax signals that it has to be the agent. While the "semantic P600" results have raised very interesting questions about the interplay of semantic and structure processing in sentence comprehension, all of the accounts agree that the P600 component reflects something about the amount of work that is required to determine and use sentence structure toward the goal of interpreting a sentence (Hagoort et al., 1993; Hahne & Friederici, 1999; Kaan & Swaab, 2003; Osterhout & Holcomb, 1992). Thus, P600 effects at the disambiguating word in garden-path sentences could provide a more specific measure of the work required to reanalyze garden path sentences, and thus might be a better predictor than reading times of responses to questions after sentences.

In Experiment 6, we took advantage of the properties of the P600 component to try to specifically examine the relationship between the amount of syntactic reanalysis work at the disambiguating verb and the likelihood of lingering misinterpretation, as indexed by question

responses. The prediction was that bigger P600 at the disambiguating verb should be associated with more correct responses after the sentences.

Method

Participants

Participants were sixty-four undergraduate student (29 males; mean age 19; range 18-22) at the University of Illinois at Urbana-Champaign. All were native speakers of English, were strongly right-handed as assessed by the Edinburgh inventory (Oldfield, 1971), had normal or corrected-to-normal vision and no neurological or psychiatric disorder. All gave written informed consent and received course credits or payment for taking part. Data from four participants (2 males) were excluded from analysis due to low response accuracy to distractor items. Data from another six participants were excluded from analysis due to problems with data collection (2 participants, both females) or excessive loss of trials to artifacts (4 participants, 1 male).

Materials and Design

Critical sentences in Experiment 6 were exactly the same as Experiment 4 and 5. The distractors from Experiment 5 were also used in Experiment 6. The questions asked at the end of the sentences were the non-explicit versions, such as *Did the man hunt the deer?*. Sentences were distributed over two lists according to the Latin Square design, and were presented to all participants in the same order as in Experiment 5.

Procedures

Participants were seated comfortably in a dimly lit and sound-attenuated booth in front of a 23-inch LCD monitor. Each trial began with a fixation point, which stayed in the center of the screen for 500 milliseconds. Because eye movements cause artifacts that contaminate the EEG signal, sentences were presented word-by-word at the center of the screen in 26-point white Arial font on a black background, at a rate of 400 ms per word (300 ms text, 100 ms blank screen).

After each sentence, a comprehension question was presented (e.g., *Did the man hunt the deer?*). Participants responded by pressing one of two buttons on a Cedrus RB-830 response box. A "Too Slow" warning was presented if no response was made within four seconds. Feedback was not given regarding response accuracy. Stimulus presentation was controlled by the Presentation® software package. Each list was divided into four blocks. Participants were given a short break after each block and were instructed to try to minimize blinking and body movement during the presentation of the sentences. They were encouraged to blink between trials when they needed to. A practice block of five trials was given at the beginning. The recording session lasted about forty-five minutes and the entire session lasted approximately two hours.

EEG Recording and Data Analysis

Continuous EEG was recorded from 27 Ag/AgCl sintered electrodes placed in an elastic cap (EasyCap, 10-10 system; Chatrian, 1985), referenced online to the left mastoid and re-referenced offline to the average of left and right mastoids: midline: Fz, Cz, Pz; lateral: AF3/4, F3/4, F7/8, FT7/8, FC3/4, C3/4, T3/4, CP3/4, T5/T6, P3/4, P5/6, PO7/8. Eye blinks and eye movements were detected with electrodes above and beneath the right eye and at the outer canthi of both eyes. EEG and EOG recordings were amplified by a Grass Model 12 amplifier and

sampled at a frequency of 200 Hz. A .01-30 Hz analog bandpass filter was applied during online recording and a .1 Hz high-pass digital filter was applied offline. Impedances were maintained below 5kΩ.

Epochs were extracted from the continuous waveforms from 100 ms before the onset of the disambiguating verb through 1100 ms later. Trials contaminated with artifacts during this epoch were rejected using the ERPLAB toolbox (Lopez-Calderon & Luck, 2014). Blinks and eye movements were detected using a moving window peak-to-peak function on the EOG channels, and non-ocular artifacts were identified using the same moving window peak-to-peak function applied to the EEG channels, with individualized thresholds determined by visual inspection of each participant's data. Data were excluded from further analyses if artifact rejection led to a loss of over 30% of the data in any of the conditions. This process removed six participants' data. Epochs contaminated with artifacts were discarded, leading to an average loss of 9.9% of the data, which did not differ across conditions (OPT: ambiguous 11.5%, unambiguous 11.6%; RAT: ambiguous 10.9%, unambiguous 12.0%).

Mean amplitudes were calculated for each channel in each condition for each participant for the conventional 600-900 ms time window to capture the P600 component, and were submitted to separate repeated-measures analyses of variance to conduct two sets of analyses. One set of analyses included all lateral electrodes and another included just midline electrodes. The ANOVA including all lateral electrodes had four within-subject factors: two levels of ambiguity (Ambiguous, Unambiguous), two levels of question accuracy (Correct, Incorrect), three levels of electrode site anteriority (Frontal, Central, Posterior) and two levels of electrode site laterality (Left, Right). The ANOVA including just midline electrodes (Fz, Cz, Pz) consisted of the same within-subject factors except that there was no laterality factor. When interactions

with electrode site in the omnibus ANOVAs motivated further analyses, analyses were conducted on six regions of interest (ROIs), each comprising four electrodes: left anterior (AF3, F3, F7, FT7), right anterior (AF4, F4, F8, FT8), left central (FC3, C3, CP3, T3), right central (FC4, C4, CP4, T4), left posterior (P3, T5, P5, PO7) and right posterior (P4, T6, P6, PO8). When interactions did not involve the laterality factor, further analyses were conducted on three ROIs: anterior (AF3, F3, F7, FT7, AF4, F4, F8, FT8), central (FC3, C3, CP3, T3, FC4, C4, CP4, T4) and posterior (AF4, F4, F8, FT8, FC3, C3, CP3, T3), rather than six ROIs. Analyses within ROIs included two within-subject factors: two levels of ambiguity and two levels of question accuracy. The Greenhouse-Geisser correction was applied wherever necessary to correct for violations of sphericity (Greenhouse & Geisser, 1959). Corrected *p*-values and original degrees of freedom are reported. Grand average ERPs were digitally low-pass filtered at 10 Hz to smooth the waveforms for display, but analyses were performed before such filtering was applied.

Results

Comprehension Accuracy

Comprehension accuracy for distractors was analyzed to evaluate whether or not participants were paying attention to the sentences. Four participants were excluded from further analyses because their response accuracy to distractors was below 75%. The average accuracy of the remaining participants was 91%.

Comprehension accuracy for critical sentences was analyzed using logit mixed-effect models with maximum random effects structure and ambiguity as a fixed effect. The analysis procedures for question accuracy were the same as in Experiments 4 and 5. Results revealed a main effect of ambiguity on comprehension accuracy in both sentences with OPT and RAT verbs

(OPT: β=0.85, *SE*=0.21, z=4.06, p<.001; RAT: β=1.19, *SE*=0.23, z=5.22, p<.001), with more incorrect responses for ambiguous than unambiguous sentences (OPT: ambiguous 58%, unambiguous 44%; RAT: ambiguous 41%, unambiguous 22%), as shown below in Figure 14.

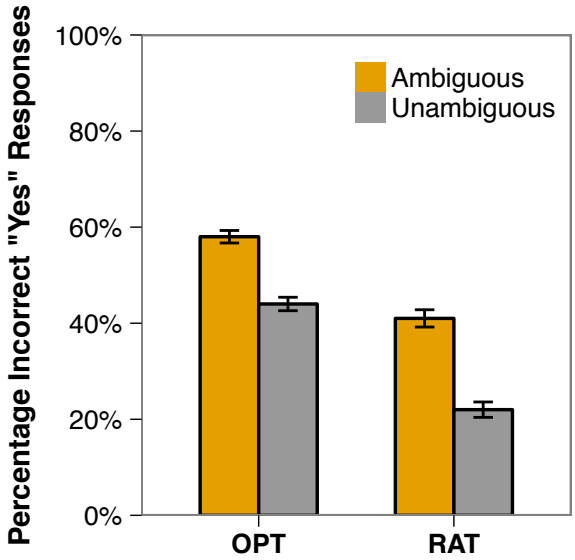

Figure 14. Error rates for question responses in Experiment 6.

ERPs

ERP data were analyzed using analysis of variance (ANOVA) rather than mixed effects models, purely for pragmatic reasons. The EEGLAB and ERPLAB analysis software packages assume that what will be submitted to statistical analyses is subject/condition means rather than individual trials, which is consistent with ANOVA but not mixed effects models. It is not impossible to use mixed effects models to analyze single-trial ERP data, but it is substantially more difficult to get the data into the required form, so that task has been postponed for now.

OPT verbs. Visual inspection revealed that the P600 time window (600-900 ms) for the disambiguating verb (e.g., *ran*) in the ambiguous condition was more positive than in the

unambiguous condition, as illustrated in Figure 15. This effect was centroparietally distributed, as is typical for the P600 component. This observation was confirmed by statistical analyses. ANOVAs over all lateral electrodes revealed a main effect of ambiguity, $F(1,53)=4.45$, $p<.05$, and an interaction between ambiguity and anteriority, $F(2,106)=11.61$, $p<.001$, which resulted because the P600 effect was significant at central sites, $F(1,53)=6.58$, $p=.01$, and posterior sites, $F(1,53)=15.91$, $p<.001$, but not at frontal sites, $F<1$. ANOVAs over the midline electrodes showed the same pattern, with a main effect of ambiguity, $F(1,53)=5.90$, $p<.05$, and an interaction between ambiguity and anteriority, $F(2,106)=8.18$, $p<.001$, because the P600 effect was significant at Cz, $F(1,53)=6.72$, $p=.01$, and Pz, $F(1,53)=11.77$, $p=.001$, but not at Fz, $F<1$.

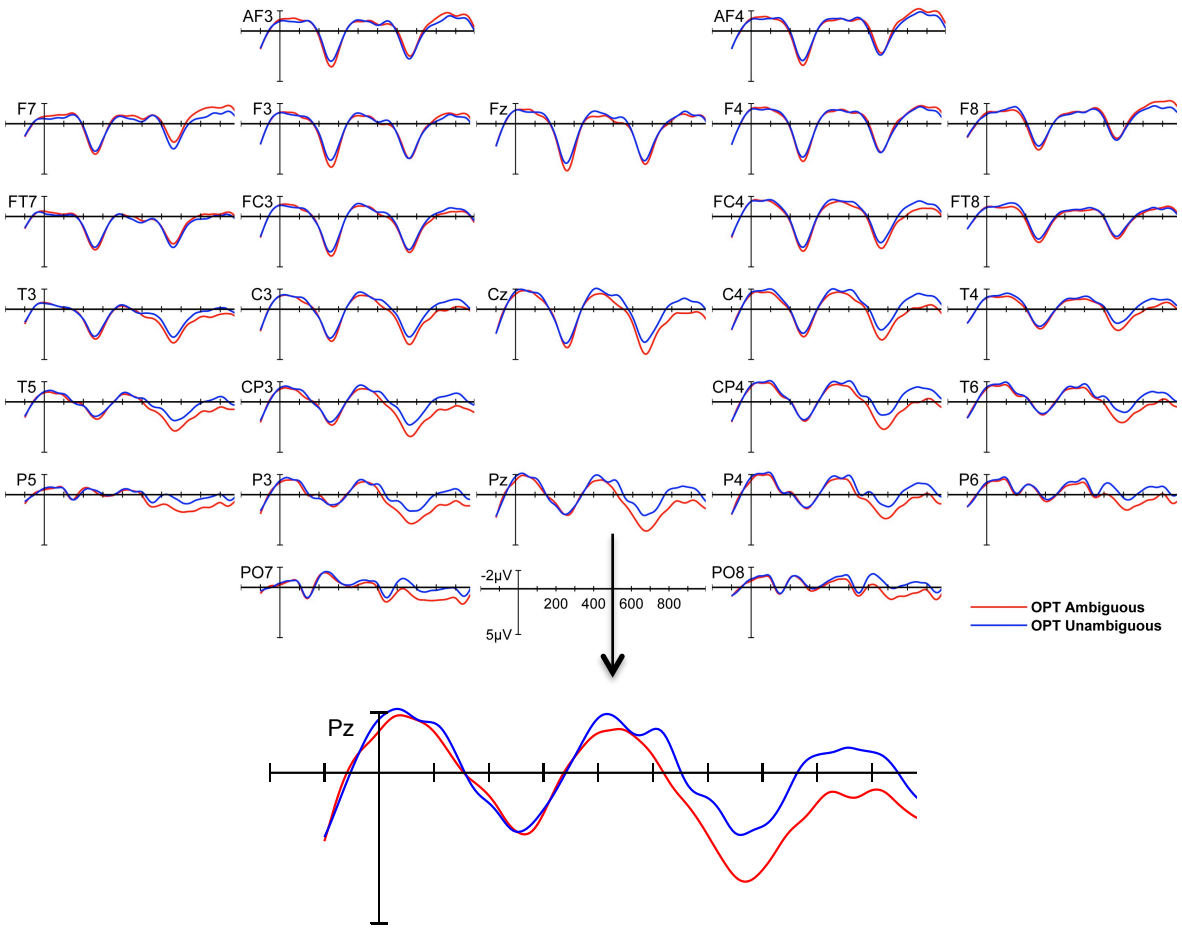

Figure 15. Grand average ERPs for the disambiguating verb at all electrodes in ambiguous and unambiguous sentences with OPT verbs in Experiment 6, baselined on 100 ms before the onset of the disambiguating verb. Y-axis position indicates onset of the disambiguating verb. Centroparietal electrodes showed a larger P600 for the ambiguous than the unambiguous condition.

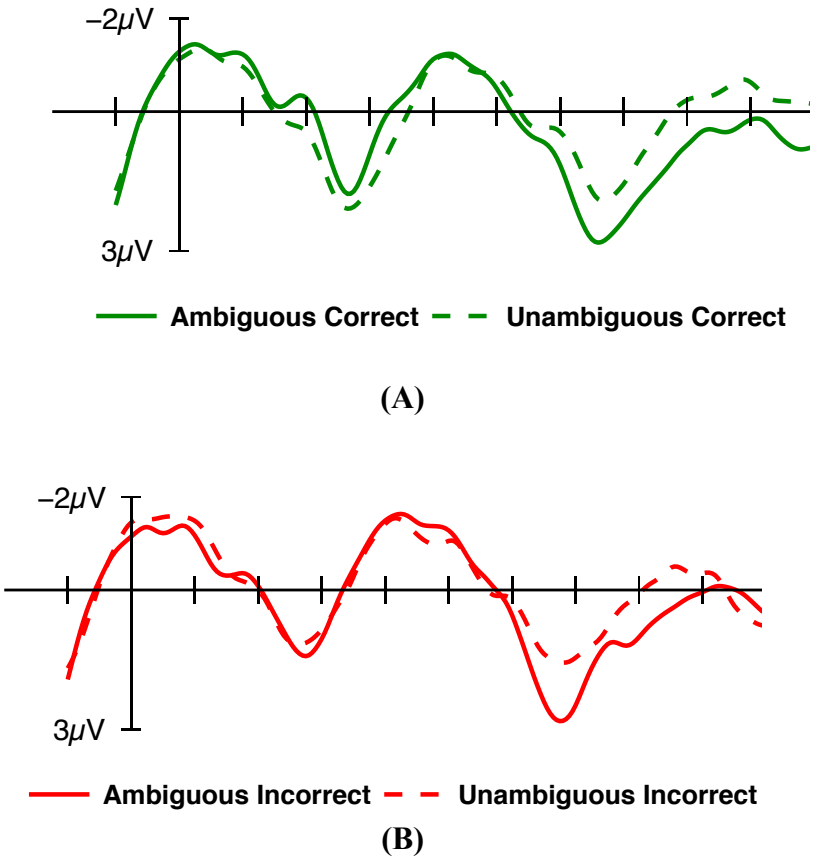

Figure 16. ERPs averaging across electrodes in the centroparietal region for items with OPT verbs, for ambiguous and unambiguous sentences with (A) correct responses and (B) incorrect responses, baselined on 100 ms before the onset of the disambiguating verb.

When the waveforms were broken down by question response accuracy, visual inspection revealed that the ambiguity effect for sentences with correct responses did not differ from that for sentences with incorrect responses, as shown in Figure 16 (ambiguous correct mean voltage: 1.35; ambiguous incorrect: 1.28; unambiguous correct: 0.48; unambiguous incorrect: 0.46), suggesting that it was not true that when readers answered the question correctly, they had worked harder to reanalyze at the disambiguating verb. This observation was confirmed by statistical analysis. In the ANOVAs over all lateral electrodes, over the midline electrodes, and

over just the centroparietal electrodes where P600 tends to be biggest, there were no main effects of question response correctness nor any interaction involving correctness, all *F*s<1. .

RAT verbs. Probably due to the smaller number of items with RAT verbs (12 trials per condition per subject), ANOVAs over all lateral electrodes and over the midline electrodes for sentences with RAT verbs showed no main effect of ambiguity nor any interactions between ambiguity and anteriority, all *F*s<1. There were also no effects involving the question response correctness factor, all *p*s>.1, with one exception. The ANOVA on midline sites showed an interaction of correctness by anteriority, $F(2,106)=3.92$, $p=.04$. Further analyses revealed that this interaction resulted because the correctness effect was marginally significant at the posterior region, $F(1,53)=2.77$, $p=.10$, but not at the frontal or central regions, *p*s>.1. However, this is likely to be a spurious effect, because it resulted mainly from the unambiguous incorrect condition being much more positive starting from the beginning of the epoch and continuing throughout it at the posterior region only. As no ERP effect related to garden-path reanalysis could begin as early as the onset of the word eliciting it, the correctness by anteriority effect was probably caused by noise due to insufficient data. Only 20% of the unambiguous trials with RAT verbs were followed by incorrect answers, as shown in Figure 14, so there were not enough trials to average out the noise in this condition.

Visual inspection of the data for the sentences with RAT verbs appeared to show a small difference between ambiguous and unambiguous versions in the N400 time window (400-600 ms). To test whether an N400 effect was elicited instead of a P600 effect, statistical analyses were also performed on measurements taken from the N400 window. Results again showed no significant main effects nor interactions involving ambiguity in any of the ANOVAs, all *p*s>.1.

Discussion

The results of Experiment 6 indicated that in sentences with OPT verbs, the disambiguating verb (e.g., *ran*) in the ambiguous condition triggered a larger P600 than in the unambiguous condition, suggesting that syntactic reanalysis took place at the disambiguation. However, the size of the P600 ambiguity effect was not modulated by question response accuracy, which is consistent with the reading time data in Experiments 4 and 5. Assuming that question response accuracy reflects whether or not the initial misinterpretation is revised at the disambiguation and that P600 amplitude indexes reanalysis effort, this result is inconsistent with the "Incomplete Reanalysis" explanation of lingering misinterpretation in the Good-Enough Processing Account.

In summary, Experiments 4, 5 and 6 all converge to show that neither reading times nor P600 amplitude at the disambiguating verb predict question response accuracy, even when answering the questions based on inference was reduced in Experiment 5 and when a more specific measure of structural reanalysis, the P600, was used in Experiment 6. In Experiment 7, the advantages of Experiments 5 and 6 were combined by asking explicit questions in an ERP experiment, with the goal of tightening the link between the reanalysis processing measure (P600) and responses to the questions.

Experiment 7

The critical sentences in Experiment 7 were the same as those in Experiments 4, 5 and 6, except that more items with OPT verbs were added to compensate for the loss of trials caused by artifacts in ERP experiments. No items with RAT verbs were added because almost all of the

limited number of available verbs were already used twice. Following each sentence, an explicit question such as *Did the sentence explicitly say that the man hunted the deer?*, was asked, as in Experiment 5.

Method

Participants

Participants were forty-two undergraduate students (24 males; mean age 19.3; range 18-23) at the University of Illinois at Urbana-Champaign. All were native speakers of English, were strongly right-handed as assessed by the Edinburgh inventory (Oldfield, 1971), had normal or corrected-to-normal vision and no neurological or psychiatric disorder. All gave written informed consent and received course credits or payment for taking part. Data from three participants (1 male) were excluded from analysis due to problems with data collection or excessive loss of trials to artifacts.

Materials and Design

As with Experiment 4, 5, and 6, critical sentences were ambiguous and unambiguous versions of garden-path sentences containing OPT verbs and RAT verbs. There were sixty items with OPT verbs, forty from Experiments 4, 5 and 6 and an additional twenty items taken from Maxfield, Lyon and Silliman (2009). The same twenty-four items with RAT verbs used in Experiments 4, 5, and 6 were used in Experiment 7. Distractors were 210 sentences, among which 120 sentences were ambiguous sentences with direct object/sentential complement ambiguity and their unambiguous versions, and 90 sentences were unambiguous sentences in

which the noun immediately following the verb was the direct object. An explicit question was asked at the end of each sentence. Among the 90 distractors, the correct responses to sixty-six of them was *yes* and twenty-four was *no*. The purpose of having more *yes* than *no* correct responses to distractors was to balance the *yes* and *no* responses across the experiment, since as shown in Experiment 5, readers tended to produce more *no* correct responses to critical sentences when explicit questions were asked.

Procedures

Procedures were exactly the same as Experiment 6. The total of 294 sentences was divided into six blocks with forty-nine sentences each. A practice block of 6 sentences was given at the beginning. The recording session lasted about seventy-five minutes and the entire session lasted 2-2.5 hours.

EEG Recording and Data Analysis

EEG recording and data analyses were the same as in Experiment 6. Artifact rejection led to loss of 4% of the data, slightly more for the sentences with RAT verbs than for those with OPT verbs (OPT: ambiguous 4.5%, unambiguous 4.3%; RAT: ambiguous 5.5%, unambiguous 6.6%; distractors: 3.3%). Visual inspection of the waveforms suggested that there might be effects other in P600 present, so mean amplitudes were measured from the N400 (300-500 ms) and the P600 time windows (600-900 ms) to capture potential N400, P600 and Sustained Frontal Negativity effects.

Results

Comprehension Accuracy

The average question response accuracy to distractors was 88%, with a range of 75% to 98%. Question response accuracy to critical sentences was analyzed using logit mixed-effect models, which included ambiguity as a fixed effect. Results revealed a main effect of ambiguity on response accuracy for items with both OPT and RAT verbs (OPT: β=1.22, *SE*=0.26, *z*=4.73, $p<.001$; RAT: β=1.98, *SE*=0.49, *z*=4.03, $p<.001$), with more incorrect responses to ambiguous than unambiguous sentences for items with both OPT verbs (50% vs 32%) sentences, and RAT verbs (39% vs 20%), as shown in Figure 17.

As was found when comparing Experiments 4 and 5, a comparison of Experiments 6 and 7 showed that question error rates for ambiguous and unambiguous sentences with OPT verbs decreased in Experiment 7, by about 10% in both ambiguous and unambiguous conditions. Thus, asking "explicit" questions seems to have reduced the likelihood of answering the questions based on easily drawn inferences. However, the question response error rates for items with RAT verbs did not decrease from Experiment 6, which is different from Experiment 5 compared to Experiment 4.

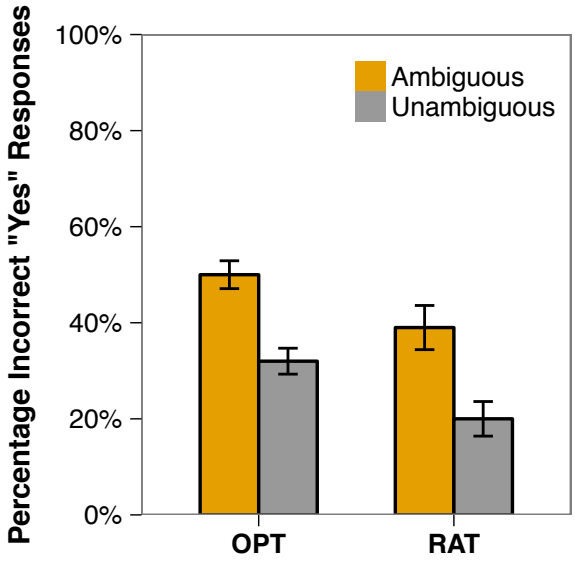

Figure 17. Error rates for question responses in Experiment 7.

ERPs

OPT verbs. Visual inspection showed that contrary to expectation, there was no P600 effect elicited by the disambiguating verb. Instead, there was a broadly-distributed negativity beginning in the N400 time window and persisting throughout the epoch that was larger for the ambiguous than the unambiguous condition, as illustrated in Figure 18. These observations were confirmed by statistical analyses. For the N400 time window measure, ANOVAs over all lateral electrodes revealed a main effect of ambiguity, $F(1,38)=11.03$, $p=.001$, but no interaction between ambiguity and anteriority, $F<1$. Analysis over midline channels showed the same pattern: a main effect of ambiguity, $F(1,38)=15.37$, $p<.001$, but no interaction with anteriority, $F<1$. Consistent with the absence of an interaction, analysis of individual ROIs showed that the ambiguity effect was significant over all ROIs: Left Frontal, $F(1,38)=8.68$, $p<.01$; Right Frontal,

$F(1,38)=4.44$, $p<.05$; Left Central, $F(1,37)=10.08$, $p<.01$; Right Central, $F(1,38)=6.90$, $p<.05$; Left Posterior, $F(1,38)=5.26$, $p<.05$; Right Posterior, $F(1,38)=4.83$, $p<.05$.

For the 600-900 ms time window measure, ANOVAs over all lateral electrodes showed that the ambiguity effect persisted in this time window, $F(1,38)=12.37$, $p<.001$, but that it was modulated by an interaction between ambiguity and anteriority, $F(2,76)=4.56$, $p<.05$ because the scalp distribution of the difference changed over time. The ambiguous condition remained more negative than the unambiguous condition at the Frontal, $F(1,38)=14.68$, $p<.001$, and the Central Regions, $F(1,38)=10.04$, $p<.01$, but not at the Posterior Region, $F(1,38)=1.16$, $p>.1$. The ANOVA over midline channels also showed a main effect of ambiguity, $F(1,38)=10.08$, $p<.01$, and a marginal interaction between ambiguity and anteriority, $F(1,38)=2.48$, $p=.09$, which was caused by the ambiguous condition being more negative than the unambiguous condition at Fz, $F(1,38)=15.54$, $p<.001$, and Cz, $F(1,38)=5.56$, $p<.05$, but only marginal at Pz, $F(1,38)=2.77$, $p=.10$. The topographical maps in Figure 18 show how the scalp distribution of the ambiguity effect changed over time. During the N400 time window, the maximum difference was in the centroparietal region but by 700-900 ms it had shifted to a frontal maximum.

When the waveforms were broken down by question response accuracy, the ambiguity effect in the N400 time window did not differ between correctly-answered trials and incorrectly-answered trials, as shown in Figure 19. In the ANOVAs with all lateral electrodes and with just midline electrodes, there were no main effects of question response correctness, all *p*s>.1, nor any interactions between ambiguity and correctness, all *p*s>.1. An ANOVA on just the centroparietal electrodes, where the N400 effect was most prominent, also revealed no effects involving question response correctness, all *p*s>.1. Thus, there was no evidence that the size of

the ambiguity effects in the waveforms predicted likelihood of responding correctly to the question for the sentences with OPT verbs.

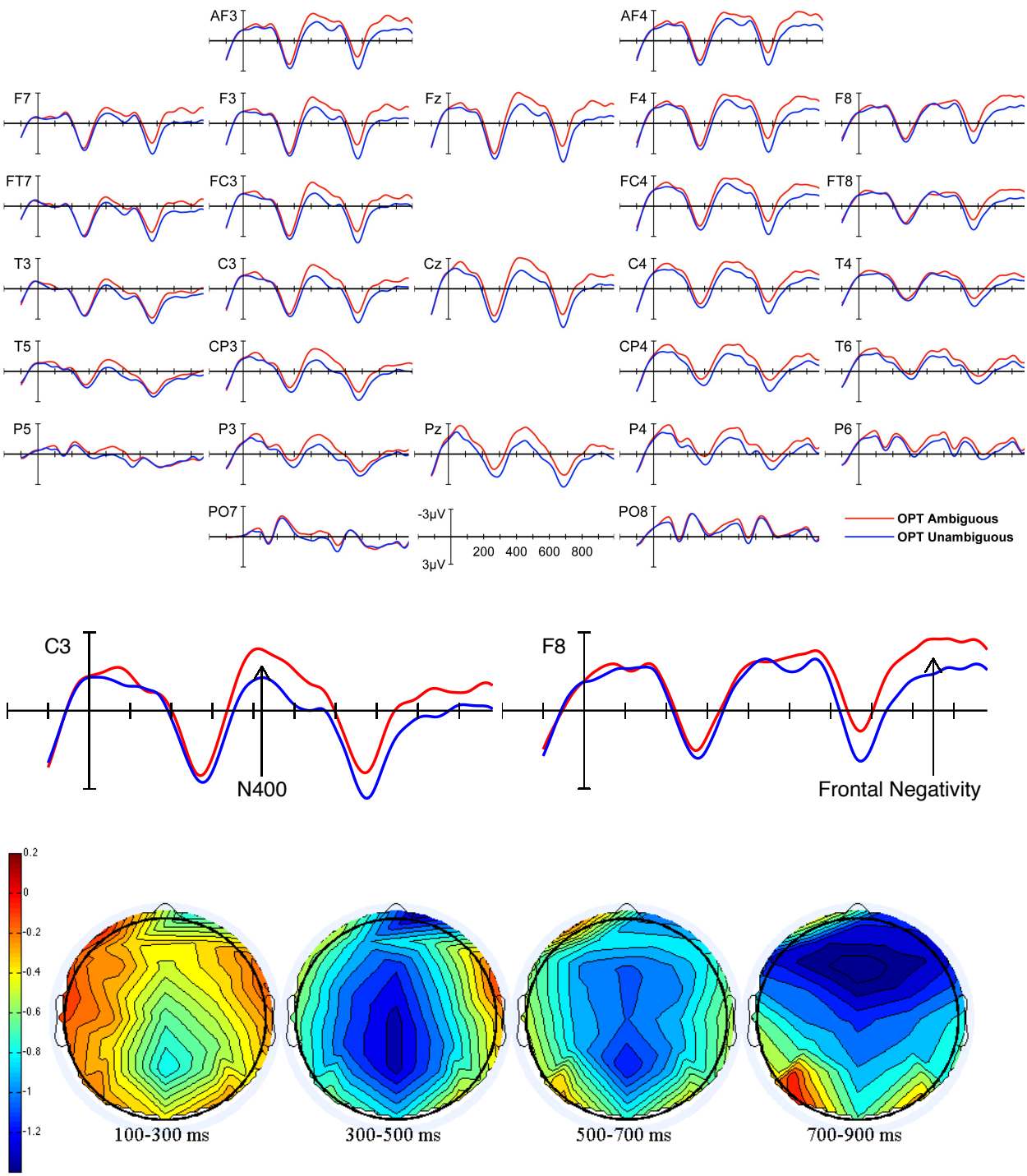

Figure 18. Grand average ERPs at the disambiguating verb at all electrodes in sentences with OPT verbs in Experiment 7, baselined on 100 ms before the onset of the disambiguating verb. Y-axis position indicates onset of the disambiguating verb. Topographical voltage maps of the ambiguity effect show that the scalp distribution of the effect changes over time.

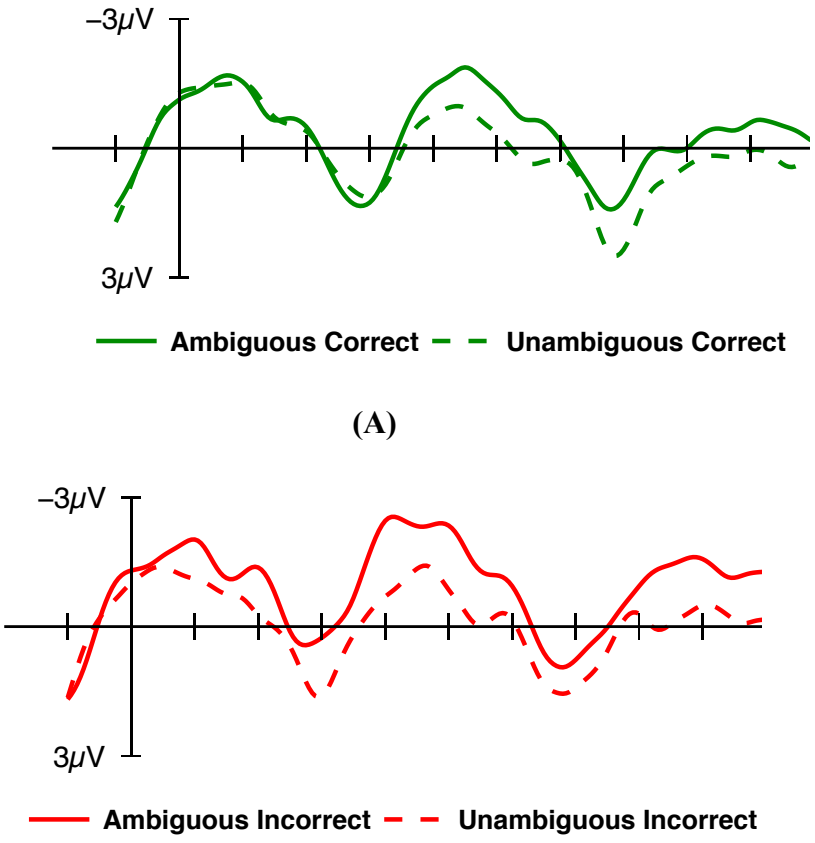

Figure 19. ERPs at the disambiguating verb averaging across the electrodes in the centroparietal region for ambiguous and unambiguous sentences with OPT verbs with (A) correct responses, and (B) incorrect responses, baselined on 100 ms before the onset of the disambiguating verb.

RAT verbs. Grand average ERPs for items with RAT verbs are shown in Figure 20. Visual inspection showed that there was a centroparietally distributed P600 effect, with more positivity for the ambiguous condition than the unambiguous condition. Thus, while in Experiments 4, 5, and 6, items with OPT and RAT verbs produced very similar results, in Experiment 7 they behaved quite differently. While the waveforms for items with OPT verbs differed between Experiments 6 and 7, as described above, the waveforms for items with RAT verbs were similar across Experiments 6 and 7.

The ANOVA over all lateral electrodes for items with RAT verbs revealed an ambiguity by anteriority interaction, $F(2,76)=3.35$, $p<.05$, which resulted because the ambiguity effect was significant at the Posterior region, $F(1,38)=9.42$, $p<.01$, marginal at the Central region, $F(1,38)=2.86$, $p=.09$, and not significant at the Frontal region, $F<1$.It appeared from visual inspection that the P600 effect might be preceded by an N400 effect at central-parietal electrodes, so ANOVAs were also done for the N400 time window. However, there were no significant effects in those analyses, all *p*s >.1.

When the waveforms for items with RAT verbs were broken down by question response accuracy, visual inspection suggested that the P600 ambiguity effect was bigger for incorrectly-answered trials than for correctly-answered trials, as shown below in Figure 21, but the difference was not reliable. ANOVAs over all lateral electrodes and over just midline electrodes showed no effects involving question response, all *p*s>.05.

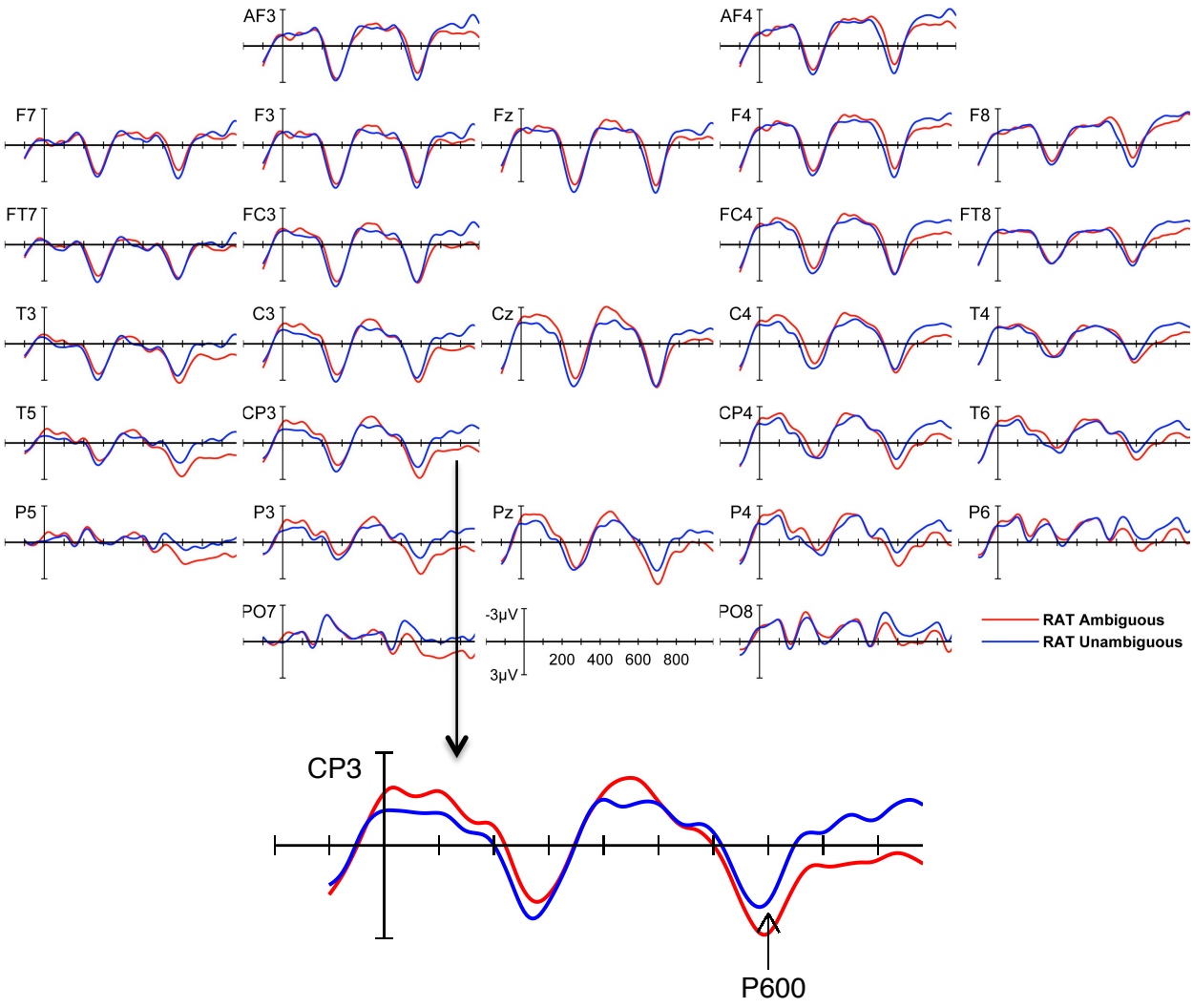

Figure 20. Grand average ERPs at the disambiguating verb at all electrodes in sentences with RAT verbs in Experiment 7, baselined on 100 ms before the onset of the disambiguating verb. Y-axis position indicates onset of the disambiguating verb. The CP3 electrode illustrates the P600 effect.

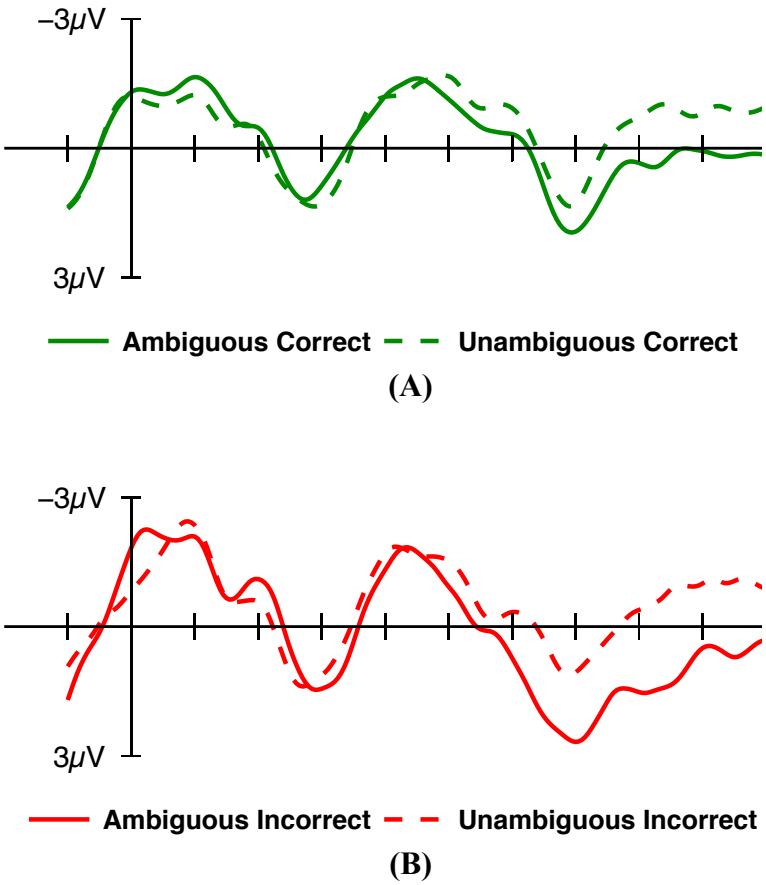

Figure 21. ERPs at the disambiguating verb averaging across the electrodes in the centroparietal region, showing ambiguity effects for correctly-answered and incorrectly-answered trials, baselined on 100 ms before the onset of the disambiguating verb, for ambiguous and unambiguous sentences with RAT verbs with (A) correct responses and (B) incorrect responses.

Discussion

In Experiment 7, ERP responses to the disambiguating verb in sentences with OPT and RAT verbs were measured to see whether the amplitude of the P600 component predicted question response accuracy. Sentences with RAT verbs elicited the expected P600 effect that was also seen in Experiment 6, and also as in Experiment 6, its amplitude was unrelated to question response accuracy.

In contrast to both the sentences with RAT verbs in Experiment 7 and the sentences with OPT verbs in Experiment 6, ambiguous sentences with OPT verbs in Experiment 7 elicited an unexpected long-lasting negativity, which had the typical scalp distribution for an N400 effect during the usual N400 time window but then evolved into a sustained negativity with a frontal maximum later in the waveform, rather than the expected P600. Before turning to possible explanations for this change in what ERP components were elicited, it is important to note that the amplitude of the elicited N400 was also unrelated to question response accuracy. Thus, in none of studies so far is there any evidence that measures of the amount of reanalysis work done at the disambiguating region has any relationship to how the post-sentence question is answered. Such a result is inconsistent with the "Incomplete Reanalysis" version of the Good-Enough Processing account, which predicts that there should be a bigger ambiguity effect associated with correctly-answered ambiguous and unambiguous sentences than incorrectly-answered ambiguous and unambiguous sentences.

P600, N400, & SFN. In Experiment 6, the expected P600 ambiguity effect was elicited by the disambiguating verb in sentences with OPT verbs when the question asked *Did the man hunt the deer?*. However, when an explicit question was asked after exactly the same sentences in Experiment 7 (e.g., *Did the sentence explicitly say that the man hunted the deer?*), what was elicited at the disambiguating verb appeared to be an N400 effect following by a sustained negativity with a frontal maximum, rather than P600. In contrast, in sentences with RAT verbs, there was a P600 effect just as in Experiment 6 (although it actually did not reach significance for RAT verbs in Experiment 6). The different patterns in sentences with the two verb types may provide an important clue to help explain the change for sentences with OPT verbs.

The goal of the explicit questions was to encourage people to respond based on what the sentence actually said had happened, rather than on inferences that could easily be drawn from the sentence, with the idea that discouraging inference-based responding would lead to a tighter link between the online processing measures and the question responses. The questions seemed to have the desired effect because incorrect question responses declined in both studies using them. However, another likely consequence of the explicit questions was that they encouraged people to be generally more careful in deciding on a response. In ambiguous sentences with RAT verbs (e.g. *While Anna dressed the baby spit up on the bed*.), it becomes clear at the disambiguating verb (*spit*) that the subordinate clause subject (*Anna*) is both subject and object (i.e., it is herself that Anna dressed) – that is the crucial property of Reflexive Absolute Verbs like *dressed*. Thus, it does not matter how good *the baby* is as the object of *dressed* because someone else (*Anna*) automatically becomes its object instead. In contrast, in ambiguous sentences with OPT verbs (e.g., *While the man hunted the deer ran into the woods.*), when it becomes clear at the disambiguating verb (*ran*) that *the deer* has to be its subject, that leaves *hunted* with no specified object. There is no automatic replacement of the object with the verb's subject, as there is for RAT verbs. Under these circumstance where the sentence does not say what the man hunted, the explicit question seems to have triggered a more thorough analysis of the plausibility of the deer as the object hunted, given that no other object is available, and that led to an increase in amplitude of the N400 component instead of P600.

The tradeoff between P600 and N400 effects depending on the type of question asked in Experiments 6 and 7 is consistent with a finding that has recently been reported at a conference but not yet published. Oines and Kim (2014) asked participants to read role-reversal sentences that typically elicit the "semantic P600" effect, which was introduced briefly earlier. Sentences

like *The hearty meal was devouring*… would be expected to elicit an N400 effect at *devouring* because it is nonsensical to say that a meal is devouring something, but a P600 effect has been observed instead. This and other similar results (e.g., Kuperberg, 2007) have been interpreted as showing that there is conflict between the outcomes of meaning-based and structure-based processing streams. Oines and Kim asked participants to perform one of two tasks while reading these types of sentences. In the structural repair task, they were asked to figure out how to fix the structure of the sentences so that they made sense, while in the semantic integration task, they were asked to try very hard to figure out the meanings of the sentences, given their structure. The structural repair group showed a P600 effect while the "semantic integration" group instead showed a Left Anterior Negativity (LAN). Thus, task determined which ERP component was observed. In the group with semantic integration task, the LAN, which has been linked to working memory load among other things (King & Kutas, 1995; Kluender & Kutas, 1993; Weckerly & Kutas, 1999), was interpreted as reflecting the need to retrieve word order information from working memory, since that is what determines the role *meal* plays in the *devouring* event. These results show that the same sentences can elicit different ERP responses when different tasks are imposed. In Oines and Kim's study, the LAN was elicited rather than the P600 when word order was a crucial factor in determining the role of a noun with respect to a verb. In Experiment 7 here, it was N400 that was elicited rather than P600 because people tried to use the plausibility of a noun as the object of a verb as the basis for answering the explicit question.

There is an alternative possible explanation of Oines and Kim's findings, since task was a between-subjects manipulation. Other recent work has found that sentences that elicit clear P600 effects in some people elicit N400 effects in others (Tanner & Van Hell, 2014). It is possible,

though rather unlikely, that Oines and Kim's results were due to inherent differences between the subjects in their two task conditions, rather than due to the tasks themselves. Inherent individual differences are even less likely to provide an explanation of the pattern observed here, since in Experiment 7 the same people showed P600 ambiguity effects in sentences with RAT verbs but N400 ambiguity effects in sentences with OPT verbs.

Sustained Frontal Negativity. In Experiment 7, a sustained frontal negativity was observed in response to the disambiguating verb in sentences with OPT verbs. It begin during the N400 time window and had the centroparietal maximum scalp distribution that is typical of the N400 at that point, but then it persisted and shifted to a frontal maximum scalp distribution. The change in scalp distribution over time provides some justification for considering it to be two different but temporally overlapping effects. Sustained frontal negativities have been found in a variety of circumstances, including sentences with ambiguity about which of two possible referents is the antecedent of an anaphor (Nref effect, Nieuwland, Otten, & Van Berkum, 2007; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003), word sequences with certain kinds of lexical ambiguity that are not resolved by context (C.-L. Lee & Federmeier, 2006, 2009; Wlotko & Federmeier, 2011, 2012), and sentences with ambiguity about which noun is the subject of a verb (E.K. Lee & Garnsey, 2015). It has been interpreted as reflecting the processing load occasioned by the need to resolve conflict among competing alternatives. It seems possible that the sustained frontal negativity arose in response to items with OPT verbs in Experiment 7 because the explicit questions caused people to evaluate more carefully both possible answers, with the result that more conflict between the two possible answers persisted longer.

Experiment 8

Experiments 4-7 converged to show that incomplete reanalysis might not be the primary reason for incorrect question responses. If the amount of reanalysis is not the deciding factor in successful comprehension of garden-path sentences, then what is? Given the fundamental role that incorrect question responses have played in the development of the Good Enough Processing account, it seems important to try to answer this question.

Across Experiments 4-7, it became apparent that questions after some items rarely got incorrect "Yes" responses (e.g., the question *Did the caricaturist draw the child?* after *While the caricaturist drew the child who was freckled and talkative stood on the sidewalk* was responded to incorrectly only 27% of the time), while others got incorrect "Yes" responses very often (e.g., The question *Did the skipper sail the boat?* after *While the skipper sailed the boat that was small and leaky veered off course.* was responded to incorrectly 87% of the time).Thus, it seemed that sentences varied in how much they led people to think that an event had been described in which the temporarily ambiguous noun still played the role of the theme of the subordinate clause verb even though it had turned out not to be its direct object in the sentence structure. Experiment 8 attempted to assess that for the whole sentence and Experiment 9 attempted to do so for particular subcomponents of the sentence. In Experiment 8, participants first read the sentences used in Experiments 4-7, presented all at once, and then answered a question asking how likely it was that the event including the misinterpretation of the temporarily ambiguous noun as direct object was. So, after reading *While the man hunted the deer that was brown and graceful ran into the woods.*, they were asked *How likely is it that the man hunted the deer?*.

Method

Participants

Fifty undergraduate and graduate students (28 males; mean age 20; range 18-28) at the University of Illinois participated in Experiment 8. All were native speakers of English, had normal or corrected-to-normal vision, gave written informed consent and received course credit for taking part.

Materials and procedures

Materials were the ambiguous and unambiguous sentence with OPT and RAT verbs that were used in Experiments 4-7, except that the twenty items with OPT verbs that were added to Experiment 7 were not included. Ambiguous and unambiguous versions of each item were distributed over two lists according to a Latin Square design, so that no participant saw both versions of the same sentence.

Sentences were presented all at once on the computer screen. Following each sentence, participants were asked to give a percentage rating to the questions such as *How likely is it that the man hunted the deer?*. Sentences were randomized for each participant. Item-by-item mean likelihood ratings were obtained by averaging across participants and were then entered into logit mixed effect models as a fixed effect to see whether they predicted the question response accuracy in other studies.

Results

Statistical analysis of the mean likelihood rating of each item averaged across all participants showed a main effect of verb type, $F(1, 124)=88.46$, $p<.001$, with items with OPT verbs items rated more likely than those with RAT verbs (OPT 69%, RAT 46%; $F(1, 124)=48.27$, $p<.001$). Ambiguous sentences were also rated as more likely than unambiguous sentences (Ambiguous 69%, Unambiguous 52%), and there was also an interaction between ambiguity and verb type, $F(1, 124)=8.67$, $p<.01$, because the difference between ambiguous and unambiguous sentences with RAT verbs was bigger than the difference for sentences with OPT verbs (OPT: Ambiguous 75%, Unambiguous 64%; RAT: Ambiguous 59%, Unambiguous 33%).

Logit mixed-effect models were used to evaluate the relationship between the likelihood ratings and question responses in Experiments 4-7, by including likelihood rating as a fixed effect. In addition, since the analysis of the ratings showed that they were affected by ambiguity, and question responses were also affected by ambiguity in Experiments 4-7, ambiguity was included as another fixed effect in the models so that the relationship between likelihood ratings and question responses could be evaluated separately from the effect that ambiguity had on both of them. The initial models all included the interaction between ambiguity and likelihood rating, but because this interaction was not significant for any of the four experiments, it was removed from the models. The results showed overall that likelihood ratings were reliably related to question responses, such that questions after items that were rated more likely were also more likely to be given incorrect *Yes* responses. Analyses were conducted separately for items with OPT and RAT verbs, but the verb types are shown collapsed together in Figure 22 for each experiment because effects were generally the same for both verb types.

In Experiment 4, for items with OPT verbs question responses were reliably predicted by both ambiguity, β=.78, *SE*=.24, *z*=3.23, *p*=.001, and likelihood ratings, β=.39, *SE*=.11, *z*=3.45, *p*<.001. Ambiguous items and items with higher likelihood ratings were both more likely to receive incorrect "Yes" responses, as shown in Figure 22. Experiment 5 showed exactly the same pattern of results: ambiguous items, β=1.71, *SE*=.58, *z*=2.95, *p*<.01, and items with higher likelihood ratings, β=.02, *SE*=.01, *z*=2.38, *p*<.05, were more likely to received incorrect "Yes" responses. The same pattern was also true for Experiment 6, where again both ambiguous items, β=.49, *SE*=.23, *z*=2.16, *p*=.03, and items with higher likelihood ratings, β=.03, *SE*=.01, *z*=3.71, *p*<.001, were more likely to be responded to incorrectly. In Experiment 7, however, there was only an effect of ambiguity on question responses, β=1.11, *SE*=0.31, *z*=3.56, *p*<.001, and no effect of likelihood ratings, β=0.08, *SE*=0.01, *z*=1.45, *p*>.1. The absence of an effect of likelihood ratings in Experiment 7 was probably related to the decrease in inferences drawn because of the explicit questions, though if that were the sole explanation it should have also have been true in Experiment 5.

The pattern of results for items with OPT verbs was mostly replicated in items with RAT verbs. In Experiment 4, both ambiguous items, β=1.09, *SE*=.52, *z*=2.07, *p*<.05, and items with higher likelihood ratings, β=1.09, *SE*=.52, *z*=2.07, *p*<.05, were more likely to have incorrect "Yes" responses. In Experiment 5, however, while ambiguous items were more likely to be responded to incorrectly, β=1.70, *SE*=.74, *z*=2.31, *p*<.05, the same was not true for items that were rated more likely, β=.02, *SE*=.02, *z*=1.42, *p*=.16. In Experiment 6, both effects were again present (ambiguity: β=.65, *SE*=.33, *z*=1.95, *p*=.05; likelihood: β=.02, *SE*=.01, *z*=2.12, *p*<.05). Finally, in Experiment 7, the likelihood ratings again predicted question accuracy, β=0.05, *SE*=0.01, *z*=5.02, *p*<.001, but ambiguity did not, β=0.39, *SE*=0.52, *z*=0.75, *p*>.1.

Overall, results for both OPT and RAT items indicated that ambiguity and the likelihood ratings had separable effects on how readers answered the questions after the sentences. Most importantly, although ambiguity affected both the likelihood ratings themselves and the question responses, there were still effects of likelihood ratings once ambiguity effects were taken into account.

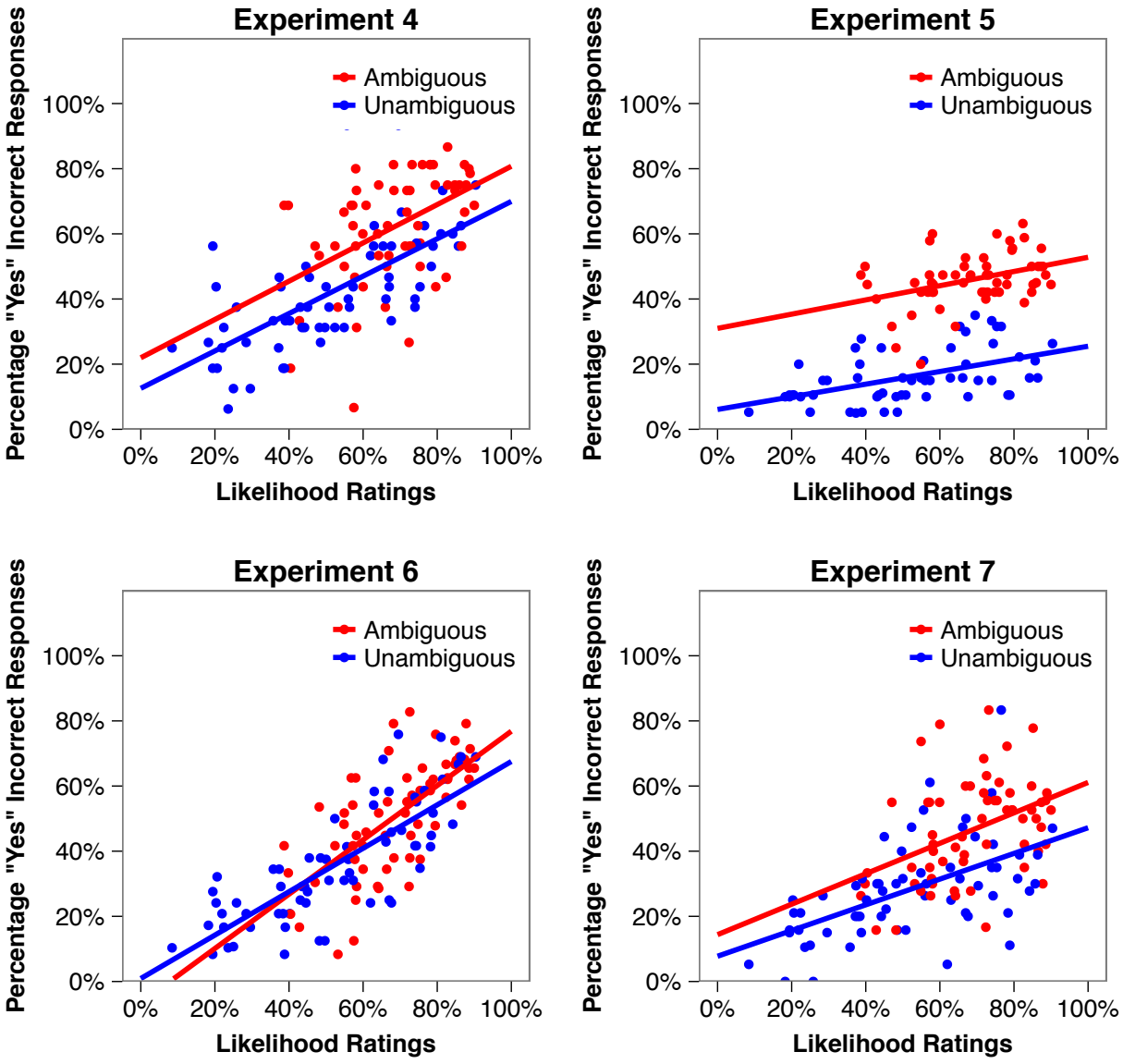

Figure 22. Scatterplots showing the relationship between the percentage of incorrect question responses an item received and the item-by-item likelihood ratings in Experiment 4, 5, 6, and 7, collapsing over items with OPT and RAT verbs.

Additional analyses were conducted to determine whether the reading times at the disambiguating region in Experiments 4 and 5 were affected by the same factors that determined the likelihood ratings. For items with both OPT and RAT verbs in Experiment 4, which used non-explicit questions, likelihood ratings did not correlate with residual reading times (OPT:

t=.70, *df*=1181, *p*>.1, *r*=.02; RAT: *t*=1.00, *df*=752, *p*>.1, *r*=.04). In Experiment 5, however, which used explicit questions, likelihood ratings were slightly but reliably correlated with residual reading times (OPT: *t*=2.34, *df*=1193, *p*<.05, *r*=.07; RAT: *t*=2.00, *df*=853, *p*<.05, *r*=.07). Overall, correlational analyses showed that there was either very small (*r*s<=.07) or no relation between likelihood ratings and reading times. This is consistent with the finding across studies that likelihood ratings predicted question response accuracy in most cases but reading times at the disambiguating verb did not. Thus, likelihood ratings of the events described in the sentences were a better predictor of question response accuracy than reading times at the disambiguating verb.

To summarize, Experiment 8 was conducted to investigate whether or not the likelihood of the events described in the garden-path sentences predicted question accuracy. The goal was to try to determine whether item-specific properties would predict question responses better than the online processing measures at the disambiguating region did. However, in retrospect, it should not be at all surprising that the likelihood ratings from Experiment 8 were so successful at predicting the question responses in Experiments 4-7, since in all cases the question was asked after the sentence was read first, and the questions were actually quite similar. The likelihood questions in Experiment 8 (*How likely is it that the man hunted the deer?*) basically asked for a graded response to almost the same questions that asked for a binary response in Experiments 4 and 6 (*Did the man hunt the deer?*). In Experiment 9, an attempt was made to evaluate the likelihood of particular subcomponents of the sentences without ever reading the whole sentence.

Experiment 9

In Experiment 8, likelihood ratings were given after the sentences were read. In Experiment 9, questions were asked about particular parts of the sentence without the whole sentence ever being seen. The idea was to examine how particular sentence constituents might have contributed to question responses in Experiments 4-7 separately from the effect of reading the whole sentence and possibly being garden-pathed when it was ambiguous. Experiment 9a attempted to examine the effect of the relative clause *that was brown and graceful* from *While the man hunted the deer that was brown and graceful ran into the woods*., since it can influence how likely the deer is to be hunted. For instance, a deer that is cute and little might be less likely to be hunted. Experiment 9b examined the effect of the main clause *the deer ran into the woods*, since it could also influence the likelihood of the event. A hunter might be less likely to hunt a deer that is asleep in the bushes (see Christianson et al., 2001, 2006 for plausibility effects on lingering misinterpretation).

Method

Materials and procedures

Two norming studies were conducted in Experiment 9. In both of them, sentence components were rated without the whole sentence ever being seen. The first of these norming studies (9a) asked participants to give a percentage rating to *How likely is it that a man would hunt a deer that was brown and graceful?*. This task will be called Adjective norming. In 9b, participants were asked to rate *How likely is it that a man would hunt a deer that ran into the woods?*, which is called Main Clause norming. Item-by-item mean likelihood ratings were

obtained by averaging across all subjects and were entered into logit mixed-effect models together to see if they predicted question response accuracy in Experiments 4-7.

Participants

Thirty undergraduate students (19 males; mean age 20; range 18-22) participated in 9a and thirty-two undergraduate students (10 males; mean age 19; range 18-23) participated in 9b. All were recruited from University of Illinois at Urbana-Champaign and were native speakers of English. They gave written informed consent and received course credits for taking part.

Results

As in Experiment 8, ambiguity was included in all statistical models to evaluate whether there were any effects of the Adjective and Main Clause likelihood ratings over and above the effect of ambiguity on question response accuracy. For Experiment 4, neither the Adjective nor the Main Clause ratings predicted question response accuracy for sentences with either OPT or RAT verbs (all *p*s>.1). For Experiment 5, the Adjective ratings did predict question response accuracy for sentences with RAT verbs only (β=.45, *SE*=.21, *z*=2.10, *p*=.04), with events rated as more likely leading to more incorrect responses to the questions. The Adjective ratings did not predict response accuracy for items with OPT verbs, and the Main Clause ratings did not predict accuracy in sentences with either type of verb (all *p*s>.1). For Experiment 6, the only effect was that the Main Clause ratings predicted response accuracy for items with OPT verbs only (β =.02, *SE*=.01, *z*=2.03, *p*=.04). In Experiment 7, there was no relation between any of the ratings and response accuracy for either verb type (all *p*s>.1). Overall, there was little or no influence of the

likelihood ratings of the relative clause modifiers and the main clauses from the original sentences on question response accuracy in Experiments 4-7.

Interestingly, however, the reading times at the disambiguating region in Experiment 4 were correlated with the Adjective ratings (the main clause has not yet appeared at the disambiguating region, so the Main Clause ratings should not have any effect at that point in reading the sentence), and the same effect was marginal for Experiment 5. For items with both OPT and RAT verbs in both experiments, disambiguation region reading times were longer the higher the Adjective ratings (Experiment 4: OPT: t=5.39, df=1181, p<.0001, r=.16; RAT: t=3.50, df=752, p<.001, r=.13; Experiment 5: OPT: t=1.70, df=1193, p=.09, r=.05; RAT: t=1.75, df=853, p=.08, r=.06). Thus, the more likely the people in the Adjective norming study found it that a man would hunt a deer that was brown and graceful, the more the different groups of people in Experiments 4 and 5 slowed down at the disambiguation, which is where they discovered that the man might not be hunting the deer after all. That is, the more plausible one group of people found the deer plus its modifying relative clause as the object of hunting, the more difficult other groups of people found it to read words that contradicted that. In other words, the more plausible the deer plus modifying relative clause was as the object of hunting, the more committed readers were to that interpretation, and thus the more difficult it was to revise that interpretation at the disambiguating region. Presumably, the more difficult the people in Experiments 4 and 5 found it to revise the initial interpretation of the deer as the thing being hunted, the more that analysis should linger and still be available to influence the question responses. In spite of that, however, the Adjective ratings were not correlated with question responses in the other experiments, just as the disambiguating region reading times were not. Since neither the modifying relative clauses nor the main clauses taken from the original sentences affected question response accuracy, the

effect of the likelihood ratings obtained in Experiment 8 on question accuracy must be due to the likelihood of the entire event described in the original sentence. The more likely an event was, the more likely the interpretation of the noun as the subordinate clause object tended to linger.

Discussion

The Good-Enough Processing Account proposed two possible explanations for why people answer questions like *Did the man hunt the deer?* incorrectly after sentences like *While the man hunted the deer that was brown and graceful ran into the woods*. According to the "Incomplete Reanalysis" version of the Good-Enough account, the initial misinterpretation lingers because the parser fails to completely reanalyze the syntactic structure of the sentences, resulting in the ambiguous noun staying in the direct object role in the subordinate clause. The other alternative is the "Lingering Interpretations" version, in which reanalysis is completed but both analyses are maintained. The present studies aimed to test the "Incomplete Reanalysis" version, which predicts that more reanalysis effort at the disambiguating verb should lead to more correct responses to the questions probing the initial misinterpretation.

The "Incomplete Reanalysis" possibility was tested by using self-paced reading times and ERP responses at the disambiguating verb as measures indexing the amount of reanalysis work done at the disambiguation, and comparing those between trials that were followed by correct responses and those that were followed by incorrect responses.

Two self-paced reading and two ERP experiments were conducted with two types of post-sentence questions. The non-explicit questions used in Experiments 4 and 6 were simply the same questions that had been used in previous studies, asking whether the temporarily ambiguous noun was the object of the subordinate clause verb (*Did the man hunt the deer?*). The

explicit questions used in Experiments 5 and 7 more specifically targeted the true content of the sentence (*Did the sentence explicitly say that the man hunted the deer?*). The goal of the explicit questions was to discourage participants from answering based on inferences they could easily draw from the sentences, i.e., that the deer was most likely what the man hunted even though the sentence did not actually say that. The idea was that explicit questions might lead to a cleaner relationship between the online processing measures at the disambiguation and question responses because question responses should be determined primarily by whether or not the initial misinterpretation was successfully reanalyzed, and the online measures should reflect that. The explicit questions apparently did succeed in discouraging inference-based responses, since the number of incorrect question responses decreased substantially. In spite of that, however, neither the reading times nor the ERP responses at the disambiguating verb predicted question responses in the predicted way in any of the studies. In the few cases where there was a relationship between the online measures and question responses, it was in the opposite direction from that predicted by the Incomplete Reanalysis version of the Good Enough Processing account. The predicted direction was that slower reading times and larger P600 amplitudes should reflect more reanalysis work and that should lead to more correct question responses, but instead slower reading times and larger P600 amplitudes tended to be associate with more incorrect question responses, suggesting more confusion in general on those trials. Thus, there was no evidence in any of the studies that when participants answered the questions incorrectly it was because they had failed to fully reanalyze the sentences.

Explicit questions also influenced how participants read the sentences, since both reading times and ERP responses at the disambiguating region were affected by which kind of question was asked. For reading times, explicit questions increased the size of the ambiguity effect, while

for ERP responses, they actually changed which ERP component showed the ambiguity effect in sentences with OPT verbs. The disambiguating verb in RAT sentences still elicited a P600, because syntactic processing was sufficient to meet the task demand for RAT sentences. Regardless of which ERP components were evoked, the amplitude of the ERP responses to the disambiguating word did not support the idea that more reanalysis effort led to more correct responses. The findings from all our experiments converged to show that "Incomplete Reanalysis" might not be the primary reason of lingering misinterpretation. A likelihood rating task was done in Experiment 8 to explore whether the likelihood of the events described in the sentences could explain question response accuracy. Results showed that the more likely the event described by the initial misinterpretation could happen, the more likely that readers answered the questions incorrectly, and that this likelihood effect had independent and additive effect to the effect of ambiguity on question response accuracy. Event likelihood appeared to be a better predictor of response accuracy than reading times or P600/N400 amplitudes at the disambiguating verb. Thus, the predictions made by the "Incomplete Reanalysis" version of the Good-Enough Processing Account were not borne out in the present study; rather, question response accuracy could be accounted for by the likelihood of the events described by the sentences.

The present study was based on the assumption that reading times and amplitudes of ERP responses at the disambiguating word indexes reanalysis, as most psycholinguistic studies did. However, it is possible that slower reading time may simply indicate that readers are confused. This is especially possible with strong garden-pathing sentences, such as *The horse raced passed the barn fell*. Slower reading time at *fell* may indicate revision processes, but in cases when readers cannot recall the words coming before *fell*, and thus can not perform reanalysis, slower reading time at *fell* may only indicate that they detect the error signal and that they are confused

by it. The idea that slowing down at the disambiguating region, or larger amplitude of the ERP component may indicate confusion rather than revision is supported by the majority of comparisons between correctly- and incorrectly-answered trials in our experiments. In Experiment 4, incorrectly-answered trials took numerically longer time to read the disambiguating verb than correctly-answered trials in unambiguous OPT sentences, ambiguous RAT and unambiguous RAT sentences. In Experiment 5, this pattern was observed in all comparisons that crossed ambiguity and verb type. Slower reading times at the disambiguating verb were associated with incorrectly-answered OPT ambiguous, OPT unambiguous, RAT ambiguous and RAT unambiguous sentences. In Experiment 7, incorrectly-answered OPT ambiguous and unambiguous sentences were associated with bigger N400 amplitude than correctly-answered ones, and the same was true for RAT sentences. Overall, except Experiment 6 and one comparison in Experiment 4, all comparisons revealed that it was the incorrectly-answered trials that were read slower or elicited bigger ERP components compared to the correctly-answered trials, suggesting that reading times and ERP responses may indicate the degree of confusion, rather than success of reanalysis. The more confused readers were, the more likely they answered the questions incorrectly.

The results from Experiment 8 also raised the issue that this type of questions that specifically probe the initial misanalysis may not provide the best source of evidence for lingering misinterpretation, because they are strongly influenced by the likelihood of events described in the sentences. Maybe this is why we did not find the relation between real-time measures at the disambiguating word and question response accuracy. Ironically, the Good-Enough Processing Account was developed because researchers were surprised at how many errors people made when answering these questions. However, we did not rule out the Good-

Enough Processing Account, as there was evidence for lingering misinterpretation from other experimental paradigms that used implicit measures of comprehension, such as syntactic priming (Christianson et al. 2010; van Gompel et al. 2006), processing of a subsequent sentence (Slattery et al, 2013), translation (Lim & Christianson, 2013), paraphrases (Patson et al. 2009) and using other types of questions (Christianson & Luke, 2011).

The present study was designed specifically to test the "Incomplete Reanalysis" version of the Good-Enough Processing Account. The results did not support the "Incomplete Reanalysis" version of the Good-Enough Processing Account, but they did not rule out the Good-Enough Account altogether, because we did not test the "Lingering Interpretations" version. The "Lingering Interpretations" version predicts lingering misinterpretation regardless of whether the reanalysis process is successful or not, which in turn predicts that question response accuracy is unrelated to what occurs at the disambiguating verb. The findings from the present study is compatible with this prediction. However, more work is needed to specifically test the "Lingering Interpretations" version of the Good-Enough Processing Account.

The likelihood of events described in the sentences appeared to be a stronger predictor of lingering misinterpretation than on-line measures at the disambiguation in the present study, which was consistent with previous studies that found the effect of plausibility on the likelihood of lingering misinterpretation. In Slattery et al. (2013), the initial misinterpretation lingered only when the ambiguous noun was a plausible direct object for the subordinate clause verb, as in *While Frank dried off the truck was peed on by a stray dog.* There was no evidence of lingering misinterpretation when the ambiguous noun and the subordinate verb formed an implausible interpretation, as in *Frank dried off the grass was peed on by a stray dog.* Similar effect was reported by Nakamura and Arai (in press) from Japanese garden-path sentences such as *The*

professor-NOM $_{RC}$*[paper-ACC wrote] pupil-ACC praised*, in which the relative clause verb *wrote* and its direct object *paper* were temporarily analyzed as part of the main clause, resulting in the initial misinterpretation *The professor wrote the paper*. Plausibility of the initial misinterpretation was manipulated, and Nakamura and Arai found that the more plausible the initial misinterpretation, the harder it was for readers to abandon that interpretation. Although the present study did not directly manipulated plausibility of the initial misinterpretation, the likelihood ratings corroborated the aforementioned two studies to indicate that the likelihood that a misinterpretation lingers is influenced by the likelihood of events described in it.

To conclude, the present study did not find evidence for the "Incomplete Reanalysis" version of the Good-Enough Processing Account, which claimed that lingering misinterpretation was resulted from the parser's failure to fully reanalyze the sentence, since in all our experiments slower reading times or larger ERP components at the disambiguating word were not associated with correct responses to comprehension questions. Rather, the likelihood of the events described by the sentences was a stronger predictor than reanalysis effort for lingering misinterpretation. The more likely an event was, the more likely the initial misinterpretation tended to linger.

CHAPTER 5

General Discussion

This dissertation set out to answer three previously unaddressed questions in the sentence processing literature: 1) how verb bias and plausibility cues are used in the processing of L2 sentences; 2) how verb bias and plausibility cues are used in the processing of Mandarin sentences; and 3) whether on-line measures at the disambiguation are good indicators of the amount of syntactic reanalysis in garden-path sentences.

Conclusions and Contributions

The answer to the first research question addresses the general question of how sentences are processed by L2 learners. A prevailing view in the L2 psycholinguistic literature argues that the L2 parser is qualitatively different from the L1 parser in that it is restricted to the use of lexical-semantic information only, whereas the L1 parser has access to both the semantic and syntactic information in the on-line computation of syntactic structures. What has not been mentioned on this view is how L2 learners use lexically-associated frequency information about syntactic structures. It is unclear how L2 learners use this type of information given that it is both lexical and structural.

Chapter 2 (Experiment 1) provided an answer to this question by examining how L1-Mandarin learners of L2-English process DO/SC sentences in English. Results showed that they were capable of learning the verb bias cues that were specific to the L2, and of using such cues fast enough to influence the processing of subsequent words. In addition, Chapter 2 also explored whether cue usage in the L2 is affected by the L1. In Mandarin, verbs appear early in sentences

and thus they are useful in predicting the upcoming structure. In contrast, Mandarin does not have a complementizer that reliably signals an upcoming embedded clause. Thus, Mandarin speakers may be able to use verb bias cues but may show insensitivity to the complementizer cue in English. Results showed that contrary to the predictions, they were able to use both cues and were able to use them combinatorily, indicating that L2 learners were capable of learning lexically-associated frequency information, even when such information was about syntactic structure. They were also able to learn linguistic cues that were absent in their L1.

Findings from Chapter 2 contradict those from Lee et al. (2013) to some extent. Lee et al. showed that L1-Korean learners of L2-English with lower proficiency must rely on the presence of the complementizer cue to make use of the verb bias cue in processing English DO/SC sentences. Chapter 2 of this dissertation, however, found that L1-Mandarin learners with lower proficiency could use both cues, and the use of any one cue did not depend on the presence or absence of the other. A possible explanation of this difference is that the complementizer cue is a reliable cue in Korean, because it is obligatory at the end of all embedded clauses and thus L1-Korean speakers are used to relying on that cue, while L1-Mandarin speakers are not. Verb bias is in general a less reliable cue because it is probabilistic, but L1-Mandarin learners of L2-English may be used to relying on it in their L1. Taken together, Lee et al. and Chapter 2 of this dissertation showed that L2 learners' acquisition of useful cues in the L2 is affected by how the cues are used in their L1.

Chapter 2 (Experiment 2) further explored the relative importance of verb bias and plausibility in the resolution of temporary structural ambiguity in the L2. This question has not been examined before, and yet it is informative to ongoing debate regarding whether the L2 parser compensates for its underuse of syntactic information by relying heavily on plausibility

cues. The findings provided evidence against such an account, since L2 learners were found to be insensitive to the plausibility information that could possibly help them avoid garden-pathing, just like native speakers.

Chapter 3 aimed to test whether the observed verb bias effect in English sentences was generalizable to other languages, rather than just a property that was specific to English. Verb bias studies have played an important role in the history of psycholinguistic research, as they have provided a window to test between the two most influential sentence processing theories, i.e., the Garden-Path Model and the constraint-based models. Thus it is important to show that previous research did not use a property that was specific to English to test the language processing mechanism that was supposed to be applicable to all languages. Chapter 3 eliminated this concern by showing that verb bias had a strong effect in Mandarin just like in English.

Chapter 3 also tested a claim that Mandarin relied more on plausibility than on syntax during sentence comprehension (Su, 2001a, 2001b, 2004). Findings from Chapter 3 challenged this claim by showing that verb bias played a dominant role in deciding syntactic structures in Mandarin, and plausibility only showed an effect when verb bias allowed it. In addition, Chapter 3 contributed to the field of Mandarin sentence processing by testing a unique structure, which is temporarily ambiguous between a blended structure (the direct object of the first clause is also the subject of the second clause) and an embedded clause structure. Plausibility is one of the ways to resolve this ambiguity. When the noun in the first clause is plausible as the direct object, a blended structure results, as in *The angry reporter revealed the truth had been...*, and when it is implausible, an embedded clause structure results, as in *The angry reporter revealed the park had been…* Chapter 3 showed that in Mandarin, the parser preferred the blended structure over the comma-disambiguated embedded clause structure and the implausible-noun-disambiguated

embedded clause structure. Such a parsing preference will provide a good test ground in the future for testing whether L2-Mandarin learners are capable of learning Mandarin-specific parsing strategies. One prediction derived from the Shallow Structure Hypothesis is that L2 learners are not capable of acquiring L2-specific parsing strategies due to their limited access to syntactic information during on-line processing. Thus testing L2 learners' use of L2-specific parsing strategies provides a way to test the predictions from the Shallow Structure Hypothesis (which is the rationale behind the studies on the L2 processing of sentences with relative clause attachment ambiguity, such as Papadopoulou & Clahsen, 2003 and Felser et al., 2003).

Previous research demonstrated that readers tend to misinterpret garden-path sentences such as *While the man hunted the deer that was brown and graceful ran into the woods* as meaning that *the man hunted the deer and the deer ran into the woods*, as evidenced by their high rate of *yes* incorrect responses to the question *Did the man hunt the deer?*. The Good-Enough Processing Account proposed that if 1) reanalysis is not completed, or 2) reanalysis is completed but both the initial misinterpretation and the correct interpretation after reanalysis linger, then readers may answer the questions incorrectly (Christianson et al., 2001). In a series of six experiments, Chapter 4 tested the "Incomplete Reanalysis" version of the Good-Enough Processing Account. Such an account predicts more evidence of reanalysis at the disambiguation for sentences that are answered correctly than those that are answered incorrectly. Chapter 4 found that slower reading times and larger amplitude P600/N400 ERP components at the disambiguation were unrelated to question response accuracy (Experiments 4 and 6), even when readers were discouraged from answering the questions based on inferences (Experiments 5 and 7), thus indicating that failure to fully reanalyze garden-path sentences might not be the primary reason for the lingering initial misinterpretation. Experiments 8 and 9 showed that the likelihood

of the events described in the initial misanalysis part of garden-path sentences was a better predictor of question response accuracy than on-line measures at the disambiguation.

Chapter 4 concluded that unsuccessful reanalysis is not the primary reason that readers interpret garden-path sentences incorrectly. However, in all experiments, whenever there was a numeric trend showing the relationship between on-line measures and off-line question response accuracy, the trend was in the direction that slower reading times or larger ERP effects at the disambiguation were associated with more incorrect responses. This relationship raised a possibility that on-line measures at the disambiguation did not reflect primarily reanalysis effort. They might instead reflect a combination of the amount of reanalysis and the amount of confusion from having multiple potential syntactic structures. Confusion here refers to the processing difficulty that readers experience after receiving competing cues that support different syntactic structures. However, more processing difficulty does not necessarily lead to more reanalysis work, since readers may simply proceed without successfully reanalyzing the syntactic structure.

The results from Chapter 4 suggest that in Chapters 2 and 3, reading times at the disambiguation might also reflect a combination of processing difficulty (or confusion) and reanalysis work, rather than just reanalysis as we had previously assumed. For L2 learners, the possibility of on-line measures indexing the amount of confusion is even bigger, since L2 learners in general are worse than native speakers at using syntactic cues to recover from garden-pathing. The studies in Chapter 2 did not probe L2 learners' final interpretation of DO/SC sentences, and therefore it was unclear whether L2 learners had successfully recovered from the initial misanalysis. Previous research and Chapter 4 in this dissertation demonstrated that native speakers often misinterpret garden-path sentences if the initial misinterpretations are consistent

with world knowledge (e.g., *The man hunted the deer*). Based on current theories on L2 sentence processing, L2 learners may be even less accurate in interpreting garden-path sentences than native speakers, for two reasons. First, L2 learners rely heavily on plausibility, and thus may find it harder to abandon sensible interpretations derived from the initial incorrect syntactic analyses; and second, L2 learners are less sensitive to syntactic cues, and thus may perform syntactic reanalysis less successfully than native speakers. Since syntactic reanalysis is the deciding factor for reaching the correct interpretation, L2 learners may commit more errors in interpreting garden-path sentences than native speakers.

Chapter 4 provided theoretical and methodological implications for investigating L2 learners' final interpretation of garden-path sentences. Whereas incomplete reanalysis has been found not to be the primary reason for incorrect final interpretations by native speakers, it may be the primary reason for L2 learners. Data from native speakers showed that the likelihood of the events described in the initial misinterpretation is the deciding factor for lingering misinterpretation, indicating that although syntactic reanalysis may have been successfully conducted, the interpretation derived from the revised syntactic structure may not be strong enough to override the interpretation derived from the initial incorrect syntactic structure. For L2 learners, in contrast, as they tend to underuse syntactic cues during on-line parsing, they may experience confusion at the disambiguation but such confusion may not trigger syntactic reanalysis. Therefore, the initial misinterpretation may linger because reanalysis is never performed. Results from Chapter 3 also raise the possibility that L2 learners' final interpretation of garden-path sentences may be affected by how similar syntactic structures are processed in their L1. For instance, Mandarin does not require syntactic revision at the first verb in the embedded clauses in DO/SC sentences because Mandarin allows null subjects, and thus L1-

Mandarin speakers of L2-English may never syntactically revise the structure when reading English sentences with early/late closure ambiguity. In summary, current L2 processing theory predicts that, unlike native speakers, incomplete reanalysis may be the primary reason that L2 learners fail to interpret garden-path sentences correctly. Future research is needed to test this prediction.

Future directions

In what follows, I outline some assumptions that were used in this dissertation and suggest possible ways to test them. I also suggest some approaches to further explore the research questions.

In Chapters 2 and 3 (Experiments 1-3), L2 learners of English and native speakers of Mandarin were found to experience processing difficulty when sentence structures turned out to be inconsistent with verbs' structural preferences (e.g., DO-bias verb is followed by an embedded clause). This result has been interpreted as showing that they were able to use verb bias cues to predict the upcoming syntactic structure. This interpretation is based on the assumption that predictive processing is available in sentence comprehension. In recent years, predictive processing has become a topic of debate in psycholinguistics. Traditionally, sentence processing has been viewed as a process that is incremental and receptive, in the sense that the language processor integrates each word as soon as it is encountered, but does not go beyond that to actively predict the following words or structures. However, evidence is accumulating to show that predictive processing is possible especially when the preceding context is strongly constraining (e.g., DeLong, Urbach, & Kutas, 2005; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Most relevant to Chapters 2 and 3 is a study by Arai and Keller (2013), who used

visual-world eye-tracking paradigm to show that readers actively predicted the following direct object after transitive verbs (e.g., *punish*) but not after intransitive verbs (e.g., *disagree*), and predicted reduced relative clauses more often after verbs that were frequently used in the past participle form (e.g., *record*) than those that were frequently used in the past tense form (e.g., *watch*). This study indicated that native speakers are able to use lexically-specific frequency information about syntactic structure to predict the following words. In Chapters 2 and 3, readers most likely have used the verb bias cues predictively. However, it would be useful to test whether prediction is involved by using the visual-world paradigm. Such future research would also provide evidence for or against a contentious view in the L2 sentence processing literature, which concerns whether L2 learners can predict during real-time sentence processing (see Kaan, 2014, for a review).

Chapter 4 used comprehension questions to directly probe whether the initial misinterpretation lingered. However, as mentioned earlier, a better way to probe the final interpretation is to use implicit measures, since the comprehension questions used in Chapter 4 and in previous studies might have reactivated the initial misinterpretation (Sturt, 2007; van Gompel et al., 2006). Future work is needed to look at the relationship between on-line measures at the disambiguation and the final interpretation, where the final interpretation is measured with implicit measures such as the reading times of a subsequent sentence that is consistent with the correct but not the incorrect analysis of the first sentence. Slattery et al. (2013) is an example of such a design. In this study, readers read two sentences, in which the first sentence contained early/late closure ambiguity and the second sentence tested the final interpretation of the first sentence by examining the reading time of a region in the second sentence where semantics was only congruent with the final correct interpretation, as in *While Fred dried off the truck that was*

dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog. Readers slowed down at *himself* in the second sentence, indicating that the initial misinterpretation persisted beyond the point of reanalysis. However, Slattery et al. did not examine the relationship between the reading time at the disambiguation in the first sentence and the reading time at *himself* in the second sentence. A correlation between the two, such that slower reading times at the disambiguation in the first sentence lead to faster reading times at *himself* in the second sentence will support the "Incomplete Reanalysis" version of the Good-Enough Processing Account.

Previous research has mostly assumed that the initial incorrect syntactic structure is successfully revised after reanalysis is performed. For instance, Sturt (2007) observed that readers only slowed down at the disambiguating words *was actually* in *The explorer found the South Pole was actually right at their feet*, but not the following words, and slowed down at both the disambiguating words *was actually* and also the words following them in *The explorer found the South Pole was actually impossible to reach*. He concluded that syntactic reanalysis was successfully completed, because slowing down was localized to the disambiguating region in the first sentence. However, semantics derived from the initial misinterpretation persisted after the initially-built structure is successfully revised, as shown in the slow reading time at *impossible to reach*, which is semantically inconsistent with the initial misinterpretation *the explorer found the South Pole*. However, findings from Chapter 4 in this dissertation indicated that processing difficulty localized to the disambiguating region did not necessarily mean that syntactic revision was successfully performed, because slowing down at the disambiguation could reflect multiple factors, including confusion. Thus, it is possible that the correct syntactic structure was not achieved after the disambiguating region was read. Future research is needed to examine whether

the syntactic structure is successfully revised after slowing down at the disambiguation has occurred. One way to do this is to examine a region in a subsequent sentence that is only syntactically consistent with the correct structure of the preceding garden-path sentence.

An important and interesting finding in Chapter 4 is that the same sentences, such as *While the man hunted the deer that was brown and graceful ran into the woods*, elicited a P600 effect at the disambiguation when the question asked *Did the man hunt the deer?* and a N400 effect when the question asked *Did the sentence explicitly say that the man hunted the deer?*. This is one of the first studies showing that different ERP components may be evoked depending on what strategies readers adopt when processing the sentences. N400 was evoked when readers thought that more work on checking the likelihood of a man hunting a deer would help them answer the comprehension question. Somewhat similar results have been found in two other ongoing studies. Oines and Kim (2014) showed that when readers were given the instruction of trying to figure out the literal meaning of sentences like *The hearty meal was devouring…*, a LAN effect was observed at *devouring*, rather than a "semantic P600" effect that was usually evoked by this type of "role-reversal" sentences. Preliminary results from another ongoing study (Garnsey, in prep) manipulating verbs' biases in DO/SC sentences suggested that presentation rate may also have an effect on brain responses elicited by the disambiguation. Existing evidence is far too limited so far to allow any strong conclusions to be drawn, but Chapter 4 suggested that sentences can be processed in different ways under slightly different tasks. Future research is needed to provide a clearer picture of what factors may fundamentally change the way sentences are processed.

To conclude, this dissertation found that incomplete reanalysis is not the primary reason for incorrect interpretation of garden-paths sentences by native speakers, and paved the way for

future work examining the final interpretation of garden-path sentences by L2 learners. Results also showed that native English speakers, L2 learners of English and native Mandarin speakers were similar in their use of the verb bias and plausibility cues, such that verb bias plays a stronger role than plausibility in guiding on-line interpretation of syntactic structures, challenging the claims that L2 learners and native Mandarin speakers rely heavily on plausibility during on-line sentence processing.

APPENDIX

Experimental stimuli for Experiment 1:

SC-bias verbs:

1. The unreliable butler admitted (that) the theft could have been prevented if he was not sleeping.
2. The ticket agent admitted (that) the mistake might be hard to correct.
3. The dedicated soldier admitted (that) the defeat might not have been completely inevitable.
4. The new receptionist admitted (that) her error should have been corrected sooner.
5. The defensive journalist argued (that) the view could have confused readers who were not experts.
6. The district attorney argued (that) the point would make a difference to everyone.
7. The divorce lawyer argued (that) the issue should be attended to very carefully.
8. The art professor argued (that) the interpretation might have been too controversial.
9. The captivated audience believed (that) the magician should be willing to explain his tricks.
10. The naive girl believed (that) the urban myth might not be a myth after all.
11. The shrewd officer believed (that) the criminal might have a concealed weapon on him.
12. The magazine editor believed (that) the article might be the best article he had ever written.
13. The murder suspect confessed (that) the crimes had gotten much worse over time.
14. The ashamed boy confessed (that) the lie might have deceived his whole family.
15. The government official confessed (that) the conspiracy could have damaged international relationships.
16. The fanatical terrorist confessed (that) the plot could be uncovered by the authorities.
17. The certified accountant figured (that) the budget should adjust to meet the increase in costs.
18. The insurance agent figured (that) the deductible should have decreased for the safe driver.
19. The delivery manager figured (that) the weight needed to decrease by several pounds.
20. The overwhelmed parents figured (that) the tuition might cost more than they could afford.
21. The gardener's assistant indicated (that) the temperature would be good for the flowers.
22. The office manager indicated (that) the problem could be worst for the new secretaries.
23. The roof inspector indicated (that) the leak would be expensive to fix.
24. The traffic officer indicated (that) the direction might be congested with many cars.
25. The sensitive boy inferred (that) the insult had been directed at him personally.
26. The church congregation inferred (that) the meaning was badly explained by the minister.
27. The rejected bachelor inferred (that) the reason could be his reluctance to make a commitment.
28. The hired investigator inferred (that) the evidence meant the suspect was not guilty.
29. The careful scientist proved (that) the theory might be difficult to explain.
30. The successful tests proved (that) the hypothesis could reveal the underlying mechanism.
31. The local detectives proved (that) the conspiracy had caused the government to crack down.
32. The birth certificate proved (that) the birthplace was not where we thought.
33. The plastic surgeon suggested (that) the operation would be too costly for the patient.
34. The swimming instructor suggested (that) the technique might be too difficult for the frightened novice.

35. The guidance counselor suggested (that) the job would help the student learn to be more responsible.
36. The writing instructor suggested (that) the book would need to be revised.
37. The ship's captain suspected (that) the mutiny would be damaging to his career.
38. The boxing referee suspected (that) the outcome had been staged right from the start.
39. The irate student suspected (that) the roommate stole the money while he was in class.
40. The wary teacher suspected (that) the cheating could cause bad feelings among the students.

DO-bias verbs

41. The admissions office accepted (that) the application did not include some of the necessary documents.
42. The annoyed professor accepted (that) the excuse had been completely made up by the student.
43. The basketball star accepted (that) the contract requires him to play every game.
44. The department head accepted (that) the proposal would be resubmitted very late.
45. The brilliant doctor discovered (that) the cure would soon be shown to work for everyone.
46. The determined biologists discovered (that) the organism had not been seen before.
47. The famous archaeologist discovered (that) the artifacts might have been very clever fakes.
48. The FBI investigator discovered (that) the plot had have improved safety in the lab.
49. The enthusiastic students established (that) the club could be a meeting place for chess matches.
50. The head referee established (that) the rules were not to be strictly enforced.
51. The new lawyer established (that) the practice aims to serve the whole community.
52. The gossipy neighbor heard (that) the story could not be further from the truth.
53. The excited children heard (that) the fireworks were being planned to be the biggest ever.
54. The marine sergeant heard (that) the explosion might have been the result of an accident.
55. The orchestra conductor heard (that) the violins were not properly in tune.
56. The astronomy buff observed (that) the comet had been approaching very quickly.
57. The bird watcher observed (that) the sparrows had been taken from the nest.
58. The clever journalist observed (that) the scene could have been tampered with by police.
59. The construction worker observed (that) the house seemed to be in great condition.
60. The accused doctor protested (that) the lawsuit should have been settled out of court.
61. The activist group protested (that) the discrimination had been covered up by the governor.
62. The elementary students protested (that) the uniforms were too uncomfortable to play in.
63. The navy veterans protested (that) the war could become too expensive to continue.
64. The commanding general revealed (that) the strategy would help the army defeat the enemy.
65. The confessing criminal revealed (that) the hideout appeared to just be an abandoned warehouse.
66. The confident magician revealed (that) the rabbit had disappeared from his cage.
67. The gallery owner revealed (that) the painting is the most expensive one he's ever sold.
68. The club members understood (that) the bylaws would be applied to everyone.
69. The disciplined lieutenant understood (that) the orders were standard for all new recruits.
70. The foreign diplomat understood (that) the translation might take longer than they had
71. anticipated.

72. The frustrated tourists understood (that) the message had never been sent.
73. The bank worker forgot (that) the policy would be implemented the very next day.
74. The college student forgot (that) the answer could be found at the back of the textbook.
75. The elderly woman forgot (that) the address had been changed since her last visit.
76. The hapless suitor forgot (that) the flowers reminded the woman of her ex-husband.
77. The angry farmer warned (that) the trespassers would not be allowed onto his fields.
78. The army general warned (that) the civilians might be in danger from the bombs.
79. The kind usher warned (that) the audience should not bring food or drink into the theater.
80. The new professor warned (that) the students should be on time for his class.

Experimental stimuli for Experiment 2:

SC-bias verbs:

1. The unreliable butler admitted (that) the theft could have been prevented if he was not sleeping.
   The unreliable butler admitted (that) the meal could have been cold by the time of serving.
2. The ticket agent admitted (that) the mistake might be hard to correct.
   The ticket agent admitted (that) the kiosk might be difficult to find.
3. The dedicated soldier admitted (that) the defeat might not have been completely inevitable.
   The dedicated soldier admitted (that) the trench might not have been very sturdy.
4. The new receptionist admitted (that) her error should have been corrected sooner.
   The new receptionist admitted (that) her phone should have been disconnected earlier.
5. The defensive journalist argued (that) the view could have confused readers who were not experts.
   The defensive journalist argued (that) the watch could have some scratches after she dropped it.
6. The district attorney argued (that) the point would make a difference to everyone.
   The district attorney argued (that) the haircut would make him look more professional.
7. The divorce lawyer argued (that) the issue should be attended to very carefully.
   The divorce lawyer argued (that) the potato should be peeled with precision and care.
8. The art professor argued (that) the interpretation might have been too controversial.
   The art professor argued (that) the artist might have used new techniques.
9. The captivated audience believed (that) the magician should be willing to explain his tricks.
   The captivated audience believed (that) the tickets should be sold at a cheaper price.
10. The naive girl believed (that) the urban-myth might not be a myth after all.
    The naive girl believed (that) the bus might not stop at all the stops.
11. The shrewd officer believed (that) the criminal might have a concealed weapon on him.
    The shrewd officer believed (that) the coat might have some evidence in the pockets.
12. The magazine editor believed (that) the article might be the best article he had ever written.
    The magazine editor believed (that) the fridge might be broken because his soda was quite warm.
13. The murder suspect confessed (that) the crimes had gotten much worse over time.
    The murder suspect confessed (that) the clothes had gotten smaller in the dryer.
14. The ashamed boy confessed (that) the lie might have deceived his whole family.
    The ashamed boy confessed (that) the car might have more damage than expected.
15. The government official confessed (that) the conspiracy could have damaged international relationships.
    The government official confessed (that) the child could have been protected better.
16. The fanatical terrorist confessed (that) the plot could be uncovered by the authorities.
    The fanatical terrorist confessed (that) the gun could be hidden in the basement.
17. The certified accountant figured (that) the budget should adjust to meet the increase in costs.
    The certified accountant figured (that) the customer should adjust his expectations about the total cost.
18. The insurance agent figured (that) the deductible should have decreased for the safe driver.
    The insurance agent figured (that) the art should have been protected much more carefully.

19. The delivery manager figured (that) the weight needed to decrease by several pounds.
    The delivery manager figured (that) the envelope needed to be sealed with tape.
20. The overwhelmed parents figured (that) the tuition might cost more than they could afford.
    The overwhelmed parents figured (that) the holidays might cost them a lot of money.
21. The gardener's assistant indicated (that) the temperature would be good for the flowers.
    The gardener's assistant indicated (that) the light would be insufficient for the plants.
22. The office manager indicated (that) the problem could be worst for the new secretaries.
    The office manager indicated (that) the party could be too much of a distraction.
23. The roof inspector indicated (that) the leak would be expensive to fix.
    The roof inspector indicated (that) the machine would be fixed tomorrow evening.
24. The traffic officer indicated (that) the direction might be congested with many cars.
    The traffic officer indicated (that) the squirrel might be responsible for the accident.
25. The sensitive boy inferred (that) the insult had been directed at him personally.
    The sensitive boy inferred (that) the milk had been left out too long.
26. The church congregation inferred (that) the meaning was badly explained by the minister.
    The church congregation inferred (that) the carpet was badly stained with grape juice.
27. The rejected bachelor inferred (that) the reason could be his reluctance to make a commitment.
    The rejected bachelor inferred (that) the computer could be helpful in finding him a date.
28. The hired investigator inferred (that) the evidence meant the suspect was not guilty.
    The hired investigator inferred (that) the officer meant the victim was still alive.
29. The careful scientist proved (that) the theory might be difficult to explain.
    The careful scientist proved (that) the researchers might be falsifying the data.
30. The successful tests proved (that) the hypothesis could reveal the underlying mechanism.
    The successful tests proved (that) the scientist could reveal his surprising results.
31. The local detectives proved (that) the conspiracy had caused the government to crack down.
    The local detectives proved (that) the pothole had caused the massive car crash yesterday.
32. The birth certificate proved (that) the birthplace was not where we thought.
    The birth certificate proved (that) the boy was not an American citizen.
33. The plastic surgeon suggested (that) the operation would be too costly for the patient.
    The plastic surgeon suggested (that) the girl would be completely satisfied with the results.
34. The swimming instructor suggested (that) the technique might be too difficult for the frightened novice.
    The swimming instructor suggested (that) the weather might be too rough to have the competition.
35. The guidance counselor suggested (that) the job would help the student learn to be more responsible.
    The guidance counselor suggested (that) the grades would help the student get into a top college.
36. The writing instructor suggested (that) the book would need to be revised.
    The writing instructor suggested (that) the storm would need a full description.
37. The ship's captain suspected (that) the mutiny would be damaging to his career.
    The ship's captain suspected (that) the moon would be covered by thick clouds.
38. The boxing referee suspected (that) the outcome had been staged right from the start.
    The boxing referee suspected (that) the match had been thrown by the expected champion.
39. The irate student suspected (that) the roommate stole the money while he was in class.

The irate student suspected (that) the lecture stole the content from a different physics class.

40. The wary teacher suspected (that) the cheating could cause bad feelings among the students.
    The wary teacher suspected (that) the lesson could cause her students to fall asleep.

DO-bias verbs:

41. The admissions office accepted (that) the application did not include some of the necessary documents.
    The admissions office accepted (that) the parade did not mean they could go home early.
42. The annoyed professor accepted (that) the excuse had been completely made up by the student.
    The annoyed professor accepted (that) the fire had been set to cause a fire alarm.
43. The basketball star accepted (that) the contract requires him to play every game.
    The basketball star accepted (that) the airport requires him to go through security.
44. The department head accepted (that) the proposal would be resubmitted very late.
    The department head accepted (that) the temperature would be lower over break.
45. The brilliant doctor discovered (that) the cure would soon be shown to work for everyone.
    The brilliant doctor discovered (that) the waitress would soon be bringing the meal he ordered.
46. The determined biologists discovered (that) the organism had not been seen before.
    The determined biologists discovered (that) the conference had not been rescheduled yet.
47. The famous archaeologist discovered (that) the artifacts might have been very clever fakes.
    The famous archaeologist discovered (that) the pants might have been stained while digging.
48. The FBI investigator discovered (that) the plot had been planned for three years.
    The FBI investigator discovered (that) the judge had been biased throughout the trial.
49. The biology class established (that) the routine could have improved safety in the lab.
    The biology class established (that) the frog could have died from a lack of oxygen.
50. The enthusiastic students established (that) the club could be a meeting place for chess matches.
    The enthusiastic students established (that) the hamster could be a good pet for biology class.
51. The head referee established (that) the rules were not to be strictly enforced.
    The head referee established (that) the kids were not allowed on the field.
52. The new lawyer established (that) the practice aims to serve the whole community.
    The new lawyer established (that) the speech aims to outline the firm's objectives.
53. The gossipy neighbor heard (that) the story could not be further from the truth.
    The gossipy neighbor heard (that) the razor could not have been the murder weapon.
54. The excited children heard (that) the fireworks were being planned to be the biggest ever.
    The excited children heard (that) the brownies were being handed out in the school auditorium.
55. The marine sergeant heard (that) the explosion might have been the result of an accident.
    The marine sergeant heard (that) the light might have helped in finding the missing sailors.
56. The orchestra conductor heard (that) the violins were not properly in tune.
    The orchestra conductor heard (that) the lights were not turning off completely.
57. The astronomy buff observed (that) the comet had been approaching very quickly.
    The astronomy buff observed (that) the afternoon had been a complete waste.

58. The bird watcher observed (that) the sparrows had been taken from the nest.
    The bird watcher observed (that) the voice had been distorted by the wind.
59. The clever journalist observed (that) the scene could have been tampered with by police.
    The clever journalist observed (that) the format could have changed the number of pages.
60. The construction worker observed (that) the house seemed to be in great condition.
    The construction worker observed (that) the morning seemed to drag on and on.
61. The accused doctor protested (that) the lawsuit should have been settled out of court.
    The accused doctor protested (that) the nurse should have done a much better job.
62. The activist group protested (that) the discrimination had been covered up by the governor.
    The activist group protested (that) the fence had been built without consulting them first.
63. The elementary students protested (that) the uniforms were too uncomfortable to play in.
    The elementary students protested (that) the concepts were too difficult to fully comprehend.
64. The navy veterans protested (that) the war could become too expensive to continue.
    The navy veterans protested (that) the ocean could become dangerous during the storm.
65. The commanding general revealed (that) the strategy would help the army defeat the enemy.
    The commanding general revealed (that) the night would help provide cover for the attack.
66. The confessing criminal revealed (that) the hideout appeared to just be an abandoned warehouse.
    The confessing criminal revealed (that) the cell appeared to be much smaller than usual.
67. The confident magician revealed (that) the rabbit had disappeared from his cage.
    The confident magician revealed (that) the institution had disappeared without a trace.
68. The gallery owner revealed (that) the painting is the most expensive one he's ever sold.
    The gallery owner revealed (that) the holiday is the most lucrative time of the year.
69. The club members understood (that) the bylaws would be applied to everyone.
    The club members understood (that) the pool would be closed on Mondays.
70. The disciplined lieutenant understood (that) the orders were standard for all new recruits.
    The disciplined lieutenant understood (that) the shoes were standard issue for every soldier.
71. The foreign diplomat understood (that) the translation might take longer than they had anticipated.
    The foreign diplomat understood (that) the car might take too long to arrive.
72. The frustrated tourists understood (that) the message had never been sent.
    The frustrated tourists understood (that) the hotel had never been remodeled.
73. The bank worker forgot (that) the policy would be implemented the very next day.
    The bank worker forgot (that) the escalator would be out of commission all day.
74. The college student forgot (that) the answer could be found at the back of the textbook.
    The college student forgot (that) the snow could be quite slippery and dangerous to drive on.
75. The elderly woman forgot (that) the address had been changed since her last visit.
    The elderly woman forgot (that) the FBI had been suspicious about her son's alibi.
76. The hapless suitor forgot (that) the flowers reminded the woman of her ex-husband.
    The hapless suitor forgot (that) the woods reminded the woman about her accident.
77. The angry farmer warned (that) the trespassers would not be allowed onto his fields.
    The angry farmer warned (that) the seeds would not grow tall without being fertilized.
78. The army general warned (that) the civilians might be in danger from the bombs.
    The army general warned (that) the resolution might be too difficult for the men.
79. The kind usher warned (that) the audience should not bring food or drink into the theater.

The kind usher warned (that) the movie should not be seen by young and impressionable children.

80. The new professor warned (that) the students should be on time for his class.
    The new professor warned (that) the textbook should be brought to the class everyday.

Experimental stimuli for Experiment 3:

SC-bias verbs:

1. 新来 的 员工 抱怨(,) 领导 总是 让 他 做 最累的 工作。
   新来 的 员工 抱怨(,) 晚上 总是 让 他 一个人 加班。
2. 普通 市民 抱怨(,) 物价 怎么 总是 在 不停地 上涨。
   普通 市民 抱怨(,) 幸福 怎么 总是 那么 遥不可及。
3. 刚工作 的 年轻人 抱怨(,) 房价 怎么 总是 那么 高。
   刚工作 的 年轻人 抱怨(,) 知识 怎么 总是 不够 用。
4. 病床上 的 病人 相信(,) 医生 一定 会 尽力 救他。
   病床上 的 病人 相信(,) 疼痛 一定 会 慢慢 减弱。
5. 有爱心 的 老师 相信(,) 学生 不 会 故意 撒谎。
   有爱心 的 老师 相信(,) 手机 不 会 是 学生 弄坏的。
6. 善良 的 妻子 相信(,) 丈夫 不会 是 小偷。
   善良 的 妻子 相信(,) 皮夹 不会 是 丈夫 偷的。
7. 经济学家 预测(,) 房价 很快 会 大幅度 下跌。
   经济学家 预测(,) 市民 很快 会 卖掉 手上的 股票。
8. 气象专家 预测(,) 气温 将 会 超过 四十度。
   气象专家 预测(,) 冰山 将 会 慢慢 融化。
9. 地质专家 预测(,) 地震 可能 会 发生 在 沿海一带。
   地质专家 预测(,) 明年 可能 会 发生 地震。
10. 获奖 的 作家 得知(,) 喜讯 早就 已经 传到了 他的单位。
    获奖 的 作家 得知(,) 电视 早就 已经 播放了 他 得奖的 消息。
11. 绝望的 病人 得知(,) 病情 可能 会 危及 生命。
    绝望的 病人 得知(,) 灰尘 可能 会 造成 术后 感染。
12. 参加了 高考 的 学生 得知(,) 成绩 将 会 在 一个月 后 公布。
    参加了 高考 的 学生 得知(,) 上海 将 会 大幅度 提高 招生 数量。
13. 细心的 助理 担心(,) 演员 可能 会 被 大雨 淋湿。
    细心的 助理 担心(,) 垃圾 可能 会 弄脏 演员的 裙子。
14. 体育老师 担心(,) 学生 可能 会 在 比赛中 受伤。
    体育老师 担心(,) 足球 可能 会 被 学生 弄丢。
15. 孝顺的 孙子 担心(,) 奶奶 有 可能 不愿 去 医院 看病。
    孝顺的 孙子 担心(,) 中药 有 可能 治不好 奶奶的 病。
16. 有经验 的 科学家 估计(,) 数据 有 可能 不 准确。
    有经验 的 科学家 估计(,) 海洋 有 可能 存在 一千万个 物种。
17. 销售部 的 经理 估计(,) 营业额 可能 会 超过 一千万 元。
    销售部 的 经理 估计(,) 洗碗机 可能 会 卖不掉。
18. 有经验 的 医生 估计(,) 病因 可能 是 劳累 过度。
    有经验 的 医生 估计(,) 牛奶 可能 是 造成 病人 肚子疼 的 原因。
19. 滑冰 运动员 怀疑(,) 裁判 可能 会 不让 她 参加 比赛。
    滑冰 运动员 怀疑(,) 大雨 可能 会 影响 比赛。

20. 认真的 统计员 怀疑(,) 数据 很 可能 错 了。
认真 的 统计员 怀疑(,) 现金 很 可能 不 够。
21. 聪明的 法官 怀疑(,) 证据 会 不会 是 假的。
聪明 的 法官 怀疑(,) 手枪 会 不会 是 假的。
22. 人事处 经理 宣布(,) 名单 很快 会 公布 出来。
人事处 经理 宣布(,) 午饭 很快 会 送 来。
23. 漂亮的 主持人 宣布(,) 结果 很快 就要 出来 了。
漂亮的 主持人 宣布(,) 电脑 很快 就要 算出 结果 了。
24. 恋爱 三年 的 同事 宣布(,) 婚事 要 回 老家 办。
恋爱 三年 的 同事 宣布(,) 夏天 要 回 老家 结婚。
25. 细心的 潜水员 发现(,) 沉船 可能 是 著名的 泰坦尼克号。
细心的 潜水员 发现(,) 大火 可能 是 这艘船 沉没的 原因。
26. 正在 实习的 大学生 发现(,) 问题 真的 很 难 解决。
正在 实习的 大学生 发现(,) 社会 真的 很 复杂。
27. 细心的 丈夫 发现(,) 秘密 已经 被 妻子 知道了。
细心的 丈夫 发现(,) 客厅 已经 被 妻子 的 衣服 堆满了。
28. 漂亮的 女演员 否认(,) 传言 已经 影响 到了 她的 工作。
漂亮的 女演员 否认(,) 天气 已经 影响 到了 她的 工作。
29. 狡猾的 罪犯 否认(,) 事实 已经 被 警察 知道了。
狡猾的 罪犯 否认(,) 汽车 已经 被 他 卖掉了。
30. 爱管闲事 的 邻居 听说(,) 这件事 其实 是 老李 捏造的。
爱管闲事 的 邻居 听说(,) 面条儿 其实 是 老李 偷吃的。
31. 新来 的 秘书 听说(,) 传言 其实 是 老板 编造的。
新来 的 秘书 听说(,) 公司 其实 是 老板娘 创立的。
32. 电影厂 的 导演 听说(,) 谣言 可能 是 真 的。
电影厂 的 导演 听说(,) 服装 可能 是 旧 的。

DO-bias verbs:

33. 愤怒 的 记者 揭露(,) 真相 已经 被 封锁 了。
愤怒 的 记者 揭露(,) 公园 已经 被 官员 用来 建 别墅 了。
34. 环境 专家 揭露(,) 问题 还 没有 引起 政府的 重视。
环境 专家 揭露(,) 空气 还 没有 达到 国家 规定的 质量 标准。
35. 航空 公司 的 空姐 揭露(,) 事实 已经 被 机长 隐瞒了 很久。
航空 公司 的 空姐 揭露(,) 行李 已经 被 偷偷地 掉包了。
36. 懂事 的 女儿 理解(,) 妈妈 为什么 会 突然 生 那么大 的 气。
懂事 的 女儿 理解(,) 每天 为什么 会 有 那么多 的 功课。
37. 聪明 的 小学生 理解(,) 这道题 为什么 可以 有 两种 解法。
聪明 的 小学生 理解(,) 洗衣机 为什么 可以 把 衣服 洗干净。

38. 优秀 的 作家 理解(,) 这篇文章 为什么 会 吸引 那么多 读者。
优秀 的 作家 理解(,) 道路两旁 为什么 会 有 那么多 家 书店。
39. 外科 护士 提出(,) 建议 没有 得到 医生的 重视。
外科 护士 提出(,) 病人 没有 得到 很好的 休息。
40. 大桥 的 设计师 提出(,) 新方案 还 可以 让 大桥 更 美观。
大桥 的 设计师 提出(,) 正前方 还 可以 修建 一条 隧道。
41. 著名 的 数学家 提出(,) 算法 可能 有 很多种。
著名 的 数学家 提出(,) 课本 可能 有 不少 错误。
42. 中国 政府 关心(,) 人民 是否 能 吃上 放心的 蔬菜。
中国 政府 关心(,) 周末 是否 能 让 高速 公路 免费。
43. 忧心忡忡 的 家长 关心(,) 孩子 能 不能 考上 名牌大学。
忧心忡忡 的 家长 关心(,) 中午 能 不能 把 孩子 接回家。
44. 热恋中 的 小陈 关心(,) 女友 能 不能 忘记 前男友。
热恋中 的 小陈 关心(,) 后天 能 不能 见到 女友的 父母。
45. 交通部 官员 调查(,) 事故 有 没有 妨碍 人们 出行。
交通部 官员 调查(,) 太阳 有 没有 妨碍 司机 的 视线。
46. 当地 警方 调查(,) 案情 是 不是 跟 罪犯 交代的 一样。
当地 警方 调查(,) 后天 是 不是 毒贩 接头的 日子。
47. 年轻 的 妈妈 惦记(,) 女儿 是 不是 已经 到了 学校。
年轻 的 妈妈 惦记(,) 电视 是 不是 已经 损伤了 宝宝 的 视力。
48. 孝顺 的 女儿 惦记(,) 妈妈 是否 已经 康复 了。
孝顺 的 女儿 惦记(,) 医院 是否 已经 治好了 妈妈 的 病。
49. 操心 的 父亲 惦记(,) 儿子 有 没有 通过 这次的 升学 考试。
操心 的 父亲 惦记(,) 大学 有 没有 破格 录取 他的 儿子。
50. 生物学 教授 读到(,) 论文 有 可能 要 经过 严格的 审阅后 才能 发表。
生物学 教授 读到(,) 蝴蝶 有 可能 在 核辐射 下 发生 变异。
51. 报社 的 编辑 读到(,) 新闻 被 证实 是 假的。
报社 的 编辑 读到(,) 市长 被 证实 逃税 了。
52. 汽车 公司 的 律师 读到(,) 声明 有 可能 不 符合 法律 规范。
汽车 公司 的 律师 读到(,) 新车 有 可能 在 设计上 侵权了。
53. 新来 的 工人 打听(,) 小道消息 是 不是 真 的。
新来 的 工人 打听(,) 下午五点 是 不是 可以 下班 回家了。
54. 考完试 的 大学生 打听(,) 成绩 会 不会 告诉 家长。
考完试 的 大学生 打听(,) 下次 会 不会 考得 更难。
55. 参赛 的 运动员 打听(,) 比赛结果 是 不是 已经 出来了。
参赛 的 运动员 打听(,) 下个星期 是 不是 还有 一场 比赛。
56. 几个 中学生 讨论(,) 作业 该 怎么 写 才能 令 老师 满意。
几个 中学生 讨论(,) 后面 该 怎么 写 才能 让 这个 故事 有个 完美的 结局。
57. 地震 灾区 的 专家们 讨论(,) 灾情 应该 怎样 向 上级 汇报。
地震 灾区 的 专家们 讨论(,) 教室 应该 怎样 用于 安置 灾民。
58. 有经验 的 法官们 讨论(,) 案情 到底 是不是 跟 罪犯 交代的 一样。

有经验 的 法官们 讨论(,) 剪刀 到底 是不是 作案 工具。

59. 内疚 的 丈夫 回想(,) 过去 都 是 自己 做得 不对。
内疚 的 丈夫 回想(,) 家务 都 是 前妻 一个人 默默地 承担。

60. 新婚 的 妻子 回想(,) 婚礼 是 她 这一生 最 甜蜜的 时刻。
新婚 的 妻子 回想(,) 工厂 是 她 和 丈夫 认识的 地方。

61. 退伍 的 老人 回想(,) 战争 是 他 这辈子 见过的 最残酷 的 事情。
退伍 的 老人 回想(,) 米饭 是 他 在 战争 年代 最渴望 的 东西。

62. 商场 的 保安 警告(,) 顾客 可能 会 蜂拥 而至。
商场 的 保安 警告(,) 台阶 可能 会 绊倒 顾客。

63. 公交车 司机 警告(,) 小偷 有 可能 已经 上车 了。
公交车 司机 警告(,) 轮胎 有 可能 会 在 中途 坏掉。

64. 美国 政府 警告(,) 朝鲜 可能 会 使用 核武器。
美国 政府 警告(,) 大桥 可能 会 遭到 恐怖 袭击。

References:

Arai, M., & Keller. (2013). The use of verb-specific information for prediction in sentence processing. Language and Cognitive Processes, 28, 525-560.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Barton, S.B., & Sanford, A.J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition, 21*, 477-487.

Bever, T.G. (1970). The cognitive basis for linguistic structures. *Cognition and the development of language, 279*, 1-61.

Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., & Carlson, G.N. (1995). Verb Argument Structure in Parsing and Interpretation: Evidence from wh-Questions. *Journal of Memory and Language, 34*, 774-806.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, *5*, e10729.

Chatrian, G.E. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *American Journal Electroencephalograph Technology, 25*, 83-92.

Christianson, K., Hollingworth, A., Halliwell, J.F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368-407.

Christianson, K., & Luke, S.G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading, 15*, 136-166.

Christianson, K., Luke, S.G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 538-544.

Christianson, K., Williams, C., Zacks, R.T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes, 42*, 205-238.

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107-126.

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Cousineau, D. (2005). Confidence Intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42-45.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2009). Constraints on L2 learners' processing of wh-dependencies: Evidence from eye movements. *Essex Research Reports in Linguistics, 58*, 1-24.

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probablistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Dussias, P.E., & Cramer Scaltz, T.R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta psychologica, 128*, 501-513.

Dussias, P.E., Marful, A., Gerfen, C., & Bajo Molina, M.T. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish-English bilinguals. *Behavior Research Methods, 42*, 1004-1011.

Dussias, P.E., & Pinar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research, 26*, 443-472.

Erickson, T.D., & Mattson, M.E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540-551.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75-84.

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571-603.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second Language Research, 23*, 9-36.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453-489.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164-203.

Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11-15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*, 3-20.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Ferreira, F., & Henderson, J.M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555-568.

Ferreira, F., & Henderson, J.M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725-745.

Ferreira, F., & Henderson, J.M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73-100). Dordrecht: The Netherlands: Kluwer Academic.

Ferreira, F., Lau, E.F., & Bailey, K.G.D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science, 28*, 721-749.

Ferreira, F., & Patson, N.D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*, 71-83.

Fodor, J.D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (Vol. 21, pp. 101-141). Kluwer: Dordrecht.

Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. *ETD Collection for University of Connecticut*, AAI7914150.

Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 559-586). Erlbaum: Hillsdale, NJ.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24*, 437-468.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C. (1998). Sentence reanalysis and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143-176). Dordrecht: The Netherlands: Kluwer Academic.

Frazier, L., & Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology, 50A*, 119-148.

Friederici, A.D. (1998). The neurobiology of language processing. In A. D. Friederici (Ed.), *Language Comprehension: A Biological Perspective* (pp. 263-301). Berlin: Springer.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition, 85*, B83-B92.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3*, 111-127.

Garnsey, S.M., Pearlmutter, N.J., Myers, E., & Lotocky, M.A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language, 37*, 58-93.

Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research, 32*, 141-166.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439-483.

Hahne, A., & Friederici, A.D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*, 193-204.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*, 369-397.

Jackson, C.N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning, 58*, 875-909.

Jackson, C.N., & Bobb, S.C. (2009). The processing and comprehension of wh-questions among second language speakers of German. *Applied Psycholinguistics, 30*, 603-636.

Jackson, C.N., & Dussias, P.E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition, 12*, 65-82.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.

Jennings, F., Randall, B., & Tyler, L.K. (1997). Graded effects of verb subcategory preferences on parsing: Support for constraint-satisfaction models. *Language and Cognitive Processes, 12*, 485-504.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning, 48*, 107-147.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? Linguistic Approaches to Bilingualism, 4, 257-282.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159-201.

Kaan, E., & Swaab, T. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*, 98-110.

Kail, M. (1989). Cue validity, cue cost, and processing type in sentence comprehension in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 77-117). New York: Cambridge University Press.

Kamas, E.N., Reder, I.M., & Ayers, M.S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition, 24*, 687-699.

Kaschak, M.P., & Glenberg, A.M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133*, 450-467.

Kennison, S.M. (2001). Limitations on the use of verb information in sentence comprehension. *Psychonomic Bulletin & Review, 8*, 132-138.

Kennison, S.M. (2009). The use of verb information in parsing: Different statistical analyses lead to contradictory conclusions. *Journal of Psycholinguistic Research, 38*, 363-378.

Kim, E., Baek, S., & Tremblay, A. (in press). The role of island constraints in second language sentence processing. *Language Acquisition*, 1-33.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205-225.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196-214.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., & Oor, P.J.W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*, 1-36.

Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: Challengers to syntax. *Brain Research, 1146*, 23-49.

Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes, 21*, 489-530.

Kutas, Marta, & Federmeier, Kara D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Kutas, Marta, & Hillyard, Steven A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science, 207*, 203-205.

Lau, E.F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes, 20*, 633-666.

Lee, C-L., & Federmeier, K.D. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. *Brain Research, 108*, 191-202.

Lee, C-L., & Federmeier, K.D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language, 61*, 538-555.

Lee, E-K., Lu, H-Y., & Garnsey, S.M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language & Cognition, 16*, 761-775.

Lee, E.K., & Garnsey, S.M. (2015). An ERP study of plural attraction in attachment ambiguity resolution: Evidence for retrieval interference. *Journal of Neurolinguistics, 36*, 1-16.

Li, C., & Thompson, S. (1976). Subject and topic: A new typology for language. In C. Li (Ed.), *Subject and topic*. New York: Academic.

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language, 32*, 169-192.

Lim, J-H., & Christianson, K. (2013a). Integrating meaning and structure in L1-L2 and L2-L1 translations. *Second Language Acquisition*, 233-256.

Lim, J-H., & Christianson, K. (2013b). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, 518-537.

Lin, Y., & Garnsey, S.M. (2011a). Animacy and the resolution of temporary ambiguity in relative clause comprehension in Mandarin. In H. Yamashita, Y. Hirose & J. L. Packard (Eds.), *Processing and Producing Head-final Structures* (pp. 241-276). Netherlands: Springer.

Lin, Y., & Garnsey, S.M. (2011b). Verb bias in Mandarin relative clause processing. *Concentric: Studies in Linguistics, 37*.

Loftus, G.R., & Masson, M.E. (1994). Using confidence intervals in within-subject design. *Psychonomic Bulletin & Review, 1*, 476-490.

Lopez-Calderon, J., & Luck, S.J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*, 127-150.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second Language sentence processing. *Studies in Second Language Acquisition, 27*, 53-78.

Maxfield, N.D., Lyon, J.M., & Silliman, E.R. (2009). Disfluencies along the garden path: Brain electrophysiological evidence of disrupted sentence processing. *Brain and Language, 111*, 86-100.

McRae, K., Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283-312.

Morey, R.D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61-64.

Nakamura, C., & Arai, M. (in press). Persistence of initial misanalysis with no referential ambiguity. *Cognitive Psychology*.

Nieuwland, M.S., Otten, M., & Van Berkum, J.J.A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 228-236.

Oines, L., & Kim, A. (2014). *Integrate or repair? ERP responses to semantic anomalies depend on choice of processing strategy.* Paper presented at the Architectures and Mechanisms for Language Processing (AMLaP) Conference, Edinburgh, Scotland.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 563-588.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., & Holcomb, P.J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*, 413-437.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 786-803.

Pan, H.Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua, 121*, 221-236.

Papadopoulou, D. (2005). Reading-Time Studies of Second Language Ambiguity Resolution. *Second Language Research, 21*, 98-120.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition, 25*, 501-528.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering Misinterpretations in Garden-Path Sentences: Evidence From a Paraphrasing Task. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*, 280-285.

Pickering, M.J., & Traxler, M.J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 940.

Pickering, M.J., & Traxler, M.J. (2003). Evidence against the use of subcategorization frequency in the processing of unbounded dependencies. *Language and Cognitive Processes, 18*, 469-503.

Pickering, M.J., Traxler, M.J., & Crocker, M.W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language, 43*, 447-475.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition, 16*, 167-182.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374.

Sabourin, L., & Stowe, L.A. (2008). Second language processing: when are first and second languages processed similarly? *Second Language Research, 24*, 397-430.

Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*, 437-442.

Sanford, A.J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*, 382-386.

Slattery, T.J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from flawed semantic processing. *Journal of Memory and Language, 69*, 104-120.

Spivey-Knowlton, M.J., & Sedivy, J.C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227-267.

Spivey-Knowlton, M.J., Trueswell, J.C., & Tanenhaus, M.K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*, 276.

Stowe, L.A., Tanenhaus, M.K., & Carlson, G.N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech, 34*, 319-340.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition, 105*, 477-488.

Sturt, P., Pickering, M.J., & Crocker, M.W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40*, 136-150.

Su, I-R. (2001a). Context effects on sentence processing: A study based on the Competition Model. *Applied Psycholinguistics, 22*, 167-189.

Su, I-R. (2001b). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics, 22*, 83-112.

Su, I-R. (2004). The effects of discourse processing with regard to syntactic and semantic cues: A competition model study. *Applied Psycholinguistics, 25*, 587-601.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*, 201-216.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning Memory and Cognition*, 431-450.

Tanner, D., & Van Hell, J.G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289-301.

Townsend, D.J., & Bever, T.G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, Matthew J. (2005). Plausibility and verb subcategorization in temporarily ambiguous sentences: Evidence from self-paced reading. *Journal of Psycholinguistic Research, 34*, 1-30.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research. *Studies in Second Language Acquisition, 33*, 339-372.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566-585.

Trueswell, J.C., & Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language, 39*, 102-123.

Trueswell, J.C., & Tanenhaus, M.K. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. *Perspectives on sentence processing*, 155-179.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language, 33*, 285-318.

Trueswell, J.C., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147-182.

Van Berkum, J.J.A., Brown, C.M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235-248.

Van Dyke, J.A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285-316.

van Gompel, R., Pickering, M.J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language, 55*, 335-362.

Van Hell, J.G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research, 26*, 43-74.

Van Herten, M., Chwilla, D.J., & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*, 1181-1197.

Van Herten, M., Kolk, H.H., & Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241-255.

Warner, J., & Glass, A.L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language, 26*, 714-738.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36*, 559-570.

Williams, J.N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausiblity constraints. *Applied Psycholinguistics, 22*, 509-540.

Wilson, M.P., & Garnsey, S.M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language, 60*, 368-392.

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics, 33*, 419.

Wlotko, E., & Federmeier, K.D. (2011). *Flexible implementation of anticipatory language comprehension mechanisms.* Paper presented at the Cognitive Neuroscience Society Meeting, San Francisco, CA.

Wlotko, E., & Federmeier, K.D. (2012). So that's what you mean! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62*, 356-366.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes, 27*, 1489-1524.